Archaeol Anthropol Sci
DOI 10.1007/s12520-015-0257-0

ORIGINAL PAPER

Validating niche-construction theory through path analysis

William A. Brock[1,2] · **Michael J. O'Brien**[3] · **R. Alexander Bentley**[4]

Received: 11 August 2014 / Accepted: 24 May 2015

Abstract Under the conventional view of evolution, species over time come to exhibit those characteristics that best enable them to survive and reproduce in their preexisting environments. Niche construction provides a second evolutionary route to establishing the adaptive fit, or match, between organism and environment, viewing such matches as dynamical products of a two-way process involving organisms both responding to problems posed by environments as well as setting themselves new problems by changing their environments through further niche construction. Not surprising, the analysis of niche construction is complicated. For example, variables of interest might contain measurement error, or some variables might not be observable. In other cases, variables might not be datable and have to be measured at the same date. A time-series generalization of path analysis, which itself can be viewed as a version of simultaneous-equation analysis, offers a means of highlighting causal relationships in complex systems of niche construction by graphically representing a hypothesis of causality between variables and, in some instances, providing an estimated weight that a hypothesized causal variable has on another variable. Path analysis forces researchers to specify how variables relate to one another and encourages development of clear and logical theories concerning the processes that influence a particular outcome. As we show through a case study—the coevolution of cattle husbandry and the tolerance for milk consumption—path analysis can also call attention to potential areas of weakness and ambiguity in data sets and how they are used in constructing inferences.

Keywords Dairying · Granger causality · Lactase persistence · Neolithic · Niche-construction theory · Path analysis

Introduction

Under the conventional view of evolution, with its emphasis on natural selection as a causal process, species over time come to exhibit those characteristics that best enable them to survive and reproduce in their preexisting environments (Godfrey-Smith 1996). Although environmental change may trigger bouts of selection, standard evolutionary theory (SET) holds that changes in organisms, rather than changes in environments, generate the organism–environment match commonly described as "adaptation." In SET, this is a one-way process.

In contrast, niche-construction theory (NCT) provides a second evolutionary route to establishing the adaptive fit, or match, between organism and environment. In NCT, such matches are treated as a two-way process involving not only organisms responding to "problems" posed by their environments (SET) but also solving some of those problems, and creating new problems, by changing their environments through further niche construction (Lewontin 1983, 2000; Odling-Smee 1988; Odling-Smee and Turner 2011). NCT emphasizes the capacity of organisms, through their metabolism, activities, and choices, to modify not only their own niche but also those of others (Erwin 2008; Levins and Lewontin 1985; Lewontin 1983; Odling-Smee et al. 2003).

✉ Michael J. O'Brien
obrienm@missouri.edu

[1] Department of Economics, University of Missouri,
Columbia, MO 65211, USA

[2] Department of Economics, University of Wisconsin,
Madison, WI 53706, USA

[3] Department of Anthropology, University of Missouri,
Columbia, MO 65211, USA

[4] Department of Comparative Cultural Studies, University of Houston,
Houston, TX 77204, USA

In NCT, organisms co-direct their own evolution, often, but not exclusively, in a manner that suits their genotypes. In the process, organisms modify patterns of selection acting back on themselves as well as on other species that inhabit their environment. Because modification of the environment continues over generations, the process is referred to as "ecological inheritance" (Erwin 2008; Odling-Smee 1988, 2010; Odling-Smee and Laland 2011; Odling-Smee and Turner 2011), although the by-products of niche-construction activities may confer no selective advantage to the constructor itself (Jones et al. 1997; Krakauer et al. 2009; Wright 2006). Multiple case studies, among human and animal communities, support the multiple-inheritance processes of NCT (e.g., Furrow et al. 2011).

Numerous mathematical models have been developed to explore the evolutionary ramifications of niche construction (e.g., Boni and Feldman 2005; Borenstein et al. 2006; Donohue 2005; Ihara and Feldman 2004; Laland et al. 1996, 1999, 2001, 2009; Lehmann 2008; Odling-Smee et al. 2003; Oyama et al. 2001; Schwilk and Ackerly 2001; Silver and Di Paolo 2006). All have concluded that niche construction is evolutionarily consequential and that ecological inheritance generates unusual evolutionary dynamics (Laland et al. 1996, 1999, 2000, 2001, 2009; Silver and Di Paolo 2006). Populations evolving in response to features of the environment modified by their ancestors exhibit lag effects, such as continuing to evolve in the same direction after selection has stopped or reversed or had a delayed evolutionary response to selection for a number of generations (Laland and Brown 2006). With respect to humans, mathematical models suggest that niche construction resulting from cultural processes is likely to be even *more* potent than gene-based niche construction (Laland et al. 2007; Odling-Smee et al. 2003; Smith 2007a). Furthermore, such models establish that cultural niche construction can plausibly modify selection on human genes and actually *drive* evolutionary events (Feldman and Laland 1996; Gerbault et al. 2011; Laland et al. 2001; Rendell et al. 2011).

These claims have created a lively debate between those possessing SET and NCT perspectives (e.g., Brodie 2005; Dawkins 2004; Scott-Phillips et al. 2014).[1] With SET as the long-established approach, the challenge for NCT advocates is to test their hypotheses robustly through empirical evidence (Odling-Smee et al. 2003; Scott-Phillips et al. 2014). One method of testing NCT has been to detect statistically how phenotypic characters (recipient traits) evolve as a consequence of feedback from prior niche-constructing traits (Harvey and Pagel 1991). Such analysis, for example, yielded only weak support for hypothesized coevolution of dog use and successful specialized reindeer economies in Late Paleolithic (ca. 14.7–11.5 thousand years ago) Scandinavia (Riede 2011).

Because NCT involves complex cause-and-effect relationships, fitting NCT predictions to real data can be challenging, especially concerning human niche construction, which often involves genetic and archaeological evidence that vary in time intervals, strength of selection, and regional variation. It is one thing to posit NCT as a driver of evolution; it is another to prove it, especially in situations where one is confronted with a large number of confounding factors such as missing data and measurement error—not uncommon occurrences in archeology. Here, we propose an explicit path-analysis approach as a way forward. Path analysis has a long history in the life sciences (Wright 1921, 1934) as well as in the social and behavioral sciences (e.g., Dunbar and Shultz 2007; Pearl 1995, 2009; Shipley 2000; Wolfle 2003).

In brief, path analysis offers a way of graphically representing a hypothesis of causality (sometimes weighted) between variables. By forcing researchers to specify how variables relate to one another, path analysis clarifies the hypothesized processes that influence a particular outcome (Lleros 2005). In simplified terms, path analysis effectively uses multiple-regression analysis to estimate the strength of effects in a hypothetical causal system (Lleros 2005; Scheiner et al. 2000). In one approach, the hypothesized causal scheme is used to make predictions about patterns of evolution, either from first principles or from previous experimental and observational studies (Scheiner et al. 2000).

Let us start with SET, which is based on a straight-line path:

$$\text{Selection}\rightarrow\text{Organisms}_t\rightarrow\text{Adaption}\rightarrow\text{Organisms}_{t+1}$$

Here, environmental change works on a population of organisms at time t, leading to differential reproduction and the appearance of an adaptation in the subsequent population at time $t+1$. Under SET, any path dependence is independent of the activities of the focal organisms. As originally formulated in the cross-sectional literature (e.g., Wright 1934), path diagrams were recursive, meaning that cause moves in one direction only, so that if X_1 affects X_2, then X_2 does not affect X_1 either directly or indirectly.

When we consider a path under NCT, however, the challenge of testing it (versus SET as the null hypothesis) becomes apparent. NCT can imply not only multiple possible pathways but also pathways whose dynamics are interdependent and therefore prone to feedback in the long-term dynamics. Thus, selective environments encountered by descendent

---

[1] To clear up possible confusion, we need to make clear that the article by Scott-Phillips et al. (2014), "The niche construction perspective: a critical appraisal," started out with four authors—Thomas Scott-Phillips, David Shuker, Tom Dickins, and Stuart West—all of whom subscribe to the position that NCT is little more than a proximate mechanism that is already accounted for by SET. They then invited Kevin Laland, a prominent NCT advocate, to join them as an author and to respond to their claims. The contexts in which we cite the article should make it clear which side we are talking about.

organisms at time $t+1$ stem not only from independent sources of natural selection pressures—climate, for example—but also from sources of natural selection that have previously been modified by ancestral niche construction. The two-way nature of NCT pathways, often featuring multiple dependent variables, makes them well-suited to modern path-diagram analysis, which deals with causality in both directions (Hauser and Goldberger 1971; Heckman and Pinto 2014).

Given that the debate between advocates of NCT and SET lies in the time-series domain, we focus here on path analysis and Granger (1969) causality testing within it. We hope to move the debate into a statistical-testing domain by introducing concepts from time-series causality analysis, such as exogeneity testing (Engle et al. 1983; Ericsson 1992; Geweke 1984) and, more recently, time-series directional-causality testing and path-diagram analysis (Eichler 2007). These concepts allow us to discuss not only what is needed to test NCT versus SET but also to discuss who should bear the burden of proof in a precise manner: What is the null hypothesis being tested? What should be the choice of significance level? What should be the alternative hypothesis? Our own inclination is that, while SET may serve as the null hypothesis, more latitude (e.g., higher p values before rejection) should be given for exploring NCT as an alternative hypothesis in order to harness the creativity of the hypotheses it generates (Laland and Sterelny 2006). As we intend to show, this is especially important because NCT places a high burden of proof on archaeological and genetic evidence and should not be rejected before we obtain the resolution of evidence to test it.

To exemplify how this approach sharpens the SET–NCT debate, we discuss the evolution of lactase persistence among humans, which has become “a flagship illustration of niche construction” (Scott-Phillips et al. 2014, p. 3). We hope to use the discussion as a prelude to identifying not only lines of further data acquisition and analysis but also the precise kinds of data that will be needed, including proxy data in instances where direct measures are impossible to attain.

NCT case study: the coevolution of cattle husbandry and lactase persistence

Despite the nutritional value of milk (Wooding 2007) in its fat, proteins, carbohydrates, vitamins, and calcium, about 65 % of the world’s population does not have the lactase persistence allele (*LP*) and cannot digest milk because of an inability to break down the disaccharide lactose into its two monosaccharides (Ingram et al. 2009; Itan et al. 2010; McCracken 1971; Simoons 1970). LP populations lie mainly in Western Africa, the Middle East, North America, and northern Europe, where LP rates can exceed 95 % (Ingram et al. 2009; Itan et al. 2009, 2010). Most LP populations have a prehistory of dairying activity (Gerbault et al. 2011), and there is strong phylogenetic evidence that adult lactose tolerance must have arisen independently at least three times in cultures that already kept animals for milk but not in nondairying cultures (Holden and Mace 1997; Mace 2009).

Despite multiple origins (Mulcare et al. 2004; Nagy et al. 2011; Schlebusch et al. 2013; Tishkoff et al. 2006), the pattern in Europe, which is tied to a single nucleotide polymorphism—*13910*T*, located 13.9 kilobases upstream of the lactase gene (Enattah et al. 2002)—appears to explain much of the modern distribution of LP (Gerbault et al. 2013; Itan et al. 2010). Estimated dates of origin of the *LP* allele range from about 2000 years ago to just over 20,000 years ago (Bersaglieri et al. 2004; Coelho et al. 2005; Mulcare 2006). The allele has not been found (so far) in ancient DNA extracted from early Neolithic Europeans of 7000–8000 years ago (Burger et al. 2007). Farming diffused from Anatolia to Britain over these same millennia, so it appears that dairying spread with the earliest Neolithic populations of Europe (Gerbault et al. 2011), although there was by no means equal coverage by such populations (Gamba et al. 2014). The *LP* allele had risen to over 70 % in west-central Europe—its more-or-less modern percentage—by around 1200 AD (Krüttli et al. 2014).

Overall, the combined genetic and archaeological evidence (discussed later) indicates that the natural-selection coefficient for LP in European populations is among the strongest of any gene in the genome, between about 0.1 and 0.2 (Bersaglieri et al. 2004; Tishkoff et al. 2006). Although “no definitive conclusions have been drawn about why LP was so advantageous” (Gerbault et al. 2013, p. 988), it seems reasonable to assume that milk provided nutrients more reliably than seasonal crops (Shennan et al. 2013) and that strong selection for the *LP* allele would have necessitated milk production being a reliably learned tradition, such that the offspring of milk drinkers were themselves milk drinkers with high probability (Aoki 1996; Feldman and Cavalli-Sforzi 1989). Both aspects are likely if there existed specialist traditions of producing cheese—manufactured in central Europe beginning 6800–7200 years ago (Bogucki 1986; Salque et al. 2013)—as it is more digestible, less perishable, and more transportable than raw milk (Lomer et al. 2008). Probably “the manufacture of cheese was an essential step in the exploitation of milk by populations who possibly had a high level of lactose intolerance” (Bogucki 1984).

