Population Size as an Explanation for Patterns in the Paleolithic Archaeological Record: More Caution Is Needed

Author(s): Mark Collard, Briggs Buchanan, and Michael J. O'Brien
Source: *Current Anthropology*, Vol. 54, No. S8, Alternative Pathways to Complexity: Evolutionary Trajectories in the Middle Paleolithic and Middle Stone Age (December 2013), pp. S388-S396
Published by: The University of Chicago Press on behalf of Wenner-Gren Foundation for Anthropological Research
Stable URL: http://www.jstor.org/stable/10.1086/673881
Accessed: 26/05/2014 09:30

Population Size as an Explanation for Patterns in the Paleolithic Archaeological Record

More Caution Is Needed

by Mark Collard, Briggs Buchanan, and Michael J. O'Brien

Recently it has become commonplace to use population size to explain patterns in the Paleolithic archaeological record. Several modeling studies support the idea that population size can affect cultural evolution, but the results of empirical studies are ambiguous. Here we report a study that used tool kit data from recent hunter-gatherers, in conjunction with correlation analysis and a global sample, a continental sample, and a regional sample. The results of the analyses do not support the hypothesis. Population size was correlated with some tool kit variables in the global sample, but these relationships disappeared when two factors that have previously been found to affect hunter-gatherer tool kits—risk of resource failure and mobility—were controlled for. Population size was not correlated with the tool kit variables in the other samples. The regression analyses also did not support the population size hypothesis. Together, these results challenge the use of population size to explain patterns in the Paleolithic archaeological record. Population size may explain some of the patterns in question, but this needs to be demonstrated through tests in which the population size hypothesis is explicitly pitted against competing hypotheses, such as adaptation to shifting ecological conditions.

Introduction

Recently a number of researchers have argued that population size might explain several long-debated patterns in the Paleolithic archaeological record. Shennan (2001), for example, has suggested that the so-called creative explosion of the late Middle Stone Age and Upper Paleolithic might have resulted from a large, climate-driven increase in population size. Similarly, Riede (2008) has argued that the emergence of the Bromme and Perstunian technocomplexes in Northern Europe during the Late Glacial period was driven by population size reduction associated with the Laacher See eruption. Powell, Shennan, and Thomas (2009) have proposed that population increase might also explain why many cultural innovations seem to have appeared, disappeared, and then reappeared during the late Pleistocene. Premo and Kuhn (2010) have argued that two key features of the Middle Paleolithic and Middle Stone Age archaeological records—an absence of directional technological change and the reappearance of previously existing cultural behaviors—might be a function of a high rate of extirpation of small, isolated groups and subsequent repopulation.

Support for these hypotheses comes from a number of formal models that suggest population size can have a significant effect on the evolution of fitness-relevant cultural traits. The earliest of these models was described by Shennan (2001). Shennan modified a population-genetics model to incorporate social learning among individuals and then carried out a series of simulation trials. He found that larger populations have a major advantage over smaller ones when it comes to adaptive cultural innovation because of the decreasing role of sampling effects as populations grow. His results suggested that when effective population size is large, there is a far greater probability of fitness-enhancing cultural innovations being maintained and deleterious ones being lost than when effective populations are small. In the latter situation, innovations that are maintained tend to be less beneficial in terms of reproduction and also less attractive for imitators.

Mark Collard is Canada Research Chair and Professor, and **Briggs Buchanan** is Postdoctoral Fellow, at the Human Evolutionary Studies Program and Department of Archaeology of Simon Fraser University (8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada [mcollard@sfu.ca]). **Michael J. O'Brien** is Dean of the College of Arts and Science, Professor of Anthropology, and Director of the Museum of Anthropology in the Department of Anthropology of the University of Missouri (107 Swallow Hall, Columbia, Missouri 65211-1440, U.S.A.). This paper was submitted 3 VII 13, accepted 4 IX 13, and electronically published 20 XII 13.

 0011-3204/2013/54S8-0016$10.00. DOI: 10.1086/673881

Other models that demonstrate that population size can affect the evolution of fitness-relevant traits have been reported by Henrich (2004); Powell, Shennan, and Thomas (2009); Mesoudi (2011); and Kobayashi and Aoki (2012). Henrich (2004) argued that population size can affect the probability of more complex skills being invented and maintained. In his model, learners preferentially copy the most skilled practitioner in their population with some amount of error. The probability distribution that determines the amount of error is such that a learner will only occasionally arrive at a behavior that gives a better result than the previous best. The likelihood of this occurring is dependent on population size because in large populations even improbable events occur now and again, and the larger the population, the more likely this is. Powell, Shennan, and Thomas (2009) implemented Henrich's (2004) model with a spatially structured metapopulation and found that contact and migration affect cultural evolution in a similar manner to increase in population size. Mesoudi (2011) showed that Henrich's (2004) results could be replicated when acquisition costs are allowed to increase as skill level increases. Kobayashi and Aoki (2012) modified Henrich's (2004) model to examine the effects of overlapping generations and found that the effects of population size on cultural evolution are amplified when generational overlap is taken into account.

The idea that population size can affect cultural evolution has also been supported by formal and agent-based models involving selectively neutral traits. For example, Neiman (1995) investigated the amount of variation to be expected in the decoration of a pottery assemblage if the motifs are neutral in terms of adaptation and showed that random loss, or "drift," destroys variation more quickly in smaller populations than in larger ones. More recently, Premo and Kuhn (2010) used an agent-based model to show that local group extinction can reduce cultural richness and complexity even when cultural traits do not affect fitness.

