

PROMISCUOUS INVENTIONS

Modeling Cultural Evolution with Multiple Inheritance

JACOB G. FOSTER AND JAMES A. EVANS

“Trillian, this is my semi-cousin Ford, who shares three of the same mothers as me . . .”

—ZAPHOD BEEBLEBROX, *THE HITCHHIKER’S GUIDE TO THE GALAXY*

IN THIS CHAPTER, we argue that ideas and inventions—like Zaphod—can have many mothers. This is not always the default assumption. Powerful tools from macroevolution have been used to reconstruct cultural phylogenies (“trees”) in a variety of spheres (O’Brien et al. 2013), including language (Gray, Drummond, and Greenhill 2009), crafts (Tehrani and Collard 2002), and lithic technology (O’Brien, Darwent, and Lyman 2001). In order for these tools to retrieve accurate genealogies, however, the underlying patterns of cultural evolution must fit the assumptions of the biological methods—above all, the predominance of vertical information transfer (from “mother” to “child”) and tree-like branching. Such phylogenetic methods treat the horizontal transfer of information from other lineages as contaminating noise. For this reason, the application of phylogenetic methods to prokaryotic taxa has been challenged. Far from being “noise,” horizontal transfer is common among these organisms (Doolittle and Baptiste 2007).

Horizontal transfer is common in human culture, too, thanks to our rich communicative capacity and increasingly frequent population movement. In fact, scholars of modern technology often assume recombinant processes and hence substantial horizontal transmission (Arthur 2009; Wimsatt 2013b). This common assumption suggests that phylogenetic methods are not

always applicable to cultural data, just as they have limited application to prokaryotes. In a further contrast with biology, contemporary cultural evolution often leaves a detailed and relatively complete “fossil record” of past forms. Standard phylogenetic methods are designed to work without fossil data and typically only use them, if at all, for calibration (Felsenstein 2004; Gray and Atkinson 2003). More sophisticated methods that fully incorporate available fossil data (Fisher 2008; Huelsenbeck and Rannala 1997) exist but are rarely deployed.

In this chapter, we sketch an exploratory framework that learns from this rich historical data to infer the possible histories and patterns of cultural inheritance, beginning with which and how many “parents” contribute to each offspring. Simple vertical transfer is now treated as a special case of combinatorial evolution (Arthur 2009; Wimsatt 2013a), in which one or more parents from distinct lineages provide the raw materials involved in spawning a new “child,” which could be an invention, an organization, or a literal developing human being. We focus on the directed acyclic graphs (DAGs) that best trace the inheritance of known features and explain their observed distribution across cultural types. Temporal and geographical constraints on the space of plausible histories allow us to enforce the time directedness and spatial localization of inheritance. We describe a formalism that assumes independence in the choice of parents, but this can be relaxed to allow non-independence and structured parent choice. Groups of parents can “mate” with each other (or be “chosen” by offspring) according to a range of criteria, including unobserved but inferred fitness (i.e., appeal), past fecundity (i.e., number of offspring), or population structure (e.g., different disciplines or craft traditions that limit interbreeding between lineages). We describe both parsimony-based and probabilistic, generative approaches. Probabilistic methods are especially valuable when dealing with historical phenomena because they allow us to encode our assumptions about the underlying process and then reason rigorously from the data to a universe of plausible historical trajectories. In other words, these methods demand and leverage “new conceptual frameworks” to tackle the “massive new data sets” that are often available to trace the trajectories of cultural evolution (see chapter 1).

We show that our approach can apply to a wide range of cultural phenomena, including the evolution of technology, organizations, genres, and art forms, as well as the changing cultural constitution of individual human beings—that is, it can be used to model the “sequential dependencies in the acquisition of cultural traits during development” (see the introduction). We

conclude by describing the relationship between the modes of cultural inheritance revealed by this approach (branching or reticulate) and the mixture of transmission-isolating mechanisms (TRIMs) (Durham 1991) and transmission-accelerating mechanisms (TRAMs) that together shape inheritance and pattern cultural evolution.

This chapter responds to the key question raised by Love and Wimsatt in their introduction to this volume: “How to characterize cultural heredity with multiple parents.” In answering that question, we embrace core insights about the distinctive internal and external structures (Wade 2016) that influence cultural evolution: “Sequential dependencies in the acquisition of cultural traits” and “the roles of external structure” like institutions, organizations, and infrastructures in setting up the population structure that shapes cultural evolution (see the introduction by Love and Wimsatt; chapter 1 by Wimsatt).

Our approach incorporates several of the elements of an “adequate theory of cultural evolution” outlined by Wimsatt (chapter 1). It explicitly analyzes the complex lineages of “ideational, behavioral, and material items, which are capable of being modularly decomposed or chunked and black boxed hierarchically.” It can be used to model the complex cultural growth of “developing biological individuals” as well as organizations. Finally, in our analysis of TRIMs and TRAMs, we show how institutions and infrastructures can work together to produce and maintain “cultural breeding populations” and structure the processes of inheritance and invention by which culture evolves. In other words, it represents a sustained conceptual and mathematical effort to think “beyond the meme.”

THE QUESTION CONCERNING PHYLOGENIES

Are cultural phylogenies possible? In other words, are there cultural units whose evolution traces tree-like topologies? As recently as 1997, this was an open question (Boyd et al. 1997). Less than twenty years later, it has been answered decisively in the affirmative. Formal methods for phylogenetic inference, including parsimony, maximum likelihood, and Bayesian inference, have been applied to a range of cultural forms, from languages (Gray, Drummond, and Greenhill 2009) to projectile points (O’Brien, Darwent, and Lyman 2001) and textiles (Tehrani and Collard 2002); see O’Brien et al. (2013) for an extensive list. Explicit tests using a standard goodness-of-fit metric (the retention index) suggest that trees fit cultural data just as well as

they do biological data (Collard, Shennan, and Tehrani 2006). What is the alternative to the branching, tree-like pattern of cultural inheritance? Reticulation: a topology in which lineages not only split but blend, join, and recombine. Given that cultural phylogenies are indeed possible, why did we (and should we) consider reticulation?

Reticulation is plausible for a simple reason: the capacity to generate particular cultural traits moves with relative ease from one living individual to another. In other words, such traits or *transmissible elements* (Wimsatt 2013a) are capable of horizontal transmission between cultural lineages. A small note on terminology: when discussing the transmission of genetic information, biologists typically refer to *transfer*. In vertical transfer, genetic information flows from parent to offspring via reproduction. This preserves the integrity of lineages and ultimately builds up tree-like, branching topologies. In horizontal transfer, genetic information flows nonreproductively from one individual to another, potentially between distinct lineages. This breaks down the integrity of lineages and, if common, produces highly reticulate, recombinant topologies.¹ In both cases, the underlying genetic information is assumed to transfer unaltered, although under special circumstances it may undergo simultaneous mutation or recombination.

When discussing the flow of cultural traits (Mesoudi 2011), evolutionary anthropologists often refer to *transmission* rather than transfer, making an analogy to epidemiology. In vertical transmission, cultural traits flow from parent to child. In horizontal transmission, cultural traits flow between a pair of individuals, who may be unrelated.² Vertical transmission helps to preserve the integrity of cultural lineages. Horizontal transmission can produce distinct and well-separated lineages, as long as the reach of horizontal transmission is limited—for example, via TRIMs (Durham 1991) that maintain separate cultural breeding populations (see chapter 1). In contrast to the genetic case, however, the “flow” of cultural traits can be much more complex than the “flow” of genetic information. While genetic information can be copied with minimal error, transmitting the capacity to manifest particular cultural traits is nontrivial. Far from mere copying (as in simple memetic pictures), it often involves detailed reconstruction and reverse engineering (Claidière, Scott-Phillips, and Sperber 2014) and can depend on the prior acquisition of other cultural traits that scaffold sequential acquisition (see the introduction by Love and Wimsatt, as well as chapter 1).

Despite the need for reconstruction and reverse engineering, the horizontal transmission of cultural traits from person to person can be relatively

easy, and hence the transmission of cultural traits from one cultural “lineage” to another becomes possible. As Stephen J. Gould (2010) quipped in an oft-cited quote, “Five minutes with a wheel, a snowshoe, a bobbin, or a bow and arrow may allow an artisan of one culture to capture a major achievement of another.” While Gould underestimates the difficulty of inferring a generative procedure from an artifact, five *years* as apprentice to an artisan from another culture probably suffices for the horizontal, cross-cultural transmission of many major technical achievements (see chapter 8). Given the right scaffolding³ and enough time, a novice will acquire the skills, knowledge, and practices that make her “infectible” by a new technology (Wimsatt 2013a).⁴ The same is true for other cultural traits, like complex beliefs. For this reason, proponents of cultural phylogenetics are careful to emphasize that tree-like structures may not always be appropriate (Cochrane and Lipo 2010). As we noted above, tree-like structures are not always suitable for biological evolution either, as in the case of bacteria, with their rampant horizontal transfer of genetic information via plasmids, transformation, or transduction (Gogarten and Townsend 2005). Whether phylogenetic trees can accurately represent a particular evolutionary history is therefore an empirical question, not a theoretical one. Trees provide a reasonable representation of particular histories of cultural transmission when TRIMs (Durham 1991) limit or prevent cross-lineage transmission for the trait in question and thus maintain branching as the dominant mode of cultural macroevolution (at a certain level of analysis).