Other cultural dynamics would have enhanced the niche-construction effects of dairying: domesticated cattle provided not only milk, meat, and labor but also sociocultural advantages such as status and wealth, which correlate positively with human reproductive success (Bogucki 1993; Conolly et al. 2012; Mace 1998). Demographic processes in early Neolithic Europe were also a factor, especially if population sizes experienced bottlenecks through substantial cycles of rise and decline (Shennan et al. 2013). Consequently, selection on LP should have varied across time and space (Gerbault et al. 2011).

Understanding LP causality through path analysis

Given that the *LP* allele spread only after domesticated dairy animals were introduced (Burger et al. 2007), both SET and NCT advocates agree that prehistoric dairying effected a selective pressure to which human genomes responded through lactase persistence (Burger et al. 2007; Curry 2013; Flatz 1987; Gerbault et al. 2011, 2013; Harvey et al. 1998; Hollox et al. 2001; Leonardi et al. 2012; Myles et al. 2005; Ulijaszek and Strickland 1993). The difference between SET and NCT lies in the pathways they invoke in explaining this evolution. In SET, "the advent of dairy farming… meant that variation in the gene(s) for lactose tolerance… correlated with variation in fitness, which was acted upon by natural selection" (Scott-Phillips et al. 2014, p. 3; see also Gerbault et al. 2011). In other words, SET sees dairying, like climate, as just another aspect of the exogenous environment that contributes to selection pressure. In NCT, the fact that dairy farming preceded genetic change is more significant in that dairy farming is what *caused* the evolutionary event of LP. Unlike environmental change, niche construction, as "the expression of genetic and acquired (semantic) information," *drives* the ecological change, often over multiple generations (Scott-Phillips et al. 2014, p. 5).

Path analysis makes the difference between SET and NCT accounts more explicit. The path diagram for NCT, shown in Fig. 1, illustrates a chain of proposed causal influences in the evolution of LP, such that the domestication of cattle triggers (1) milk consumption, which (2) favors the spread of LP, (3) which promotes further milk consumption, which (4) elicits further milk-product manufacture and consumption, which (5) leads to selective breeding of cattle, which (6) selects for alleles conferring high milk yield in dairy cattle. In addition, cattle farming and dairy-product consumption (7) lead to population growth, which (8) triggers dispersal into new environments.

In the LP example, matters are complicated by using proxies for actual variables, such as cattle raising inferred from archaeological faunal remains (with specially bred dairy cattle inferred from the age–sex structure, which itself is inferred), milk consumption inferred from genetic markers and milk residues on pottery (supposed nonceramic containers such as gourds are used), and population growth inferred from the size and number of Neolithic communities through time and/or the fraction of adolescents among human skeletons recovered. The chronology of all these changes is based on radiocarbon dating, which has a substantial uncertainty during the Neolithic period because of "flat" sections of the radiocarbon calibration curve.

At any point in the analysis, we might be confronted with problems that seem intractable, such as large measurement error, variables that are not separable, or variables with uncertain dates. We maintain, however, that path analysis offers a way forward and here turn to simple methods—Granger causality testing and linear vector autoregressions—that can be used with limited data to test for causality as well as to estimate the strength of a path within a causal-path diagram.

Granger causality

Because niche construction involves bi-directionality, we use Eichler's (2007) notion of path diagrams that was developed specifically for nonrecursive, time-series analysis (Granger 1969). We refer to this as *Granger causality.* This general linear modeling approach involves a variety of structural-equation modeling methods (Geweke 1984; Maruyama 1998). We stress here that Granger "causality" is not "true causality" in the sense of Heckman and Pinto (2014). Rather, it is only a statement about incremental prediction. We say that time series X "linear Granger causes" time series Y if past values of X and Y incrementally predict future values of Y better than past values of Y alone.

Although the coupled dynamics of NCT are nonlinear (Han and Hui 2014; Laland et al. 1999; Leonardi et al. 2012), we nonetheless start by imagining linear stochastic dynamics. Linear methods enable us to use some of the simplest concepts from time-series analysis to begin to scratch the surface of this difficult identification-and-quantitative-estimation problem in settings where data availability is lean and data quality is low. We can consider Granger causality testing as applied to the niche-construction context in a minimalist dynamic linear system of regression equations (nonparametric and nonlinear versions can be found in Diks and Panchenko (2006) and Bahadori and Liu (2013). We restrict ourselves to linear vector autoregressive systems as a workable approximation (at the initial stage where the frequency of the LP_t allele is close to zero), but we point out that a variable such as allele frequency, LP_t, lies in the interval [0,1], so clearly, we must use a general nonlinear Granger causality approach that respects this basic nonlinearity[2] in a more realistic approach.

---

[2] In the nonlinear case, an example of Granger causality testing in the context of a nonlinear analog of the bi-variate linear example (1b) is to test for the cross partial derivative, $\partial f_2(M_t,LP_t,e_{2,t+1})/\partial LP_t=0$, in the pair of nonlinear Eq. (1a),

$$M_{t+1}=f_1\left(M_t,LP_t,e_{1,t+1}\right),LP_{t+1}=f_2\left(M_t,LP_t,e_{2,t+1}\right) \quad (1a)$$

Testing $\partial f_2(M_t,LP_t,e_{2,t+1})/\partial LP_t=0$, instead of $\alpha_{21}=0$, in the linear Eq. (1b) can be done using the methods of Diks and Panchenko (2006) and Bahadori and Liu (2013). Lee et al. (1993) discuss testing for the presence of neglected nonlinearity. Please see the appendix for more discussion of nonlinear dynamical systems and the NCT–SET issue.

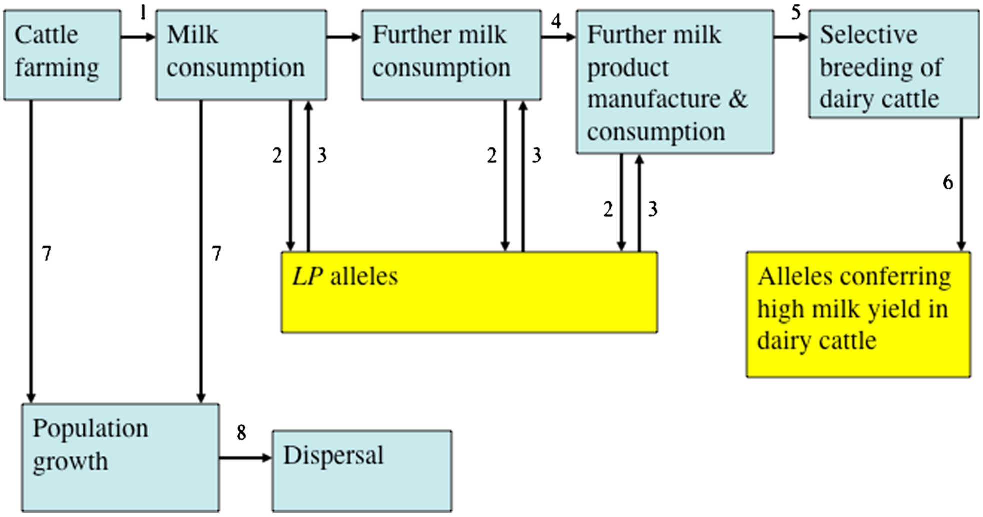

Fig. 1 Path diagram depicting the chain of causal influences following a cultural niche-constructing practice for dairy farming (from O'Brien and Laland 2012). Cultural processes are shown in *light-blue boxes*, and genetic change is shown in *yellow boxes*. The domestication of cattle triggers (*1*) milk consumption, which (*2*) favors the spread of lactase persistence (LP), (*3*) promoting further milk consumption, which (*4*) elicits further milk-product manufacture and consumption, which (*5*) leads to selective breeding of cattle, which (*6*) selects for alleles conferring high milk yield in dairy cattle. In addition, cattle farming and dairy-product consumption (*7*) lead to population growth, which (*8*) triggers dispersal into new environments

Initially, we thought that, for simplicity of an initial illustration, we could restrict ourselves to the bivariate linear autoregressive system (1) below,

$$\begin{aligned} M_{t+1} &= \alpha_{10} + \alpha_{11}M_t + \alpha_{12}LP_t + e_{1,t+1}, t = 1, 2\ldots \\ LP_{t+1} &= \alpha_{20} + \alpha_{21}M_t + \alpha_{22}LP_t + e_{2,t+1}, t = 1, 2\ldots \end{aligned} \tag{1a}$$

where *LP* denotes fractions of people with the allele, which we call the "gene pool," and *M* denotes numbers of dairy animals, which we call the "environment." This formulation is quite restrictive, however, because it allows only direct influences between *LP* and *M* (via coefficients α_{ij}) rather than the indirect influences of the more complex niche-construction pathways that we set out below.

We assume the system (1a) as well as (1b) below is initialized with a very small value of LP_1. We emphasize that all series that we write down are assumed to be stationary and to satisfy the relevant assumptions needed for valid regression analysis. Greene (2003, Chapter 20) contains an excellent discussion of problems that arise in time-series analysis. In a more realistic nonlinear system, in which the probability that the *LP* allele "invades" the population depends on a critical density of dairy animals, we could assume $\alpha_{22}=\alpha_{22}(M_t)$ is zero until M_t is large enough that it becomes positive.

That dairy animals were present before emergence of the *LP* allele is not decisive in the SET–NCT debate, as both can incorporate this into their accounts (Scott-Phillips et al. 2014). The difference is that SET sees domestication as "the environment"—no different in evolutionary perspective than climate, for example—whereas NCT sees domestication as the start of the human-constructed niche (O'Brien and Laland 2012). In other words, an SET hypothesis would let the density of dairy animals at date *t*, M_t, "cause" the frequency of lactase persistence at date *t*+1, LP_{t+1}. Under SET, this is one-way causation, as LP_t does not then "cause" more dairy animals, M_{t+1}, at date *t*+1 any more than it would "cause" the state of the climate at date *t*+1.

For NCT skeptics, at the risk of repeating, the change in the number of dairy animals, unless there was evidence of active domestication of those animals (e.g., evidence of cheese or yogurt consumption), is analogous to a change in climate (Scott-Phillips et al. 2014). In Eq. 1a, this implies $\alpha_{12}=0$ and $\alpha_{21}\neq 0$. In other words, all feedback is from the presence of dairy animals at date *t* to the number of people having the LP allele at date *t*+1. We will refer to the hypothesis $\alpha_{12}=0$ as "only the earlier and current environments cause the future gene pool" and consider it as the null hypothesis.

To make their case, NCT advocates would like to reject this null hypothesis, $H_0{:}\alpha_{12}=0$, in favor of $\alpha_{12}>0$, which would be consistent with the NCT view that Neolithic groups with the *LP* allele at date *t* constructed a "niche" of more dairy-producing animals at date *t*+1 than would have been present if α_{12} were zero. We doubt, however, that SET advocates would be very impressed with an NCT study that rejected this particular null hypothesis in this bivariate system. They would likely point out that a population with a higher fraction of the LP_t allele would surely be associated with more dairy-producing animals at period *t*+1 and, hence, testing $H_0{:}\alpha_{12}=0$ is not a "fair" test of their position. SET advocates might argue that we need to add a third equation for the dynamics of the population and the effect of those dynamics on the number of dairy animals, e.g.,

$$\begin{aligned} M_{t+1} &= \alpha_{10} + \alpha_{11}M_t + \alpha_{12}LP_t + \alpha_{13}N_t + e_{1,t+1} \\ LP_{t+1} &= \alpha_{20} + \alpha_{21}M_t + \alpha_{22}LP_t + \alpha_{23}N_t + e_{2,t+1} \\ N_{t+1} &= \alpha_{30} + \alpha_{31}M_t + \alpha_{32}LP_t + \alpha_{33}N_t + e_{3,t+1} \end{aligned} \tag{1b}$$

We suspect that an NCT study that now rejected $H_0\colon\alpha_{12}=\alpha_{32}=0$ would be more convincing to SET advocates. This rejection would be consistent with the view that the gene pool at time *t* had an effect on the number of dairy animals at period $t+1$ above and beyond the effect of the number of people at date *t*. Adding population to the system as in (1b) raises the potential issue of time trends in N_t and maybe even the number of dairy animals, M_t, over the period of interest, which began with the introduction into Europe of milkable domestic species from Anatolia some 8000 years ago (Leonardi et al. 2012). It is likely that the Neolithic population grew over time, as did the cattle population, so the time-series measurements may need to be adjusted for time trends in order to correct for the "spurious regressions" problem (Greene 2003, Section 20.3.2).