The situation with regard to empirical support for the population size hypothesis is more complicated. To date, only Powell, Shennan, and Thomas (2009) have attempted to test the hypothesis with Paleolithic archaeological data. They used molecular data to estimate when different regions of the world would have reached the same population density as Europe at the start of the Upper Paleolithic and then compared those estimates with the timing of the appearance of markers of modern behavior in the regions. Their results were mixed. They found a reasonable correspondence between the timing of the crossing of the density threshold and the timing of the appearance of markers of modern behavior in sub-Saharan Africa, North Africa, and the Levant, but there was a considerable gap between the timing of the crossing of the density threshold and the appearance of markers of modern behavior in southern, northern, and central Asia. As such, Powell, Shennan, and Thomas's (2009) results only partially support the population size hypothesis.

A number of other empirical studies have a bearing on the hypothesis (Collard, Kemery, and Banks 2005; Collard et al. 2013*a*, 2013*b*; Kline and Boyd 2010; Neiman 1995; Nelson et al. 2011). Some studies support it. Neiman (1995) investigated the amount of variation to be expected in the decoration of a pottery assemblage if the motifs are neutral in terms of adaptation. He then analyzed rim decoration variation among seven successive phases of the Woodland period in Illinois, United States, and found that it matched the expectations of his model. He concluded that the patterns of variation depended on changing levels of intergroup contact, which started low, increased, and then declined again. Kline and Boyd (2010) examined the effect of population size on marine foraging tool kits of 10 recent nonindustrial farming populations from Oceania and found that population size had a significant effect on both the number of tools and the average number of parts per tool. Collard et al. (2013*a*) applied simple linear and stepwise multiple regression analysis to data from 45 nonindustrial farming and pastoralist groups to test the hypothesis. Results of the analyses were consistent with the predictions of the hypothesis: both the number of tools and the number of tool parts were positively and significantly influenced by population size in the simple linear regression analyses. The multiple regression analyses demonstrated that these correlations were independent of the effects of risk of resource failure. Collard et al. (2013*a*) concluded from this that population size influences cultural evolution in recent nonindustrial food-producing populations.

Other empirical studies do not support the population size hypothesis. Collard, Kemery, and Banks (2005) included population size in a study designed to shed light on the drivers of food-getting tool kit structure among recent hunter-gatherers. The core of their data set comprised counts of the number of tools and tool parts in the tool kits of a worldwide sample of 20 recent hunter-gatherer populations. Their results did not support the population size hypothesis. The only variables that had a significant effect on the tool kit structure measures were measures of risk of resource failure, effective temperature, and net aboveground productivity. Read (2008) also tested the hypothesis as part of an investigation into factors that drive variation in tool kit structure among recent hunter-gatherers. Like Collard, Kemery, and Banks (2005), he found no support for the hypothesis. Nelson et al. (2011) used archaeological data to examine the relationship between regional population density and the number of pottery wares in the U.S. Southwest between 1000 CE and 1600 CE. They found population density and pottery richness to be inversely correlated and argued that this indicates that social conformity becomes increasingly important as population density increases. Although Nelson and colleagues do not discuss the population size hypothesis, their results are clearly not consistent with it. Finally, Collard et al. (2013*b*) tested the hypothesis as part of a study that focused on the drivers of technological richness among 85 recent hunter-gatherer groups from western North America. They found that the total number of material items and techniques was correlated with both a proxy for environmental risk—mean rainfall for the driest month—and population size. However, the direction of the relationship was the opposite of the one predicted by the population size hypothesis (it was negative rather than positive).

Currently, then, the situation with respect to empirical support for the population size hypothesis is mixed. Several studies support it, several refute it, and one study has yielded ambiguous results. Given this, and given the potential importance of the hypothesis for our understanding of the Paleolithic and cultural evolution in general, we decided to revisit the relationship between population size and the number and intricacy of the food-getting tools used by recent hunter-gatherer populations. Collard et al. (2011) recently reported an analysis that suggests the effect of risk of resource failure on hunter-gatherer food-getting technology is dependent on the scale of risk differences among populations. They found that when there are large differences in risk of resource failure among populations, risk has a significant effect on the number and intricacy of the food-getting tools used by hunter-gatherers. When differences in risk of resource failure are small, in contrast, risk does not have a significant effect on the structure of hunter-gatherers' food-getting tool kits. This finding suggests that the conflicting results of studies that have tested the population size hypothesis with ethnographic and archaeological data may be more apparent than real. Specifically, it raises the possibility that the studies that have failed to support the hypothesis have done so because they have employed samples in which there are large risk differences, and therefore the impact of population size on tool kit structure has been obscured by the impact of risk. With this in mind, we tested the hypothesis with samples spanning three levels of among-population risk difference: a global sample consisting of populations from several continents, a continental sample comprising populations from North America, and a regional sample made up of populations from the Pacific Northwest.

Material and Methods

Oswalt (1973, 1976) developed the method we used in the study. The method focuses on tools employed directly in the acquisition of food, which Oswalt termed "subsistants." Oswalt divided subsistants into four categories: instruments, weapons, tended facilities, and untended facilities. Instruments are used to procure food that cannot run away or threaten its pursuer, such as plants or sessile animals. A digging stick is an example of an instrument. Weapons are designed to kill or maim potential prey that can escape or may harm its pursuer. Weapons include boomerangs, crossbows, and harpoons. Facilities are structures that control the movement of animals or protect them to a human's advantage, such as a fish weir or a livestock pen. Tended facilities require continuous monitoring while in use (e.g., a fishhook), whereas untended facilities are capable of functioning without a human present and require only occasional monitoring (e.g., a deadfall trap). Oswalt created a further distinction between simple and complex subsistants. Simple subsistants do not change structurally during use, whereas complex subsistants have multiple parts that change position relative to one another during use.