As Mesoudi (2011) notes, however, the TRIMs that apply to projectile points and textiles (e.g., language, limited intergroup contact, and ethnocentrism) are unlikely to apply directly to the evolution of scientific ideas and technological inventions—although related social mechanisms might, along with the need for scaffolded skill acquisition (Goodwin 2017).⁵ Instead, the picture presented by the literature on science and technology is positively promiscuous, with recombination an essential and often primary process (Fleming and Sorenson 2001, 2004; Uzzi et al. 2013; Arthur 2009; Foster, Rzhetsky, and Evans 2015). This implies that reticulation should be common; that ideas and inventions—like Zaphod in our epigraph—can have many mothers.

Before proceeding, an important caveat. We note that vertical and horizontal transmission at the level of people is logically independent from vertical and horizontal transmission at the level of specific cultural traits or products. For example, the degree to which a biological individual’s repertoire of cultural traits or products (e.g., ideas, beliefs, practices, technologies)

emerges within a social lineage or across them is independent of whether specific, novel instances of cultural traits or products are produced through conservative, vertical tinkering or liberal, horizontal recombination. We focus here not on the vertical or horizontal transmission of cultural traits from person to person (though see section 7) but on the vertical or horizontal transmission of elements from one cultural trait to another. Still, the two may be empirically related. If the capacity for horizontal transmission from person to person were limited, then possibilities for the combinatorial generation of new culture would be highly constrained. As a result, promiscuous horizontal transmission between people is a necessary but not sufficient condition for promiscuous *combinatorial invention* in which cultural artifacts have multiple parents.

TACKLING MULTIPLE INHERITANCE

This picture of promiscuous combinatorial invention suggests that standard methods of phylogenetic inference will often produce distorted pictures of the pattern of cultural evolution in contemporary science and technology.⁶ It is worth noting that some extensions of phylogenetic methods permit horizontal transmission between lineages (Nicholls and Gray 2006). Inference can be robust to moderate levels of horizontal transmission (Greenhill, Currie, and Gray 2009), and network-based methods can detect signals of reticulation directly (Lipo 2006; Gray, Bryant, and Greenhill 2010; Huson Rupp, and Scornavacca 2010). Ignoring reticulation, however, runs the risk of distorting a true history by forcing multiple inheritance into a branching tree. Network-based methods, on the other hand, are essentially exploratory. Because they lack an underlying generative model (but see Wen, Yu, and Nakhleh 2016), they can neither draw on existing knowledge about the inventive process nor represent uncertainty or improve with additional data (Ghahramani 2015). Nevertheless, cultural evolution is sufficiently complex (see chapter 1) that all models are distortions (Wimsatt 2002), and there is no one “true” representation for every trajectory. As Doolittle and Baptiste (2007) note, “Different evolutionary models and representations of relationships will be appropriate, and true, for different taxa or at different scales or for different purposes.” We thus embrace pattern and process pluralism in cultural just as in biological evolution.

Pattern pluralism aside, traditional phylogenetic methods do not take advantage of a distinctive feature of contemporary technological evolution: the

incredibly rich “fossil” record of past forms, which often possesses detailed information about timing, sequence, and spatial location (Evans and Foster 2011). In biology, such information is often sparse and always hard to come by. It usually involves a lot of digging and scraping. For this reason, phylogenetic methods are generally designed to make inferences in the *absence* of extensive evidence about past forms. Available information can be used to “root” trees with an out-group or to calibrate particular branching points (Felsenstein 2004). The latter application is relatively common in the reconstruction of linguistic phylogenies, where fossil traces, for example, written materials that have persisted to the present, are exceedingly rare (see, for example, Gray and Atkinson 2003). Biologists have also developed parsimony (Fisher 2008) and likelihood-based (Huelsenbeck and Rannala 1997) methods that penalize trees if they infer ancestral states with no trace in the fossil record.

We set these methods aside for two reasons. First, they are fundamentally tree-based and hence suffer the same problem of forcing multiple inheritance onto branching topologies.⁷ Second, they depend on the inference of past forms. Yet past forms are densely documented in scientific and technological data sets (see chapter 6), thanks to ongoing incentives to publish (Merton 1973) and patent (Owen-Smith and Powell 2001). This rich history is ripe for analysis, thanks to the increasing electronic availability of data (Evans and Foster 2011). Scholars of technological evolution should use it, and phylogenetic methods are simply not designed for situations in which history is as richly and densely documented as it is in science and technology. Taking this record into account will dramatically improve our characterization of the process that generated it, as well as our prediction of what will come next.

Finally, ancestral cultural forms can influence the present in a way that ancestral biological forms cannot. When a species goes extinct, its distinct genetic information is lost forever, along with its phenotype, behavior, and entailed ecological interactions, the loss of which can tip other species toward extinction. When an idea or technology “goes extinct”—in the sense that it no longer occupies any minds or has a physical presence in the contemporary population—it can nevertheless contribute to a new technology or idea. If an artifactual or textual trace of the extinct idea remains, contemporary inventors can draw components, features, or inspiration from it (Tëmkin and Eldredge 2007). Because any act of cultural transmission always involves some inference and reconstruction (Claidière, Scott-Phillips,

and Sperber 2014), even limited traces of an earlier idea or technology can contribute ingredients to a novel cultural unit decades or even centuries later.

These considerations suggest that if we take the possibility of multiple inheritance seriously, we should develop models that allow both reticulate patterns of multiple inheritance and tree-like patterns of descent with modification. In other words, we should go beyond the inference of phylogenetic trees. In the rest of this chapter, we introduce exploratory and model-based methods that can detect and describe multiple inheritance in densely sampled cases of cultural evolution. These methods themselves have multiple parents; in addition to the phylogenetic tradition, they draw on ideas from latent variable modeling (Blei 2014), Bayesian nonparametric models (Gershman and Blei 2012; Ghahramani 2013), probabilistic machine learning (Ghahramani 2015), and network analysis (Newman 2003). In the following sections, we provide a high-level description of these methods and the underlying ideas and intuitions. The appendix gives specific mathematical descriptions of several methods. We also discuss the philosophy behind inference using probabilistic models. We close with a reflection on the role of entrenchment and scaffolding in the tempo and mode of technological evolution.

Before proceeding, we describe in words and a little notation our general picture of the evolution of ideas and technologies. Imagine that we observe an invention at time t_j . Given our observation, we know that at some earlier time $(t_j - \epsilon)$ a “creative unit” (which could be an individual inventor/scientist or a team) must have assembled (consciously or not) a set of influences P_j . For example, the Bessemer process of steel production involved removing impurities from pig iron with oxidation by blowing air through the molten metal (Birch 1967). “Parental influences” here include pig iron, the oxidation process, and, ultimately, the use of dolomite or limestone linings for the Bessemer converter. Taken together, these influences P_j provide the raw material from which the invention was assembled; hence the set P_j contains the parents of the new invention j . If cultural evolution in this particular domain is dominated by vertical transmission and descent with modification (i.e., tinkering), then P_j may only have one member, and the invention j only has one parent (for example, the inventor slightly adjusts the technique typically used to process a particular stone). If cultural evolution is dominated by horizontal transmission and is combinatorial, then P_j may have several members, and the invention will have multiple parents. Note that, in principle, any invention that precedes invention j in time is a possible

parent. The set of all possible parents is denoted \mathcal{P}_j . It is time ordered (each earlier invention is time stamped). Depending on the typical length of the inventive process and the typical difficulty of mastering a new invention, it may take some time before a given invention p can become a parent. This implies a lower bound on the difference between the time of invention t_j and the time of observation of an allowable parent t_p , such that some time Δt_{jp} must have passed before p is a possible parent of j .

Any idea or technology may be coarsely characterized by its elementary building blocks; as Wimsatt notes in his discussion of transmissible or replicable elements (chapter 1), TREs can be “modularly decomposed.” The outcome of this modular decomposition might be the components that make up an invention or the concepts that make up the idea of a scientific paper. We refer to these parts as *features*, and the set of all features as \mathcal{F} .⁸ This characterization is a necessary precondition for analysis, but it is more than a useful fiction: for any particular community of practice, the relevant coarse-graining—the “principles of vision and division” (Bourdieu 1990; Foster, Rzhetsky, and Evans 2015)—will be relatively consistent.⁹

For any invention j , each of its parents $p \in P_j$ is characterized by its own set of features $F_p \in \mathcal{F}$. To create a new invention j , the inventor selects its features from the set of features possessed by its parents. This set is simply the union of all the parental feature sets: $\bigcup_{p \in P_j} F_p$. Occasionally, an invention introduces an entirely novel feature rather than just drawing on the features of the past. At other times, an invention may “bundle” together several preexisting features (from one or several parents) into an effective, integrated unit. This new unit becomes a “feature” available to future inventions—its constituent subfeatures are henceforth sampled together. This process is called *black boxing* (Latour 1987). To embrace these generative possibilities, our model must allow inventors to black box or introduce a novel feature with some probability (which will typically be small in cases where invention is largely combinatorial).

Note that this inventive process can generate a range of inheritance pathways and hence topologies. It can describe unilineal inheritance, in which a single (cultural) parent is selected and (perhaps) slightly modified. It can also describe multilineal inheritance, in which multiple (cultural) parents are selected and their features recombined. Starting from a picture of invention that is entirely agnostic about tinkering versus recombination is essential if we are to let the rich traces of inventive activity reveal the underlying modes of inheritance and pattern(s) of technological evolution. Such an agnostic

analysis can also provide data-driven hints as to the modes and mechanisms of cultural evolution.