Linear systems 1a and 1b are oversimplified versions of, for example, the di-allelic two-locus dynamic system, where the frequency of alleles at the E locus governs a population's capacity for niche construction (Han and Hui 2014; Laland et al. 1999). In one example, Laland et al. (1999) have four state variables to represent the genetics alone, and the pathway in Fig. 1 has even more state variables. One can imagine assuming that the fitness functions of the Laland et al. system are functions of environmental-state variables (e.g., the stock of dairy animals) and then linearizing that system around a convenient base point and treating unobserved genetic-state variables as unobserved state variables. We hope to illustrate what a simple linear time-series model would look like of systems 1a and 1b, which can be viewed as a linearization of a one-locus, two-allele system, where the fitnesses are influenced by environmental variables. With one state variable for the genetics and several state variables for the environment below, systems 1a and 1b are intended to illustrate what kinds of hypothesis testing could be done.

The bi-variate system (1a) without the explicit population channel forces the effect of the population size at period *t* on the number of dairy animals at period $t+1$ to go through the LP_t channel; hence, this would likely lead to a spurious rejection. NCT advocates might argue, however, that the playing field of statistical testing is loaded against them if SET advocates force them to reject $H_0\colon\alpha_{12}=\alpha_{32}=0$ at a standard significance level of $p=.05$. If using SET as the null hypothesis inhibits discovery of subtler NCT pathways, we think it is fair to level the playing field by writing down time-series-model representations of NCT and SET that specify testable hypotheses for each.

In order to estimate (1) and obtain the estimated path-link strengths, $\{a_{ij}, i=1,2,3, j=1,2,3\}$, we must use available data to construct proxies for M_t, LP_t, N_t and discuss what the empirical counterparts of the time periods $t=1,2\ldots$ might be in any data set that is potentially available. Furthermore, we must discuss whether enough dated measures for M_t, LP_t, N_t can be constructed so that the regression analysis in Eq. (1) can estimate the parameters $\{a_{ij}, i=1,2,3, j=1,2,3\}$ precisely enough to be useful. Because we focus here on cases where the relative frequency of the LP_t allele is initially small—ca. 7000 years ago, when adults were uniformly lactose intolerant (Burger et al. 2007)—the linear approximation of Eq. (1) should be useful for expositional simplicity (see note 1 concerning nonlinear methods).

In what follows, we translate the path diagram in Fig. 1 by using the linear-dynamic-system assumption of Eq. (1) but with more variables and equations. This allows us to test for the presence of that particular path by using the ordinary least-squares (OLS) method. We begin by imagining an "ideal" data situation, which shows what parameters can be empirically identified in a causal-path diagram and highlights the limitations that exist even with infinite data (so much for an "ideal" situation). We then consider two problems that exist with most data sets: varying degrees of measurement error and the omission of key variables.

The "ideal" case

Let us suppose that we have excellent data for all variables in our LP example. What kinds of data might those be? As background, the coevolution of European LP and dairying probably originated between the northern Balkans and central Europe sometime between ca. 6250 and 8700 years ago (Gerbault et al. 2013; Itan et al. 2009). It is assumed that LP frequency increased slowly in these Balkan populations through selection but then spread rapidly into central and north-central Europe through the rapid demographic expansion of the subsequent Linearbandkeramik, or LBK, culture from around 8000 to 7500 years ago (Edwards et al. 2007). This led to the establishment of highly developed cattle-based dairying economies of Neolithic Europe around 6500 years ago.

Faunal remains from radiocarbon-dated Neolithic sites show that farmers began domesticating cattle, goats, and sheep between 10,000 and 11,000 years ago in the Middle East (Helmer et al. 2005; Peters et al. 2005; Troy et al. 2001; Zeder and Hesse 2000) and that milk production in Mediterranean Europe and the Middle East began about 2000 years later (Conolly et al. 2011, 2012; Manning et al. 2014; Vigne 2008; Vigne and Helmer 2007). Milk residues on pottery fragments (Dudd and Evershed 1998) indicate early milk-production exploitation as early as 7000–8000 years ago in northwestern Anatolia, about 7000 years ago in central Europe (Salque et al. 2013), and 6000 years ago in Britain (Copley et al. 2003, 2005; Craig et al. 2005a, b; Evershed et al. 2008).

Returning to Fig. 1, we can label the variables this way:

C_t	Dairy animal husbandry
N_t	Population
M_t	Milk-product consumption
LP_t	LP alleles
FM_t	Further consumption of milk products
FMP_t	Further milk-product manufacture and consumption
SB_t	Selective breeding of dairy animals
AM_t	Alleles conferring high milk yield in dairy animals
D_t	Dispersal at date t

We start by assuming that there is only one observable measure—which may be archaeological, genetic, or archaeogenetic—for each of variables at time t. Table 1 presents potential data sources for construction of proxies for the variables listed above except for the four variables that to our knowledge have never been quantified for the early Neolithic: FM_t, FMP_t, SB_t, and AM_t.

As Table 1 suggests, the main constraint is the temporal resolution of time series for each variable. Some variables, where data exist, are characterized only for one or two millennia-scale eras of the past. Among the best data are estimates of population densities, N_t, using the relatively new proxy of total number of radiocarbon dates per region per chronological bin, corrected for an assumed exponential decay in probability of site discovery with age (e.g., Collard et al. 2010; Gamble et al. 2005). Shennan et al. (2013) used 7944 radiocarbon dates from across western Europe as a proxy for population size, from which they substantiated the radical interpretation (Shennan 2000) that Neolithic populations did not just grow steadily but underwent substantial rise-and-decline cycles. Their data, as a proxy for N_t, are the best-resolved time series currently available for Table 1. Figure 2 shows an example from one of their subregions, southern Germany, with 841 dates binned into 246 time slices over a period of about 4000 years. This is about a bin every 16 years, or about one per human generation. If all the variables in Table 1 were well-resolved chronologically, we might be in business.

A substantial aggregated dataset, from 133 cemeteries during the transition from foraging to farming, shows the proportion of human skeletons from individuals 5 to 19 years old among all skeletons in the assemblage (Bocquet-Appel 2011). For the period prior to the Neolithic, there are about 23 datum points over 4000 years, or about one every 160 years. For the Neolithic, there are about 100 data points in 4000 years, or about one every 40 years. Although these data indicate the general population growth that took place in farming societies (Bocquet-Appel 2011), the data are from across the Northern Hemisphere and again show the current data-resolution limits

Table 1 Potential data sources for construction of proxies

Variable	Description	Evidence/proxy	Date(s)	Best resolution	Location	References
C_t, D_t	Domestic-animal density	400,000 animal bones recovered from 114 archaeological sites	ca. 12,000–7500 BP	1000 years	SW Asia, SE Europe	Connoly et al. 2011, 2012
C_t	Domestic animals	Proportion of domestic fauna in the assemblages of Early LBK sites	7500 BP	0	Central Europe	Jeunesse and Arbogast 1997; Chaix 1997
C_t	Domestic animals	6-household village: 40 cattle and 40 caprines using 6 km^2 land	ca 7000 BP	0	SW Germany	Gregg 1988:165-167
N_t	Human population	7944 radiocarbon dates from across western Europe	ca. 8000–5000 BP	16 years	Europe	Shennan et al. 2013 (and refs. 33–41 within)
N_t	Population	LBK longhouses in the Merzbach Valley	7300–6900 BP	25 years	Germany	Stehli 1989
D_t	Population	Radiocarbon-dated sites	9000–4000 BP	100 years	Europe	Gkiasta et al. 2003; Pinhasi et al. 2005; Shennan et al. 2013; Timpson et al. 2014
dN_t/dt	Rate of change in population	Proportion of adolescent skeletons in the assemblage	n.d.	40 years	Europe	Bocquet-Appel 2011
M_t	Milk-product consumption	Fraction of dated pottery sherds with milk residues	8500 BP	0	Anatolia, southeastern Europe, and the Middle East	Evershed et al. 2008; Gerbault et al. 2013; Salque et al. 2013
LP_t	LP alleles	Frequency of the (−13,910*T) allele across Europe	20,000–2000 BP	0	Central Europe	Burger et al. 2007; Bersaglieri et al. 2004; Coelho et al. 2005; Gerbault et al. 2013; Itan et al. 2009; Mulcare 2006

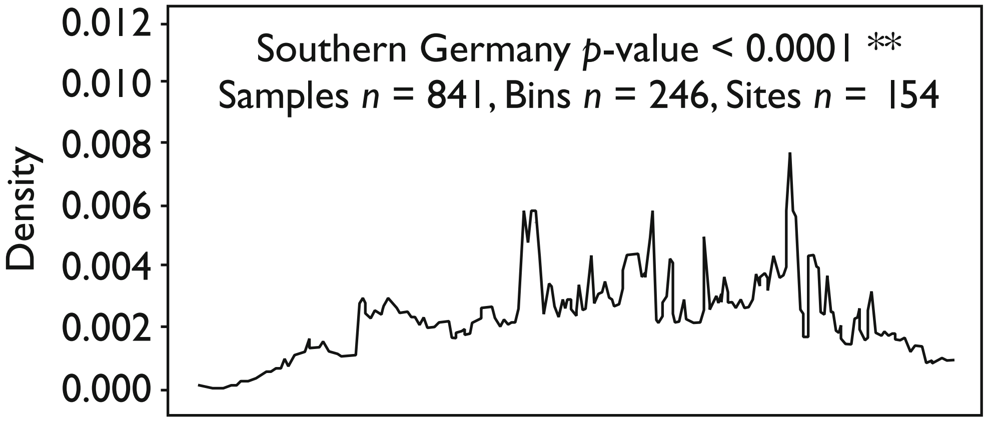

Fig. 2 An example of summed radiocarbon dates as a proxy for population size. Data from southern Germany—841 dates binned into 246 time slices from 8000 to 4000 calBP (from Shennan et al. 2013)

in establishing a well-resolved time series for N_t and D_t in testing the NCT hypothesis.

Bocquet-Appel (2011) estimates that, compared to the Mesolithic, the Neolithic fertility rate increased by about two births per woman, followed by an increase in mortality rate, such that pre-industrial farming populations grew about 0.1 % per year. On the local scale, however, growth in N_t might have been much faster. A good case study is LBK settlement of the Merzbach Valley in western Germany, where about 160 longhouses were constructed along a 1-km stretch of the Merzbach River over a period of four centuries, 5300–4900 BC (Ammerman and Cavalli-Sforza 1984; Stehli 1989). Assuming population growth was logistic (Ammerman and Cavalli-Sforza 1984), an initial growth rate, $dN_t/dt/N_t$, of about 2 % per year fits the archaeological sequence (Stehli 1989) fairly well, except for the sharp drop in occupied houses after 5000 BC. Again, these data are from but a single region.

Regarding the other variables in Table 1, the next best in terms of resolution would probably be the density of domestic animals, C_t. One of the most highly resolved data sets concerning C_t has been reported by Conolly et al. (2011, 2102) from southeastern Europe and Southwest Asia, with 400,000 animal bones recovered from 114 archaeological sites. As impressive as this is, the temporal resolution is reduced considerably once the minimum number of identified specimens per site has been estimated from these data and the results put into temporal bins such as the five shown in Fig. 3 that refer to the percentages of bones of domestic and wild cattle pooled from sites across the Levant that date between the 12th and 8th millennia calBP.

These kinds of data, most of which are publically archived at places such as the Archaeological Data Service at York University, yield excellent geographic representations of the relative dispersal of people (e.g., Gkiasta et al. 2003; Pinhasi et al. 2005) or domestic- versus wild-animal dispersal (Conolly et al. 2012) and would appear to yield hope for good data for the dispersal variable D_t, but again, the data are binned into time slices, with the resolution of the time series being one or two datum points per millennium. This is not the kind of temporal resolution one would want in an ideal situation.

In the Neolithic of central and northern Europe, where we currently have the strongest selection for the LP allele (Itan et al. 2009), C_t might be less variable in time but more variable in space. In Early LBK sites in the upper Rhine and upper Rhône valleys, for example, domestic fauna constitute over 90 % of the assemblage (Chaix 1997; Jeunesse and Arbogast 1997). Around the same time, however, early LBK sites in the Paris Basin range from 60 to 97 % domestic fauna (Jeunesse and Arbogast 1997), and at sites in the Alps south of Geneva, there may be less than 10 % domestic animals (Chaix 1997). For illustration, Fig. 4 shows the variability in proportion of domestic fauna in France and southern Germany ca. 7500 calBP (Jeunesse and Arbogast 1997).

