ANCESTRY MATTERS: DESCENT LINE GROWTH AND EXTINCTION*

Xi SONG
University of California, Los Angeles

Cameron D. CAMPBELL
The Hong Kong University of Science and Technology, Shanghai Jiaotong University (visiting)

James Z. LEE
The Hong Kong University of Science and Technology, Shanghai Jiaotong University (visiting)

April 2013

Keywords: descent line; population composition; extinction; multigenerational inequality;

Word count (including text, abstract, footnotes, references): 14,878

* Correspondence to Xi Song, Department of Sociology, University of California, Los Angeles, 264 Haines Hall, Los Angeles, CA 90095-1551, USA; email: songxi@ucla.edu. We are grateful to Shuang Chen, Dwight Davis, Hao Dong, Joseph Ferrie, Margot Jackson, Benjamin Jarvis, Byung-Ho Lee, Robert Mare, Matthew Noellert, Judith Seltzer, Hongbo Wang, Linlan Wang and Yu Xie for their suggestions. Versions of this paper were presented at the Population Association of America Annual Meeting, San Francisco, CA, May 3-5, 2012; the University of Pennsylvania Population Studies Center, Philadelphia, PA, October 8, 2012; the Social Science History Association Annual Meeting, Vancouver, BC, November 1-4, 2012; and the California Institute of Technology Division of the Humanities and Social Sciences, Pasadena, CA, March 1, 2013. While conducting this analysis, the authors received support from NICHD R01 grant HD070985-01. Preparation and documentation of the China Multigenerational Panel Dataset-Liaoning (CMGPD-LN) for public release via ICPSR Data Sharing for Demographic Research (DSDR) was supported by NICHD R01 HD057175-01A1 "Multi-Generation Family and Life History Panel Dataset" with funds from the American Recovery and Reinvestment Act.

ANCESTRY MATTERS: DESCENT LINE GROWTH AND EXTINCTION

ABSTRACT

We analyze representation by descendants in subsequent generations as a new outcome of social stratification. Our key question is whether and how a descent line founder's social status affected the size and growth rate of his descent line in later generations. By analyzing two large prospective, multigenerational demographic databases that provide complete pedigrees for males in more than twenty thousand descent lines in 18th and 19th century China, we show that the status of a descent line founder influences his representation in later generations not only because he has more children, but because his children and their descendants also have more children. Descendants of high status males eventually account for a disproportionately large share of the population. Contrary to the common belief that families in the past maximized their total number of offspring, our results suggest the socioeconomic status of founders has a greater and more enduring effect by minimizing the chances of descent line extinction, not maximizing the growth rate of the surviving descent lines.

INTRODUCTION

We examine how differential reproduction by socioeconomic groups change population composition over time. Specifically, we assess whether the male descendants of men of high status account for a disproportionate share of the population in later generations, and identify the demographic mechanisms that underlie such processes. By itself, this is not a novel topic. Relationships between reproductive differentials and trends in the distribution of traits in a population have long been a concern of population geneticists, who have developed elaborate mathematical models of underlying processes. And yet, in spite of the theoretical importance of socioeconomic differentials in reproduction to the long-term evolution of patterns of stratification and inequality, simulation studies and indirect evidence, data limitations until recently have precluded direct, empirical measurement of the influence of socioeconomic differences in reproductive success on population composition over the long term. The best previous assessments of the long-term implications of socioeconomic differences in demographic behavior and status transmission decompose component outcomes separately and then integrate them back together with a simulation model (Mare 1997; Preston and Campbell 1993; Wachter, Hammel and Laslett 1978) or make creative use of indirect evidence (Clark 2007). In contrast, we take advantage of two recently compiled large prospective multi-generational populations to measure directly the influence of descent line founder's social status on their representation in subsequent generations.

We conceptualize representation by descendants in the population many generations later not just as a demographic outcome of interest to population geneticists and biologists, but as a stratification outcome that should be of general interest to researchers interested in the long-term implications of stratification on patterns of inequality. According to evolutionary theory, the animating force in the competition for status in each generation is the primal desire to be represented in succeeding generations. From the standpoint of a descent line founder, all of the outcomes normally considered in studies of the implications of status, including marriage, fertility, health, mortality, and the transmission of status, are simply the means to an end: representation, or overrepresentation, by descendants in later generations. By focusing on representation in later generations instead of the various demographic and social outcomes that interact to determine it, we introduce a new perspective that emphasizes the total implication of social and economic status, not the separate, partial implications.

Understanding the determinants of representation in later generations is also crucial because of the role of kin networks in shaping the demographic behavior and socioeconomic attainment of individuals, especially in historical societies. Descent line size is not only a stratification outcome for the descent line founder, but a stratifying variable for their descendants.[1] For example, Campbell and Lee (2008ab, 2011) show that in historical China, characteristics of kin networks, including the number of kin and their socioeconomic status, influence individual social and demographic outcomes. Such research treats the characteristics of kin networks as exogenous, and interprets associations that persist after the introduction of controls for individual and parental characteristics as evidence of a causal influence of kin network characteristics on individual outcomes. A finding that socioeconomic status influenced numbers of living descendants many generations later would challenge the common assumption that kin network characteristics are exogenous by showing that the composition of an individual's kin network was influenced by the same characteristics of a possibly distant ancestor that influenced his or her parents' characteristics by more traditional mechanisms of intergenerational transmission of status, Markovian or not.

By analyzing two large prospective, multigenerational demographic databases that describe populations at opposite ends of the Qing (1644-1911) social spectrum, the Qing Imperial Lineage who lived largely in Beijing and populations of farmers organized under the Qing Imperial Household who lived in the northeast Chinese province of Liaoning (Lee, Campbell and Wang 1993; Lee, Campbell and Chen 2010; Wang, Lee and Campbell 2010), we show that the status of a descent line founder influences his representation in later generations not only because he has more children, but because his children and their descendants also have more children. In other words, descent lines with high status founders not only have higher growth rates in the initial generation, but also in later generations. Descendants of high status males therefore eventually account for a disproportionately large share of the population. Contrary to the common belief that families in the past maximized their total number of offspring, our results suggest that the effect of founders' status on descent line size and growth rate operated by minimizing the chances of descent line extinction. The socioeconomic status of founders has a greater and more enduring effect on the chances of descent lines becoming extinct than it does on their growth rates.

Although many demographers and historians discuss the implication of differentials in demographic behavior for long-term social and population change, they rely on projections or simulations from two-generation models, or indirect or fragmentary evidence such as changes in the frequency of rare surnames associated with high status founders (Bongaarts, Burch and Wachter 1987; Clark 2007; Mare 1997; Preston and Campbell 1993; Wachter et al. 1978). To our knowledge, this study is one of the first to use high-quality multi-generational historical data to trace specific descent lines over multiple generations.[2] Since our data are prospective and the populations are closed, they are free of the survivor bias that confronts studies based on retrospective data. They allow us to follow not only those descent lines that flourish, but those descent lines that become extinct.[3] Even though these two populations are not representative of all historical populations or even historical Chinese populations in a formal, statistical sense, their positions at nearly opposite ends of the social spectrum make it reasonable to suggest that observed similarities are indicative of basic processes common to historical Chinese populations, and perhaps other historical populations as well.

Through this analysis, we extend the study of social stratification from the two-generation parent-child dyad to the multigenerational descent line. We apply regression-based methods to relate descent line size and growth rate many generations later to founder's characteristics. We distinguish between several possible mechanisms that could underpin a link between founder's socioeconomic status and subsequent demographic growth. This new view suggests that social inequality in one generation has implications not only for the composition of the next generation, but for all subsequent generations.

BACKGROUND

Research on long-term population dynamics and intergenerational social mobility rarely intersect. Only a few empirical studies (e.g., Lam 1986; Mare 1997; Mare and Maralani 2006; Matras 1961; Musick and Mare 2004; Preston 1974; Preston and Campbell 1993) examine how social inequality and demographic differentials interact to shape long-term processes of population renewal. They mostly model long-term change in demographic composition based on assumptions about fertility and mortality differentials (e.g., Keyfitz 1968; Lee and Tuljapurkar 1994; Pollard 1973; Preston, Heuveline and Guillot 2001; Rogers 1975; Schoen 1988; Stoto 1983). Studies of intergenerational social mobility typically examine transmission of social status from parents to offspring independently of mortality and fertility patterns at the population level

(Blau and Duncan 1967; Duncan 1966; Erikson and Goldthorpe 1992; Featherman and Hauser 1978; Grusky and Fukumoto 1989; Hauser et al. 1975; Hout 1983; Long and Ferrie forthcoming).

Intergenerational transmission of inequality interacts with population composition because in most societies, socioeconomic status is transmitted from parents to offspring, and simultaneously influences survival and reproduction chances. At the individual level, the social resources and inherited characteristics associated with social position influence marriage, mortality and fertility chances in the current and next generation (Campbell and Lee 2010b). Conversely, demographic behaviors and outcomes such as sibship size, assortative mating, and composition of kin network, maintain and modify life chances of social mobility (Campbell and Lee 2008a; Guo and VanWey 1999; Hauser and Sewell 1985; Mare and Maralani 2006). At the family level, individuals inherit not only their parents' genetic traits, but perhaps more importantly, their parents' attitudes, preferences, aspirations, and knowledge, as well as specific social and economic advantages or disadvantages.

The implications of status for reproductive differentials depend on the strategies that families apply to manage their reproduction. In our empirical analysis, we assess the relative importance of two possible strategies. On the one hand, families may seek to maximize their representation in the next generation by having as many children as possible. This is analogous to what ecologists call an *r* strategy, and was until recently the common, implicit assumption about historical Chinese reproductive behavior (Lee and Wang 1999; Wolf 1995; Wolf and Huang 1980; Chuang et al. 2006). On the other hand, families may seek trade-offs between the quantity and quality of children, following what ecologists call a *K* strategy. While this may result in fewer births in the short term, it might increase descent line survival chances over the long term because even though there are fewer children, they are each more likely to survive to adulthood, marry, and reproduce. Recent studies of fertility in historical China suggest that families exercised control over the pacing and number of their births (Campbell and Lee 2010a; Wang et al. 1995). Given that families appear to have approached reproduction in a calculated and deliberate fashion, they very well may have been capable of pursuing even more complex strategies.