Before defining a model-based approach for studying multiple inheritance, we describe some exploratory methods for tracing multiple inheritance in densely sampled, time-ordered data. These methods are much simpler than the model-based approach but require further assumptions about the inventive process.

GENERALIZING PARSIMONY

Parsimony-based methods provide powerful exploratory tools for constructing possible phylogenies (O'Brien et al. 2013). Here we describe methods than can reconstruct possible reticulated histories (directed graphs) in the case of multiple inheritance. As a modeling strategy, parsimony emphasizes simplicity; it seeks the minimal explanation for the observed facts. In phylogenetic reconstruction, the present distribution of features (genetic or morphological) provides the observed facts; the phylogenetic tree provides a potential explanation for those facts. In seeking the simplest explanation, parsimony methods minimize the number of genetic changes implied by the proposed tree. A tree that can explain the present distribution of features with five changes is preferable to a tree that requires six. The intuition underlying this simplicity criterion is quite plausible: not only are fewer changes “simpler” in an absolute sense, but every change (mutation) is a low-probability event. Hence, we should generally seek explanations for present facts (i.e., trees) that minimize the number of such events (note that this basic idea is also used in some approaches to phylogenetic networks; see Huson and Scornavacca [2011]).

What is the analog of this parsimony principle in our generalized model of inheritance? Recall that in our model, new inventions sample over the features of past inventions. All else being equal, it probably takes more time and effort to sample from three past inventions than from two. Hence, the most parsimonious or simplest explanation might minimize the number of past inventions needed to account for the features of the present invention. On the other hand, consider the following scenario. A new invention has six features. Four of those six features can be found in a single predecessor; the remaining two can be found separately in several possible predecessors. This history would lead to three “parents.” Alternately, three of the features can be found in a single predecessor and the remaining three in another predecessor.

This history would lead to two parents. Which history provides the simplest explanation? Naive parsimony (i.e., minimizing the number of parents) ignores the fact that a single parent can account for the majority of features in the new invention; in such ambiguous cases, it's quite likely that there are many ways to account for the residual features, whereas there might be only one history that splits all features across two parents. We balance these various nuances of the parsimony principle through a *greedy* inventive process. In essence, we assume that at any stage of invention inventors draw as many features from a parent as possible.¹⁰ In particular, inventors tend to draw a large number of features from one parent and a small number from several others, rather than drawing a moderate number from two or three. This greedy assumption also makes the problem more computationally tractable, as we do not have to search over different combinatorial histories.¹¹ At every step, we pick the simplest explanation—the parent that accounts for the most features.¹²

We now describe how to implement this parsimony principle in practice. Before implementing the following algorithm, we first reduce the feature sets of all inventions by removing any novel features—features that appear for the first time in that particular invention. These features cannot be accounted for by the past. In cases where the same novel feature appears in multiple inventions in the same time slice, we treat those features as novel for all inventions. Now, we execute the following algorithm for each invention j . Note that this algorithm can be executed independently for every invention.

PARSIMONY ALGORITHM

1. Establish the set of possible parents. This may be the set of all earlier inventions, or it may have some restrictions (e.g., all inventions more than six months older than the focal invention).
2. For each possible parent, count up the number of features in the focal invention that could have been inherited from that parent.
3. Add to the set of j 's parents whichever prior invention explains the most features (and remove that invention from the set of possible parents). If there are multiple equally explanatory inventions, we can implement additional principles of simplicity as desired. For example, we can prefer the most recent ancestor or the ancestor with the smallest spatial distance, social distance (e.g., as computed in a social network), or cognitive distance (e.g., as computed in a network or space of skills,

ideas, etc.). If *all* parsimony principles have been exhausted and multiple possible parents remain, choose at random.

4. Eliminate from the feature set of the focal invention all features that have been explained by the parent set.
5. If features remain to be explained, go back to step 2. Otherwise, stop.

We repeat this procedure for all j to reconstruct a parsimonious history for our observed technologies. This history will be a directed acyclic graph (DAG); we establish the convention that arcs run *to* an invention j from its parents p , k , m , and so on to represent the flow of ideas from past to present.¹³ This directed graph can be weighted, with arc weights counting the number of features inherited from each parent. Because of the random steps in the construction of the DAG, we should construct multiple complete histories for a given data set and look at properties of the ensemble, which is only necessary if the random number generator is called. Because the reconstruction process is independent for each distinct invention, it can be trivially parallelized. See appendix for mathematical details.

Parsimonious Insights

What can we learn from the DAGs reconstructed via parsimony? First, remember that we are trying to make inferences about an unobserved history of invention and inheritance from richly sampled but incomplete evidence. In general, we will not have traces of the inheritance process (i.e., that the inventors of technology j drew on technologies p , k , and m); even when we do, those traces are incomplete and potentially biased. What we do have is a record of what technologies with what features exist at what times. Our inferences are also shaped by assumptions about the inventive process that has generated this record—namely, that it is a greedy local search, in which inventors sequentially sample the space of possible parents and prefer to extract as much as possible from each parent along the way, until their new invention is complete.¹⁴

We can mine several insights from the weighted DAG that represents a parsimonious reconstruction of cultural inheritance under this model of technological evolution and the inventive process. First, we can ask what fraction of “explicable” features in each invention is inherited from each parent. For a particular invention, this tells us whether most of its features come from a single parent or whether its features can be better explained by even

sampling from several parents. Consider the largest such fraction for each invention; call this the *primary inheritance*. The frequency distribution of primary inheritance reveals how many inventions are largely explicable with a single parent and how many require multiple parents to explain their features (see Bedau, chapter 6, for an empirical analysis of multiple parentage in U.S. patents). From this distribution, we can obtain a good guess at the dominant mode of cultural inheritance in a given domain and make inferences about the dominant mode of cultural evolution. If the distribution is peaked at large values (close to one), then the mode of inheritance is primarily unilineal, and evolution proceeds via descent with modification. We can turn the DAG into a tree by retaining the highest-weight incoming arc for each node. This tree likely represents a good first approximation of the inheritance pattern. At the very least, it suggests that vertical transfer and descent with modification together provide a parsimonious explanation for the facts. If the distribution is spread out across possible fractions—or even peaked at lower values—then we have evidence that the inheritance pattern is reticulate, involving multiple parents, and that the mode of evolution may be combinatorial. Given that our greedy reconstruction process is biased toward trees, a broad distribution of primary inheritance provides substantial evidence for reticulate cultural evolution and multiple parentage.

Second, consider the number of features present in a given invention p . These are the features that could be passed on to any descendants. In a given reconstruction, we can compute the fraction of such features that are actually passed on to each descendant; call this the *primary contribution*. The mean and mode of this quantity, computed over all descendants, can tell us whether the features of p are typically inherited as a block or whether they are separable and used as a selective smorgasbord. While block inheritance of features happens in biology (e.g., genes are bundled together into chromosomes), cultural and especially technological evolution is distinctive in its capacity to create such building blocks from more primitive pieces; this is a key part of the internal or *endogenetic* structure of cultural inheritance (see the introduction by Love and Wimsatt).¹⁵ We briefly discuss how such black-boxing events can be detected in parsimonious DAGs.

Black Boxing

We can use the two measures described above to identify potential moments of modularization or black boxing (Latour 1987). In practice, black boxing may involve miniaturization (allowing a bundle of features to “fit” in new

places); compression (simplifying components, removing redundancies, and integrating parts to maximize efficiency); autocatalysis (relations of mutual dependence across parts that sustain and reproduce coparticipation; see chapter 11 of this book); and the streamlining and/or standardizing of input and output (making it easier for the set of features to recombine; see chapter 2 for the importance of standardization to combinatorial processes in genomics and proteomics). The key signature of black boxing for a particular invention p is the relative size of its (average) primary contribution, compared to all other inventions. If recombinant evolution is typical, the average size of the average primary contribution will be relatively low. When an invention has an above-average primary contribution across descendants, this strongly suggests that its components are black boxed and drawn upon as whole units rather than as a set of parts. Now consider the primary inheritance of a specific invention k . If the primary inheritance is low, then invention k has sampled from several sources. When the primary inheritance is low and the primary contribution is high, this suggests that k has drawn on several parents and bundled the parts together into a unit with emergent value.¹⁶ There is a synergistic, nonadditive, epistatic interaction among the parts, which leads others to select the whole black box. We can validate this intuition using related traces. For example, black-boxing events will likely correspond to cases in which the citations to a black-boxing patent supersede and largely replace citations to the patents (and separable components) on which it draws (Funk and Owen-Smith 2016).

PROBABILISTIC MODELS, POSSIBLE HISTORIES

Before discussing model-based approaches to the study of cultural genealogies, we pause to discuss the role of models, uncertainty, and evidence. Despite the rich electronic record of inventive activity in science, technology, and other cultural domains, much remains unknown. Specific influence pathways may be discoverable, but only after considerable effort—for example, using traditional historical methods. Hence, the principled integration of model and evidence is important and the rigorous representation of uncertainty essential. Probabilistic models provide a comprehensive framework for such integration (Ghahramani 2015).

Insights from the previous section were limited in two ways. First, our model of the discovery process was implicit and narrow: greedy search. While this model provided a useful parsimony principle allowing us to construct

well-defined, parsimonious “explanations” of observed histories (i.e., DAGs), it may distort inference insofar as it misspecifies the inventive process. By biasing reconstruction toward tree-like structures, parsimony provides a conservative test for multiple inheritance. Data that support a parsimonious explanation with multiple inheritance are quite likely to have been generated by some kind of recombinant process, but the details are likely to be wrong, and we learn nothing about the inventive process from the data. This leads to the second limitation. We have no idea how much certainty to have in our reconstructed cultural genealogy.