Oswalt (1973, 1976) devised three measures of tool kit structure. One is the total number of subsistants, which Oswalt suggested is an indicator of the size of a tool kit. Other researchers have referred to the total number of subsistants as tool kit "diversity" (Collard, Kemery, and Banks 2005; Collard et al. 2011; Shott 1986; Torrence 1983, 1989), but this term is potentially confusing. In ecology, "diversity" has two dimensions: "richness" and "evenness." The former refers to the number of taxa in a community, landscape, or region; the latter refers to how close the taxa in a community, landscape, or region are in terms of numbers of individuals (Colwell 2009). The dimension of species diversity that the variable "total number of subsistants present in a tool kit" is akin to is clearly "species richness." Thus, to reduce the potential for confusion, here we refer to the total number of subsistants as "tool kit richness" rather than "tool kit diversity." Another of Oswalt's measures of tool kit structure is the total number of techno-units. Formally, a techno-unit is an "integrated, physically distinct, and unique structural configuration that contributes to the form of a finished artifact" (Oswalt 1976:38), but in simpler terms, techno-units are the different kinds of parts of a tool. The total number of techno-units included in a tool kit is a measure of its "complexity" (Collard, Kemery, and Banks 2005; Collard et al. 2011; Oswalt 1973, 1976; Torrence 1983, 1989). Oswalt's third measure of tool kit structure is the average number of techno-units per subsistant, which is calculated by dividing the total number of techno-units in a tool kit by its richness. Again, this is a measure of tool kit complexity.

Using Oswalt's (1973, 1976) method, we generated values for total number of subsistants (STS), total number of techno-units (TTS), average number of techno-units per tool (AVE), and population size (POP) for a sample of 49 contact-era hunter-gatherer populations. Thirty of the populations are from North America, five are from South America, five are from Africa, five are from Asia, and four are from Oceania. The names and locations of the populations are given in table 1. The majority of the tool kit data was taken from previous studies that have used Oswalt's (1973, 1976) method to quantify tool kit structure (Collard, Kemery, and Banks 2005; Collard et al. 2011). These data were supplemented with STS, TTS, and AVE values generated specifically for this study. The sources from which the latter data were extracted vary in age from the late 1800s to the mid-20th century. The values for POP were taken from Binford (2001). The POP data were transformed to base e because the POP hypothesis predicts a concave relationship between POP and tool kit richness and complexity. Log-transforming POP made the expected relationship between POP and each tool kit measure a linear one.

Table 1. Names and locations of hunter-gatherer groups included in the samples

Name	Location
Mbuti	Africa
Hadza	Africa
!Kung San	Africa
Nharo	Africa
G/Wi	Africa
Punan	Asia
Great Andamanese	Asia
Veddas	Asia
Chenchu	Asia
Yukaghir	Asia
Copper Inuit	North America, arctic
Iglulik	North America, arctic
Netsilik	North America, arctic
Angmagsalik	North America, arctic
Tareumiut	North America, arctic
Twana	North America, Pacific Northwest
Nootka	North America, Pacific Northwest
Quinalt	North America, Pacific Northwest
Upper Stalo	North America, Pacific Northwest
Coast Salish	North America, Pacific Northwest
Makah	North America, Pacific Northwest
Kwakiutl	North America, Pacific Northwest
Tlingit	North America, Pacific Northwest
Klamath	North America, plateau
Lillooet	North America, plateau
Coeur D'Alene	North America, plateau
Okanagan	North America, plateau
Sanpoil-Nespelem	North America, plateau
Shuswap	North America, plateau
Owens Valley Paiute	North America, Southwest
Surprise Valley Paiute	North America, Southwest
Fort Nelson Slave	North America, subarctic
Kaska	North America, subarctic
Carrier	North America, subarctic
Lower Koyukon	North America, subarctic
Chipewyan	North America, subarctic
Ingalik	North America, subarctic
Nabesna	North America, subarctic
Tanaina	North America, subarctic
Caribou Inuit	North America, subarctic
Tiwi	Oceania
Groote-eylandt	Oceania
Northern Arenda	Oceania
Tasmanians	Oceania
Yaruro	South America
Siriono	South America
Botocudo	South America
Ona	South America
Yahgan	South America

Some researchers contend that the technological variables should also be logged when testing the population size hypothesis. We tried this approach as well, and the results were not qualitatively different.

In addition to generating technological and POP data, we obtained values for three measures of risk of resource failure and two measures of residential mobility. We did so because, as mentioned earlier, some previous tests of the POP hypothesis found that POP did not affect tool kit structure when measures of risk of resource failure and measures of mobility were included in the analysis (Collard, Kemery, and Banks 2005; Read 2008). The proxies for risk of resource failure were effective temperature (ET), net aboveground productivity (NAGP), and mean rainfall for the wettest month of the year (RHIGH). Also known as "warmth," ET was developed to better understand the effect of temperature on the distribution of living and fossil plants (Bailey 1960). It is defined as the temperature characteristic of the start and finish of the period in which plant growth occurs (Bailey 1960). NAGP is the amount of new cell life that is added to a given location by photosynthesis and growth in a year (measured in grams per square meters per year; Binford 2001). The measures of residential mobility we included were number of residential moves per year (NOMOV) and total distance moved per year during residential moves (DISMOV). The values for ET, NAGP, RHIGH, NOMOV, and DISMOV were obtained from Binford (2001).