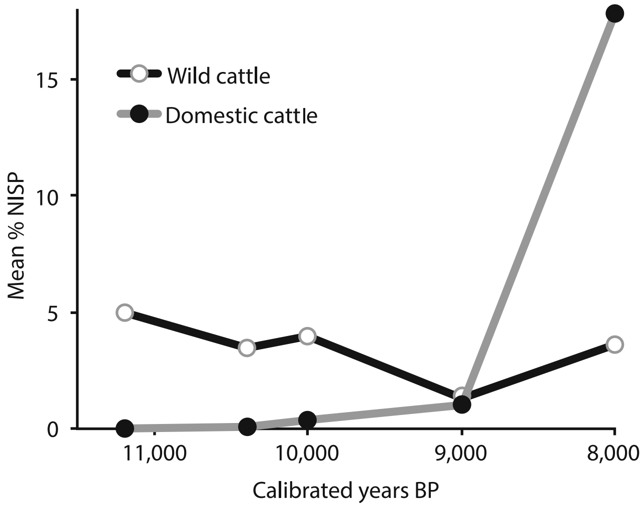

Fig. 3 Percentages of wild and domestic cattle bones (based on number of identified specimens), pooled from sites across the Levant that date between the 12th and 8th millennia calBP. Note that from about 11,800 to 9000 calBP, the use of domestic cattle increased slowly and that after 8500 calBP—corresponding to the earliest introduction and development of the Neolithic in southeastern Europe—domestic cattle increased exponentially. This shows that it took as long as 2000 years before domestic cattle began to fix within early agricultural communities as the dominant domestic animal species (from Conolly et al. 2012

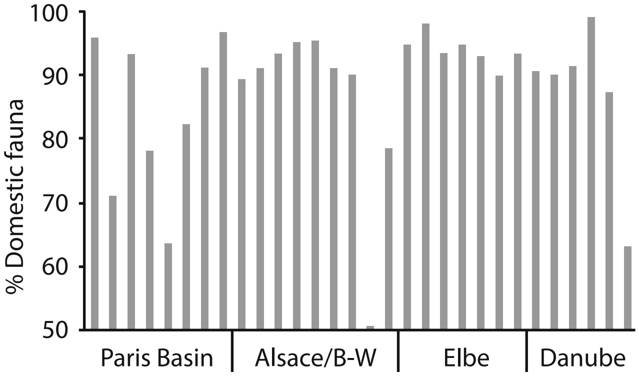

Fig. 4 Variability in proportion of domestic animals from Neolithic sites in France and southern Germany that date ca. 7500 calBP (after Jeunesse and Arbogast 1997)

Let us say, then, that we want to focus an NCT model on an area with majority domestic fauna. What can we say about the dynamics of C_t? Different lines of evidence—clustering of Neolithic houses, preserved plant remains, and isotopic analysis of human skeletons—suggest that land ownership developed in the early LBK (Bentley et al. 2012; Bogaard 2004; Bogaard et al. 2011), and Gregg (1988) estimated (based partly on assumptions of optimality) that a six-household Neolithic village might have held about 40 cattle and 40 sheep/goats and used perhaps 6 km^2 of land for cultivation and grazing. Perhaps climate records could be used to extend this estimate over time. Conolly et al. (2012), for example, discovered a significant positive correlation in the Levant between the ratio of domestic/wild cattle and annual precipitation, and a negative correlation between that ratio and maximum annual temperature. Are there similar correlations elsewhere?

In the ideal case, in which we have time-series data available from proxies of all the above variables (Fig. 1), we can write the estimating equations as

$$\begin{aligned}
M_{t+1} &= a_{10} + b_{12}C_t + a_{13}LP_t + a_{12}N_t + e_{1t+1}\\
N_{t+1} &= a_{20} + b_{22}C_t + a_{21}M_t + a_{22}N_t + e_{2,t+1}\\
LP_{t+1} &= a_{30} + b_{32}C_t + a_{31}M_t + a_{33}LP_t + a_{34}FM_t + a_{35}FMP_t + e_{3,t+1}\\
FM_{t+1} &= a_{40} + a_{41}M_t + a_{43}LP_t + a_{42}N_t + e_{4,t+1}\\
FMP_{t+1} &= a_{50} + a_{53}LP_t + a_{54}FM_t + a_{52}N_t + e_{5,t+1}\\
SB_{t+1} &= a_{60} + a_{65}FMP_t + a_{62}N_t + e_{6,t+1}\\
AM_{t+1} &= a_{70} + a_{76}SB_t + e_{7,t+1}\\
D_{t+1} &= a_{80} + a_{82}N_t + e_{8,t+1}
\end{aligned} \tag{2}$$

We treat the cattle dynamics, $C_{t+1}=c_0+c_1C_t+e_{C,t+1}$, as exogenous to the system (2) in order to give a faithful translation of the O'Brien and Laland (2012) path diagram reproduced in Fig. 1. O'Brien and Laland associate the paths a_{13},a_{43},a_{53} with NCT, so it would be natural to set up a null hypothesis for NCT that all these path coefficients are zero. More will be said about this below.

Let the column vector, X_{t+1}, denote the variables of (2) that appear on the left-hand side of the equation, which are numbered by rows, i.e., M is row 1, N is row 2, LP is row 3, and so on. Note that the coefficients $a_{i,j}$ quantify, from the path diagram (Fig. 1), the relationships between the variables in rows i and j. Then we can write (2) in the equivalent vector–matrix form,

$$X_{t+1} = a_0 + AX_t + b\otimes C_t + e_{t+1} \tag{3}$$

where X is an 8×1 column vector, A is an 8×8 matrix, C is a scalar, b is an 8×1 column vector, and $b\otimes C_t$ denotes the Kronecker product, which is an 8×1 vector. Note that only the first three elements of the 8×1 column vector b are nonzero and that each multiplies C_t. Now that we have an explicit time-series path-analytic model, we can define SET and NCT in the context of the model. The quantity of dairy animals, C_t, that are present at time t are the "environment," which influences the fraction of DNA sequences in the population having the LP allele, i.e., LP_{t+1} at period $t+1$.

In the representation above, in which SET is nested within NCT, the difference between SET and NCT lies in the estimations of coefficients a_{ij}. Specifically, we could resolve that SET implies that $a_{13}=a_{43}=a_{53}=0$, meaning that no gene/allele variable at time t influences "environmental" variables at $t+1$. Advocates of NCT would then look to reject the null hypothesis $H_0{:}a_{13}=a_{43}=a_{35}=0$, but at what confidence level?

Although $p=.05$ is typical, we believe that NCT deserves a higher confidence level, say, of $p=.25$ because NCT offers such novel, rich potential insights. As NCT is a multivariate system, we might make an analogy to a crossword puzzle, in which one may guess at words with much less than 95 % certainty in order to generate a critical mass of evidence to either to solve the puzzle or make it obvious that those initial guesses were incorrect.

If NCT advocates failed to reject such a null, would this force them to give up NCT in favor of SET? Perhaps it would be better to take a Bayesian model uncertainty approach (e.g., Cogley et al. 2011), such that one could put a prior of 0.5 on the NCT model and the SET model, estimate them on the data, and compute the posterior probability of each model given the data. This might enable both sides to advance their research agendas by formulating and estimating models that reflect both NCT and SET.

Equation (3) describes a vector auto-regression system with one lag, for which it is standard to assume the square matrix A is stable, in that its eigenvalues are all inside the unit circle of the complex plane (Greene 2003). If the time-series analog of the standard condition (Greene 2003, Eq. 1b)

$$E\{e_{t+1}|X_t, C_t\} = 0, t = 1, 2, 3, \ldots \tag{4}$$

holds, then, under additional regularity assumptions in the time-series case, we may estimate systems (2) and (3) by

OLS, row by row. To put it another way, if the regression errors, $\{e_{i,t+1}, i = 1, 2, \ldots, 8\}$, are conditionally independent of the regressors at each date t, then we may estimate each equation of (2) by OLS, equation by equation, and obtain, under modest additional sufficient conditions, consistent estimates of the parameters. In this case, consistent estimation means that the estimates of the parameters of the matrix A in Eq. 2 converge to their true values as the number of observations of the Xs gets larger and larger (Greene 2003).

In addition to consistency of estimation, the practical understanding of experienced researchers must ensure that specification of the equations shown in (1b) is at least approximately correct (White 1996). For example, we are using a linear approximation at the initial stage of diffusion of the *LP* allele, for the period when field research indicates that the frequency should be approximately zero.

Condition (4) may be violated in the case of omitted variables, which can lead to bias in the estimated regression coefficients (Greene 2003, Eq. 8.4). For example, in the bi-variate context of Eq. 1, the estimated coefficients are likely to be biased as a result of the omission of other variables that should be in the equations in addition to M_t and LP_t. Even in the context of (2), one might be tempted to just add more variables to try to correct for omitted variable bias. As Clarke (2005) shows in the context of linear regression analysis, however, this strategy of "bloated specifications" will not work. Concerning the uncertainty of any parameter estimate using finite sample data, we are avoiding the more technical discussions of how to handle the uncertainty in estimating that uncertainty itself (Greene 2003).

The case of measurement error

We now turn to a less-ideal situation, one where all variables in the path-diagram Eq. (1b) are measured, but with error. Suppose we observe $X=X^*+u, C=C^*+u_C$ and try to estimate the parameters A, b of the true system,

$$X^*_{t+1} = a_0 + AX^*_t + bC^*_t + e_{t+1}, E\{e_{t+1}|X^*_t, C^*_t\} = 0 \tag{5}$$

using the observed data $X=X^*+u, C=C^*+u_C$, where variables with asterisks are the true values and those without asterisks are observed but have large measurement error, i.e., the variances of u, u_C are large relative to the variances of X^*, C^*. Using only observed variables, we are reduced to attempting to estimate the system,

$$\begin{aligned} X_{t+1} &= a_0 + A(X_t - u_t) + b(C_t - u_{C,t}) + e_{t+1} + u_{t+1} \\ &= a_0 + AX_t + bC_t - Au_t - bu_{C,t} + e_{t+1} + u_{t+1} \end{aligned} \tag{6}$$

to obtain unbiased estimates of (A,b). If the condition

$$E\{u_{t+1}|X_t, C_t\} = 0 = E\{e_{t+1}|X_t, C_t\} \tag{7}$$

holds, then if $\{u_t, u_{C,t}\}$ were absent, OLS could be applied to obtain unbiased estimates of (A,b). If (7) holds but $\{u_t, u_{C,t}\}$ is present, then for unbiased estimates of (A,b), we need the following condition to hold:

$$E\{Au_t|X_t, C_t\} = E\{Au_t|X^*_t + u_t, C^*_t + u_{C,t}\} = 0 \tag{8}$$

In more detail, the relationship between the true and observed values, $X^*_t = X_t - u_t$, causes a dependence between the measurement error and the observed variable, which biases the OLS estimates of the matrix, A, and the vector, b.

These measurement errors change the system of equations from (2) as follows:

$$\begin{aligned} & M_{t+1} = a_{10} + b_{12}(C_t - u_{C,t}) + a_{13}(LP_t - u_{3,t}) + a_{12}(N_t - u_{2,t}) + u_{1,t+1} + e_{1t+1} \\ & N_{t+1} = a_{20} + b_{22}(C_t - u_{C,t}) + a_{21}(M_t - u_{1,t}) + a_{22}(N_t - u_{2,t}) + u_{2,t+1} + e_{2,t+1} \\ & LP_{t+1} = a_{30} + b_{32}(C_t - u_{C,t}) + a_{21}(M_t - u_{1,t}) \\ & + a_{33}(LP_t - u_{3,t}) + a_{34}(FM_t - u_{4,t}) + a_{35}(FMP_t - u_{5,t}) + u_{3,t+1} + e_{3,t+1} \\ & FM_{t+1} = a_{40} + b_{41}(M_t - u_{1,t}) + a_{43}(LP_t - u_{3,t}) + a_{42}(N_t - u_{2,t}) + u_{4,t+1} + e_{4,t+1} \\ & FMP_{t+1} = a_{50} + a_{53}(LP_t - u_{3,t}) + a_{54}(FM_t - u_{4,t}) + a_{52}(N_t - u_{2,t}) + u_{5,t+1} + e_{5,t+1} \\ & SB_{t+1} = a_{60} + b_{65}(FMP_t - u_{5,t}) + a_{62}(N_t - u_{2,t}) + u_{6,t+1} + e_{6,t+1} \\ & AM_{t+1} = a_{70} + b_{76}(SB_t - u_{6,t}) + u_{7,t+1} + e_{7,t+1} \\ & D_{t+1} = a_{80} + a_{82}(N_t - u_{2,t}) + u_{8,t+1} + e_{8,t+1} \end{aligned} \tag{9}$$

This gives us a context in which to discuss the effects of measurement error among the proxies for LP. If we can rank the relative sizes of measurement errors in (9), we can then expect that the equations with small measurement errors in the regressors on the RHS should yield less-biased estimates of their coefficients (Greene 2003). Among linked equations, measurement error in one variable typically biases all the other estimates (Greene 2003), but in the last three equations of (9), which have only one regressor on the right-hand side, only the error in those single regressors should bias the respective estimates (Greene 2003). In this case, one might even work out explicit expressions for biases in estimated coefficients in terms of measurement errors in the one right-hand side regressor, and if the constant term, a_{82}, is restricted to be zero, then the bias in the slope term a_{82} is easy to write down. Generally, measurement error biases the estimate of a_{82} toward zero, and the amount of attenuation toward zero is larger as the variance of the measurement error increases (Greene 2003, equations 5.29 and 5.30). The problem is that measurement error induces correlation between the regressors and the disturbance, which then biases the estimations.