Probabilistic model-based inference has neither of these difficulties. First, we can construct a much more flexible model than greedy search. This flexible model allows us to specify what we know about cultural inheritance (e.g., from qualitative or historical investigations of innovation, of which there are several examples in this book)—and what we do not know. This lack of knowledge is represented by parameters in the model: we may have a general sense of the underlying generative process, but different model parameters realize different generative scenarios. Any hunches we have about the generative process can be further specified through priors on the parameters. For example, our model might have a parameter controlling the average number of parents that contribute to a new invention (it will). If we have a strong reason to suspect that the average number of parents is two, then we can put a prior on that parameter concentrated around two. If not, we may choose a totally uninformative prior to represent our uncertainty about its value. But most of the action in probabilistic modeling does not take place in the priors; it takes place in inference. Inference is simply a process of learning from the data. The rules of probability (specifically, Bayes’ rule) allow us to use data to update our uncertainty. Doing so avoids the second limitation of parsimony methods: we can precisely quantify our certainty in the reconstructed genealogy.

Flexibility is important when reconstructing cultural genealogies, because of our agnostic position on patterns and processes of cultural evolution. It is very likely that cultural evolution follows different patterns in different domains. We know that stone tools and some features of language (to pick two examples) follow branching patterns. We strongly suspect that some areas of high technology follow combinatorial, reticulate patterns (Fleming and Sorenson 2001, 2004; Arthur 2009). Model-based inference allows us to discover different patterns and processes of cultural evolution in different domains. We do not claim that our models perfectly describe the

world (even after inference). We do claim, however, that they give a relatively precise sense of the generative processes and historical trajectories that could explain available evidence. Crucially, our models can focus attention on the most plausible or informative influence pathways that merit detailed, costly historical or ethnographic investigation (Wimsatt 2013b); in other words, they can provide structure to the larger problem agenda of understanding cultural evolution in specific domains and guide the attention of relevant disciplinary partners to maximize the value of their contributions (see the introduction by Love and Wimsatt).

LEARNING ABOUT MULTIPLE INHERITANCE

With model-based inference, we allow the data to reduce our uncertainty about the nature of the inheritance process (Ghahramani 2015). As input data, we again have a set of types ordered in time. These types could be patents, publications, products, or other complex cultural entities (e.g., organizations or people—anything decomposable into documented building blocks; see chapter 1). Each type j is characterized by a unique set of features.¹⁷ A type may correspond to multiple entities, insofar as these entities are “indistinguishable” from the perspective of these features. The more refined the set of elementary features (i.e., the larger the number of distinct features), the more types there will be. Consider, for example, the description of patents using a few classification codes, as opposed to more detailed descriptions extracted and normalized from full text. In the former case, many patents might correspond to the same type; in the later case, a single type might correspond to just a few patents, or even a unique one.

Types are ordered in time; we can retain time as a component of the model to account for time intervals, as in parsimony. We can also use temporal information to model the probability that the recent and the ancient past are considered as sources of potential parents. If we have information about the spatial, social, and cognitive “place” of invention, we can learn from the data whether there are similar “local” biases (Adams 2002). For example, we might have detailed information about the time and place of invention. Since recent inventions are generally easier to retrieve than much older inventions and local knowledge is easier to access than distant knowledge (e.g., due to the institutional or organizational structuring of cultural breeding populations), we might modulate the probability of choosing a particular

parent by a decaying function of temporal separation and geographic distance. But since we do *not* know how much more likely inventors are to retrieve recent or local knowledge over ancient or distant knowledge, we characterize that function with unknown parameters. We can learn from the data a reasonable range of possible parameter values.

For simplicity and concreteness, we describe the model as a generative process. From the generative model, we can construct a joint probability distribution over the observed types \mathbf{F} , the DAG of parentage assignments \mathbf{P} , and the parameters Θ that control the number of parents and the sampling of features from parents. Given the joint distribution, we can construct the posterior probability over DAGs and parameters conditional on the observations using Bayes' rule. The posterior probability is our ultimate target: given our modeling assumptions, our priors, and (most importantly) our available evidence, we can sample from the posterior probability distribution to discover which DAGs are more (and less) likely and which parameter values are more probable, given the evidence.

This is conceptually identical to the standard Bayesian approach to phylogenetic tree reconstruction (Felsenstein 2004; Bergstrom and Dugatkin 2012). In that case, we want to construct (or at least sample from) the posterior distribution over trees, conditioned on available data D . In principle, we may have a prior over trees; in practice, a flat prior is usually used so that every tree is equally probable, *a priori*. Using Bayes' rule, we can construct the posterior:

$$\Pr(\text{Tree} | D) = \frac{\Pr(D | \text{Tree})\Pr(\text{Tree})}{\Pr(D)} \quad (1)$$

where parameters of the model of character or sequence evolution have been suppressed. The likelihood $\Pr(D | \text{Tree})$ is well defined and can be easily computed; it is the probability of the observed data given a particular tree, a particular model of evolution, and particular parameter values characterizing that model.

Our generative model for multiple inheritance can be quickly summarized by listing its steps. The model begins at the earliest observation and iterates over the following:

1. Choose the number of parents.
2. Choose the identity of the parents.
3. Choose features from the set of parents.

This is, of course, the same basic picture that guided our parsimony method above. In a fully Bayesian approach, we would begin the generative process by drawing the parameters from prior distributions (Gershman and Blei 2012). Given these parameters, we would then iterate the steps above. Note that even here some assumptions are baked into the model; for example, the number of parents is not influenced by the identity of the parents nor can the features selected from one particular parent influence the selection of subsequent parents.¹⁸ We now (briefly) describe each step; we provide mathematical details in the appendix.

Number of Parents

For a given observation, the generative process begins by choosing the number of parents. The number of parents is drawn from a distribution controlled by one or more parameters θ_p . The simplest such distribution would be a Poisson, in which case the parameter would control the average number of parents. This picture is similar to the so-called Indian buffet process, in which customers sample dishes from the buffet until they have chosen a number of dishes drawn from a Poisson distribution (Griffiths and Ghahramani 2011). Using a Poisson distribution, however, assumes that there is a typical number of parents and that the distribution of parents is tightly peaked around that number. That might not be the case—another instance in which model specification will shape inference. Ideally, one would explore models with alternative distributions (and mixtures of distributions) and check them using techniques for model criticism, as through predictive sample reuse or posterior predictive checks (Blei 2014). In full generality, one might permit the parameter(s) controlling the number of parents to change over the course of evolutionary history, allowing one mode of cultural evolution (e.g., branching) to dominate earlier portions of the DAG and another mode (e.g., reticulation) to dominate later parts. See Silvestro et al. (2014) for an inspiring approach to capturing such shifts.

Choosing Parents

Once the number of parents has been selected—equivalently, once we have selected the in-degree of the node j in the directed acyclic graph representation—we must choose *specific* parents. There are many ways to formalize this choice process. For simplicity, we assume that each parent is chosen independently. If each parent, in turn, has an equal probability of

being chosen (a highly unrealistic assumption), then each invention will have an asymptotically Poisson number of offspring (i.e., out-degree). A slightly more complicated model, imitating the Indian buffet process, assumes that inventors start with the most recent potential parents and then work backward in time. Each potential parent is considered; it is selected as an ancestor with a probability proportional to its popularity (i.e., its current number of offspring or, equivalently, out-degree). This process repeats until the full complement of parents is chosen. In this model, preferential attachment (which asymptotically produces a power-law distribution of out-degrees) competes with recency bias. Although older nodes may have given birth several times, they are less likely to be selected as they get older; more nodes must be “skipped” to get to them. In general, the probability of choosing a particular parent can depend on many different factors. Parents might have an intrinsic “fitness.” This fitness could be drawn from some distribution when the parent is initially created. More realistically, the fitness could be determined by the constellation of features present in the invention (thus allowing for inventions with similar features to have similar fitness). Preferential attachment (rich-get-richer dynamics) could play a role, reflecting prestige bias, conformist bias, or both (Boyd and Richerson 1988; Mesoudi 2011). Parent choice may be shaped by explicit markers of social identity, such as disciplinary, professional, or institutional affiliation (see chapters 1 and 12), as well as by temporal, spatial, social, and conceptual distance. Finally, we could explicitly model choice-set formation so that inventors make a cognitively plausible choice across a small number of possibilities, rather than implicitly considering the entire universe of possible parents (Swait and Ben-Akiva 1987; Bruch, Feinberg, and Lee 2016). These more complex models of parent choice would allow researchers to test important assumptions. For example, we could discover that inventors are more likely to select a set of parents from the “same” cultural breeding population (e.g., scientific discipline or technology area).

Number of Features

We assume that the number of features sampled from the parents is independent of the number or identity of the parents. This is, again, a simplifying assumption; it could be that inventions with more parents tend to sample more features or that inventions with high-fitness or popular parents sample more. As with the number of parents, the simplest choice for this distri-

bution is Poisson, though this could be generalized to admit more complex distributions.