After compiling the data set, we used the Kolmogorov-Smirnov test to assess how closely the variables approximate a normal distribution. In the global sample, ET, NAGP, RHIGH, and DISMOV were found to have distributions that departed significantly from normal and thus were transformed. We transformed NAGP, RHIGH, and DISMOV using the natural log. To transform ET, we applied the Box-Cox power transformation in Minitab. Because a negative value was selected as the λ parameter (-1.910), all values were then subtracted from 1. In the North American sample, the only variable whose distribution departed significantly from a normal distribution was DISMOV. We transformed it using the natural log. None of the distributions were significantly different from normal in the Pacific Northwest sample.

After completing the transformations, we carried out three sets of analyses. In the first, we used simple correlation analysis to assess the direction and strength of the correlation between population size and each of the three tool kit variables (STS, TTS, AVE). Here, as in the other tests, we began with the global sample, then analyzed the North American sample, and then the Pacific Northwest sample. The test prediction was that the relationships between tool kit variables and population size should be both positive and statistically significant. Because multiple tests were conducted, Benjamini and Yekutieli's (2001) method of significance level correction was used to reduce Type I error rates. We employed this method rather than the better-known Bonferroni correction because it has been shown to balance the reduction of Type I and Type II error rates better than the Bonferroni correction (Narum 2006).

In the second set of analyses, partial correlation analysis was used to assess the direction and strength of the correlation between population size and each of the three tool kit variables while controlling for the risk variables (ET, NAGP, RHIGH) and the mobility variables (NOMOV, DISMOV). The test prediction was the same as the one in the previous set of

Table 2. Summary of results of simple correlation analyses carried out to assess the strength of the relationship between population size (POP) and tool kit richness and complexity

Sample and variables correlated	*r*	*P*
Global (*n* = 49):		
STS, POP	.353	.013[a]
TTS, POP	.361	.011[a]
AVE, POP	.179	.219
North America (*n* = 30):		
STS, POP	.318	.086
TTS, POP	.383	.037
AVE, POP	.230	.221
Pacific Northwest (*n* = 14):		
STS, POP	.455	.102
TTS, POP	.525	.054
AVE, POP	.204	.484

Note. STS = total number of subsistants; TTS = total number of techno-units; AVE = average number of techno-units per tool.
[a] Significant correlation using Benjamini and Yekutieli's (2001) alpha correction (the critical value for three tests is $\alpha = .027$).

analyses: relationships between the tool kit variables and population size should be both positive and statistically significant.

In the third set of analyses, we used standard multiple regression analysis to assess the importance of population size as an influence on tool kit richness and complexity compared with the risk and mobility variables. The tool kit variables were the dependent variables, and population size, ET, NAGP, RHIGH, DISMOV, and NOMOV were the independent variables. When the variance inflation factor (VIF) for two or more variables exceeded 10, the variable with the highest VIF was removed and the analysis rerun. PASW (SPSS) 19 was used to carry out all the analyses.

Results

Results of the simple correlation analyses are summarized in table 2. In the analyses of the global sample, STS and TTS were significantly correlated with POP, but AVE was not. None of the tool kit variables were significantly correlated with POP in the analyses of the North American sample or for the Pacific Northwest sample.

Table 3 summarizes results of the partial correlation analyses. None of the tool kit variables were significantly correlated with POP in the analyses of the global sample. The results of the analyses using the North American sample were similar: None of the tool kit variables were significantly correlated with POP. The results of the analyses that focused on the Pacific Northwest sample were consistent with the results of the analyses of the other two samples. Once again, none of the tool kit variables were significantly correlated with POP.

The results of the regression analyses are summarized in tables 4–6. The only significant influences on tool kit richness and diversity were the risk variables ET and RHIGH and the mobility variables NOMOV and DISMOV. POP was not a significant influence on any of the tool kit variables in any of the samples; it consistently had one of the lowest standardized beta coefficients; and it was always either the fourth or fifth lowest of the six independent variables.

Discussion and Conclusions

POP was correlated with two of the three tool kit variables in the global sample, but these relationships disappeared when partial correlation analysis was used to control for risk of resource failure and mobility, both of which have previously been found to influence tool kit richness and complexity among hunter-gatherers. POP was not correlated with any of the tool kit variables in the North American and Pacific Northwest samples regardless of which form of correlation analysis was used. The regression analyses were consistent with

Table 3. Summary of results of partial correlation analyses carried out to assess the strength of the relationship between population size (POP) and tool kit richness and complexity while controlling for variables that have previously been found to influence the richness and complexity of hunter-gatherer tool kits

Sample, variables correlated, and variables controlled for	*r*	*P*
Global (*n* = 49):		
STS, POP:		
ET,[a] NAGP,[b] DISMOV,[b] RHIGH,[b] NOMOV	.246	.107
TTS, POP:		
ET,[a] NAGP,[b] DISMOV,[b] RHIGH,[b] NOMOV	.248	.104
AVE, POP:		
ET,[a] NAGP,[b] DISMOV,[b] RHIGH,[b] NOMOV	−.017	.911
North America (*n* = 30):		
STS, POP:		
ET, NAGP, DISMOV,[b] RHIGH, NOMOV	.289	.160
TTS, POP:		
ET, NAGP, DISMOV,[b] RHIGH, NOMOV	.349	.087
AVE, POP:		
ET, NAGP, DISMOV,[b] RHIGH, NOMOV	.178	.395
Pacific Northwest (*n* = 14):		
STS, POP:		
ET, NAGP, DISMOV, RHIGH, NOMOV	.428	.250
TTS, POP:		
ET, NAGP, DISMOV, RHIGH, NOMOV	.179	.491
AVE, POP:		
ET, NAGP, DISMOV, RHIGH, NOMOV	.339	.372