There is an urgent need for direct, empirical investigation of the implications of reproductive differentials for long-term population composition, because results from the application of mathematical or simulation models to two-generation data are heavily dependent on assumptions included in the model. For example, simplistic models like the ones popular with eugenicists at the beginning of the 20th century suggest that persistent high rates of fertility for individuals with desirable or undesirable traits along with intergenerational transmission of these traits should lead the population to be dominated by individuals with those traits (Herrnstein and Murray 1994). However, if there is intergenerational mobility, so that offspring do not automatically inherit the exact same set of traits as their parents, the equilibrium population composition may include individuals with a variety of traits (Sibley 1942; Mare 1997; Preston 1974). More sophisticated models may yield even more complex predictions for the implications of status differences in reproduction (Preston and Campbell 1993). Results could conceivably differ according to the assumptions made in the model about status inheritance processes and determinants of demographic differentials.

Systematic efforts to study the long-term growth and extinction of descent lines date back at least to Francis Galton and H.W. Watson's (1874) application of branching theory to study aristocratic surnames. Lotka (1929, 1931, 1941) subsequently estimated family extinction probabilities for male descent lines in the United States white population in 1920. Wachter and

Laslett (1978) analyzed the long-term influences of such demographic behaviors as marriage and reproduction on the extinction of British elite patrilines. Recent studies in population genetics move beyond extinction to examine descent line growth and eventual dominance. Some studies find that ten lineages account for greater than 95% of the 1,007 European Y chromosomes studied (Semino et al. 2000). Other studies, using micro-simulations, show that all humanity may share exactly the same set of genealogical ancestors (Chang 1999; Lachance 2009; Murphy 2004; Rohde, Olson and Chang 2004), implying that over the long-term, the overwhelming majority of descent lines become extinct, while a small group of descent lines become dominant. However, a lack of appropriate prospective, multi-generational data has until recently precluded direct, empirical investigations of descent line extinction and growth processes in human populations (Yasuda et al. 1974).

Changes over time in the distribution of surnames provide more direct evidence of processes of descent line growth and extinction (Matsen and Evans 2008). In societies experiencing little immigration, surname distributions become highly concentrated and the distribution of their sizes highly skewed (Cavalli-Sforza 1969; Cavalli-Sforza et al. 2004; Piazza et al. 1987; Colantonio et al. 2003; Yasuda et al. 1974). For example, almost 40% of the population in Vietnam has the same family name (i.e., Nguyễn) and 90% share the 15 most common family names. Similarly, in Korea 45% of the population shares three family names (Kim, Lee and Park). In Mexico, almost 40% of the population shares five family names (Hernández, García, Martínez, González and López).[4] In China, less than 5% of the surnames in use account for 85% of the population (Du et al. 1992; Colantonio et al. 2003).

Surname studies by population geneticists suggest that even if the evolution of surnames has no association with traits that are transmissible and related to reproduction, random drift will still generate inequality in descent line size that has nothing to do with founder's characteristics (Lande 1976; Masel 2011; Slatkin 1977). The Wright-Fisher model predicts that, over time, some genes, surnames, or other highly heritable traits will inevitably disappear because the gene pool or surname distribution of each new generation is a random sample from the last generation (Wright 1929). Naturally, if heritable traits affect reproductive differentials, groups that have favorable distributions of those traits in a founder generation should have more descendants than would be predicted solely by random drift.

One recent empirical study relates inequality in the long-term growth rates of descent lines to the social status of descent line founders, and suggests that such mechanisms contribution to long-term social and demographic change. Clark (2010) studies how changes in the frequency of unusual surnames in England from the 13th century to the present are associated with the economic attainment of the founders in the earliest generations. Clark concludes that founders who had high economic status had more descendants in later generations. He suggests that founders transmitted the traits that made them successful to their offspring, which in turn increased their chances of attaining status, reproducing, and transmitting these traits to descendants in the next generation. Through this vertical process of diffusion, traits conducive to economic success became widespread. Provocatively, Clark (2010) also suggests that these processes were unique to Britain, and the diffusion of traits conducive to success in a market economy help account for why the Industrial Revolution started in Britain, not elsewhere. In a related vein, Unz (2013) suggests that key features of modern Chinese society that contribute to its recent economic success are a legacy of centuries of selection for relevant traits conducive to success in a sophisticated but resource-constrained agrarian economy.

Long-term implications of interactions between status transmission and status differentials in reproduction depend on whether underlying multi-generational processes are Markovian or non-Markovian, transient or cumulative, and time-invariant or variable (Bartholomew 1982; Boudon 1973; Duncan 1966; Hodge 1966; Mare 2011; McGinnis 1968; Singer and Spilerman 1976). If status transmission is Markovian, so that founder's status only influences the number and status of his or her own offspring, founder's status will have an immediate, multiplicative effect on the share of the population accounted for by his or her descent line, but will not have any lasting impact on the growth rate of his or her descent line. If status transmission is non-Markovian, and founder's status affects not only the number and status of his or her own children, but also the number and status of their descendants' offspring in later generations, then founder's status could have a long-term effect on descent line's growth rate. To account for the possibility of non-Markovian processes, the next section draws on cumulative advantage theory to propose three different mechanisms for the influence of founders' characteristics on descent line growth.

Understanding the determinants of long-term descent line growth is of additional substantive importance because of the role that characteristics of kin networks play in shaping the social and demographic outcomes of individuals. Studies of kin or household contextual influences on individual outcomes mostly treat the size, social composition, and other characteristics of the kin network as exogenous. Associations with individual outcomes are mostly assumed to be causal. While some studies apply fixed- or random-effects models to account for the possibility of correlated demographic or social outcomes among kin due to shared but unobserved environmental or other factors (Campbell and Lee 2008ab; Sear et al. 2002, 2003), none address the possibility that among the right-hand side variables, kin network characteristics are actually endogenous to individual and parental characteristics in the sense that they are jointly determined by the characteristics of distant ancestors. A finding that founder characteristics shape the size, growth rate and other characteristics of a descent line many generations later would indicate a need to address this issue.

THEORETICAL FRAMEWORK

To assess effects of founder's characteristics on descent line growth, we begin by considering the classic population growth equation (Preston et al. 2001:11):

$$N(T) = N(0)e^{\int_0^T r(t)\mathrm{d}t} \qquad (1)$$

Here, $N(0)$ and $N(T)$ refer to descent line size at times 0 and T respectively. $r(t)$ is the instantaneous growth rate at time t. In this one-sex model, each descent line has one male founder, and $N(0)$ is defined to be the number of that founder's male offspring. The effects of descent line founders' characteristics on the total number of descendants $N(T)$ can work through either the initial reproduction of the founders, i.e., $N(0)$, the growth rate of the descent line over time, $r(t)$, or both. We identify three compatible mechanisms—Initial Advantage, Accelerated Advantage, and Advantage Dissipation —to account for changes over time in the size and growth rate of the descent line (see Table 1). We also summarize potential implications of the three mechanisms for the growth trajectory of descent lines in Figure 1.

Table 1

Figure 1

In Table 1, we assume that Initial Advantage affects descent line growth entirely through the reproduction of the founders, namely, $N(0)$ in Equation (1). In the Initial Advantage scenario, represented in Figure 1 with a grey dashed line, high fertility on the part of the founder multiplies the initial size of the descent line $N(0)$ by c but has no effect on the growth rate $r(t)$ in later generations. In other words, founder's characteristics have an effect on $r(t)$ that is transitory and doesn't last beyond the first generation. The number of high-origin descendants will always be c times the number of low-origin descendants. At the population level, this mechanism implies that high-origin descent line's share increases immediately in the generation after the founder, and then remains stable afterward.

Accelerated Advantage and Advantage Dissipation allow for changes in the growth rate $r(t)$. In the Accelerated Advantage scenario, represented as the black dashed line in Figure 1, founder's characteristics trigger a permanent increase in $r(t)$. Because the growth rate of the descent line experiences a permanent increase, the descent line accounts for a steadily increasing share of the population. Eventually, high-origin descendants dominate the population, and the share of low-origin descendants declines to insignificance.

Advantage Dissipation is an intermediate process between the Initial Advantage and Accelerated Advantage mechanisms. It is represented by the dotted line in Figure 1. It assumes that the effect of founder's characteristics on the growth rate $r(t)$ after the first generation is neither null as assumed by the Initial Advantage mechanism nor permanent as assumed by the Accelerated Advantage mechanism. Instead, high-origin descent lines will experience a higher growth rate for some number of generations, but their growth rate will decline over time and converge with that of low-origin descent lines. The share of the population accounted for by high-origin descent lines will grow until the effect of founder's characteristics on the growth rate $r(t)$ dissipates.

Of course, the actual impact of founder's characteristics on descent line growth may include a combination of these three mechanisms. For example, if Initial Advantage and

Advantage Dissipation both operate, we would expect to observe higher reproduction for the founders followed by an increase in the ratio of high-origin descendants over low-origin descendants for the next few generations.[5]

Initial Advantage

The sociological literature on stratification frequently invokes Initial Advantage as part of cumulative advantage theory to explain the evolution of inequality over time. In this conception, the advantage of one group over another depends on *initial* positions, and the subsequent growth of inequality is path-dependent. Initially minor and possibly random disparities may widen over time because success begets success. Empirical studies have tested this theory by making use of evidence from a variety of areas, including academic publication records (Allison et al. 1982; Merton 1968, 1988), cognitive development (Guo 1998), and health (Pampel and Rogers 2004; Dannefer 2003; Ross and Wu 1996). Merton (1988) describes this phenomenon as, "the ways in which initial comparative advantage of trained capacity, structural location, and available resources make for successive increments of advantage such that the gaps between the haves and the have-nots…widen" (p.606). For Merton, Initial Advantage is the essential characteristic of a cumulative advantage process. From this perspective, any exogenous events that generate an initial advantage can have long-term consequences on patterns of inequality (DiPrete and Eirich 2006).