What is the implication of measurement error for attempting to test the null hypothesis, $H_0 : a_{13} = a_{43} = a_{53} = 0$, which we had earlier established as one-way NCT advocates could generate potentially convincing evidence if this null hypothesis could be strongly rejected by the data? The well-known "attenuation bias" Greene (2003, Section 5.6.1) could potentially bias the estimated coefficients toward zero, and these estimated coefficients are needed to test the null. Hence, measurement error is potentially a serious problem in detecting NCT effects in the data when NCT effects are actually there.

The case of missing data

What about the impact of data limitations? Suppose we cannot obtain observations on each component of the vector, X_t. In this case, even with no measurement error, we are forced to estimate the system,

$$X_k = a_0 + AX_k + bC_k + e_k \tag{10}$$

for each observation $k=1,2,\ldots,$N. One might view Eq. (10) as a steady-state version of Eq. (2), which solves for the vector X_k, given (C_k, e_k).

Even if $E\{e|X,C\}=0$, we face difficulties in estimating the parameters of the matrix A. Suppose A is n x n, i.e., there are n variables. For each $r=1,2,\ldots,n$, we can solve (10) for X_{rk} and run an OLS regression of X_{rk} on the remaining regressors in row r of (10). Thus, for each row r, OLS estimates the coefficients as

$$(1-a_{rr})^{-1} a_{rs}, (1-a_{rr})^{-1} b_{rs}, s = 1, 2, \ldots, n, s \neq r \tag{11}$$

Recall that in our two-dimensional example, the terms $(1-a_{rr})^{-1} b_{rs}$ are nonzero only for $r=1,2$ and that $s=2$. The basic problem here is that we are short one regressor for each row $r=1,2,\ldots n$, so only the ratios in (11) may be estimated. The individual components of each ratio cannot be separated.

To return to our case example, say we have measures of milk consumption only from bone fragments and pottery. This means we do not have independently dated measures of $M(t)$, $FM(t)$, and $FMP(t)$. Suppose we obtain only one measure of all three,

$$\begin{aligned} FM(t) &= M(t), \\ FMP(t) &= M(t) \end{aligned} \tag{12}$$

In this case, (2) becomes

$$\begin{aligned}
M_{t+1} &= a_{10} + b_{11}C_t + a_{13}LP_t + a_{12}N_t + e_{1,t+1} \\
N_{t+1} &= a_{20} + b_{21}C_t + a_{21}M_t + a_{22}N_t + e_{2,t+1} \\
LP_{t+1} &= a_{30} + b_{32}C_t + a_{33}LP_t + (a_{12} + a_{35})M_t + e_{3,t+1} \\
M_{t+1} &= a_{40} + b_{41}C_t + a_{43}LP_t + a_{42}N_t + e_{4,t+1} \\
M_{t+1} &= a_{50} + a_{53}LP_t + a_{54}M_t + a_{52}N_t + e_{5,t+1} \\
SB_{t+1} &= a_{60} + a_{65}M_t + a_{62}N_t + e_{6,t+1} \\
AM_{t+1} &= a_{70} + a_{76}SB_t + e_{7,t+1} \\
D_{t+1} &= a_{80} + a_{82}N_t + e_{8,t+1}
\end{aligned} \tag{13}$$

It is clear from the third equation of (13) that only the sum $a_{34}+a_{35}$ can be identified, which means that we cannot identify all the parameters of the original model (2). Furthermore, notice that the fourth and fifth equations are the same regression equation with different coefficients from matrix A, which forces the parameters to satisfy the restriction, $a_{41}=a_{54}, a_{43}=a_{53}, a_{42}=a_{52}$. Finally, suppose unobserved variables, X_t^u, C_t^u, are present in the matrix–vector equation:

$$X_{t+1} = a_0 + AX_t + bC_t + A^u X_t^u + b^u C_t^u + e_{t+1} \tag{14}$$

The omitted variable formula of Greene (2003, eq. 8.4) may be adapted and applied to each row of (14) in order to obtain the bias in the estimated coefficients of each row of (14) and to obtain sufficient conditions for that bias to be zero. Prolifically adding "control" variables will not fix this problem and may make it worse (Clarke 2005).

Discussion

We have used the example of LP evolution to illustrate the basic method of testing for causal feedback paths as a step forward in testing two evolutionary models, SET and NCT. It would be taxonomic to go through all the possible causal paths in the path diagram in Fig. 1 that could be tested using the statistical time-series methods we advocate. The NCT literature has strongly emphasized the two-way feedback loops involved in niche-construction dynamics and has pointed out that those loops are outside the dynamics in SET. To advance past the debate stage, we suggest employing the kinds of methods that have the ability to test for the statistical significance of any feedback path. The use of explicit dynamic linear-estimating equations to represent the path diagram in Fig. 1 enables us to apply Granger causality testing to the potential two-way feedback implied by NCT. Actually, any empirical testing of niche construction must use a time-series causality framework because NCT implies the possibility that phenotypic features (O) at time *t* may cause genes (G) at *t*+1, that O_{t+1} may cause G_{t+2}, and so on, as well as causation running from genes at *t* to organismal characteristics at *t*+1 for all times.

The Granger causality framework is a convenient method for testing for the directionality of potential causal paths. We say that "G Granger causes O" if adding past measurements of genes incrementally helps predict future organismal characteristics in addition to usage of past measures of organism characteristics. To test that hypothesis, we would test the null hypothesis, namely, that O fails to Granger cause G. More formally, adding past measurements of organism characteristics, including measures of niche-construction activity, fails to incrementally predict future measured genes beyond usage of past measured genes.

We illustrate in the simplest possible system the hypothesis formulation and testing strategy for the context of LP and milk production by removing all the feedback channels in Fig. 1 and concentrating on only the minimalist linear system of regression equations shown in (1). Recall that *LP* and *M* denote frequencies of the *LP* allele and milk usage, respectively. NCT argues that there is feedback from the *LP* allele in previous periods to the spread of dairy products consumption in future periods (Fig. 1). Note that the bi-variate model aggregates all three paths from LP to milk-product consumption into only one path. Hence, we set up the null hypothesis, α_{12}=0, to be tested. How might this be done? Suppose we have measures of the frequency of *LP* alleles in populations *p*=1,2,…P at two periods of time *t*=1,2. Suppose we also have measures of milk consumption for these same P populations for the same two periods of time *t*=1,2. We may then run OLS regressions,

$$\begin{aligned} M_{p2} &= \alpha_{10} + \alpha_{11} M_{p1} + \alpha_{12} LP_{p1} + e_{1,p,2} \\ LP_{p2} &= \alpha_{20} + \alpha_{21} M_{p1} + \alpha_{22} LP_{p1} + e_{2,p,2} \end{aligned} \tag{15}$$

for *p*=1,2,…P and get OLS estimates $\hat{\alpha}_{ij}$, $\hat{\sigma}_{\hat{\alpha}_{ij}}$ of the regression coefficients and their standard errors.

How might we use these estimates to test the null hypothesis? A common convention is that the estimate $\hat{\alpha}_{ij}$ is statistically significant if the ratio $\hat{\alpha}_{ij}|/\hat{\sigma}_{\hat{\alpha}_{ij}} > 2$. Under this convention, one would reject the null hypothesis if $|\hat{\alpha}_{12}|/\hat{\sigma}_{\hat{\alpha}_{12}} > 2$. In other words, if $|\hat{\alpha}_{12}\Big|/\hat{\alpha}_{\hat{\alpha}_{12}} > 2$, we would say that the evidence is consistent with LP Granger causing M. In plain English, we would say (1) there is evidence of feedback from the spread of the *LP* allele to milk-product usage, i.e., through dairy husbandry, and (2) that this is indirect evidence of niche construction. We emphasize that this bi-variate model is being used only to illustrate statistical testing of a path direction in the simplest possible way. It does not specify, for instance, other benefits of dairy husbandry—such as the production of hides, meat, and rudimentary cheese and yogurt—that probably also helped prepare the niche for relatively rapid spread of *LP* alleles.

Even in this simple linear formulation, we are facing a general model-selection problem, i.e., finding a model process that is closest in some sense to the real data-generating process. Vuong (1989) has a particularly clear exposition of methods that use the Kullback–Leibler information criterion (KLIC) distance measure for estimating closeness of model processes to the real data-generating process. These methods are all based on likelihood ratio-based tests that the competing models are equally KLIC close to the real process against the alternative that one of the models is KLIC closer. As Vuong (1989) shows, these methods are closely related to the work of Akaike (1973) based on the KLIC criterion as well as Neyman and Pearson (1928) and methods based on quasi-maximum-likelihood estimators (see White 1996). We have focused on simple linear methods here and have avoided discussion of methods, for example, of optimal selection of the number of lags in order to keep the presentation as simple as possible.

Regarding the debate between NCT and SET (Scott-Phillips et al. 2014), Bayesians would probably avoid it and simply let data speak by reporting the posterior odds of SET and NCT. Of course, a Bayesian approach raises the issue of how to choose the prior. The Bayesian model uncertainty approach (Cogley et al. 2011) weights each model in a way that keeps a model with a small weight, but with important implications, in case such a transformative model turns out to be "true." This approach could frame the SET–NCT debate around the uncertainty of different time-series path-analytic models, one in which organisms actively construct the environment that impacts the future dynamics of their gene pool (NCT) and the other in which environmental dynamics exogenously drives the dynamics of DNA sequences (SET).

To further the debate, these analytical statistical approaches necessitate an explicit dynamic time-series model of NCT and of SET, as in (2), that allows for explicitly testing for the presence/absence of feedbacks that distinguish NCT from SET. It may be difficult to do this, however, because NCT and SET are hard to discriminate using real data, via their respective causal path channels in an explicit time-series model.

Conclusion

The NCT research program is gaining considerable traction in archeology and anthropology (e.g., Broughton et al. 2010; Collard et al. 2011; Gerbault et al. 2011; Kendal et al. 2011; Laland and Brown 2006; Laland and O'Brien 2010, 2011; Laland et al. 2001, 2010; O'Brien and Bentley 2015; O'Brien and Laland 2012; Odling-Smee and Turner 2011;

Riede 2011; Rowley-Conwy and Layton 2011; Shennan 2011; Smith 2007a, b, 2009, 2011a, b, 2015; Wollstonecroft 2011; Zeder 2012, 2015), which is appropriate, given that humans are the "ultimate niche constructors" (Odling-Smee et al. 2003, p. 28). As Laland and O'Brien (2010, 2012) point out, NCT would appear to offer a broad, biologically and culturally informed conceptual framework suited to the human sciences—one that recognizes the active agency of humans as part causes of their own development, history, and evolution. It also recognizes ecological inheritance as a second general legacy that organisms inherit from their ancestors, thereby providing researchers with additional explanatory mechanisms. These mechanisms would seem particularly relevant to archaeologists, given that (1) human niche construction is frequently a manifestation of acquired characters (Laland and O'Brien 2010, 2011) and (2) human ecological inheritance includes a rich material culture (O'Brien and Bentley 2015; Wollstonecroft 2011). Furthermore, NCT offers a highly fertile way of thinking about problems and brings with it a suite of novel hypotheses, both directly from an emphasis on the agency of the organism and indirectly by suggesting general patterns and rules.

Despite being enthusiastic, advocates of NCT must face the fact that testing NCT versus SET is not easy. The difficulty lies in the statistical-identification problem of a dynamic population-genetics model, where the fitnesses are functions of environmental variables, not all of which are observable. Advocates of NCT must deal with unobserved "confounders, " which might misleadingly look like NCT effects, and statistically identify whether the dynamics of the environmental variables are coupled (nonlinearly) to the dynamics of the gene pool.

So, what have we learned, and what, if anything, can path analysis and time-series analysis contribute to the debate between advocates of NCT and advocates of SET? It seems to us that our analysis and discussion are useful for structuring and focusing the debate in more precise directions and focusing discussion on what kinds of data need to be generated and what kinds of statistical-inference problems must be faced in order to advance the discussion on causal directionality of path links. We believe the kind of analysis we advocate is also useful for identifying path links that might have been overlooked without the road map provided by path analysis. Our discussion is also useful for identifying potential "confounders" that might lead one to think that causality is in a particular direction when it is not actually there. In settings where quality data are hard to come by, especially in the social sciences, where disputes about direction and size of causation are particularly strong, some researchers have advanced a partial-identification approach, whereby we simply agree to, and accept, the reality of large uncertainties about directionality and size of causal paths. This approach reports "partial identification sets," which can be narrowed only by better data or by certain kinds of experiments (Manski 2005). Path-analysis discussions are useful for focusing attention on where effort should be allocated toward reducing the size of the partial identification set.