Note. STS = total number of subsistants; TTS = total number of techno-units; AVE = average number of techno-units per tool; ET = effective temperature; NAGP = net aboveground productivity; RHIGH = rainfall for the wettest month of the year; NOMOV = number of residential moves per year; DISMOV = total distance moved per year during residential moves.
[a] Transformed with Box-Cox method and reciprocal taken before analysis; see "Material and Methods" section for details.
[b] Converted to natural logarithm before analysis; see "Material and Methods" section for details.

Table 4. Summary of results of standard multiple regression analyses using the global sample ($n = 49$) carried out to assess the relative importance of various variables as drivers of tool kit richness and complexity

Dependent variable	Full model	POP	ET	NAGP	RHIGH	NOMOV	DISMOV
STS	$F = 6.199$, df = 6, 42, $P = .000$,[a] $r^2 = .470$	$\beta = .195$, $P = .107$, VIF = 1.116	$\beta = -.658$, $P = .002$,[b] VIF = 3.310	$\beta = .204$, $P = .379$, VIF = 4.183	$\beta = -.264$, $P = .039$,[b] VIF = 1.215	$\beta = -.265$, $P = .143$, VIF = 2.495	$\beta = .011$, $P = .954$, VIF = 2.940
TTS	$F = 7.745$, df = 6, 42, $P = .000$,[a] $r^2 = .525$	$\beta = .187$, $P = .104$, VIF = 1.116	$\beta = -.740$, $P = .000$,[b] VIF = 3.310	$\beta = .271$, $P = .220$, VIF = 4.183	$\beta = -.298$, $P = .015$,[b] VIF = 1.215	$\beta = -.226$, $P = .186$, VIF = 2.495	$\beta = -.146$, $P = .428$, VIF = 2.940
AVE	$F = 5.469$, df = 6, 42, $P = .000$,[a] $r^2 = .439$	$\beta = -.014$, $P = .911$, VIF = 1.116	$\beta = -.544$, $P = .013$,[b] VIF = 3.310	$\beta = -.012$, $P = .959$, VIF = 4.183	$\beta = -.222$, $P = .088$, VIF = 1.215	$\beta = -.097$, $P = .598$, VIF = 2.495	$\beta = -.409$, $P = .045$,[b] VIF = 2.940

Note. STS = total number of subsistants; POP = population size; TTS = total number of techno-units; AVE = average number of techno-units per tool; ET = effective temperature; NAGP = net aboveground productivity; RHIGH = rainfall for the wettest month of the year; NOMOV = number of residential moves per year; DISMOV = total distance moved per year during residential moves; VIF = variance inflation factor.
[a] Significant correlation using Benjamini and Yekutieli's (2001) alpha correction (the critical value for three tests is $\alpha = .027$).
[b] Significant at $P \leq .05$.

the results of the partial correlation analyses: POP was not a significant influence on any of the tool kit variables in any of the three samples and consistently had one of the lowest standardized beta coefficients. Thus, the analyses did not support the POP hypothesis. Even when the influence of the factor that previously has been found to most affect tool kit richness and complexity among hunter-gatherers—risk of resource failure—was minimized, there was no evidence that POP influenced tool kit richness and complexity.

This means that there are now four empirical studies that support the POP hypothesis (Collard et al. 2013*a*; Kline and Boyd 2010; Neiman 1995; Powell, Shennan, and Thomas 2009) and four that do not (Collard, Kemery, and Banks 2005; Collard et al. 2013*b*; Nelson et al. 2011; this study). There are two basic potential explanations for this disagreement. One is that the studies that have failed to support the hypothesis suffer from shortcomings that are sufficiently serious to have resulted in Type II errors; that is, the studies' failure to support the hypothesis is a false negative. The other is that the results of the studies that have failed to support the hypothesis are reliable and the hypothesis needs modification.

Regarding the first possibility, there are three potential shortcomings that need to be evaluated. One is the accuracy of the population estimates. Henrich (2006), Kline and Boyd (2010), and Boyd, Richerson, and Henrich (2013) have argued that the studies by Collard, Kemery, and Banks (2005) and Read (2008) failed to support the hypothesis because they did not take into account cultural transmission among populations and therefore did not accurately measure the effective POP for cultural traits. This is unlikely. Population values for both studies were generated in the same way as the values used by Collard et al. (2013*a*), which supported the hypothesis. That Collard et al.'s (2013*a*) study supported the hypothesis implies that the method of collecting population data is adequate. Additionally, one of the other studies that have failed to support the POP hypothesis (Collard et al. 2013*b*) cannot be criticized for not taking into account cultural transmission among populations. It controlled for cultural transmission and still failed to find support for the hypothesis. Thus, use of inadequate estimates of POP seems unlikely to explain the failure of Collard, Kemery, and Banks's (2005), Read's (2008), Collard et al.'s (2013*b*), and this study to support the POP hypothesis.