For our outcome of interest, descent line size in later generations, the Initial Advantage mechanism is analogous to what the genetics literature refers to as 'founder effects' (Falconer 1960). In the first generation of a hypothetical population, assume that a high-status male fathers three children and a low-status male fathers only two children. In later generations, there are no reproductive differentials between the two descent lines. Each male fathers two children, regardless of which of the two lines he belongs. The high-status founder would have 3 descendants in generation 1, 6 in generation 2, 12 in generation 3 and so on. For the low-status founder, the corresponding generation sizes would be 2, 4, 8, and so on. Over time, the arithmetic difference in the sizes of two descent lines grows, but the ratio of their sizes remains constant (the grey dashed line in Figure 1). In this scenario, differences in lineage size according to the status of founders reflects a path-dependent process driven by differences in the reproduction of the founders themselves, even if there are no differences in reproduction and social status among their descendants.

Accelerated Advantage

Disparities in descent line size according to the socioeconomic status of founders may continue to widen after the first generation if the founder's status affects the reproduction of members of later generations, whether directly or indirectly. We refer to this scenario as Accelerated Advantage, in which 'accelerated' refers to the possibility that high status origin descent lines continue to grow faster than other descent lines. Our inspiration is Allison et al.'s (1982) observation that cumulative advantage does not produce additional changes in patterns of inequality later in time unless the rate of accumulation continues to vary between population subgroups.

The relevance of Accelerated Advantage depends on whether or not a founder transmits traits to his offspring that affect their reproduction. For example, if a high status founder not only has more children, but also in turn transmits high status to these children, the share of the population accounted for by the descent lines of high status founders will expand steadily over time. Imagine a modification of the example in our introduction of Initial Advantage such that not only the high-status founder but also his descendants have three children. The number of descendants in the next four generations for the high-status founder would be 3, 9, 27, and then 81. If the descendants of the low-status founder only had two children each, their lineage size in the next four generations would be 2, 4, 8, and then 16. The ratio of the sizes of the high and low origin lineages would steadily increase, until the high status lineage accounts for nearly the entire population (the black dashed line in Figure 1). By contrast, in Initial Advantage, the ratio of descent line sizes is constant after the second generation.

Advantage Dissipation

Advantage Dissipation addresses the reality that founder characteristics are unlikely to trigger the permanent increases in growth rates assumed in the *Accelerated Advantage* scenario. Due to limitations in the carrying capacity of the system, growth rate differentials may narrow as advantages become increasingly difficult to sustain (DiPrete and Eirich 2006). More concretely, in the multigenerational process of descent line growth, resource dissipation and downward social mobility may serve as a 'brake' that eventually attenuates differentials and prevents inequality in descent line size from growing indefinitely. Thus the dotted line in Figure 1 that represents Advantage Dissipation assumes that when the status and demographic behavior of descendants of high status founders finally become indistinguishable from the rest of the population, relative sizes of descent lines stabilize.

Previous literature that draws attention to the phenomenon of "regression to the mean" in family advantage suggests the likely importance of advantage dissipation in understanding the influence of founder's status on subsequent descent line size. Becker (1991: 273) argues that "almost all earnings advantages and disadvantages of ancestors are wiped out in three generations. Poverty would not seem to be a 'culture' that persistent for several generations". Similarly, centuries before Becker, a common folk expression in China was that "wealth doesn't last for three generations" ("富不过三代"). In an empirical study based on genealogies, biographies and local histories, Ho (1964) found that even under extremely lenient standards for social status the average descent line fell into complete oblivion, or at least mediocrity, in some eight generations. All these observations are in line with Galton's early conceptualization of "regression toward mediocrity" (Zimmerman 1992).

Our data and methods allow us to discern the relative importance of these three mechanisms. Prior sociological theory suggests that Initial Advantage, Accelerated Advantage, and Advantage Dissipation alone or in combination will generate different patterns of results in empirical studies (DiPrete and Eirich 2006). Because we can measure the influence of founder's characteristics on descent line size and growth rate in each generation, we can assess the relative importance of the three processes. Our approach therefore advances on the indirect one applied by Clark (2010) in his analysis of the long-term prevalence of rare surnames in Britain from the 13th century to the present. Clark's (2010) reliance on inference from surname distributions at specific points in time and fragmentary evidence on status transmission and demographic differentials preclude a formal analysis of underlying processes like the one in this study.

ANALYTIC APPROACH

Modeling Stock and Flow

To distinguish the roles of Initial Advantage, Accelerated Advantage, and Advantage Dissipation in explaining the effect of founder's status on subsequent descent line growth, we apply a stock-flow analytic framework. We use the *stock* of a descent line to refer to the number of descendants alive at time *t* and the *flow* of a descent line to refer to the instantaneous rate of change in the number of descendants at time *t*. In the discrete-time model that we adopt for our empirical analysis, flow refers to growth from time *t*-1 to *t*. We assess the relative importance of the three proposed mechanisms by estimation four sets of regression models. Each set makes different assumptions about the relationships of stocks and flows to the right-hand side variables: linear, exponential, Poisson, and negative binomial. The linear stock and flow models shown below assume that the number of descendants at time *t* and the changing number of descendants from time *t*-1 to *t* are linear in founder's characteristics *X*.

$$N_i(t) = \boldsymbol{X}_i\beta + \varepsilon_i$$

$$N_i(t) - N_i(t-1) = \boldsymbol{X}_i\beta + \varepsilon_i$$

The exponential stock and flow models assume an exponential relationship of founder's characteristics *X* to the number of descendants at time *t* and the change in the number of descendants from time *t*-1 to *t*:

$$\log(N_i(t)) = \boldsymbol{X}_i\beta + \varepsilon_i$$

$$\log(N_i(t) - N_i(t-1)) = \boldsymbol{X}_i\beta + \varepsilon_i$$

In an early model of descent line growth, Fisher (1922) applied the Poisson distribution and assumed that the probability of extinction is determined by the average number of offspring per individual. Following this tradition, our third set of models assumes that the number of descendants at time *t* for the *i*th descent line, $N_i(t)$, follows a Poisson distribution with parameter $\mu > 0$.

$$\mathrm{P}(N_i(t)|\mu_{it}) = \frac{\exp(-\mu_{it})\mu_{it}{}^{N_i(t)}}{N_i(t)!} \qquad (2)$$

In equation (2), μ_{it} is the expected number of descendants at time *t*. The assumption of the Poisson distribution also requires that the expected number of descendants equals the variance of the number of descendants, ∀ i: $E(N_i(t)) = Var(N_i(t))$.

The Poisson stock model assumes that the conditional mean of the number of descendants at time t is a function of independent variables **X's** that describe time-invariant founder's characteristics:

$$\mu_{it} = E(N_i(t)|\boldsymbol{X_i}) = \exp(\boldsymbol{X}_i\beta) \tag{3}$$

The Poisson flow model represented in (4) is similar except that it introduces a control for the count of descendants at time t-1.

$$\mu_{it} = E(N_i(t)|\boldsymbol{X_i}, N_i(t-1)) = \exp(\log(N_i(t-1)) + \boldsymbol{X}_i\gamma)^6 \tag{4}$$

Equivalently,

$$E\left(\frac{N_i(t)}{N_i(t-1)}\middle|\boldsymbol{X_i}, N_i(t-1)\right) = \exp(\boldsymbol{X}_i\gamma) \tag{5}$$

To account for the possibility that the Poisson model's assumption of equality of the variance and mean of the number of descendants is violated because of over-dispersion, we also estimate a fourth set of models: negative binomial regression models. In that situation, the relationship between founder's characteristics and the expected number of descendants still follow equations (3) and (4). However, the variance in the number of descendants is assumed to follow a gamma distribution with a parameter that is estimated separately.

In our analysis, we use 'stock effect' to refer to the ratio of expected mean number of descendants of high-status founders to estimated expected mean number of descendants from low-status founders. Along these lines, we use 'flow effect' to refer to the ratio of the expected growth rate of descent lines with high-status founders to the expected growth rate of descent lines with low-status founders. In either case, if the ratio is 1, the founder's characteristics do not influence descent line size (stock) or growth rate (flow).

Initial Advantage, Accelerated Advantage, and Advantage Dissipation predict different patterns of effects of founder's characteristics on stock and flow. If Initial Advantage is present, we expect to observe an effect on stock in every generation. However, we only expect to see an effect on flow in the first generation. This is because the growth rate of the descent line is elevated only in the founding generation, and then reverts to be the same as other descent lines in later generations. If Accelerated Advantage is present, both the stock effect and the flow effect should be apparent in every generation, because high status not only increases the total number of the descendants but also their growth rate. If Advantage Dissipation is present, the stock effect may keep increasing until the flow effect disappears, because founder's effect on growth rate fades away with time.

Modeling Extinction and Growth

Previous studies of descent line growth and decline (e.g. Campbell and Lee 2010b) assume that the probability of extinction is a byproduct of the distribution of the number of offspring. For example, the Poisson model introduced above requires that the proportion of observed zeroes (extinctions) in the empirical data matches the proportion of zeroes predicted by the Poisson distribution. This implies an assumption that the same underlying mechanism accounts for the influence of founder's status on the probability of extinction, and conditional on avoiding extinction, the probabilities of having different numbers of descendants.

However, there are strong reasons to expect that in each generation, the factors that influence a descent line's probability of extinction are different from the ones that influence its growth rate, conditional on avoiding extinction. Most importantly, descent lines face hurdles in each generation that must be overcome for them to have any offspring at all, or if any of their members marry, for any of them to have children who survive to the next generation.

The most obvious of these hurdles is marriage. At least one man in each generation must marry for there to be any hope of the descent line being represented in the next generation. Empirically, the determinants of marriage differed substantially in the past from the determinants of the fertility of married men. In historical China, the social and economic determinants of male marriage differed from the determinants of fertility. Most importantly, father's and own social and economic status had stronger and more consistent effects on a male's chances of marriage than they did on his fertility within marriage (Campbell and Lee 2008b). Kinship was much more important for marriage chances than it was for fertility within marriage: correlations in marriage chances among men in the same descent line were much stronger than the correlations in their marital fertility (Campbell and Lee 2011). Because of the correlation in marriage chances among male kin, the proportion of descent lines in which nobody in a generation married should have been higher than it would be if the marriage chances of related males were independent.

Descent lines may have sought to minimize the chances of extinction, not maximize the expected number of descendants in the next generation. As we noted earlier, families may have pursued a variant of what population biologists refer to as a *K* strategy, not an *r* strategy. Instead of maximizing the total number of sons and investing relatively little in each one, they may have traded quantity for quality, having fewer sons, but investing more in each son to maximize the chances that at least one would survive to adulthood, marry, and reproduce.