Increasing Complexity

As currently described, this generative model has a major limitation. It cannot easily deal with cultural evolution in which features *accumulate*. Yet this is an incredibly common mode, both in technological evolution (Arthur 2009) and in the sequential acquisition of skills by developing biological individuals (Love and Wimsatt, the introduction to this book; see also Wimsatt, chapter 1). Building blocks already characterized by many features can be used to assemble an even larger invention, such as airplanes and boats combined into an aircraft carrier (Arthur 2009). If our types are defined by features at a consistent granularity, then later inventions may have more features, on average. We can capture this by allowing the average number of features to grow over time; this growth rate can be controlled by one or more parameters subject to inference. A more interesting approach would allow the data to “suggest” bundles of elemental features that should be treated as a single feature because of frequent copresence. This compression or dimensionality reduction of the feature space implements a form of parsimony; it attempts to simplify the explanation of observed facts by reducing the number of components. There are a range of approaches to so-called feature or representation learning. Matrix factorization (Bengio, Courville, and Vincent 2013) could be applied periodically or continuously to update the feature space confronting inventors; the compression schedule could be optimized so that the number of compressed features in any given invention remains relatively constant. A latent aggregation–fragmentation process provides a purely probabilistic alternative. Features can aggregate into a bundle with some small probability, and bundles can disaggregate into constituent features with another (Ghahramani 2013; Blei 2014). This would provide an explicit probabilistic model of the black-boxing process. A more radical alternative would replace surface features with latent *feature generators*, emulating topic modeling (Blei 2014). However it is implemented, such chunking (Wimsatt 2013a) is consistent with both plausible limits on working memory (Miller 1956) and the robustness of modular assembly (Simon 1969; Latour 1987; Arthur 2009). As in the exploratory analysis, the consistent chunking of several features into a bundle suggests a black-boxing event and could be used to detect such moments in the unfolding cultural-evolutionary process.¹⁹

Choosing Features

Once the number of features has been selected, we must choose specific features. As with parent selection, it is simplest to assume that features are selected independently. Indeed, the simplest version of feature selection would look very much like parent selection, moving through features in some order and selecting them proportional to their popularity, either over the entire past history of the system or over the set of parents.²⁰ Unlike parent selection, however, we allow the creative unit (the inventive individual or team) to introduce some number of new features unobserved in the parent set—and possibly never yet observed in the history of the system. This step allows for the introduction of radical novelty to the inventive system; not just the novel combination of features but the addition of entirely new features (Foster, Rzhetsky, and Evans 2015).²¹ On a more mundane level, this modeling assumption allows any invention to be generated from any parent set, albeit with very small probability. This is useful computationally. It is also important substantively: it may be that inventors introduce a particular feature by plucking it from the inventive zeitgeist, rather than drawing on a particular parent. The capacity to generate new features also connects this generative model to Bayesian nonparametric processes more generally, as the number of potential features is not determined a priori in the model, although it is obviously given by the data.

Inference

Although somewhat nonobvious from the generative description, the model outlined above is remarkably close to standard phylogenetic inference in structure. Instead of a tree, the parentage assignment \mathbf{P} describes a DAG that respects the time ordering of inventions \mathbf{F} . Earlier inventions point toward later inventions that draw on them for features. For a given parentage assignment and values of the generative parameters Θ (i.e., the two explanatory parts of the model), we can calculate the probability of the data $\Pr(\mathbf{F} \mid \mathbf{P}, \Theta)$ directly. We have priors on the model parameters $\Pr(\Theta)$, which may be informative or uninformative. Given the parameters, the probability of any particular DAG \mathbf{P} is determined. We can combine all these parts using Bayes' rule to compute the posterior distribution over the space of DAGs (i.e., explanatory histories) and parameters (i.e., explanatory processes). It is

$$\Pr(\mathbf{P}, \Theta \mid \mathbf{F}) = \frac{\Pr(\mathbf{F} \mid \mathbf{P}, \Theta) \Pr(\mathbf{P} \mid \Theta) \Pr(\Theta)}{\Pr(\mathbf{F})}. \quad (2)$$

The denominator cannot be calculated because computing the probability of the data requires a sum over all possible DAGs \mathbf{P} . We can approximate the posterior, as in phylogenetic inference, using standard methods like Markov chain Monte Carlo (Gershman and Blei 2012) to draw from or otherwise approximate the posterior distribution.

MODELING THE CULTURAL EVOLUTION OF DEVELOPING INDIVIDUALS, ORGANIZATIONS, AND INSTITUTIONS

While our approach was inspired by the challenges of modeling multiple inheritance in technological evolution, it can be applied to any cultural data for which there is dense sampling and information about the sequence of observations. One particularly exciting application concerns data in which well-defined units with temporal duration but malleable features (e.g., individual humans, organizations, genres, or states) are observed repeatedly. In this case, we can view an observation of unit j at time t as a recombination of its state at last observation with features drawn from other available “parents.”

This strategy emulates the approach suggested in Boyd and Richerson (1988) for theoretical models of horizontal transmission during the life span. In other words, unit j selects its characteristics at time t by sampling from its previous state as well as from its contemporaries and predecessors. The astute reader will have noticed that this model is very close to models of social contagion; given this similarity, we must be vigilant against the possible confounding of social contagion with latent homophily (Shalizi and Thomas 2011). That said, the adopted feature(s) must come from *somewhere*, and it is possible that specific contagion versus diffuse adoption driven by latent homophilous traits can be distinguished by the presence or absence of specific influence paths in the posterior distribution over DAGs.

In practice, capturing the known features of human cultural development, such as the sequential nature of skill acquisition, would require relaxing many of the assumptions outlined above. The features retained by unit j from its past state would affect its selection of “parents” for cultural updating, as well as the features chosen from them (Foster 2018). For example, it is almost surely the case that someone who knows single-variable calculus at time t and multivariable calculus at time $t + 1$ retained his knowledge of single-variable calculus and learned the multivariable version from his teacher and/or textbook. It is also likely that this teacher is someone close in physical, social, and organizational space. Incorporating geographic or

social proximity in the choice of cultural parents, including evolving markers of social identity (see chapter 12), would allow us to deal directly with cultural population structure (Wimsatt 2013a). Note that a model of “parent” selection incorporating cumulative advantage is very similar to prestige bias, an important mechanism in cultural microevolution (Boyd and Richerson 1988; Mesoudi 2011). In our running example, this imaginary student is more likely to select a popular model known for her excellent pedagogy. The student may pick up other cultural traits as a by-product of this learning relationship, such as a specific story or preference for a certain mode of investigation. This same framework would provide a powerful and precise technique for studying the evolution of organizations and institutions more broadly, as there are often repeated observations of these units.

In other words, this formal trick extends the range of our framework from cultural macroevolution to the microevolutionary dynamics of cultural change. We thereby provide an intriguing twist on Wimsatt’s observation in chapter 1 of this volume that heredity and development “interchange roles in the study of biology and culture,” with cultural development being more transparent to investigation and hence helping to illuminate cultural heredity. In our framework, long-term patterns of cultural heredity and short-term patterns of cultural development are treated in the same way!

TESTING MODELS OF MULTIPLE INHERITANCE

In describing our approach to the study of multiple inheritance, we have emphasized that studying cultural macroevolution requires uncertain inference of unknown processes from rich data. How might we validate these models? We briefly mentioned internal checks using model criticism, as through predictive sample reuse or posterior predictive checks (Blei 2014). Such checks are important, but they are unlikely to persuade the obdurate skeptic. Thus, we note that, just as biologists rely on paleontologists to validate the presence of particular extinct organisms at particular times, so too can students of computational cultural evolution turn to historians, sociologists, anthropologists, and archaeologists to validate particular claims about particular influence pathways and inventive events; they might also turn to cognitive scientists to validate the detailed cognitive mechanisms or processes implied by their inferences. Because such validating steps are expensive in time, labor, and expertise, validation should start with inferences that show the least uncertainty (e.g., an assembly process that shows up in 99 per-

cent of the DAGs sampled from the posterior distribution), although weaker inferences can give provocative hypotheses as well. In this way, large-scale computational studies of cultural evolution depend on and inform a wide range of rich disciplinary perspectives and methodologies. In other words, our approach both scaffolds and is scaffolded by a much larger research agenda. It provides a way to analyze unprecedented new data sets (Evans and Foster 2011) as important model organisms for the large-scale quantitative study of cultural evolution without embracing the limiting conceptual vocabulary of a single discipline (see the introduction by Love and Wimsatt).²²

TRIMS, TRAMS, AND THE MODE OF CULTURAL EVOLUTION

In this chapter, we described the foundations of an agnostic approach to reconstructing cultural lineages—one general enough to identify both patterns dominated by branching and patterns dominated by reticulation.²³ It is worth reflecting briefly on when and why we might expect to see these two archetypal modes of cultural inheritance. Approaches based on phylogenetic inference have leaned on the assumption that Transmission Isolating Mechanisms (TRIMs) like geopolitical boundaries, ethnocentrism, and language barriers limit the mixture and recombination of cultural components across lineages (Durham 1991; Mesoudi 2011). In the language of Wimsatt (2013a), these TRIMs mostly appeal to population structure—they are mechanisms that prevent culturally distinct populations from mixing and create distinct cultural breeding populations (see chapter 1). For example, language barriers could be viewed as institutionally induced cultural population structure. TRIMs make the pattern of cultural evolution branch-like, with a relatively slow pace—novelty is just harder to come by when new components and combinations must be produced within a cultural lineage. Hence, TRIMs create patterns of cultural evolution perfectly suited for detection by existing methods of inference that assume a single dominant inheritance pathway for each observed entity.