Acknowledgments We thank Lisa Hildebrand and two anonymous reviewers for numerous helpful comments on earlier drafts. We are also grateful to Nick Conard for his kind assistance during the review process.

Appendix

Freedman (2004) is an excellent source for the warnings and difficulties of trying to infer causality from regression analysis. Sensible theory and interpretation must always be used, even if the "fit" is good and all the usual specification tests are passed. The theory in the LP example is quite convincing about the qualitative directions of the causal arrows in the path diagram, but suppose we want to estimate the quantitative size of those paths of causality and conduct quantitative statistical inference. How would we do it, what problems are likely to arise, and how can we fix them?

The main text of the paper focuses on how to statistically test for two-way feedback effects that distinguish NCT from SET using the expository device of linear time-series analysis. The text illustrated this problem using the example of lactase persistence (LP), which was modeled using a linear time-series dynamical system with several environmental variables and a one-locus, two-allele system. In some sense, the LP example is not a good choice to illustrate the approach because the causal mechanisms behind the spread (and absence) of LP are quite well understood (e.g., Ingram et al. 2009).

Population-genetics modeling, however, inherently generates nonlinear dynamical systems. Here, we construct some simple models of NCT versus SET, including a simple nonlinear dynamic population-genetics model where the fitness functions depend not only on allelic frequencies but also on environmental stock variables (e.g., stock of animals capable of producing dairy products), which may have been augmented by activities in the past. Examples of mathematical models of niche construction are Laland et al. (1999), Creanza et al. (2012), and Han and Hui (2014).

We begin by writing down a discrete time-simple dynamic model of interactions between organisms and environment by adapting Laland et al.'s (2009) Eqs. (1a, 1b) and (2a, 2b) and executing a "small noise" expansion to get a linear time-series model. We then use the resulting linear model to discuss some statistical-identification issues in distinguishing NCT from SET.

We have as a mathematical metaphor for SET, where all feedback is from E to O and not vice versa,

$$O_{t+1} = f\left(O_t, E_t, \varepsilon e_{1,t+1}\right),\ E_{t+1} = g\left(E_t, \varepsilon e_{2,t+1}\right) \tag{A1}$$

Differentiating (A1) with respect to ε and evaluating it at $\varepsilon=0$, we get

$$dO_{t+1}/d\varepsilon = f_O dO_t/d\varepsilon + f_E dE_t/d\varepsilon + f_e e_{1,t+1},\ dE_{t+1}/d\varepsilon = g_E dE_t/d\varepsilon + g_e e_{2,t+1} \tag{A2}$$

Here, partial derivatives are denoted by subscripts, partial derivatives are evaluated at $\varepsilon=0$, and O_t,E_t denote the state of the organism and the state of the environment at date t. Equation (A2) is a linear time-series model for the perturbations.

A "mathematical metaphor" for NCT is

$$O_{t+1} = f\left(O_t, E_t, \varepsilon e_{1,t+1}\right),\ E_{t+1} = g\left(O_t, E_t, \varepsilon e_{2,t+1}\right) \tag{A3}$$

with derivative equation

$$dO_{t+1}/d\varepsilon = f_O dO_t/d\varepsilon + f_E dE_t/d\varepsilon + f_e e_{1,t+1},\ dE_{t+1}/d\varepsilon = g_O dO_t/d\varepsilon + g_E dE_t/d\varepsilon + g_e e_{2,t+1} \tag{A4}$$

Here, there is feedback from organisms to environment because of niche construction. If we had data on proxies for the perturbations, $\{dO_s/d\varepsilon, dE_s/d\varepsilon, s=1,2,\ldots\}$, we could potentially design a statistical test of (A2) against (A1) by estimating the slope coefficient, g_O, and testing the null hypothesis that g_O is zero against the alternative that it is not zero.

If we put the state of the organism to be the fraction of organisms—humans in this case—having the *LP* allele at date t, i.e., $O_t=LP_t$, and define the environmental-state variable, $E_t=M_t$, then the linear system (1a) in the text is the same as equation (A4), provided the partial derivatives are constant in time. If the regression condition is $E\{e_{it+1}|LP_t,M_t\}=0, i=1,2$ for all dates t, then we can run simple ordinary least squares (OLS) to test the null hypothesis $H_0{:}\alpha_{12}=0$. Rejection of the null is consistent with NCT being true. Discussion is in the main text. If we replace the scalar E_t with the vector $E_t=(M_t, N_t)$, then Eq. (A4) would correspond to Eq. (1b) in the main text, provided the partial derivatives in (A4) are constant in time. The main text discusses the relevant null hypothesis in this case.

When might the partial derivatives be constant in time? One example is when the perturbations occur around a deterministic steady state of the system. What should one do if the underlying system is too nonlinear for linear approximations to be of workable quality? In this case, we would recommend a nonlinear testing approach (e.g., Bahadori and Liu 2013; Diks and Panchenko 2006; Lee et al. 1993), as explained in the main text.

A potentially useful analogy to testing NCT versus SET is testing endogenous growth theory (EGT) against conventional exogenous growth theory (CGT) in economics (Jones 1995). EGT argues that increases in human research-and-development (R&D), much like increases in niche construction, create a permanent effect on growth, whereas CGT would argue that the effect of increased R&D is transient and wears off over time. Jones argued that statistical methods for testing for "unit roots" could be used to test for EGT effects if one has an explicit measure for the amount of R&D conducted at each date. His methods shifted the debate between advocates of EGT and CGT when he found little evidence for the persistent effect on productivity of the economy that was predicted by EGT.

By analogy, if we had an explicit measure for the amount of niche construction being done at each date, we could adapt Jones's methods to test NCT versus SET. Unfortunately, finding data on quantitative measures of niche construction at different time periods is difficult. Thus, if we use this route, we must have a proxy for the actual amount of niche construction. It might be difficult to detect niche-construction effects even if they are there. As an example, consider the following modification of Eqs. (A3) and (A4):

$$\begin{aligned} &LP_{t+1} = f\left(LP_t, E_t, \varepsilon e_{1,t+1}\right), \\ &E_{t+1} = \mu_E + \rho_E E_t + \beta_{LP} LP_t + \beta_{-LP}(1-LP_t) + e_{2,t+1} \\ &\beta_{LP} > \beta_{-LP} \geq 0 \end{aligned} \tag{A5}$$

The idea here is that humans with and without the *LP* allele can indulge in environmental modification—for example, by the husbandry of animals with potential dairy-production capability for whatever reasons, some of which may be totally independent of consumption of dairy products. In (A5), we assume $\beta_{LP}>\beta_{-LP}\geq 0$ to capture the idea that humans with the *LP* allele should be more interested in adding and caring for animals with dairy-product potential than humans not carrying the allele. However, humans without the *LP* allele may have $\beta_{LP}>\beta_{-LP}\geq 0$. Suppose we cannot observe the actual amount of dairy husbandry but have observations on LP_t. We can look at parameter values in (A5) that are consistent with feedback from LP_t to E_{t+1} but where this feedback is hard to detect statistically. Note that we have not required the function f(.) in (A5) to be linear. Nonlinearity in f(.) allows one to treat threshold phenomena—for example, *LP* being an advantageous allele only when the population of dairy animals reaches a threshold level.

Even though (A5) is a simple, essentially trivial dynamic model, it directs our attention to two potential problems in statistically detecting NCT effects, even though the effects are present. The first is where the difference, $\beta_{LP}-\beta_{-LP}$, is small. A regression of an observable measure of E_{t+1} on E_t would deliver estimates of the constant term, $\mu_E+\beta_{-LP}$, the slope effect, $\beta_{LP}-\beta_{-LP}$, and the persistence parameter, ρ_E. If $\beta_{LP}-\beta_{-LP}$ is close to zero, one would conclude there is no NCT effect, even though it is present. If the persistence parameter, $\rho_E>0$, is close to one, the NCT effect on the ultimate state of the environment could be very large, even though it would be extremely difficult to detect.

The second problem is the case where $\beta_{-LP}=0$ and $\beta_{LP}>0$ is small. It is clear from the discussion above that it would be empirically difficult to detect the NCT effect, $\beta_{LP}>0$. However, the NCT effect could be very important if $\rho_E>0$ is very close to one. Indeed, if $\mu_E=0$, so that the environment would decay to nothing if there were no niche construction, a very small $\beta_{LP}>0$, coupled with $\rho_E>0$ very close to one, would lead to a large ultimate effect on the environment, E, even if only a small fraction of the population has the *LP* allele.

A crude approach to detecting NCT effects in the two cases above might be to take the current state of the environment as a proxy for niche-construction activity and examine how fast a surprise increase in the environment—perhaps a result of some kind of observable positive shock to the environmental-state variable—wears off on the measure of organism state once we have a measure of the state of the organisms. In other words, we could plot the empirical impulse-response function (Jones 1995) of the system to a positive shock to the state of the environment. A rapid decay rate would be consistent with the absence of NCT effects, and a much slower decay rate would be consistent with the presence of NCT effects. Admittedly, this would be a crude, indirect way of dealing with the problems exposed in the two cases above, but perhaps it is better than nothing.

Here is another example where NCT effects might be difficult to detect when present. Suppose niche construction is not "free" to an organism, meaning it might take energy, time, or some other resource away from the organism that it could have used for something else that would increase fitness. A simple model of this effect is

$$O_{t+1} = f\left(O_t, E_t, A_t, \varepsilon e_{1,t+1}\right),\ E_{t+1} = g\left(E_t, A_t, \varepsilon e_{2,t+1}\right) \quad \text{(A6)}$$

where an increase in A_t increases g(.) but where f(.) is increasing then decreasing in A_t. This assumption on f(.) reflects diminishing returns and increasing marginal costs to the organism in terms of its reproductive success for the next period.

Suppose there is selection for organisms to solve the problem

$$\max_{A_t} \int \left\{ f\left(O_t, E_t, A_t, \varepsilon e_{1,t+1}\right) \right\} dF_{e_1} \quad \text{(A7)}$$

This maximization problem implies that A_t is a function of (O_t,E_t), call it $A(O_t,E_t)$. If $A(O_t,E_t)$ is a function of E_t alone, and if we do not observe A_t, then clearly we will not detect NCT-type feedback into g(.) by trying to relate E_{t+1} to E_t by any method. When might this happen? A sufficient condition is

$$\partial^2 f / \partial A \partial O = 0 \quad \text{(A8)}$$

Examples of functions $f(O,E,A,e)$ that satisfy (A8) are

$$\begin{aligned} f(O,E,A,e) &= f_1(O,E,e) f_2(A,E,e) \\ f(O,E,A,e) &= f_1(O,E,e) + f_2(A,E,e) \end{aligned} \quad \text{(A9)}$$

When might NCT effects be detected in an empirical study but in reality are not actually there, i.e., a false positive occurs? Consider the dynamics

$$O_{t+1} = f\left(O_t, E_t, Z_t, \varepsilon e_{1,t+1}\right),\ E_{t+1} = g\left(E_t, Z_t, \varepsilon e_{2,t+1}\right) \quad \text{(A10)}$$

where Z_t is an unobserved state variable at period t. If the dynamics were linear with constant coefficients, we could write them in the form of a bi-variate auto-regression:

$$\begin{aligned} E_{t+1} &= \alpha_{10} + \alpha_{11} E_t + \alpha_{12} O_t + \alpha_{13} Z_t + e_{1,t+1},\ t = 1, 2, \ldots, T \\ O_{t+1} &= \alpha_{20} + \alpha_{21} E_t + \alpha_{22} O_t + \alpha_{23} Z_t + e_{2,t+1},\ t = 1, 2, \ldots, T \end{aligned} \quad \text{(A11)}$$

We could use the omitted variable formula (Greene 2003, Section 8.2.1) to indicate that typically there will be bias in the estimate of α_{12}, even when the true value is zero. Because it is difficult to exclude omitted variables, Z_t, unless one has strong theory, omitted variables are likely to be present and cause false positives of NCT effects even when they are not really there. Using theory to argue for the presence (or absence) of omitted variables may help sharpen theoretical discussions in adducing arguments in support for (or against) NCT.