To allow founder's characteristics to have separate effects on extinction probabilities, and conditional on avoiding extinction, the growth rate, we introduce a mixture Poisson distribution that models processes of extinction and growth jointly (Johnson, Kemp and Kotz 2005). Suppose that π and $1-\pi$ are the probabilities of failure and success for overcoming a 'hurdle' that conditions success at reproduction, and thereby avoiding extinction. For example, π might be the probability that no sons survived to adulthood and married, and $1-\pi$ the probability that at least one son survived and married. Again, the assumption is that these are independent of the distribution assumed for the number of offspring for those who did marry. Let the probability of having *j* descendants in a truncated distribution be written as p_j. Then we have

$$\mathrm{P}[N_i(t)=0|Z_i]=\pi, \tag{6}$$

$$\mathrm{P}[N_i(t)=j|\boldsymbol{X}_i,\ N_i(t)>0]=\tfrac{(1-\pi)p_j}{1-p_0},\quad i,j=1,\ 2,\dots$$

where $\mathrm{P}[N_i(t)=j]$ is the probability that the number of descendants for the *i*th descent line at time *t* is *j*. $\boldsymbol{Z}$ is the set of covariates to explain extinction and $\boldsymbol{X}$ is the set of covariates to explain descent line growth.

For the sake of simplicity, we use a logistic model to predict $\mathrm{P}[N(t)=0]$ and assume the truncated part $\mathrm{P}[N(t)=j\mid N(t)>0]$ still follows a Poisson distribution. Then

$$\mathrm{P}[N_i(t)=0|\boldsymbol{Z}_i]=\frac{1}{1+\exp(\boldsymbol{Z}_i\gamma)}$$

$$\mathrm{P}[N_i(t)=j|\boldsymbol{X}_i,N_i(t)>0]=\frac{(1-\mathrm{P}[N_i(t)=0|\boldsymbol{Z}])\exp(-\mu_{it}){\mu_{it}}^{N_i(t)}}{N_i(t)!\,[1-\exp(-\mu_{it})]},\qquad i,j=1,2,\dots \tag{7}$$

The mixture Poisson flow and stock models handle extinct descent lines differently. The stock models treated extinction as an absorbing state. Once a descent line is extinct, it is included as a zero in the models for the remaining time periods. In the flow model, we estimate the probability of extinction from time *t*-1 to *t*. That is, given that the total number of descendants at *t*-1 is greater than 0, we estimate the probability of extinction by time *t*. Since the flow model includes descent line size at time *t*-1 as an offset or control, it accounts for the possibility that a higher probability of extinction might simply be an artifact of smaller descent line size 25 years ago.

DATA AND MEASURES

Data

Our data are derived from the China Multi-Generational Panel Dataset-Imperial Lineage (CMGPD-IL, 1652-1936) and the China Multi-Generational Panel Dataset-Liaoning (CMGPD-LN, 1749-1909). The CMGPD-IL records 83,256 males in the Aixin Jueluo imperial lineage. The lineage originated in northeast China and founded the Qing dynasty (1644-1912).[7] Specifically, the CMGPD-IL records Takeshi (塔克世, born around 1550), his four brothers, and their male descendants down through 1936 (Lee et al. 1993; Wang et al. 2010). Takeshi's grandson, Huangtaiji (皇太极), established the Qing dynasty and became its first emperor in 1644. At the beginning of the Qing Dynasty, the imperial lineage was a small, very elite group. Many of its members held official positions or noble titles. As the lineage grew, a steadily larger proportion of men in the lineage was very distant relatives of the emperors and had neither official position nor noble title. All the original data were recorded prospectively by the Qing Office of the Imperial Lineage and have been well preserved in the national First Historical Archive in Beijing and the Liaoning Provincial Archive. [8] In contrast with traditional Chinese family genealogies that are constructed retrospectively, and tend to omit low status, never-married, or infecund ancestors (Campbell and Lee 2002; Harrell 1987; Telford 1990), the CMGPD-IL is unusually complete. The prospective design guarantees minimal loss to follow up. Published studies of fertility (Campbell and Lee 2010a; Wang et al. 1995), mortality (Lee et al. 1994), and other outcomes have already established the suitability of the CMGPD-IL as a source for the analysis here.

Our other analytical sample, the China Multigenerational Panel Dataset-Liaoning (CMGPD-LN), is derived from triennial household registers of farming populations who produced for the Qing Imperial Household Office (Lee et al. 2010). Because the suitability of the CMGPD-LN for the analysis here is already established in published studies (Campbell and Lee 2008ab, 2011; Lee and Campbell 1997), and the complete dataset and documentation are now public and available for download at the Interuniversity Consortium for Political and Social Research, our introduction is brief. [9] Whereas the imperial lineage is an elite urban population concentrated in Beijing and Shenyang, the CMGPD-LN covers a large, rural population spread over a very large area in what is now Liaoning province in northeast China. The farmers covered by the CMGPD-LN were descended from Han-Chinese immigrants who migrated from Shandong and other locations into Liaoning in the late 17th and early 18th century. The data consist of 29 sets of triennial household registers with 1.5 million observations of more than 260,000 unique individuals between 1749 and 1909 (Lee et al. 2010).

The Liaoning household registers provide far more comprehensive and accurate demographic and sociological data than other available household registers for China before the twentieth century (Lee and Campbell 1997; Lee et al. 2010). The format and organization of the data closely resemble a linked triennial census. Entries in each register were grouped first by village, then by household group and then by household. In contrast with most historical censuses, the triennial registers allow for linkage of the records of an individual in successive registers. Given that the population is closed, in the sense that the registers follow families that moved from one village to another within the region, the registers are uniquely suitable for prospective reconstruction of descent lines through intergenerational record linkage, and prospective study of the predictors of descent line growth, decline, and extinction (Campbell and Lee 2010b).

Measures

In both the CMGPD-IL and the CMGPD-LN, we specify founders, and then reconstruct their descent lines through straightforward automated record linkage. To create our sample of descent lines in the CMGPD-IL, we define *descent line founders* as all men born between 1675 and 1725, who survived at least to age 25. In specifying a definition of founders to use in the construction of the analytic samples, there is a tradeoff between the numbers of founders and the number of generations for which they can be followed. The definition we apply yields 3,314 founders whose descent lines were followed for the next 150 years, or about six generations. We also experimented with alternative definitions for descent line founders and the results are consistent with the ones reported here. Similarly, in the CMGPD-LN we define descent line founders to be all men born between 1715 and 1765[10]. We treat the year of the founder's birth as the founding year of the descent line, and track each descent line for 125 years, or about five generations.

For our basic outcome variable, descent line size, we count the numbers of living male biological descendants of each founder every 25 years. To do this in the CMGPD-IL, we transform it from a person file in which each male was represented by one record into a person-year file in which each record was an observation describing a male in a specific calendar year. For the CMGPD-LN, we transform the original triennial observations into a genealogical file like the CMGPD-IL in which each entry recorded one person, including years of birth and last observation, and then used that entry to produce person-year files like the one for the CMGPD-IL in which each record described a male in a specific calendar year. We attach founder identifiers to the resulting person-year records, and then count up the numbers of male descendants at 25-year intervals. At time zero, defined as the founder's year of birth, descent line size was always zero. 25 years later, it would include any surviving sons of the founder. The descent line size could be still be zero if the founder had not yet had any sons. 50 years later, the count would include any living sons and grandsons.[11]

We construct dichotomous variables to indicate whether descent line founder's status was high or low. For the CMGPD-IL, high-status origin descent lines are defined to consist of the ones founded by men who worked in the Qing bureaucracy (e.g., ministers, military generals) or held bestowed or inherited noble titles (e.g., princes and dukes). The low-status origin descent lines refer to the ones founded by men who did not have any such distinguished status. For the CMGPD-LN, we defined the high status founders to include holders of salaried official positions, examination titles, purchased and honorary titles, and unsalaried heads of household groups (*zu zhang*). These constituted the local elite (Campbell and Lee 2010b). The high-status founders in the CMGPD-LN who held salaried positions typically had much more mundane ones than their counterparts in the CMGPD-IL, most commonly soldier, scribe, or artisan. We include men who held honorary and purchased titles because they indicate substantial personal or family resources. *Zu zhang* were heads of household groups. Though *zu zhang* held the lowest ranked position in the administrative hierarchy and were unsalaried, they were ostensibly selected by the households themselves on the basis of ability, thus selection to serve as a *zu zhang* may have reflected advantage in the form of talent or possession of useful skills.

We also include control variables to distinguish the lines according to their administrative status. For the CMGPD-IL, we used a flag to distinguish between the members of the Main Line and Collateral Line. Men descended from Takeshi are the *Zongshi* or Main Line. Men descended from Takeshi's brothers are the *Jueluo,* or Collateral Line (Wang et al. 2010). We control for membership in the Main or Collateral Lines because men in the former were accorded more privileges than men in the latter. Similarly, we include a control variable to divide the lines in the CMGPD-LN into two status groups according to which register series they were recorded in: regular populations, most of whom were hereditary tenants farming state-owned land, and specialized populations, which provided designated services to the state such as collecting honey, raising bees, fishing, picking cotton, and tanning and dyeing. These specialized populations had a lower status than the regular populations (Lee et al. 2010).

We include seniority among siblings and total number of male siblings as additional controls in the regression analysis of effects' of founder's status on stock and flow. Founder's total number of siblings is intended to account for the tendency in historical China for men with more male siblings to have higher chances of both marriage and attainment (Campbell and Lee 2008b). With this control, we hope to account for the possibility that unmeasured characteristics simultaneously affected a founder's chances of attaining high status, and his chances of having more offspring. The control for whether or not a founder was an eldest brother is intended to account for advantages that first sons had in terms of attainment and marriage.

RESULTS

Descriptive Statistics

We begin with descriptive statistics that provide a general picture of descent line dynamics in our data. Figure 2 presents the numbers of men in the CMGPD-IL according to whether they are in high- or low-origin descent lines. From 1725 to 1875, the descent lines with high-status founders experienced rapid growth. By contrast, growth in the numbers of males in low-status descent lines was negligible. As a result of this disparity, even though men in high-status lines originally accounted for less than one third of the population, by the end of the period they accounted for more than one half of the population. This trend lends some preliminary support to the Initial Advantage and Advantage Dissipation mechanisms illustrated in Table 1 and Figure 1. The growth of the population is also accompanied by the growth in cumulative probability of extinction, shown in Figure 3. After 150 years, only half of high-status origin descent lines are extinct, in contrast with nearly three-quarters of low status origin descent lines.