A second potential shortcoming that needs to be considered is sample size. In principle, it is possible that the studies that have failed to support the hypothesis have done so because the samples they used were too small to pick up the influence of POP, but this seems unlikely. Samples employed in the studies that have tested the hypothesis with ethnographic data and found support for it comprised 10 and 45 populations, respectively (Collard et al. 2013*a*; Kline and Boyd 2010). The majority of the samples that have failed to support the hypothesis are larger than Kline and Boyd's (2010) sample, and some of them are larger than Collard et al.'s (2013*a*) sample. The samples used by Collard, Kemery, and Banks (2005) and by Read (2008) comprised 20 hunter-gatherer populations. Collard et al.'s (2013*b*) sample consisted of 85 populations. Samples employed in the study reported here range in size from 14 populations to 49 populations. As such, it is unlikely that small sample size explains the failure of the studies of Collard, Kemery, and Banks (2005), Read (2008), Collard et al. (2013*b*), and the one reported here to support the POP hypothesis.

A third potential shortcoming concerns sample bias. In addition to suggesting that Collard, Kemery, and Banks's (2005) population estimates are inaccurate, Kline and Boyd (2010) and Boyd, Richerson, and Henrich (2013) claim that Collard and colleagues' results are unreliable because North

Table 5. Summary of results of standard multiple regression analyses using the North American sample ($n = 30$) carried out to assess the relative importance of various variables as drivers of tool kit richness and complexity

Dependent variable	Full model	POP	ET	NAGP	RHIGH	NOMOV
STS	$F = 1.526$, df = 5, 24, $P = .219$, $r^2 = .241$	$\beta = .239$, $P = .218$, VIF = 1.125	$\beta = -.251$, $P = .319$, VIF = 1.923	$\beta = .810$, $P = .094$, VIF = 6.813	$\beta = -.479$, $P = .241$, VIF = 5.026	$\beta = -.116$, $P = .583$, VIF = 1.371
TTS	$F = 3.351$, df = 5, 24, $P = .020$,[a] $r^2 = .411$	$\beta = .283$, $P = .102$, VIF = 1.125	$\beta = -.441$, $P = .054$, VIF = 1.923	$\beta = .675$, $P = .112$, VIF = 6.813	$\beta = -.291$, $P = .416$, VIF = 5.026	$\beta = -.337$, $P = .079$, VIF = 1.371
AVE	$F = 3.892$, df 5, 24, $P = .010$,[a] $r^2 = .448$	$\beta = .146$, $P = .372$, VIF = 1.125	$\beta = -.493$, $P = .028$,[b] VIF = 1.923	$\beta = .117$, $P = .771$, VIF = 6.813	$\beta = .151$, $P = .662$, VIF = 5.026	$\beta = -.500$, $P = .010$,[b] VIF = 1.371

Note. DISMOV was excluded from the analysis because of multicollinearity with NOMOV (see "Material and Methods" for details). STS = total number of subsistants; POP = population size; TTS = total number of techno-units; AVE = average number of techno-units per tool; ET = effective temperature; NAGP = net aboveground productivity; RHIGH = rainfall for the wettest month of the year; NOMOV = number of residential moves per year; DISMOV = total distance moved per year during residential moves; VIF = variance inflation factor.

[a] Significant using Benjamini and Yekutieli's (2001) alpha correction (the critical value for three tests is $\alpha = .027$).

[b] Significant at $P \leq .05$.

American populations dominate their sample. Collard et al.'s (2013*b*) sample consists solely of North American populations, and North American populations also dominate the sample used in the study reported here. Thus, Kline and Boyd's (2010) and Boyd, Richerson, and Henrich's (2013) concerns can be extended to the other studies that have failed to support the hypothesis, but there are reasons to think their concerns are unwarranted. To begin with, there is an important corollary to the idea that the studies that have failed to support the hypothesis have done so because North American samples dominate the samples. The corollary is that the hypothesis does not apply to North American populations but rather to populations from other regions of the world. Thus, even if the sample-bias argument were correct, it would require us to revise the POP hypothesis to explain its failure to apply to North America populations. In other words, the sample-bias argument simply changes the scope of the refutation of the hypothesis rather than explaining away the failure to support the hypothesis.

Another reason to reject the claims of Kline and Boyd (2010) and Boyd, Richerson, and Henrich (2013) is that we obtained similar results with a balanced global sample to our original global sample. We used a random-number generator to select one population from each of five culture regions represented among the North American populations in the sample. We then deleted the other 25 North American populations. This left us with 24 populations: five from North America, five from South America, five from Africa, five from Asia, and four from Oceania. We then repeated the partial correlation analyses in which we correlated STS, TTS, and AVE with POP while controlling for ET, NAGP, RHIGH, NOMOV, and DISMOV. STS was not positively and significantly correlated with POP ($r = 0.193$, $P = .428$), nor was TTS ($r = 0.185$, $P = .448$) or AVE ($r = -0.100$, $P = .683$). Thus, the balanced global sample did not support the POP hypothesis. This indicates that the failure of this study to support the hypothesis cannot be explained away as a consequence of sample bias and suggests the same holds for the other studies that have not supported it.

It appears, then, that the disagreement between the empirical studies that support the hypothesis and those that do not is not a consequence of the latter studies suffering from shortcomings that are sufficiently serious to have resulted in Type II errors. Rather, it appears that the disagreement is substantive.