Estimating selection strength

We have been asked on occasion to compare our approach to others, such as that of Itan et al. (2009). We can do that by considering their Eq. (3):

$$\begin{aligned} p' &= \left[(1+s)p^2 + (1+s)pq\right] \Big/ \left[1 + s\left(p^2 + pq\right)\right] \\ N' &= N\left(1 + s\left(p^2 + 2pq\right)\right) \end{aligned} \quad \text{(A12)}$$

where their p,p' and N,N' are, in our notation,

$$p \equiv p_t = LP_t, p' \equiv p_{t+1} = LP_{t+1}, q \equiv 1 - p, N \equiv N_t, N' \equiv N_{t+1} \quad \text{(A13)}$$

The Itan et al. (2009) population-genetics system (A12) is a one-locus, two-allele diploid model, where fitnesses are specified as $w_{11}=w_{12}=1+s$ for homozygotes and heterozygotes with the *LP* allele and $w_{22}=1$ for homozygotes without the *LP* allele, and where $s>0$ denotes the relative fitness of those individuals with the *LP* allele. Populations containing individuals with the *LP* allele are assumed to grow faster.

The analogy to (A12) in our model is

$$LP_{t+1} = a_{30} + b_{32}C_t + a_{31}M_t + a_{33}LP_t + a_{34}FM_t + a_{35}FMP_t + e_{3,t+1}$$
$$N_{t+1} = a_{20} + b_{22}C_t + a_{21}M_t + a_{22}N_t + e_{2,t+1} \tag{A14}$$

which can be viewed as a linearization of the Itan et al. (2009) model, where their selection coefficient, "s," is not constant but rather a function:

$$s_{t+1} \equiv S(C_t, M_t, LP_t, FM_t, FMP_t, e_{t+1}) \tag{A15}$$

If we inserted (A15) in place of s in Itan et al.'s Eq. (A12) and linearized them around p_t=0, we would obtain a system much like (A14) except the coefficients would be time dependent.

References

Akaike H (1973) Information theory and an extension of the likelihood-ratio principle. In: Petrov BN, Csaki F (eds) Proceedings of the Second International Symposium on Information Theory. Akademia Kiado, Budapest, pp 257–281

Ammerman AJ, Cavalli-Sforza LL (1984) The Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton

Aoki KA (1996) Stochastic model of gene-culture coevolution suggested by the "culture historical hypothesis" for the evolution of adult lactose absorption in humans. Proc Natl Acad Sci 83:2929–2933

Bahadori T, Liu Y (2013) An examination of practical Granger causality. SIAM Conference on Data Mining, (SDM'13), available at http://www-bcf.usc.edu/~liu32/granger.pdf and http://www-bcf.usc.edu/~liu32/sdm_theory.pdf

Bentley RA, Bickle P, Fibiger L, Nowell GM, Dale CW, Hedges REM, Hamilton J et al (2012) Communty differentiation and kinship among Europe's first farmers. Proc Natl Acad Sci 109:9326–9330

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN (2004) Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74:1111–1120

Bocquet-Appel JP (2011) When the world's population took off: the springboard of the Neolithic demographic transition. Science 333: 560–561

Bogaard A (2004) Neolithic farming in central Europe. Routledge, London

Bogaard A, Strien HC, Krause R (2011) Towards a social geography of cultivation and plant use in an early farming community: Vaihingen an der Enz, south-west Germany. Antiquity 85:395–416

Bogucki P (1984) Linear pottery ceramic sieves and their economic implications. Oxf J Archaeol 3:15–30

Bogucki P (1986) The antiquity of dairying in temperate Europe. Expedition 28:51–58

Bogucki P (1993) Animal traction and household economies in Neolithic Europe. Antiquity 67:492–503

Boni MF, Feldman MW (2005) Evolution of antibiotic resistance by human and bacterial niche construction. Evolution 59:477–491

Borenstein E, Kendal J, Feldman MW (2006) Cultural niche construction in a metapopulation. Theor Popul Biol 70:92–104

Brodie DE III (2005) Caution: niche construction ahead. Evolution 59: 249–251

Broughton JM, Cannon MD, Bartlelink EJ (2010) Evolutionary ecology, resource depression, and niche construction theory: applications to central California hunter–gatherers and Mimbres–Mogollon agriculturalists. J Archaeol Method Theor 17:371–421

Burger J, Kirchner M, Bramanti B, Haak W, Thomas MG (2007) Absence of the lactase-persistence-associated allele in Early Neolithic Europeans. Proc Natl Acad Sci 104:3736–3741

Chaix L (1997) La transition Meso-Neolithique: quelques donnees de l'archeozoologie dans les Alpes du Nord et le Jura. In: Jeunesse C (ed) Le Neolithique Danubien et ses marges entre Rhin et Seine. Cahiers de Association pour la Promotion de la Recherche Archeologique en Alsace. Supplement no. 3, Alsace-Lorraine, pp 191–196

Clarke K (2005) The phantom menace: omitted variable bias in econometric research. Conflict Manage Peace Sci 22:341–352

Coelho M, Luiselli D, Bertorelle G, Lopes AI, Seixas S, Destro-Bisol G, Rocha J (2005) Microsatellite variation and evolution of human lactase persistence. Hum Genet 117:329–339

Cogley T, de Paoli B, Matthes C, Nikolov K, Yates T (2011) A Bayesian approach to optimal monetary policy with parameter and model uncertainty. J Econ Dyn Control 35:2186–2212

Collard M, Edinborough K, Shennan S, Thomas MG (2010) Radiocarbon evidence indicates that migrants introduced farming to Britain. J Archaeol Sci 37:866–870

Collard M, Buchanan B, Ruttle A, O'Brien MJ (2011) Niche construction and the toolkits of hunter–gatherers and food producers. Biol Theor 6:251–259

Conolly J, Colledge S, Dobney K, Vigne JD, Peters J, Stopp B, Manning K, Shennan S (2011) Meta-analysis of zooarchaeological data from SW Asia and SE Europe provides insight into the origins and spread of animal husbandry. J Archaeol Sci 38:538–545

Conolly J, Manning K, Colledge S, Dobney K, Shennan S (2012) Species distribution modeling of ancient cattle from early Neolithic sites in SW Asia and Europe. Holocene 22:997–1010

Copley MS, Berstan R, Dudd SN, Docherty G, Mukherjee AJ, Straker V, Payne S, Evershed RP (2003) Direct chemical evidence for widespread dairying in prehistoric Britain. Proc Natl Acad Sci 100:1524–1529

Copley MS, Berstan R, Dudd SN, Aillaud S, Mukherjee AJ, Straker V, Payne S, Evershed RP (2005) Processing of milk products in pottery vessels through British prehistory. Antiquity 79:895–908

Craig OE, Chapman J, Heron C, Willis LH, Bartosiewicz L, Taylor G, Whittle A, Collins M (2005a) Did the first farmers of central and eastern Europe produce dairy foods? Antiquity 79:882–894

Craig OE, Taylor G, Mulville J, Collins M, Parker Pearson M (2005b) The identification of prehistoric dairying activities in the western Isles of Scotland: an integrated biomolecular approach. J Archaeol Sci 32:91–103

Creanza N, Fogarty L, Feldman MW (2012) Models of niche construction with selection and assortative mating. PLoS ONE 7(8):e42744

Curry A (2013) Archaeology: the milk revolution. Nature 500:20–22

Dawkins R (2004) Extended phenotype—but not too extended. A reply to Laland, Turn and Jablonka. Biol Philos 19:377–396

Diks C, Panchenko V (2006) A new statistic and practical guidelines for nonparametric Granger causality testing. J Econ Dyn Control 30: 1647–1669

Donohue K (2005) Niche construction through phonological plasticity: life history dynamics and ecological consequences. New Phytol 166:83–92

Dudd SN, Evershed RP (1998) Direct demonstration of milk as an element of archaeological economies. Science 282:1478–1481

Dunbar RIM, Shultz S (2007) Understanding primate brain evolution. Philos Trans R Soc B 362:649–658

Edwards CJ, Bollongino R, Scheu A, Chamberlain A, Tresset A, Vigne JD, Baird JF (2007) Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. Proc R Soc B 274:1377–1385

Eichler M (2007) Granger causality and path diagrams for multivariate time series. J Econometrics 137:334–353

Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I (2002) Identification of a variant associated with adult-type hypolactasia. Nat Genet 30:233–237

Engle RF, Hendry DF, Richard JF (1983) Exogeneity. Econometrica 51: 277–304

Ericsson NR (1992) Co-integration, exogeneity, and policy analysis: an overview. J Policy Model 14:251–280

Erwin DH (2008) Macroevolution of ecosystem engineering, niche construction and diversity. Trends Ecol Evol 23:304–310

Evershed RP, Payne S, Sherratt AG, Copley MS, Coolidge J, Urem-Kotsu D, Kotsakis K et al (2008) Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. Nature 455:528–531

Feldman MW, Cavalli-Sforzi LL (1989) On the theory of evolution under genetic and cultural transmission with application to the lactose absorption problem. In: Feldman MW (ed) Mathematical evolutionary theory. Princeton University Press, Princeton, pp 145–173

Feldman MW, Laland KN (1996) Gene-culture co-evolutionary theory. Trends Ecol Evol 11:453–457

Flatz G (1987) Genetics of lactose digestion in humans. Adv Hum Genet 16:1–77

Freedman DA (2004) On specifying graphical models for causation and the identification problem. Eval Rev 28:267–293

Furrow RE, Christiansen FB, Feldman MW (2011) Environment-sensitive epigenetics and the heritability of complex diseases. Genetics 189:1377–1387

Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, Domboróczki KI, Pap I, Anders A, Whittle A, Dani J, Raczky P, Higham TFG, Hofreiter M, Bradley DG, Pinhasi R (2014) Genome flux and statis in a five millennium transect of European prehistory. Nat Commun 5:5257

Gamble C, Davies W, Pettitt P, Hazelwood L, Richards M (2005) The archaeological and genetic foundations of the European population during the late glacial implications for ‘agricultural thinking.’ Camb Archaeol J 15:193e223

Gerbault P, Liebert A, Itan Y, Powell A, Currat M, Burger J, Swallow DM, Thomas MG (2011) Evolution of lactase persistence: an example of human niche construction. Philos Trans R Soc B 366:863–877

Gerbault P, Roffet-Salque M, Evershed R, Thomas M (2013) How long have adult humans been drinking milk? Int Union Biochem Mol Biol 65:983–990

Geweke J (1984) Inference and causality in economic time series. In: Griliches Z, Intriligator MD (eds) Handbook of econometrics, vol. 2. North-Holland, Amsterdam, pp 1101–1144

Gkiasta M, Russell T, Shennan SJ, Steele J (2003) Neolithic transition in Europe: the radiocarbon record revisited. Antiquity 77:45–62

Godfrey-Smith P (1996) Complexity and the function of mind in nature. Cambridge University Press, Cambridge

Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37:424–438

Greene WH (2003) Econometric analysis (fifth ed). Prentice Hall, Upper Saddle River

Gregg S (1988) Foragers and farmers: population interaction and agricultural expansion in prehistoric Europe. University of Chicago Press, Chicago

Han X, Hui C (2014) Niche contruction on environmental gradients: the formation of fitness valley and stratified genotypic distributions. PLoS ONE 9(6):e99775

Harvey PH, Pagel MD (1991) The comparative method in evolutionary biology. Oxford University Press, Oxford

Harvey CB, Hollox EJ, Poulter M, Wang Y, Rossi M, Auricchio S, Iqbal TH et al (1998) Lactase haplotype frequencies in Caucasians: association with the lactase persistence/non-persistence polymorphism. Ann Hum Genet 62:215–223

Hauser RM, Goldberger AS (1971) The treatment of unobserved variables in path analysis. Sociol Methodol 3:81–117

Heckman J, Pinto R (2014) Causal analysis after Haavelmo. Economet Theor 31:1–37

Helmer D, Gourichon L, Monchot H, Peters J, Sana Segui M (2005) Identifying early domestic cattle from pre-pottery Neolithic sites on the Middle Euphrates using sexual dimorphism. In: Vigne JD, Peters J, Helmer D (eds) The first steps of animal domestication: new archaeozoological approaches. Oxbow, Oxford, pp 86–95

Holden C, Mace R (1997) Phylogenetic analysis of the evolution of lactose digestion in adults. Hum Biol 69:605–628

Hollox EJ, Poulter M, Zvarik M, Ferak V, Krause A, Jenkins T, Saha N et al (2001) Lactase haplotype diversity in the Old World. Am J Hum Genet 68:160–172

Ihara Y, Feldman MW (2004) Cultural niche construction and the evolution of small family size. Theor Popul Biol 65:105–111

Ingram CJ, Mulcare CA, Itan Y, Thomas MG, Swallow DM (2009) Lactose digestion and the evolutionary genetics of lactase persistence. Hum Genet 124:579–591

Itan Y, Powell A, Beaumont MA, Burger J, Thomas MG (2009) The origins of lactase persistence in Europe. PLoS Comput Biol 5(8): e1000491

Itan Y, Jones BL, Ingram CJ, Swallow DM, Thomas MG (2010) A worldwide correlation of lactase persistence phenotype and genotypes. BMC Evol Biol 10:36