Figure 2

Figure 3

Table 2

Descriptive statistics in Table 2 describe basic differences between the high- and low-status descent lines in the two samples. Of the 3,314 descent lines in in the CMGPD-IL, 1,082 are of high-status origin and 2,232 are of low-status origin. High-status founders are more likely to come from larger families and to be affiliated with the Main Line (*zongshi*) of the imperial lineage.[12] At every subsequent point in time, the average sizes of high-origin descent lines are larger than those of low-origin descent lines. In the CMGPD-LN sample, founders tend to be more similar on the distributions of the control variables. Once again, however, the average sizes of high-origin descent lines are larger at every point in time than for low-origin descent lines. Descent lines in the CMGPD-LN were in general smaller than in the CMGPD-IL, reflecting its lower status, and likely higher proportions of males who married later, or not at all, and therefore had more limited opportunities for reproduction.

Determinants of Stock and Flow

We begin our presentation of results by examining the influence of founder's characteristics on descent line size (*stock*) and growth rates (*flow*) at 25-year intervals. In these results, we assume that the process determining descent line size and growth also accounts for extinction probabilities. In other words, extinction probabilities are assumed to be a function of the parameters of the distribution for descent line size used in the regression model. In addition to estimating the model described in Equation (2) that assumes that numbers of descendants follow a Poisson distribution, we estimate additional models that represent alternative assumptions about the shape of the distribution of the number of descendants. These include a linear model that assumes that the number of descents and their growth rate are linear functions of the right-hand side variables, an exponential model in which the log of the size and growth rate are assumed to be linear functions of the right-hand side variables, and a negative binomial model that accounts for the possibility of over-dispersion in the distribution of outcomes by relaxing the assumption in the Poisson model that the variance and mean of the outcomes is the same. From the results for these models, we assess the relative importance of the three possible mechanisms (i.e., Initial Advantage, Accelerated Advantage and Advantage Dissipation) that may account for differences in the dynamics of the high- and low-origin descent lines.

Our results summarized in Table 3 demonstrate that for all models in both the CMGPD-IL and the CMGPD-LN, the influences of founder's social status on the total number of descendants increase over time. These results are extracted from the full models presented in Appendix A and hold true regardless of the assumptions about the appropriate distribution for total number of descendants. For example, under the assumption of linear effects of founder's status on descent line size, the disparity in the average sizes of high- and low-origin descent lines in the CMGPD-IL increases from 0.286 persons at the 25 year mark to 3.2 at the 150 year mark.[13] Alpha tests in Appendix A suggest that in general, the negative binomial models provide better estimates than the Poisson models.[14] Results from the CMGPD-LN are broadly similar: the gap in the sizes of the high- and low- origin descent lines widens with time.

Table 3

Founder's status has long-term effects on the growth rate of descent lines. More than one century later, differences in the growth rates of high and low origin descent lines are statistically significant. Table 4 summarizes effects of founder's high or low status on the growth rate (flow) of the number of descendants. These coefficients are extracted from the results for the full models presented in Appendix B. Results for the CMGPD-IL from the negative binomial model show that the ratio of growth rates between the high and low origin descent lines after 25 years is 1.54(= $e^{0.432}$). For the CMGPD-LN, the corresponding ratio is 1.73(= $e^{0.548}$). After 125 years, the ratio is 1.17 (= $e^{0.159}$) and 1.19 (= $e^{0.172}$) for the two populations. After 150 years, the ratio of growth rates is 1.12 (= $e^{0.110}$) in the CMGPD-IL. While the differences in growth rate are small, they remain statistically significant. These results are not simply an artifact of the founder's success at transmitting his high status to his sons, grandsons, or later descendants, and their resulting higher fertility: they persist after the introduction of controls for the number of descendants in later generations who were themselves of high status.[15]

Table 4

Graphic presentation of these results highlights the similarity of trends and patterns in the stock and flow effects across the two datasets, and helps illuminate the roles of Initial Advantage, Accelerated Advantage, and Advantage Dissipation. Figure 4 and Figure 5 present results in Table 3 and Table 4 as figures. These figures plot the high-origin/low-origin ratios of descent line size and growth rate. The horizontal line at 1 in both figures corresponds to a null founder effect, indicating an equal descent line size or growth rate between the high and low origin descent lines. Solid black lines represent the stock effect estimated from the negative binomial models. They show that the initial advantage of the high-origin descent lines emerges as early as

25 years after the founding year. The solid gray lines represent flow effects discussed below, and the dashed lines represent stock and flow effects in mixture Poisson models in Table 5 discussed in the next section of the paper.

Figure 4

Figure 5

In both populations, time trends in stock effects are also consistent with Accelerated Advantage or Advantage Dissipation. If the only mechanism is Initial Advantage, founder's characteristics should have no additional effects on stock after the first generation, and the ratios of descent line size should stabilize. In fact, the results show increasing stock effects over the course of the period. Inequality in the size of high- and low-origin descent lines grows over time. By themselves, of course, the stock effects are not sufficient to distinguish between Accelerated Advantage and Advantage Dissipation.

Inspection of trends in flow effects in both populations also shows that Advantage Dissipation was more important than Accelerated Advantage. To repeat, flow effects are represented as solid gray lines in Figure 4 and Figure 5. In both populations, the magnitude of effects of founder's status on descent line growth rates declines steadily over time. Though in both the CMGPD-LN and CMGPD-IL an effect of founder's status on growth rates is still discernible and statistically significant at the end of the observation period, it is much smaller than the effects in the earlier periods. Extrapolation from the trends in the figures suggests that within another generation or two, the growth rates of the high and low origin descent lines would be equal. Afterward, the ratio of the sizes of the descent lines would be constant. If the Accelerated Advantage predominated, the ratios of growth rates would remain above one indefinitely. Taken together, the results from the stock and the flow models are most consistent with the Initial Advantage and Advantage Dissipation mechanisms, and least consistent with Accelerated Advantage.

Extinction and Growth as Separate Processes

We next examine whether the mechanisms that govern descent line growth differ from the ones that govern extinction. Results from models based on Equation 6 allow for the determinants of extinction probabilities to differ from the determinants of descent line growth. Specifically, we apply a mixture Poisson model to fit the extinction probability and the growth rate simultaneously.[16] We extract relevant results and present them in Table 5. Coefficients from the logistic regression represent effects on the probability of descent line not being extinct, that is, its size being non-zero. Positive coefficients imply higher odds of descent line survival, or lower chances of extinction. Coefficients from the truncated Poisson regression represent effects on descent line size or growth rate, conditional on the descent line not being extinct. The stock models treat extinction as a cumulative process, so after a descent line is extinct it is still included in the analyses for later periods, during which it continues to be recorded as extinct. The flow models include controls for descent line size twenty-five years previous, and exclude descent lines that are already extinct, and thus account for the possibility that higher extinction probabilities now might be an artifact of smaller descent line size twenty five years ago. The complete results are in Appendix C.

Table 5

Results from the stock models reveal that in both populations, high-origin lines were much more likely to avoid extinction, and that differential extinction probabilities account for much of the differences in mean descent line size in Table 3. To help clarify implications of the results from the mixture Poisson regressions in Table 5 and facilitate comparisons with results in

Table 4 and Table 3, we plot them as dashed lines in Figure 4 and Figure 5. The heights of the lines are exponentiated coefficients from Table 5. Coefficients for effects of founder's status on extinction chances reach 1.1 by year 50, and decline only slightly afterwards. High status descent lines, in other words, only have about one-third (= $e^{-1.1}$) the odds of extinction of low status descent lines. The truncated Poisson portion of the stock models suggests that in the CMGPD-IL, the effect of founders' social status on the total number of descendants increases slowly. 150 years after founding, surviving high-origin descent lines are about 1.65 times (= $e^{0.501}$) the size of surviving low-origin descent lines. Comparison with stock results in Table 3 suggest that much of the difference in mean descent line size after 150 years was due to differential extinction: when extinct descent lines were included, high-origin descent lines were on average 2.48 times (= $e^{0.909}$) the size of low-origin descent lines after 150 years. Extinction processes were perhaps even more important in the CMGPD-LN. Whereas in Table 3, high-origin descent lines in the CMGPD-LN were 2.02 (= $e^{0.701}$) times the size of low-origin descent sizes overall after 125 years, according to Table 5, non-extinct high-origin descent lines were only 1.15 (= $e^{0.139}$) times the size of non-extinct low-origin descent lines.

Results for extinction in the flow models in Table 5 reveal that at each point in time, extant high-origin descent lines were more likely to survive to the next time period than low-origin descent lines of the same size. In the CMGPD-IL, comparison of same-sized high- and low-origin descent lines 125 years after founding reveals that the high-origin descent lines had 1.70 (= $e^{0.533}$) times the odds of surviving to year 150. In the CMGPD-LN, a high-origin descent line 100 years after founding has about 1.25 (= $e^{0.22}$) times the odds of surviving another 25 years.

The processes that governed growth among surviving lines were very different. According to the flow results for descent line size in Table 5, the effects of founder's status on descent line growth declined over time. For both the CMGPD-LN and CMGPD-IL, the growth rates of surviving high- and low-origin descent lines were indistinguishable by the end of the period of observation. The coefficients were small and statistically insignificant. The contrast with the comparatively stable effects of founder's original status on extinction probabilities suggests that toward the end of the period for which data are available, most if not all of the small but persistent and statistically significant effect of founder's characteristics on descent line growth rates in Table 4 was due to effects on extinction chances.

One plausible interpretation of these results is that later generations minimized the chances of having no surviving descendants in the next generation, rather than maximizing the total number of births. This is most apparent in the contrast between the trends in the effects on founder's status on extinction probabilities and descent line growth in the flow models in Table 5. Toward the end of the period under observation, high status continues to reduce the probability of extinction by the next time period, even though its effect on the number of descendants among the descent lines that persist to the next time period has disappeared.