What might account for the disagreement? So far, we have been able to identify four potential answers to this question. One concerns the mode of production. Samples that have supported the hypothesis comprise populations that were heavily dependent on domesticated species (Collard et al. 2013*a*; Kline and Boyd 2010; Neiman 1995), whereas the majority of samples that have refuted the hypothesis consist of populations that relied primarily on wild resources (Collard, Kemery, and Banks 2005; Collard et al. 2013*b*; Read 2008). Consequently, it could be that mode of production mediates the effect of POP on cultural evolution such that the technology of food producers is more affected by POP than by risk, whereas the technology of hunter-gatherers is more affected by risk than by POP. The problem with this proposal is that Nelson et al.'s (2011) data relate to small-scale farming groups. This makes the idea that mode of production mediates the effect of POP on cultural evolution less plausible given that it means that groups with a food-producing mode of production both support and refute the hypothesis.

A second possibility is that there is a threshold effect in the influence of POP on technology. Populations that support

Table 6. Summary of results of standard multiple regression analyses using the Pacific Northwest sample ($n = 14$) carried out to assess the relative importance of various variables as drivers of tool kit richness and complexity

Dependent variable	Full model	POP	ET	NAGP	RHIGH	NOMOV	DISMOV
STS	$F = 1.753$, df = 6, 7, $P = .240$, $r^2 = .600$	β = .344, $P = .250$, VIF = 1.323	β = −.181, $P = .535$, VIF = 1.354	β = .348, $P = .553$, VIF = 5.479	β = .085, $P = .883$, VIF = 5.444	β = −1.028, $P = .160$, VIF = 7.488	β = 1.620, $P = .059$, VIF = 9.034
TTS	$F = 1.289$, df = 6, 7, P .370, $r^2 = .525$	β = .447, $P = .179$, VIF = 1.323	β = .196, $P = .539$, VIF = 1.354	β = .839, $P = .211$, VIF = 5.479	β = −.345, $P = .588$, VIF = 5.444	β = −.803, $P = .297$, VIF = 7.488	β = 1.063, $P = .217$, VIF = 9.034
AVE	$F = 1.862$, df = 6, 7, $P = .217$, $r^2 = .615$	β = .258, $P = .372$, VIF = 1.323	β = .608, $P = .061$, VIF = 1.354	β = .780, $P = .198$, VIF = 5.479	β = −.610, $P = .302$, VIF = 5.444	β = .210, $P = .753$, VIF = 7.488	β = −.681, $P = .366$, VIF = 9.034

Note. STS = total number of subsistants; POP = population size; TTS = total number of techno-units; AVE = average number of techno-units per tool; ET = effective temperature; NAGP = net aboveground productivity; RHIGH = rainfall for the wettest month of the year; NOMOV = number of residential moves per year; DISMOV = total distance moved per year during residential moves; VIF = variance inflation factor.

the hypothesis and those that do not overlap in terms of size, but several of the former are much larger than the largest of the latter. Thus, it could be that POP does not have a significant effect on cultural evolution until it is greater than a value close to or above the upper end of the populations that support the POP hypothesis, which is ~12,000 people. However, the modeling work of Shennan (2001) and Henrich (2004) suggests that the effect of POP on cultural evolution should be greatest when POP is less than a few thousand, so a threshold effect where the POP holds for larger populations but not for smaller ones also seems unlikely to be the explanation for the disagreement.

The other two potential explanations involve social factors. Henrich (2010) has argued that norms and institutions that foster sharing can positively affect the spread of inventions within a population. Thus, it could be that sharing norms and institutions can mediate the effects of POP on cultural evolution such that a small population with numerous and/or strong sharing norms and institutions is equivalent or even better in terms of its ability to retain beneficial inventions than a large population with few and/or weak sharing norms and institutions. If this is the case, then it is possible that the disagreement among the studies is the result of populations that support the hypothesis having fewer and/or weaker sharing norms and institutions than populations that do not support it. Another possibility is that the disagreement is a consequence of differences in degree of task specialization. Recently, Bentley and O'Brien (2011) argued that the effect of POP documented by Henrich (2004) and Powell, Shennan, and Thomas (2009) depends on two strong assumptions: (1) the skill level of the most skilled member of the group is several times greater than the skill level of the average group member, and (2) all learners can identify and copy the most skilled member of the group. Bentley and O'Brien demonstrate that the POP effect is reduced if skill level is normally distributed within a group and/or if people are less selective about whom they copy. Indeed, they show that in certain circumstances (e.g., if individuals copy the most popular behavior or copy from each other at random), cultural complexity can increase or decrease regardless of POP. One corollary of their findings is that the POP effect is likely to be mediated by degree of task specialization. Given that skill level is primarily a result of practice time (Ericsson and Charness 1994), task specialization can be expected to increase the difference in skill level between the most skilled individual within a group and the majority of group members. This means that the POP effect should be more pronounced in populations with more task specialization than in populations with less task specialization. Thus, it is possible that the disagreement among the studies that have tested the POP hypothesis is a consequence of populations that support the hypothesis having more task specialization than populations that do not support it. At the moment, we are not in a position to determine which, if either, of these hypotheses is correct. Doing so will require further modeling work and cross-cultural studies.