Jeunesse C, Arbogast RM (1997) A propos du statut de la chasse au Neolithique moyen. La faune sauvage dans les dechets domestiques et dans les mobiliers funeraires. In: Jeunesse C (ed) Le Neolithique Danubien et ses marges entre Rhin et Seine. Cahiers de Association pour la Promotion de la Recherche Archeologique en Alsace. Supplement no. 3, Alsace-Lorraine, pp 81–102

Jones CI (1995) Time series tests of endogenous growth models. Q J Econ 110:495–525

Jones CG, Lawton JH, Shachak M (1997) Positive and negative effects of organisms as physical ecosystem engineers. Oikos 69:373–386

Kendal JR, Tehrani JJ, Odling-Smee FJ (2011) Human niche construction in interdisciplinary focus. Philos Trans R Soc B 366:785–792

Krakauer DC, Page KM, Erwin D (2009) Diversity, dilemmas, and monopolies of niche construction. Am Nat 173:26–40

Krüttli A, Bouwman A, Akgül G, Della Casa P, Rühli F, Warinner C (2014) Ancient DNA analysis reveals high frequency of European lactase persistence allele (T-13910) in Medieval central Europe. PLoS ONE 9(1):e86251

Laland KN, Brown G (2006) Niche construction, human behavior, and the adaptive-lag hypothesis. Evol Anthropol 15:95–104

Laland KN, O’Brien MJ (2010) Niche construction theory and archaeology. J Archaeol Method Theor 17:303–322

Laland KN, O’Brien MJ (2011) Cultural niche construction: an introduction. Biol Theor 6:191–202

Laland KN, Sterelny K (2006) Seven reasons (not) to neglect niche construction. Evolution 60:1751–1762

Laland KN, Odling-Smee FJ, Feldman MW (1996) On the evolutionary consequences of niche construction. J Evol Biol 9:293–316

Laland KN, Odling-Smee FJ, Feldman MW (1999) The evolutionary consequences of niche construction and their implications for ecology. Proc Natl Acad Sci 96:10242–10247

Laland KN, Odling-Smee FJ, Feldman MW (2000) Niche construction, biological evolution, and cultural change. Behav Brain Sci 23:131–175

Laland KN, Odling-Smee FJ, Feldman MW (2001) Cultural niche construction and human evolution. J Evol Biol 14:22–23

Laland KN, Kendal JR, Brown GR (2007) The niche construction perspective: implications for evolution and human behavior. J Evol Psychol 5:51–66

Laland KN, Odling-Smee FJ, Feldman MW, Kendal J (2009) Conceptual barriers to progress within evolutionary biology. Found Sci 14:195–216

Laland KN, Odling-Smee FJ, Myles S (2010) How culture shaped the human genome: bringing genetics and the human sciences together. Nat Rev Genet 11:137–148

Lee T, White H, Granger C (1993) Testing for neglected nonlinearity in time series models: a comparison of neural network methods and alternative tests. J Econometrics 56:263–290

Lehmann L (2008) The adaptive dynamics of niche constructing traits in spatially subdivided populations: evolving posthumous extended phenotypes. Evolution 62:549–566

Leonardi M, Gerbault P, Thomas MG, Burger J (2012) The evolution of lactase persistence in Europe: a synthesis of archaeological and genetic evidence. Int Dairy J 22:88–97

Levins R, Lewontin RC (1985) The dialectical biologist. Harvard University Press, Cambridge

Lewontin RC (1983) Gene, organism, and environment. In: Bendall DS (ed) Evolution from molecules to men. Cambridge University Press, Cambridge, pp 273–285

Lewontin RC (2000) The triple helix: gene, organism and environment. Harvard University Press, Cambridge

Lleros C (2005) Path analysis. In: Kempf-Leonard K (ed) Encyclopedia of social measurement (vol. 3). Elsevier, New York, pp 25–30

Lomer MC, Parkes GC, Sanderson JD (2008) Review article: lactose intolerance in clinical practice–myths and realities. Aliment Pharmacol Ther 27:93–103

Mace R (1998) The co-evolution of human fertility and wealth inheritance strategies. Philos Trans R Soc 353:389–397

Mace R (2009) Update to Holden and Mace's "Phylogenetic analysis of the evolution of lactose digestion in adults" (1997): revisiting the coevolution of human cultural and biological diversity. Hum Biol 81:621–624

Manning KM, Downey SS, Colledge S, Conolly J, Stopp B, Dobney K, Shennan S (2014) The origins and spread of stock-keeping: the role of cultural and environmental influences on early Neolithic animal exploitation in Europe. Antiquity 87:1046–1059

Manski CF (2005) Social choice with partial knowledge of treatment response. Princeton University Press, Princeton

Maruyama G (1998) Basics of structural equation modeling. Sage, Thousand Oaks

McCracken RD (1971) Lactase deficiency: an example of dietary evolution. Curr Anthropol 12:479–517

Mulcare CA (2006) The evolution of the lactase persistence phenotype. Ph.D. dissertation, University of London, London

Mulcare CA, Weale ME, Jones AL, Connell B, Zeitlyn D, Tarekegn A, Swallow DM, Bradman N, Thomas MG (2004) The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (LCT) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. Am J Hum Genet 74:1102–1110

Myles S, Bouzekri N, Haverfield E, Cherkaoui M, Dugoujon JM, Ward M (2005) Genetic evidence in support of a shared Eurasian–North African dairying origin. Hum Genet 117:34–42

Nagy D, Tömöry G, Csányi B, Bogácsi-Szabó CA, Priskin K, Bede O, Bartosiewicz L, Downes S, Raskó I (2011) Comparison of lactase persistence polymorphism in ancient and present-day Hungarian populations. Am J Phys Anthropol 145:262–269

Neyman J, Pearson CS (1928) On the use and interpretation of certain test criteria for purposes of statistical inference. Biometrika 20A:175–240

O'Brien MJ, Bentley RA (2015) The role of food storage in human niche construction: an example from Neolithic Europe. Environ Archaeol (in press)

O'Brien M, Laland KN (2012) Genes, culture and agriculture: an example of human niche construction. Curr Anthropol 53:434–470

Odling-Smee FJ (1988) Niche-constructing phenotypes. In: Plotkin HC (ed) The role of behavior in evolution. MIT Press, Cambridge, pp 73–132

Odling-Smee FJ (2010) Niche inheritance. In: Pigliucci M, Müller GB (eds) Evolution: the extended synthesis. MIT Press, Cambridge, pp 175–207

Odling-Smee FJ, Laland KN (2011) Ecological inheritance and cultural inheritance: what are they and how do they differ? Biol Theor 6:220–230

Odling-Smee FJ, Turner JS (2011) Niche construction theory and human architecture. Biol Theor 6:283–289

Odling-Smee FJ, Laland KN, Feldman MW (2003) Niche construction: the neglected process in evolution. Princeton University Press, Princeton

Oyama S, Griffiths PE, Gray RD (2001) Cycles of contingency: developmental systems and evolution. MIT Press, Cambridge

Pearl J (1995) Causal diagrams for empirical research. Biometrika 82:669–710

Pearl J (2009) Causality: models, reasoning, and inference, 2nd edn. Cambridge University Press, New York

Peters J, von den Dreisch A, Helmer D (2005) The upper Euphrates–Tigris basin: cradle of agro-pastoralism? In: Vigne JD, Peters J, Helmer D (eds) The first steps of animal domestication: new archaeozoological approaches. Oxbow, Oxford, pp 96–124

Pinhasi R, Fort J, Ammerman AJ (2005) Tracing the origin and spread of agriculture in Europe. PLoS Biol 3(12):e410

Rendell L, Fogarty L, Laland KN (2011) Runaway cultural niche construction. Philos Trans R Soc B 366:823–835

Riede F (2011) Adaptation and niche construction in human prehistory: a case study from the southern Scandinavian Late Glacial. Philos Trans R Soc B 366:793–808

Rowley-Conwy P, Layton R (2011) Foraging and farming as niche construction: stable and unstable adaptations. Philos Trans R Soc B 366:849–862

Salque M, Bogucki P, Pyzel J, Sobkowiak-Tabaka I, Grygiel R, Szmyt M, Evershed RP (2013) Earliest evidence for cheese making in the sixth millennium BC in northern Europe. Nature 493:522–525

Scheiner SM, Mitchell RJ, Callahan HS (2000) Using path analysis to measure natural selection. J Evol Biol 13:423–433

Schlebusch CM, Sjödin P, Skoglund P, Jakobsson M (2013) Stronger signal of recent selection for lactase persistence in Maasai than in Europeans. Euro J Human Genet 21:550–553

Schwilk DW, Ackerly DD (2001) Flammability and serotiny as strategies: correlated evolution in pines. Oikos 94:326–336

Scott-Phillips TC, Laland KN, Shuker DM, Dickins TE, West SA (2014) The niche construction perspective: a critical appraisal. Evolution 68:1231–1243

Shennan SJ (2000) Population, culture history and the dynamics of culture change. Curr Anthropol 41:811–835

Shennan SJ (2011) Property and wealth inequality as cultural niche construction. Philos Trans R Soc B 366:918–926

Shennan S, Downey SS, Timpson A, Edinborough K, Colledge S, Kerig T, Manning K, Thomas MG (2013) Regional population collapse followed initial agriculture booms in mid-Holocene Europe. Nat Commun 4:2486

Shipley B (2000) Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference. Cambridge University Press, Cambridge

Silver M, Di Paolo E (2006) Spatial effects favour the evolution of niche construction. Theor Popul Biol 70:387–400

Simoons FJ (1970) Primary adult lactose intolerance and the milking habit: a problem in biological and cultural interrelations. 2. A culture historical hypothesis. Am J Dig Dis 15:695–710

Smith BD (2007a) The ultimate ecosystem engineers. Science 315:1797–1798

Smith BD (2007b) Niche construction and the behavioral context of plant and animal domestication. Evol Anthropol 16:188–199

Smith BD (2009) Resource resilience, human niche construction, and the long-term sustainability of pre-Columbian subsistence economies in the Mississippi River valley corridor. J Ethnobiol 29:167–183

Smith BD (2011a) General patterns of niche construction and the management of 'wild' plant and animal resources by small-scale pre-industrial societies. Philos Trans R Soc B 366:836–848

Smith BD (2011b) A cultural niche construction theory of initial domestication. Biol Theory 6:260–271

Smith BD (2015) A comparison of niche construction theory and diet breadth models as frameworks of explanation for the initial domestication of plants and animals. J Archaeol Res 23:215–262

Stehli P (1989) Merzbachtal—Umwelt und Geschichte einer bandkermischen Siedlungskammer. Germania 67:51–76

Timpson A, Colledge S, Crema E, Edinborough K, Kerig T, Manning K, Thomas MG, Shennan S (2014) Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: a new case-study using an improved method. J Archaeol Sci 52: 549–557

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K et al (2006) Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 39:31–40

Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P et al (2001) Genetic evidence for Near-Eastern origins of European cattle. Nature 410:1088–1091

Ulijaszek SJ, Strickland SS (1993) Nutritional anthropology: prospects and perspectives. Smith-Gordon, London

Vigne JD (2008) Zooarchaeological aspects of the Neolithic diet transition in the Near East and Europe, and their putative relationships with the Neolithic demographic transition. In: Bocquet-Appel JP, Bar-Yosef O (eds) The Neolithic demographic transition and its consequences. Springer, Berlin, pp 179–205

Vigne JD, Helmer D (2007) Was milk a "secondary product" in the Old World Neolithisation process? Its role in the domestication of cattle, sheep and goats. Anthropozoologica 42:9–40

Vuong Q (1989) Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57:307–333

White H (1996) Estimation, inference, and specification analysis. Cambridge University Press, Cambridge

Wolfle LM (2003) The introduction of path analysis to the social sciences, and some emergent themes: an annotated bibliography. Struct Equ Model 10:1–34

Wollstonecroft M (2011) Investigating the role of food processing in human evolution: a niche construction approach. Archaeol Anthropol Sci 3:141–150

Wooding SP (2007) Following the herd. Nat Genet 39:7–8

Wright S (1921) Correlation and causation. J Agric Res 20:557–585

Wright S (1934) The method of path coefficients. Ann Math Stat 5:161–215

Wright J (2006) The concept of organisms as ecosystem engineers ten years on: progress, limitations, and challenges. Bioscience 56:203–209

Zeder MA (2012) The broad spectrum revolution at 40: resource diversity, intensification, and an alternative to optimal foraging explanations. J Anthropol Archaeol 31:241–264

Zeder MA (2015) Core concepts in domestication research. Proc Natl Acad Sci 112:3191–3198

Zeder MA, Hesse B (2000) The initial domestication of goats (Capra hircus) in the Zagros Mountains 10,000 years ago. Science 287: 2254–2257