At this point, we can only speculate as to how descent lines minimized extinct chances, and whether their behavior was deliberate. It may be that the lines with high-status founders tended to be families that practiced a variant of a *K* strategy according to which they had fewer offspring than physiology would allow, but invested more in each of them, and raised the chances that at least one of them would survive to adulthood, marry, and have children. Having a high-status founder may simply have been an observable proxy for the existence of such a strategy, or that the high-status founder introduced such a strategy, which persisted in later generations. Regardless, over-representation by descendants of high-status founders is evident in both our urban noble and rural peasant populations.

DISCUSSION

Using empirical evidence from two Chinese multi-generational population databases, each of which spans over 150 years, we have shown that social status in one generation has long-term implications in later generations not only for population composition, but for the reproductive behavior of descendants. First, men of high status are especially successful in the competition for representation by descendants in later generations, and their descendants also especially successful at reproduction. While the effect of founder's social status on the growth rates of descent lines is not permanent, it nevertheless lasts for many generations, and is large enough to imply a substantial effect on population composition in later generations. Nevertheless, the descendants of high status males in one generation never come to completely dominate the population, since their share of the population stabilizes as their growth rate advantage declines. In our samples, the share of the population accounted for by the descendants of high-status males in one generation stabilizes after 150 years, or roughly six generations. Of course, "new" high-status men must emerge in each generation, thus the share of the population accounted for by individuals who have at least one high-status ancestor in the last five or six generations may be much higher.

Second, the key long-term effect of high status on the reproductive behavior of descendants was that it reduced their chances of having no surviving offspring, rather than increasing their average number of offspring, In other words, high social status among descent line founders had more enduring effects on the chances of descent line extinction than it did on the growth rates of surviving descent lines. While the effect of founder's social status on the growth rate of surviving descent lines becomes negligible after 150 years, effects on the chances of extinction are still apparent.

We can rule out indirect effects on descent line growth rates in later generations via intergenerational transmission of the forms of status recorded in our data. Calculations not shown here that control for the social status of men in later generations ruled out the possibility that the advantages of high-origin lines reflected success in transmitting positions and titles to sons, grandsons, and later descendants. Tentatively, we can also rule out polygyny as a key mechanism in accounting for differential growth rates, since it was extremely rare in the largely rural CMGPD-LN population, and became rare in the CMGPD-IL. Adoption was also unlikely to have played a role because it was usually between closely related male kin, and would have had little net effect on the total number of descendants. The CMGPD-IL has detailed records of adoption, and in calculations not shown here we found little or no difference in the total number of descendants according to whether we applied a biological or social definition of ancestry when constructing pedigrees.

One likely explanation for the long-term effect of a founder's status on the growth rate of his descent line is that the attainment of the official titles and positions that we use as our indices of status is but a manifestation of latent family characteristics that affect attainment chances across a wider variety of domains than our data describe and which are much more strongly heritable. Even though father-son correlations in attainment of official positions and titles were relatively small (Campbell and Lee 2008a), there may have been much stronger father-son correlations in traits that influence attainment chances across a wider variety of domains. Campbell and Lee (2011) demonstrated correlations in the relative statuses of lineages in rural northeast China from the late Qing and the last half of the 20th century, and suggested that this long-term stability reflected transmission within families of knowledge, attitudes, aspirations, and other forms of social or cultural capital that were conducive to attainment, and which persisted even in the face of social leveling that sought to eliminate differences in wealth or income.[17] The persistence effect of founder's characteristics on reproduction, in other words, may simply reflect that intergenerational correlations in latent or unobserved traits that affect reproduction indirectly through socioeconomic attainment across a variety of domains, including ones not measured here, are far stronger than intergenerational correlations in attainment of the specific official titles and positions recorded in our data.

These findings are of course subject to a variety of caveats. Even though the populations analyzed here are from nearly opposite ends of the social spectrum in late Imperial China, the possibility remains that their experience is atypical of historical Chinese populations, and that the other populations that lay between them in the social spectrum were characterized by other dynamics. It is possible that the processes reported here are limited to China, or societies like China, and don't generalize to human populations in general. The findings may also be less germane to contemporary populations, where variances in reproduction may be smaller than those in the past, and where the chances of not having any children at all may be less tied to social status.

The processes reported here very well may have been common to preindustrial societies in which access to reproduction was conditioned by socioeconomic status, and knowledge, aspirations, orientations, and other traits conducive to success were transmitted within families. The very similar results reported in Clark's (2007) analysis of changes in the relative frequencies of rare surnames associated with high status founders indicate that in a very different social and historical context, attainment high status in one generation translated into overrepresentation in the population many, many generations later. Indeed, the results here suggest that Clark's (2007) claim that the association between high status and large numbers of descendants in later generations was unique to Britain and explained why the Industrial Revolution started there and not elsewhere may have been incorrect. Conversely, and in contrast with Unz (2013), we doubt that processes of selection and transmission operated with more force in China than elsewhere, or account for distinctive features of contemporary Chinese society.

The most important next steps are to produce an accounting of the specific demographic components of the link between founder's status and subsequent descent line growth, locate and analyze more encompassing measures of status than the official titles and positions used here, and identify the traits that affect reproductive success and are transmitted within families. First, we need to parcel out the roles of differentials in component demographic behaviors such as fertility, mortality, marriage, remarriage, polygyny, and adoption in accounting for these patterns. A very basic question is whether the advantage of the high-origin lines in avoiding extinction reflected higher marriage chances of descendants. Alternatively, it may be that married men in high-origin lines were more successful at ensuring that at least one of their sons survived to adulthood. A refined analysis would examine effects of high status in one generation on the fertility, mortality and marriage of descendants many generations later. Second, we need to replicate these analyses in data with broader measures of status. This will help assess whether the long-term effects of founder's status really are indirect ones that work through transmission of latent traits that influence attainment in domains not measured here.

Third, we need to identify the traits that are being transmitted within families that also affect reproductive success, and the mechanisms by which they are being transmitted. We suspect that the patterns reported here reflect the social transmission within descent lines of knowledge, orientations, aspirations and other traits conducive to attainment (Campbell and Lee 2011). Social transmission of knowledge and behaviors within families would have been especially important before the advent of widespread public education. Meanwhile, extrapolation from the results of research to locate genes responsible for complex neurological disorders makes us skeptical that the patterns here have a genetic basis. Even after decades of searching, few specific genes other than the ones for Huntington's Disease have been identified that are at the same time common, and have very strong effects on the chances of developing complex neurological disorders. Genes that have strong effects tend to be rare in the population, and genes that are common in the population tend to have weak effects. Along these lines, while we expect that individual genes may be located that have small influences on behaviors relevant to socioeconomic attainment, we suspect that they will either be too rare in the population or their effects too weak to drive long-term processes like the ones reported here.

CONCLUSION

The results here underscore the importance of Mare's (2011) call for studies of stratification not only to take a multi-generational approach, but also account for interaction between status transmission and demographic behavior (Mare 1997; Mare and Maralani 2006; Mare and Song 2012). Social differentiation in one generation alters the composition of the population in future generations. Presumably at least some of the effects reported here worked through transmission across multiple generations of traits that determine status, and are not recorded in the data we use. While a purely Markovian perspective in which multi-generational processes were dominated by rapid "regression towards mediocrity" would suggest that social differentials in reproduction in one generation would have only limited and short-term term impact on population composition, the results here suggest much a longer-term impact. In fact, inequality in one generation affects descent line size and population composition more broadly many generations later.

Our ancestors were highly unrepresentative of the populations of their day in terms of social status. Their less successful counterparts are underrepresented or unrepresented in the contemporary population. One practical implication is that like data on ancestral social status in family genealogies, retrospective questions about the social status of grandparents or more distant ancestors may substantially overestimate the proportions of the population in past times who were of high status, not because individual responses are incorrect, but because we sample descendants, not ancestors.

Substantively, the most important implication of these results is that characteristics of kin networks that are often assumed to be exogenous are not. The size, composition, and other characteristics of the kin networks in which we are embedded are themselves shaped by the characteristics of ancestors who lived many generations ago. If we inherit any of the characteristics that also shaped our kin networks, the number, status attainment, and other characteristics of our kin are endogenous, not exogenous. This is a more complex and subtle problem than the correlations with kin on unobserved characteristics that studies typically address by application of fixed- or random-effects models. When such studies include number, attainment or other characteristics of kin as right-hand side variables, they assume that they are independent of the characteristics of the index individual that affect the outcome of interest. This assumption may not be valid. Studies may need to consider alternative approaches, for example, instrumenting for the characteristics of the kin network.

Analysis of relationships between social status and reproduction in human populations clearly needs to account for a much more complex relationship between status in the current generation and representation in later generations than is commonly assumed. Status in the current generation not only influences the mean numbers of descendants of an individual, but the probability of not having any descendants at all. If founders' characteristics have a more important effect on the probability of avoiding extinction in subsequent generations than on the growth rates of surviving descent lines, studies based on retrospective surveys, surnames of surviving descent lines, or data from bequests may underestimate founder effects because of selection bias.

These results suggest that members of descent lines pursued complex strategies not to maximize the total number of offspring, but to minimize the chances of having no offspring who survived to adulthood, married, and had children of their own. Many generations later, the surviving descendants of high status descent line founders may not have had higher mean numbers of offspring, but were less likely to have no surviving offspring at all. We can only speculate as to what strategies they may have pursued. Possibilities include coordination among surviving descendants to maximize the chances that at least one son survived to adulthood, married, and had children, or efforts by individual descendants to maximize survival and marriage chances of specific sons, even if it meant not having as many sons as physiologically possible. If such speculations are correct, studies such as Lavely and Wong (1998) that investigate historical Chinese population dynamics through simulation models may need to reconsider assumptions about fertility. If the more complex strategies for descent line continuity suggested by the results were typical of historical populations more generally, the implication is that simulation studies of kinship networks over the long term such as Murphy (2004) and Wachter (1987) need to consider a wider range of scenarios for distributions of numbers of offspring in each generation.

The past influences the present. The patterns in two large, multigenerational databases from a 'world we have lost' form the context we experience now by shaping not only our kin networks, and the composition of the populations in which we are embedded. New, very different socio-demographic patterns have undoubtedly emerged in the last century or two, and will shape the world that our descendants experience. We eagerly await the results of similar empirical analyses from large, contemporary, multi-generational population databases to ascertain the saliency of these historic patterns.