Although the precise TRIMs that are commonly invoked in cultural phylogenetics are much less common in the modern era of science and technology, their analogs nevertheless exist. For example, the citation of patents is slower, and radiates more slowly outward in space from the focal patent, than the citation of scientific articles (Adams, Clemmons, and Stephan 2006). Scientific communities can be largely cut off from one another by geopolitical

boundaries (as in the cladogenesis that resulted in a distinctive tradition of Soviet mathematics in the mid-twentieth century) or by jargon (Vilhena et al. 2014). And population structure, whether imposed by geography, disciplines, schools of thought, or status, can substantially slow the spread of new scientific or technical knowledge, especially when it is difficult to codify (Kaiser 2009). Whenever transmissible units depend on extensive previous training or time-consuming pedagogy for reliable transmission (Kaiser 2009), their spread across populations will be slower, and cultural evolution is more likely to manifest a branching mode on some levels of analysis (Boyd et al. 1997; Wimsatt 2013a). This should be true whether the transmissible unit is crafting a stone tool or crafting an elegant proof. Thus, organizationally enabled scaffolding, while facilitating cumulative cultural evolution within a particular lineage (e.g., a discipline), promotes the development of distinct cultural breeding populations. Careers can be strongly canalized within an existing cultural population, such as when departments only hire faculty with training in their specific discipline or with degrees from a select range of similar departments (Clauaset, Arbesman, and Larremore 2015). This canalization limits cumulative cultural evolution across lineages (see chapter 1).

Nevertheless, the system of modern science and technology also contains *Transmission Accelerating Mechanisms* or TRAMs, which increase the rate of horizontal transmission and recombination. These TRAMs range across the “relevant units of the cultural system” described by Wimsatt in this book. TRAMs most obviously include infrastructure such as modern transportation and communication technologies. They also include institutional conventions, like the increased dominance of English as scientific lingua franca; indeed, spoken language and writing have been powerful TRAMs and TRIMs at different scales throughout human history (see chapters 9 and 10). Classic Mertonian norms (like universalism) promote the free flow and exchange of ideas (Merton 1973), as do the explicit references, patent subclasses, and article key words associated with the publication process itself—all conventions that make information easier to find and retrieve. International conferences break down geographic population structure, while interdisciplinary meetings aim to break down the population structure created by discipline, training, and school of thought. Interdisciplinary hiring redirects careers across multiple cultural breeding populations, facilitating recombination across cultural lineages. Most intriguingly, technologies and some ideas can internalize their scaffolding so that they have easily discernible affordances. This process of black boxing allows the technology to move and

recombine more easily, as described by Bruno Latour (1987) and Michel Callon (1986). In a sense, these black-boxed artifacts actually scaffold their own recombination (Wimsatt 2013a), and we hypothesize that such auto-scaffolding is the crucial TRAM driving rapid, recombinant, and cumulative technological evolution.

Our models and exploratory methods are designed precisely to allow a system of artifacts, ideas, institutions, or individuals to reveal its dominant mode of cultural evolution, whether that be branching, reticulation, or some mixture of the two. In revealing the varying tempo and mode of cultural evolution across many contexts, these methods will help us understand in detail the competition between the TRIMs and the TRAMs that together pattern the evolution of technology, ideas, and human culture more broadly. We hope that our methods, and the underlying conceptual apparatus, can accelerate the move “beyond the meme” toward the integrated, interdisciplinary, multimethod study of cultural evolution.

MATHEMATICAL APPENDIX

Here, we provide concrete mathematical details and illustrations for the methods outlined above. This appendix is best read in parallel with the main text.

Parsimony

For each invention j , let $F'_j = F_{j/\text{novel}}$ be the set of all features in invention j , once any novel features have been removed. The set of all possible parents of j is denoted \mathcal{P}_j . For each potential parent $p \in \mathcal{P}_j$, we compute the intersection of the set of features in j that *could have been* inherited (F'_j) and the set of features in p (F_p); call this

$$W_{jp} = F'_j \cap F_p. \quad (3)$$

We select as the first or “prime” parent the prior invention k with the maximum W_{jk} . This is the prior invention that explains the most features. If there are multiple equally explanatory inventions, we can select the invention with the smallest Δt_{jk} (recency bias), the smallest Δd_{jk} (local bias), and so forth. If all parsimony principles are exhausted, choose at random.

Now define $F'_{j/k}$ as the set of all features in invention j that remain to be explained, given that k is one of the parents. We iterate the procedure above, defining

$$W_{jp|k} = F'_{j|k} \cap F_p \quad (4)$$

and selecting as the next parent the invention m with the maximum $W_{jml|k}$ (i.e., the one that explains the most features not explained by k). We repeat this procedure until *all* heritable features of j have been explained by one or more parents. We repeat this procedure for all j to reconstruct a parsimonious history for our observed technologies. Note that this history will be a directed graph; we establish the convention that arcs run *to* invention j from parents p, k, m , and so on, so that W_{jp} counts the number of components that flow from p to j , $W_{jml|p}$ counts the number of components that flow from m to j , and so on. For notational simplicity, we will refer to $W_{jml|p}$ as W_{jm} , $W_{jq|pm}$ as W_{jq} , and so on, unless the “conditioning” is important.

Black Boxing Measures

Define $W_j = |F_j|$ —that is, the number of features in the j -th invention. Now define

$$\widehat{W}_{jp} = \frac{W_{jp}}{W_j} \quad (5)$$

as the fraction of all components in j inherited from ancestor p . The primary inheritance is just the largest \widehat{W}_{jp} over all ancestors p . Call this \widehat{W}_{jp}^{max} . Properties of the frequency distribution of primary inheritance $\Pr(\widehat{W}_{jp}^{max})$ can provide suggestive evidence for branching or reticulate evolution.

With slight abuse of notation, we can define the primary contribution

$$\widetilde{W}_{jp} = \frac{W_{jp}}{W_p} \quad (6)$$

as the fraction of components in p passed on to its descendent j in a given reconstruction, where we look at cases in which p is the primary, secondary, tertiary ancestor, and so on. The mean $\widetilde{W}_p = \frac{1}{N_j} \sum_j \widetilde{W}_{jp}$ and mode over all descendants j give us an idea of whether the components of p are typically taken as a block or whether they are separable.

The key signature for black boxing is a modal \widetilde{W}_{jk} for a given k that is significantly higher than the typical mode of \widetilde{W}_{jp} over the population of p 's or, equivalently, a mean primary contribution \widetilde{W}_k that is significantly higher than the mean \widetilde{W}_p over the population of p 's. When the mean primary contribution \widetilde{W}_k is high but the mean fraction of components that k inherits

from its ancestors $\frac{\sum_{p \in P_k} \hat{W}_{kp}}{|P_k|}$ is low (or, equivalently, when the primary inheritance is low), it is likely that invention k has sampled from several sources and black boxed the parts.

Probabilistic Models of Multiple Inheritance

As input, we again have a set of types ordered in time. Each type j is characterized by a unique set of features F_j . We can equivalently represent this as a binary feature vector \mathbf{f}_j of length M , where types have M possible features. Thus, we observe a time-ordered collection of N types $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \dots, \mathbf{f}_N\}$.²⁴

Number of Parents

For a given observation \mathbf{f}_j , the generative process begins by choosing the number of parents n_j^P . The number of parents is drawn from a distribution controlled by one or more parameters θ_p . The simplest such choice would be a Poisson distribution, $\text{Poisson}(\alpha_p)$, with α_p controlling the average number of parents. This is similar to the so-called Indian buffet process, where customers stop after they have sampled $\text{Poisson}(\alpha)$ dishes (Griffiths and Ghahramani 2011).²⁵ This distribution, however, would assume that there is a typical number of parents and that the distribution of n_j^P is tightly peaked around that number. This could be relaxed.

Choosing Parents

There are many ways to formalize parent choice. For simplicity, we assume that the probability of assembling a particular collection of n parents factorizes

$$\Pr(P_j = \{a_1, a_2, a_3, \dots, a_n\} \mid n_j^P = n) \sim \prod_1^n \Pr(a_1 \in P_j) \dots \Pr(a_n \in P_j). \quad (7)$$

Number of Features

We assume that the number of features to be sampled from the parents is independent of the number or identity of the parents. As with the number of parents, the simplest choice for this distribution is $\text{Poisson}(\alpha_f)$, though this could be generalized to admit more complex distributions.

Increasing Complexity

Our generative model, as proposed, cannot easily deal with cultural evolution in which features accumulate. One way to deal with this is to make the

parameter controlling the number of features time dependent. For example, α_f grows with t at a rate β that is also subject to inference. A more interesting approach would allow the data to “suggest” relevant bundles of elemental features that should themselves be treated as features because of frequent copresence—a compression of the feature space. This could be done in a number of ways, as described in the main text.

Choosing Features

Once the number of features has been selected, we must choose specific features. Unlike parent selection, however, we allow the creative unit to introduce some number ($0 - m$) of new features $\sim \text{Poisson}(\alpha_{\text{novel}})$ unobserved in the parent set—and possibly never yet observed in the history of the system. This allows any observation \mathbf{f}_j to be generated from any parent set while also allowing true novelty through the creation of entirely new features.