REFERENCES

Data

Lee, James Z., and Cameron D. Campbell. China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749-1909 [Computer file]. ICPSR27063-v6. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2011-09-02. doi:10.3886/ICPSR27063.v6

Publications

Allison, Paul D., J. Scott Long, and Tad K. Krauze. 1982 "Cumulative Advantage and Inequality in Science." *American Sociological Review* 47 (5): 615-25.

Bartholomew, David J. 1982. *Stochastic Models for Social Processes (3rd edition)*. New York: Wiley.

Becker, Gary S. 1991. *A Treatise on the Family (2nd edition)*. Cambridge: Harvard University Press.

Blau, Peter M. and Otis Dudley Duncan. 1967. *The American Occupational Structure*. New York: Wiley.

Bongaarts, John, Thomas K. Burch, and Kenneth W. Wachter. 1987. *Family Demography: Methods and Their Application*. Oxford: Clarendon Press

Boudon, Raymond. 1973. *Mathematical Structures of Social Mobility*. Jossey-Bass Inc., Publisher.

Campbell, Cameron and James Z. Lee. 2002. "State Views and Local Views of Population: Linking and Comparing Genealogies and Household Registers in Liaoning, 1749-1909." *History and Computing* 14 (1+2):9-29.

Campbell, Cameron and James Z. Lee. 2008a. "Kin Networks, Marriage, and Social Mobility in Late Imperial China." *Social Science History* 32 (2): 175-214.

Campbell, Cameron and James Z. Lee. 2008b. "Villages, Descent Groups, Households and Individual Outcomes in Rural Liaoning, 1789-1909." Pp. 73-104 in *Kinship and Demographic Behavior in the Pas*t, edited by T. Bengtsson and G Mineau. Springer.

Campbell, Cameron and James Z. Lee. 2010a. "Fertility Control in Historical China Revisited: New Methods for an Old Debate." *History of the Family* 15:370-85.

Campbell, Cameron and James Z. Lee. 2010b. "Social, Economic, and Demographic Determinants of Descent Line Growth and Extinction over the Long Term in Historical China." *CCPR* Working Paper.

Campbell, Cameron and James Z. Lee. 2011. "Kinship and the Long-Term Persistence of Inequality in Liaoning, China, 1749-2005." *Chinese Sociological Review* 44 (1):71-104.

Cavalli-Sforza, Luigi Luca. 1969. "Genetic Drift in an Italian Population." *Scientific American* 221 (2): 30-37

Cavalli-Sforza, Luigi Luca, Antonio Moroni, and Gianna Zei. 2004. *Consanguinity, Inbreeding, and Genetic Drift in Italy*. NJ: Princeton University Press.

Chang, Joseph T. 1999. "Recent Common Ancestors of All Present-day Individuals." *Advances in Applied Probability* 31 (4), 1027-38.

Clark, Gregory. 2007. *A Farewell to Alms: A Brief Economic History of the World*. Princeton University Press.

Clark, Gregory. 2010. "Regression to Mediocrity? Surnames and Social Mobility in England, 1200-2009." Available at SSRN: http://dx.doi.org/10.2139/ssrn.1651904

Colantonio, Sonia E., Gabriel W. Lasker, Bernice A. Kaplan, and Vicente Fuster. 2003. "Use of Surname Models in Human Population Biology: A Review of Recent Developments." *Human Biology* 75 (6): 785-807.

Dannefer, Dale. 2003. "Cumulative Advantage/Disadvantage and the Life Course: Cross-Fertilizing Age and Social Science Theory." *Journal of Gerontology: Social Sciences* 58 (6): 327-37.

DiPrete, Thomas A. and Gregory M. Eirich. 2006. "Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments." *Annual Review of Sociology* 32: 271-97.

Du, Ruofu, Yuan Yida, Juliana Hwang, Joanna Mountain, and L. Luca Cavalli-Sforza. 1992. "Chinese Surnames and the Genetic Differences between North and South China." *Working paper series (Morrison Institute for Population and Resource Studies); no. 27.*

Duncan, Otis Dudley. 1966. "Methodological Issues in the Analysis of Social Mobility." Pp. 51-97 in *Social Structure and Mobility in Economic Development*, edited by N. J. Smelser and S. M. Lipset. Chicago: Aldine.

Erikson, Robert and John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.

Falconer, Douglas S. 1960. *Introduction to Quantitative Genetics*. New York: Ronald Press.

Featherman, David L. and Robert M. Hauser. 1978. *Opportunity and Change*. New York: Academic Press.

Fisher , Ronald A. 1922. "On the Dominance Ratio." *Proceedings of the Royal Society of Edinburgh*, 42: 321-41.

Galton, Francis and Henry W. Watson. 1874. “On the Probability of the Extinction of Families.” *Journal of the Royal Anthropological Institute* 6: 138-44.

Grusky, David B. and Ivan K. Fukumoto. 1989. “Social History Update: A Sociological Approach to Historical Social Mobility.” *Journal of Social History* 23 (1): 221-32.

Guo, Guang. 1998. “The Timing of the Influences of Cumulative Poverty on Children’s Cognitive Ability and Achievement.” *Social Forces* 77 (1): 257-87.

Guo, Guang and Leah K. VanWey. 1999. “Sibship Size and Intellectual Development: Is the Relationship Causal?” *American Sociological Review* 64 (2): 169-87.

Harrell, Stevan J. 1987. “On the Holes in Chinese Genealogies.” *Late Imperial China* 8 (2): 53–79.

Hauser, Robert M., John N Koffel, Harry P. Travis, and Peter J. Dickinson. 1975. “Temporal Change in Occupational Mobility: Evidence for Men in the United States.” *American Sociological Review* 40 (3): 279-97.

Hauser, Robert M. and William H. Sewell. 1985. “Birth Order and Educational Attainment in Full Sibships.” *American Educational Research Journal* 22: 1-23.

Herrnstein, Richard J. and Charles Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: the Free Press.

Ho, Ping-ti. 1964. *The Ladder of Success in Imperial China: Aspects of Social Mobility, 1368-1911*. New York: Science Editions.

Hodge, Robert W. 1966. “Occupational Mobility as a Probability Process.” *Demography* 3: 19–34.

Hout, Michael. 1983. *Mobility Tables*. Beverly Hills, CA: Sage Publications.

Johnson, Norman L., Adrienne W. Kemp, and Samuel Kotz. 2005. *Univariate Discrete Distributions* (*3rd edition*). New Jersey: John Wiley & Sons, Inc..

Keyfitz, Nathan.1968. *Introduction to the Mathematics of Population*. Addison-Wesley Publishing Company

Lachance, Joseph. 2009. "Inbreeding, Pedigree Size, and the Most Recent Common Ancestor of Humanity." *Journal of Theoretical Biology* 261 (21): 238-247.

Lam, David. 1986. "The Dynamics of Population Growth, Differential Fertility, and Inequality." *American Economic Review* 76: 1103-16.

Lande, Russell. 1976. "Natural Selection and Random Genetic Drift in Phenotypic Evolution." *Evolution* 30 (2): 314-334.

Lavely, William and R. Bin Wong. 1998. "Revising the Malthusian Narrative: The Comparative Study of Population Dynamics in Late Imperial China." *Journal of Asian Studies* 57:714-48.

Lee, James Z. and Cameron Campbell. 1997. *Fate and Fortune in Rural China: Social Organization and Population Behavior in Liaoning, 1774-1873*. Cambridge University Press.

Lee, James Z., Cameron Campbell, and Shuang Chen. 2010. *China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN) 1749-1909. User Guide*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Lee, James Z., Cameron Campbell, and Feng Wang. 1993. "An Introduction to the Demography of the Qing Imperial Lineage, 1644-1911." Pp. 361-82 in *Old and New Methods in Historical Demography*, edited by R. Schofield and D. Reher. Oxford: Oxford University Press,.

Lee, James Z. and Feng Wang. 1999. *One Quarter of Humanity: Malthusian Mythology and Chinese Realities, 1700-2000*. Harvard University Press.

Lee, James Z., Feng Wang, and Cameron Campbell. 1994. "Infant and Child Mortality among the Qing Nobility: Implications for Two Types of Positive Check." *Population Studies* 48 (3):395-412.

Lee, Ronald D. and Shripad Tuljapurkar. 1994. "Stochastic Population Forecasts for the United States: Beyond High, Medium, and Low." *Journal of the American Statistical Association* 89 (428): 1175-189.

Long, Jason and Joseph P. Ferrie. [forthcoming]. "Intergenerational Occupational Mobility in Britain and the U.S. Since 1850." *American Economic Review*.

Lotka, Alfred J. 1929. "The Spread of Generations." *Human Biology* 1 (3): 305-20

Lotka, Alfred J. 1931. "The Extinction of Families." *Journal of the Washington Academy of Sciences* XXI 377-453.

Lotka, Alfred J. 1941. "The Progeny of an Entire Population." *The Annals of Mathematical Statistics* 13 (2): 115-26

Mare, Robert D. 1997. "Differential Fertility, Intergenerational Educational Mobility, and Racial Inequality." *Social Science Research* 26: 263-91.

Mare, Robert D. 2011. "A Multigenerational View of Inequality." *Demography* 48:1-23.

Mare, Robert D. and Vida Maralani. 2006. "The Intergenerational Effects of Changes in Women's Educational Attainments." *American Sociological Review* 71 (4): 542-64.

Mare, Robert D. and Xi Song. 2012. "Social Stratification in Multiple Generations." Paper presented at the Annual Meeting of Population Association of America (PAA), San Francisco.

Masel, Joanna. 2011. “Genetic Drift.” *Current Biology* 21 (20): 837-38

Matras, Judah. 1961. “Differential Fertility, Intergenerational Occupational Mobility, and Change in Occupational Distribution: Some Elementary Interrelationships.” *Population Studies* 15: 187-97.

Matsen, Frederick A. and Steven N. Evans. 2008. “To What Extent Does Genealogical Ancestry Imply Genetic Ancestry.” *Journal of Theoretical Biology* 74 (2): 182-90.

McGinnis, Robert. 1968. “A Stochastic Model of Social Mobility.” *American Sociological Review* 33 (5): 712-21

Merton, Robert K. 1968 “The Matthew Effect in Science.” *Science* 159:56-63.

Merton, Robert K. 1988. “The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property.” *ISIS*, 79, 606–23.

Murphy, Michael. 2004. “Tracing Very Long-term Kinship Networks Using SOCSIM.” *Demographic Research* 10 (7): 171-96:

Musick, Kelly and Robert D. Mare. 2004. “Family Structure, Intergenerational Mobility, and the Reproduction of Poverty: Evidence for Increasing Polarization?” *Demography* 41: 629-48.

Piazza, Alberto, Sabina Rendine, Gianna Zei, Antonio Moroni, and Luigi Luca Cavalli-Sforza. 1987. “Migration Rates of Human Populations from Surname Distributions.” *Nature* 329: 714-16.

Pampel, Fred C. and Richard G. Rogers. 2004. “Socioeconomic Status, Smoking, and Health: A Test of Competing Theories of Cumulative Advantage.” *Journal of Health and Social Behavior* 45 (3): 306-21.

Pollard, John H. 1973. *Mathematical Models for the Growth of Human Populations*. Cambridge University Press.

Preston, Samuel H. 1974. “Differential Fertility, Unwanted Fertility, and Racial Trends in Occupational Achievement.” *American Sociological Review* 39 (4): 492-506.

Preston, Samuel H. and Cameron Campbell. 1993. “Differential Fertility and the Distribution of Traits: The Case of IQ.” *American Journal of Sociology* 98 (5): 997-1019.

Preston, Samuel H., Patrick Heuveline, and Michel Guillot. 2001. *Demography: Measuring and Modeling Population Processes*. MA: Blackwell Publishing

Rogers, Andrei. 1975. *Introduction to Multiregional Mathematical Demography*. New York: Wiley

Rohde, Douglas L. T., Steve Olson, and Joseph T. Chang. 2004. “Modelling the Recent Common Ancestry of All Living Humans.” *Nature* 431: 562-66.

Ross, Catherine E. and Wu Chia-ling. 1996. “Education, Age, and the Cumulative Advantage in Health.” *Journal of Health and Social Behavior* 37:104–20.

Schoen, Robert. 1988. *Modeling Multigroup Populations*. New York: Plenum Press.

Semino, Ornella, Giuseppe Passarino, Peter J. Oefner, Alice A. Lin, Svetlana Arbuzova, Lars E. Beckman, Giovanna De Benedictis, Paolo Francalacci, Anastasia Kouvatsi, Svetlana Limborska, Mlladen Marcikiæ, Anna Mika, Barbara Mika, Dragan Primorac, A. Silvana Santachiara-Benerecetti, L. Luca Cavalli-Sforza, and Peter A. Underhill. 2000. “The Genetic Legacy of Paleolithic Homo sapiens in Extant Europeans: A Y Chromosome Perspective.” *Science* 290: 1155-59.

Sibley, Elbridge. 1942. “Some Demographic Clues to Stratification.” *American Sociological Review* 7 (3): 322-30.

Stoto, Michael A. 1983. “The Accuracy of Population Projections.” *Journal of the American Statistical Association* 78 (381): 13-20.

Sear, Rebecca, Fiona Steele, Ian A. McGregor, and Ruth Mace. 2002. “The Effects of Kin on Child Mortality in Rural Gambia.” *Demography* 39 (1):43-63.

Sear, Rebecca, Ruth Mace, and Ian A. McGregor. 2003. “The Effects of Kin on Female Fertility in Rural Gambia.” *Evolution and Human Behavior* 24 (1):24-42.

Singer, Burton, and Seymour Spilerman. 1976. “The Representation of Social Processes by Markov Models.” *American Journal of Sociology* 82 (1): 1-54.

Slatkin, Montgomery. 1977. “Gene Flow and Genetic Drift in a Species Subject to Frequent Local Extinctions.” *Theoretical Population Biology* 12 (3): 253-62.

Telford, Ted A. 1990. “Patching the Holes in Chinese Genealogies: Mortality in the Lineage Populations of Tongcheng County, 1300–1880.” *Late Imperial China* 11 (2): 116–36.

Unz, Ron. 2013. “How Social Darwinism Made Modern China.” *The American Conservative*. March/April: 16-27.

Vuong, Quang H. 1989. “Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses”. *Econometrica* 57 (2): 307-33.

Wachter, Kenneth W. 1987. “Microsimulation of Household Cycles” Pp. 215-27 in *Family Demography: Methods and Their Application*, edited by J. Bongaarts, T. K. Burch, and K. W. Wachter. Oxford: Clarendon Press

Wachter, Kenneth W. and Peter Laslett. 1978. “Measuring Patriline Extinction for Modeling Social Mobility in the Past” Pp. 113-135 in *Statistical Studies of Historical Social Structure*, edited by K. W. Wachter, E. A. Hammel and P. Laslett. London: Academic Press.

Wachter, Kenneth W., Eugene A. Hammel, and Peter Laslett (eds.) 1978. *Statistical Studies of Historical Social Structure*. London: Academic Press.

Wang, Feng, James Z. Lee, and Cameron Campbell. 1995. “Marital Fertility Control among the Qing Nobility: Implications for Two Types of Preventive Check.” *Population Studies* 49 (3): 383-400.

Wang, Linlan, James Z. Lee, and Cameron Campbell. 2010. “Institutions and Inequality: Comparing the *Zongshi* and the *Jueluo* in the Qing Imperial Lineage.” *Sungkyun Journal of East Asian Studies* 10 (1): 33-61.

Wolf, Arthur P. 1995. *Sexual Attraction and Childhood Association: A Chinese Brief for Edward Westermarck*. Stanford, CA: Stanford University Press.

Wolf, Arthur P. and Chieh-shan Huang. 1980. *Marriage and Adoption in China: 1845-1945*. Stanford, CA: Stanford University Press.

Wright, Sewall. 1929. “The Evolution of Dominance.” *The American Naturalist* 63 (689): 556–61.

Yasuda, N., L. L. Cavalli-Sforza, M. Skolnick, and A. Moroni. 1974. “The Evolution of Surnames: An Analysis of Their Distribution and Extinction.” *Theoretical Population Biology* 5 (1): 123-42.

Chuang Ying-chang, Theo Engelen, and Arthur P. Wolf (eds.). 2006. *Positive or Preventive? Reproduction in Taiwan and the Netherlands, 1850-1940*. Amsterdam: Aksant.

Zimmerman, David J. 1992. “Regression Toward Mediocrity in Economic Stature.” *American Economic Review* 82 (3): 409-29.

Notes

[1] We are grateful to Robert Mare for drawing our attention to this point.
[2] While most genetics studies reconstruct pedigrees for earlier generations by analysis of the current generation, some studies also use prospective bishopric records and surnames as proxies for genetic markers (Cavalli-Sforza et al 2004).
[3] Survivor bias means that family members at early generations are recalled and recorded only if they have descendants in the current generation. This leads to overrepresentation of successful descent line founders in retrospective surveys and genealogies that depend on the recollections of their descendants in the current generation (Campbell and Lee 2002).
[4] Downloaded from http://familypedia.wikia.com/wiki/List_of_most_common_surnames.
[5] In principle, it is also possible for inequality in descent line sizes to narrow, if founder's status eventually becomes negatively associated with the growth rate of the descent line. For instance, revolution may lead to persecution of descendants of previous elites that increases their mortality or reduces their fertility.
[6] For the linear flow models, the dependent variable for extinct descent lines is zero. For extinct lines in the exponential, Poisson and negative binomial flow models, we replace the logged previous descent line size with -1. This allows us to keep the numbers of cases the same across different estimations. We also experimented with other imputations, including log(N(*t*-1)+0.1) or setting log(N(*t*-1)) to 0, and with dropping extinct descent lines. The results are consistent with those reported in Table 4. These results are available upon request.
[7] The Qing dynasty ended with the Xinhai Revolution which began October 10, 1911 and ended February 12, 1912 with the abdication of the 'last emperor' Puyi.
[8] The Qing dynasty established the Office of the Imperial Lineage in 1652 to register imperial lineage members, supervise lineage activities, and maintain the lineage genealogy (Wang et al. 1995).
[9] See http://www.icpsr.umich.edu/icpsrweb/CMGPD/ for CMGPD-LN data and documentation.
[10] We did not restrict the sample to founders who survived to age 25 or above, because occasional missing populations preclude estimation of the exact age of death for some men (Lee et al. 2010: 19-20). Instead, we restrict the sample to adult founders who have a social status recorded. In our final sample, roughly 94% of the founders survived at least to age 16 *sui*, and 90% survived at least to age 25 *sui*.
[11] Because the CMGPD-IL specifies the identities of biological and adoptive fathers for each boy who was adopted between families, we calculated numbers of descendants separately based on biological descent and on social descent. The results were similar, as most adoption was between closely related male kin, and had little or no net effect on a descent line founder's total number of descendants.
[12] When calculating the number of founder's male siblings, we include all the males born to the descent line founders' father, regardless of how long they lived.
[13] We plot the actual growth from the empirical data and the predicted growth of descent lines based on different models in Figure A1 and A2.
[14] To compare the models across time, we compare the predicted and observed means of the outcome variables when all the other independent variables are fixed at their means. The results (shown in the Appendix Figure A1 and A2) suggest that the Poisson and the negative binomial model provide the best estimates for descent line growth trajectories, though they may yield biased estimates for the mean number of descendants toward the end of the period, perhaps because the models do not account for excess zeroes (i.e., extinction) in later years.
[15] Results not shown here, but available from the authors upon request.
[16] To compare the mixture Poisson model to the regular Poisson model, Appendix C provides Vuong likelihood ratio tests (Vuong 1989) comparing all models in the same year. Failure of the likelihood ratio test for alpha=0 indicates zeroes are generated by the same process as non-zero counts. Substantively, this implies that founder's socioeconomic status has the same effects on the extinction and growth processes. Overall, the Vuong likelihood ratio tests prefer the mixture Poisson over the regular Poisson. For both the stock and flow models at all the years, test statistics show a p-value <0.01.
[17] Greg Clark makes a related point in recent unpublished manuscripts. He argues that based on indirect results from analysis of changes over time in the surname distributions of elite populations that parent-child correlations in overall social status are much higher than typically suggested by parent-child correlations

income, social class, or occupational prestige. He suggests that the reason for the higher correlations suggested by his method is that it uncovers correlations in latent status, whereas conventional methods based on individual-level data are limited to correlations in individual outcomes that are influenced by latent status.