Inference

Although somewhat nonobvious from the generative description, the model defined above is remarkably close to standard phylogenetic inference in structure. The parentage assignment \mathbf{P} is just a directed acyclic graph that respects the time ordering of $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \dots, \mathbf{f}_N\}$. For a given \mathbf{P} and values of the generative parameters—for example, α_p and $\alpha_F \in \Theta$ —we can calculate $\Pr(\mathbf{F} | \mathbf{P}, \Theta)$ quite directly. Then

$$\Pr(\mathbf{P}, \Theta | \mathbf{F}) = \frac{\Pr(\mathbf{F} | \mathbf{P}, \Theta) \Pr(\mathbf{P} | \Theta) \Pr(\Theta)}{\Pr(\mathbf{F})} \quad (8)$$

where the denominator cannot be calculated because of the required sum over all possible \mathbf{P} . Thus, we can approximate the posterior, as in phylogenetic inference, using standard methods such as Markov chain Monte Carlo (Gershman and Blei 2012).

NOTES

1. The frequency of horizontal gene transfer among prokaryotic taxa (Doolittle and Baptiste 2007) has created an urgent need for methods to study reticulation in biology, for example, Kunin et al. (2005). Although computational biologists have answered the call (Huson, Rupp, and Scornavacca 2010), these methods are either too generic (i.e., they are essentially

clustering) or involve too many specific biological processes (e.g., gene deletion or insertion) to provide a useful starting point. Until very recently, there were no Bayesian, generative model-based approaches to reticulation, though Wen, Yu, and Nakhleh (2016) may provide a way forward.

2. In some cases, horizontal transmission is reserved for trait flows within a generation (peer-to-peer), and oblique transmission is used when traits flow from nonparental individuals in an earlier generation to individuals in a later generation. We will not make this distinction here.

3. Defined by Wimsatt as “structure-like dynamical interactions with performing individuals that are means through which . . . competencies are constructed or acquired by individuals or organizations.”

4. Note the epidemiological language.

5. Note here the role of several factors explored at length in this volume as TRIMs; e.g., language (see chapter 9) and identity (see chapter 12).

6. As we will argue later, the methods we propose extend unproblematically to some other cultural items and could even be used to model sequential skill acquisition by developing biological individuals (see the introduction by Love and Wimsatt; see also chapter 1). For concreteness, we focus the discussion on science and technology, but the reader should keep implicit generalizations in mind throughout.

7. For many problems, a tree-based simplification could be illuminating as an initial analysis of data.

8. We assume here that the features are already given, as in patent classes, PACS (physics and astronomy classification scheme) codes, or MeSH (medical subject heading) terms. In cases in which features must be constructed by the analyst from scratch, one can draw upon a well-developed literature in feature engineering.

9. The cleaned, curated features given by patent classes, PACS codes, and MeSH terms are useful insofar as they approximate, in some fashion, the principles of vision and division that characterize the relevant communities of practice. We leave aside the very interesting question of how different communities of practice might break up the same invention into different elementary building blocks; this would require detailed thinking about the sequentially dependent and organizationally scaffolded skill acquisition (Goodwin 2017) that would yield different ways of seeing the same invention (Love and Wimsatt, the introduction to this book; Wimsatt, chapter 1), that is, different modes of “professional vision” (Goodwin 1994). Data science techniques for feature engineering may be useful for heuristic feature

construction where expert taxonomies are incomplete or nonexistent (Scott and Matwin 1999; Anderson et al. 2013).

10. While this may seem like a strong assumption, note that it has a certain plausibility in terms of search. If there are many possible histories that enrich the “primary” parent with residual features (the *three-parent* history) but only one history that pairs the right two inventions (the *two-parent* history), then we are more likely to observe someone start from the primary parent and then enrich than to observe an inventor who lands on exactly the right pair of parents.

11. Of course, this introduces a bias into our reconstruction, but absent strong evidence to the contrary, we think that the greedy assumption tends to capture more probable pathways. It is also consistent with approaches to human cognition, like case-based reasoning (Aamodt and Plaza 1994).

12. If we wish to weaken this assumption privileging significant inheritance from a single parent p , we can search over the space of pairs, triples, tetrads, etc., for the combination that contributes the most features. The computational cost for this exploration is high, however. Rather than searching through n possible parents for the single most explanatory parent (so the overall search is $\mathcal{O}(n)$), we would have to search through $\binom{n}{2}$ pairs, $\binom{n}{3}$ triples, $\binom{n}{4}$ tetrads, etc. The computational cost grows exponentially: $\mathcal{O}(n^2)$, $\mathcal{O}(n^3)$, $\mathcal{O}(n^4)$, etc.

13. It will be acyclic—i.e., have no loops $i \rightarrow j \rightarrow k \rightarrow i$ —because the future cannot influence the past by construction.

14. This method of constructing a parsimonious evolutionary explanation is not assured to recover the actual inheritance pattern of cultural traits. Moreover, the adaptive, evolutionary significance of inheriting a particular feature may only be minimally associated with the primary inheritance on which parsimony focuses.

15. Indeed, Wimsatt notes that “black boxing is a crucial feature of most complex sequential skill acquisition” in his contribution to this book.

16. Note that, on this account invention, k (with low primary inheritance and high primary contribution) creates the black box, which persists as a packet into the next generation. In principle, persistence across multiple generations would provide stronger evidence for true black boxing.

17. The need to characterize types with discrete features or “building blocks” is an obvious limitation and a potential source of bias. These methods work best for entities that have already been characterized with features. As discussed previously, it is certainly possible to induce features when they

are not already available, but we must be especially cautious about inferences from these induced features. Independent validation of the features is a necessity. And even when features have been developed for other reasons (e.g., search or classification), these may not always correspond to the features that are relevant to the inventive process, introducing bias.

18. This will need to be relaxed in section 7 to model the sequential skill acquisition common in developing biological individuals or organizations.

19. As new, complex features are discovered—by inventors through black boxing and by analysts through feature reduction—even *old* artifacts could “acquire” new, heritable features as bundles of components are reinterpreted as coherent units. Note that the routine combination of components could take place immediately following their initial combination, could increase gradually, or could follow a discontinuous trajectory as an old combination becomes *fit* to a new environment. For example, consider the explosive rise in the use of Bayesian methods following the advent of computers, which scaffolded and explicitly catalyzed their application (see chapter 11).

20. This simplifying assumption runs roughshod over the sequential dependence of features—or even their functional interdependence.

21. The frequency of entirely new features will scale inversely with the resolution of existing features. For example, a new feature in a coarse-grained scheme might include a custom-built molecule, but this would simply be a new combination of *existing* features if atoms or molecular motifs were components (Arthur 2009).

22. Here, we are thinking especially of patents (see chapter 6).

23. While our approach allows panmixia—i.e., the selection of arbitrary parents to produce offspring (see chapter 1)—it also allows distinct cultural breeding populations to emerge from the data. It also allows researchers to encode distinct hypotheses about factors like geographic, social, or cultural distance that structure cultural breeding populations and violate panmictic assumptions.

24. We largely follow the notation and presentation of Gershman and Blei (2012) here.

25. If the feature set describing all parents is finite, we could model this by a simple Beta-Bernoulli process. There are many ways to set up a conceptually similar generative model; the important ingredients are (1) a process controlling the number of parents, (2) a process selecting the parents, and (3) a process choosing features from the set of parents.

REFERENCES

- Aamodt, A., and E. Plaza. 1994. "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches." *AI Communications* 7 (1): 39–59.
- Adams, J. D. 2002. "Comparative Localization of Academic and Industrial Spillovers." *Journal of Economic Geography* 2 (3): 253–78.
- Adams, J. D., J. R. Clemmons, and P. E. Stephan. 2006. *How Rapidly Does Science Leak Out?* Technical report. National Bureau of Economic Research.
- Anderson, M. R., D. Antenucci, V. Bittorf, M. Burgess, M. J. Cafarella, A. Kumar, F. Niu, Y. Park, C. Ré, and C. Zhang. 2013. "Brainwash: A Data System for Feature Engineering." In *Proceedings of Conference on Innovative Data Systems Research (CIDR)*.
- Arthur, W. B. 2009. *The Nature of Technology: What It Is and How It Evolves*. New York: Simon and Schuster.
- Bengio, Y., A. Courville, and P. Vincent. 2013. "Representation Learning: A Review and New Perspectives." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (8): 1798–828.
- Bergstrom, C. T., and L. A. Dugatkin. 2012. *Evolution*. New York: W. W. Norton.
- Birch, A. 1967. *The Economic History of the British Iron and Steel Industry, 1784–1879: Essays in Industrial and Economic History with Special Reference to the Development of Technology*. London: Frank Cass.
- Blei, D. M. 2014. "Build, Compute, Critique, Repeat: Data Analysis with Latent Variable Models." *Annual Review of Statistics and Its Application* 1, 203–32.
- Bourdieu, P. 1990. *The Logic of Practice*. Palo Alto, Calif.: Stanford University Press.
- Boyd, R., M. Borgerhoff-Mulder, W. H. Durham, and P. J. Richerson. 1997. "Are Cultural Phylogenies Possible?" In *Human by Nature: Between Biology and the Social Sciences*, edited by P. Weingart et al., 355–84. Mahwah: Lawrence Erlbaum.
- Boyd, R., and P. J. Richerson. 1988. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.
- Bruch, E., F. Feinberg, and K. Y. Lee. 2016. "Extracting Multistage Screening Rules from Online Dating Activity Data." *Proceedings of the National Academy of Sciences* 113 (38): 10530–35.
- Callon, M., 1986. "The Sociology of an Actor-Network: The Case of the Electric Vehicle." In *Mapping the Dynamics of Science and Technology*, edited by Michel Callon, John Law, and Arie Rip, 19–34. London: Palgrave Macmillan.

- Claidière, N., T. C. Scott-Phillips, and D. Sperber. 2014. "How Darwinian Is Cultural Evolution?" *Philosophical Transactions of the Royal Society B* 369 (1642): 20130368.
- Clauset, A., S. Arbesman, and D. B. Larremore. 2015. "Systematic Inequality and Hierarchy in Faculty Hiring Networks." *Science Advances* 1 (1): e1400005.
- Cochrane, E. E., and C. P. Lipo. 2010. "Phylogenetic Analyses of Lapita Decoration Do Not Support Branching Evolution or Regional Population Structure during Colonization of Remote Oceania." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1559): 3889–902.
- Collard, M., S. J. Shennan, and J. J. Tehrani. 2006. "Branching, Blending, and the Evolution of Cultural Similarities and Differences among Human Populations." *Evolution and Human Behavior* 27 (3): 169–84.
- Doolittle, W. F., and E. Bapteste. 2007. "Pattern Pluralism and the Tree of Life Hypothesis." *Proceedings of the National Academy of Sciences* 104 (7): 2043–49.
- Durham, W. H. 1991. *Coevolution: Genes, Culture, and Human Diversity*. Palo Alto, Calif.: Stanford University Press.
- Evans, J. A., and J. G. Foster. 2011. "Metaknowledge." *Science* 331 (6018): 721–25.
- Felsenstein, J. 2004. *Inferring Phylogenies*. Sunderland, Mass.: Sinauer Associates.
- Fisher, D. C. 2008. "Stratocladistics: Integrating Temporal Data and Character Data in Phylogenetic Inference." *Annual Review of Ecology, Evolution, and Systematics* 39:365–85.
- Fleming, L., and O. Sorenson. 2001. "Technology as a Complex Adaptive System: Evidence from Patent Data." *Research Policy* 30 (7): 1019–39.
- Fleming, L., and O. Sorenson. 2004. "Science as a Map in Technological Search." *Strategic Management Journal* 25 (8–9): 909–28.
- Foster, J. G. 2018. "Culture and Computation: Steps to a Probably Approximately Correct Theory of Culture." *Poetics* 68: 144–54.
- Foster, J. G., A. Rzhetsky, and J. A. Evans. 2015. "Tradition and Innovation in Scientists' Research Strategies." *American Sociological Review* 80 (5): 875–908.
- Funk, R. J., and J. Owen-Smith. 2016. "A Dynamic Network Measure of Technological Change." *Management Science* 63 (3): 791–817.
- Gershman, S. J., and D. M. Blei. 2012. "A Tutorial on Bayesian Nonparametric Models." *Journal of Mathematical Psychology* 56 (1): 1–12.
- Ghahramani, Z. 2013. "Bayesian Non-parametrics and the Probabilistic Approach to Modelling." *Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences* 371 (1984): 20110553.
- Ghahramani, Z. 2015. "Probabilistic Machine Learning and Artificial Intelligence." *Nature* 521 (7553): 452–59.

- Gogarten, J. P., and J. P. Townsend. 2005. "Horizontal Gene Transfer, Genome Innovation, and Evolution." *Nature Reviews Microbiology* 3 (9): 679–87.
- Goodwin, C. 1994. "Professional Vision." *American Anthropologist* 96 (3): 606–33.
- Goodwin, C. 2017. *Co-operative Action*. Chicago: Cambridge University Press.
- Gould, S. J. 2010. *An Urchin in the Storm: Essays about Books and Ideas*. New York: W. W. Norton.
- Gray, R. D., and Q. D. Atkinson. 2003. "Language-Tree Divergence Times Support the Anatolian Theory of Indo-European Origin." *Nature* 426 (6965): 435–39.
- Gray, R. D., D. Bryant, and S. J. Greenhill. 2010. "On the Shape and Fabric of Human History." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1559): 3923–33.
- Gray, R. D., A. J. Drummond, and S. J. Greenhill. 2009. "Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement." *Science* 323 (5913): 479–83.
- Greenhill, S. J., T. E. Currie, and R. D. Gray. 2009. "Does Horizontal Transmission Invalidate Cultural Phylogenies?" *Proceedings of the Royal Society B: Biological Sciences* 276 (1665): 2299–306.
- Griffiths, T. L., and Z. Ghahramani. 2011. "The Indian Buffet Process: An Introduction and Review." *Journal of Machine Learning Research* 12:1185–224.
- Huelsenbeck, J. P., and B. Rannala. 1997. "Maximum Likelihood Estimation of Phylogeny Using Stratigraphic Data." *Paleobiology* 23 (2): 174–80.
- Huson, D. H., R. Rupp, and C. Scornavacca. 2010. *Phylogenetic Networks*. Cambridge: Cambridge University Press.
- Huson, D. H., and C. Scornavacca. 2011. "A Survey of Combinatorial Methods for Phylogenetic Networks." *Genome Biology and Evolution* 3:23–35.
- Kaiser, D. 2009. *Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics*. Chicago: University of Chicago Press.
- Kunin, V., L. Goldovsky, N. Darzentas, and C. A. Ouzounis. 2005. "The Net of Life: Reconstructing the Microbial Phylogenetic Network." *Genome Research* 15 (7): 954–59.
- Latour, B. 1987. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, Mass.: Harvard University Press.
- Lipo, C. P. 2006. "The Resolution of Cultural Phylogenies Using Graphs." In *Mapping Human History: Phylogenetic Approaches in Anthropology and Prehistory*, edited by C.P. Lipo et al., 89–107. New York: Routledge.
- Merton, R. K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.

- Mesoudi, A. 2011. *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*. Chicago: University of Chicago Press.
- Miller, G. A. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* 63 (2): 81.
- Newman, M. E. 2003. "The Structure and Function of Complex Networks." *SIAM Review* 45 (2): 167–256.
- Nicholls, G. K., and R. D. Gray. 2006. "Quantifying Uncertainty in a Stochastic Model of Vocabulary Evolution." In *Phylogenetic Methods and the Prehistory of Languages*, edited by Peter Forster and Colin Renfrew, 161–71. Cambridge: McDonald Institute for Archaeological Research.
- O'Brien, M. J., M. Collard, B. Buchanan, and M. T. Boulanger. 2013. "'Trees, Thickets, or Something in between?' Recent Theoretical and Empirical Work in Cultural Phylogeny." *Israel Journal of Ecology and Evolution* 59 (2): 45–61.
- O'Brien, M. J., J. Darwent, and R. L. Lyman. 2001. "Cladistics Is Useful for Reconstructing Archaeological Phylogenies: Palaeoindian Points from the Southeastern United States." *Journal of Archaeological Science* 28 (10): 1115–36.
- Owen-Smith, J., and W. W. Powell. 2001. "To Patent or Not: Faculty Decisions and Institutional Success at Technology Transfer." *Journal of Technology Transfer* 26 (1–2): 99–114.
- Scott, S., and S. Matwin. 1999. "Feature Engineering for Text Classification." In *International Conference on Machine Learning* 99: 379–88.
- Shalizi, C. R., and A. C. Thomas. 2011. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies." *Sociological Methods and Research* 40 (2): 211–39.
- Silvestro, D., J. Schnitzler, L. H. Liow, A. Antonelli, and N. Salamin. 2014. "Bayesian Estimation of Speciation and Extinction from Incomplete Fossil Occurrence Data." *Systematic Biology* 63 (3): 349–67.
- Simon, H. A. 1969. *The Sciences of the Artificial*. Cambridge, Mass.: MIT Press.
- Swait, J., and M. Ben-Akiva. 1987. "Incorporating Random Constraints in Discrete Models of Choice Set Generation." *Transportation Research Part B: Methodological* 21 (2): 91–102.
- Tehrani, J., and M. Collard. 2002. "Investigating Cultural Evolution through Biological Phylogenetic Analyses of Turkmen Textiles." *Journal of Anthropological Archaeology* 21 (4): 443–63.
- Tëmkin, I., and N. Eldredge. 2007. "Phylogenetics and Material Cultural Evolution." *Current Anthropology* 48 (1): 146–54.

- Uzzi, B., S. Mukherjee, M. Stringer, and B. Jones. 2013. "Atypical Combinations and Scientific Impact." *Science* 342 (6157): 468–72.
- Vilhena, D. A., J. G. Foster, M. Rosvall, J. D. West, J. Evans, and C. T. Bergstrom. 2014. "Finding Cultural Holes: How Structure and Culture Diverge in Networks of Scholarly Communication." *Sociological Science* 1:221–38.
- Wade, M. J. 2016. *Adaptation in Metapopulations: How Interaction Changes Evolution*. Chicago: University of Chicago Press.
- Wen, D., Y. Yu, and L. Nakhleh. 2016. "Bayesian Inference of Reticulate Phylogenies under the Multispecies Network Coalescent." *PLoS Genetics* 12 (5): e1006006.
- Wimsatt, W. C. 2002. "Using False Models to Elaborate Constraints on Processes: Blending Inheritance in Organic and Cultural Evolution." *Philosophy of Science* 69 (S3): S12–S24.
- Wimsatt, W. C. 2013a. "Articulating Babel: An Approach to Cultural Evolution." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (4): 563–71.
- Wimsatt, W. C. 2013b. "Entrenchment and Scaffolding: An Architecture for a Theory of Cultural Change." In *Developing Scaffolds in Evolution, Culture, and Cognition*, edited by L. R. Caporael, J. R. Griesemer, and W.C. Wimsatt, 77–107. Cambridge: MIT Press.