Together, the study reported here and the other studies that have failed to support the POP hypothesis have implications for interpreting the Paleolithic archaeological record. As we noted earlier, it has become commonplace to use POP to explain patterns in the record, but given that the record was produced exclusively by hunter-gatherers, the failure of the study reported here and the studies of Collard, Kemery, and Banks (2005), Read (2008), and Collard et al. (2013*b*) to support the hypothesis challenges these interpretations. If the richness and complexity of the technology of recent hunter-gatherers are not affected by POP, there would seem to be little reason to expect changes in POP to be a broadly useful hypothesis for explaining patterns in the Paleolithic archaeological record given that it was produced exclusively by hunter-gatherers. The same holds for stability in POP. POP change/stability may explain some of the patterns the Paleolithic archaeological record, but this needs to be demonstrated on a case-by-case basis through tests in which the POP hypothesis is pitted against competing hypotheses. As mentioned earlier, several studies have suggested that environmental risk is the primary driver of technological richness and complexity among recent hunter-gatherers (Collard, Kemery, and Banks 2005; Collard et al. 2013*b*; Read 2008; Torrence 1983, 1989). Thus, adaptation to environmental conditions is one hypothesis that should be included in such tests. Another factor that should probably be taken into account is social conformity, given that it has the capacity to affect collective action (Nelson et al. 2011). Regardless of which competing hypotheses are considered, simply attributing patterns in the Paleolithic archaeological record to POP is not a defensible course of action.

Acknowledgments

We thank Steven Kuhn and Erella Hovers for the invitation to participate in the "Alternative Pathways to Complexity: Evolutionary Trajectories in the Middle Paleolithic and Middle Stone Age" symposium and Leslie Aiello and Laurie Obbink of the Wenner-Gren Foundation and the other participants in the symposium for making it such an enjoyable and intellectually stimulating experience. We also thank Michael Kemery and Jesse Morin for assistance with data collection. Mark Collard is supported by the Canada Research Chairs Program, the Social Sciences and Humanities Research Council, the Canada Foundation for Innovation, the British Columbia Knowledge Development Fund, and Simon Fraser University. Briggs Buchanan is supported by the University of Missouri and Simon Fraser University. The opinions expressed in this paper do not necessarily reflect the views of the aforementioned funding bodies.

References Cited

Bailey, Harry P. 1960. A method of determining the warmth and temperateness of climate. *Geografiska Annaler* 42:1–16.

Benjamini, Yoav, and Daniel Yekutieli. 2001. The control of false discovery rate under dependency. *Annals of Statistics* 29:1165–1188.

Bentley, R. Alexander, and Michael J. O'Brien. 2011. The selectivity of social learning and the tempo of cultural evolution. *Journal of Evolutionary Psychology* 9:125–141.

Binford, Lewis R. 2001. *Constructing frames of reference: an analytical method for theory building using ethnographic and environmental data sets.* Berkeley: University of California Press.

Boyd, Robert, Peter J. Richerson, and Joseph Henrich. 2013. The cultural evolution of technology: facts and theories. In *Cultural evolution: society, technology, language, and religion.* Peter J. Richerson and Morten H. Christiansen, eds. Pp. 119–142. Cambridge, MA: MIT Press.

Collard, Mark, Briggs Buchanan, Jesse Morin, and Andre Costopolous. 2011. What drives the evolution of hunter-gatherer subsistence technology? a reanalysis of the risk hypothesis with data from early contact era populations in the Pacific Northwest. *Philosophical Transactions of the Royal Society B* 366:1129–1138.

Collard, Mark, Briggs Buchanan, Michael J. O'Brien, and Jonathan Scholnick. 2013*a*. Risk, mobility, or population size? drivers of technological diversity among recent western North American hunter-gatherers. *Philosophical Transactions of the Royal Society B* 368:20120412.

Collard, Mark, Michael Kemery, and Samantha Banks. 2005. Causes of toolkit variation among hunter-gatherers: a test of four competing hypotheses. *Canadian Journal of Archaeology* 29:1–19.

Collard, Mark, April Ruttle, Briggs Buchanan, and Michael J. O'Brien. 2013*b*. Population size and cultural evolution in non-industrial food-producing societies. *PLoS ONE* 8:e72628.

Colwell, Robert K. 2009. Biodiversity: concepts, patterns and measurement. In *The Princeton guide to ecology.* Simon A. Levin, ed. Pp. 257–263. Princeton, NJ: Princeton University Press.

Ericsson, K. Anders, and Neil Charness. 1994. Expert performance: its structure and acquisition. *American Psychologist* 49:725–747.

Henrich, Joseph. 2004. Demography and cultural evolution: why adaptive cultural processes produced maladaptive losses in Tasmania. *American Antiquity* 69:197–218.

———. 2006. Understanding cultural evolutionary models: a reply to Read's critique. *American Antiquity* 71:771–782.

———. 2010. The evolution of innovation-enhancing institutions. In *Innovation in cultural systems: contributions from evolutionary anthropology.* Michael J. O'Brien and Stephen J. Shennan, eds. Pp. 203–223. Cambridge, MA: MIT Press.

Kline, Michelle A., and Robert Boyd. 2010. Population size predicts technological complexity in Oceania. *Proceedings of the Royal Society B* 277: 2559–2564.

Kobayashi, Yutaka, and Kenichi Aoki. 2012. Innovativeness, population size and cumulative cultural evolution. *Theoretical Population Biology* 82:38–47.

Mesoudi, Alex. 2011. Variable cultural acquisition costs constrain cumulative cultural evolution. *PLoS ONE* 6:e18239.

Narum, Shawn R. 2006. Beyond Bonferroni: less conservative analyses for conservation genetics. *Conservation Genetics* 7:783–787.

Neiman, Fraser D. 1995. Stylistic variation in evolutionary perspective: inferences from decorative diversity and interassemblage distance in Illinois Woodland ceramic assemblages. *American Antiquity* 60:7–36.

Nelson, Margaret C., Michelle Hegmon, Stephanie R. Kulow, Matthew A. Peeples, Keith W. Kintigh, and Ann P. Kinzig. 2011. Resisting diversity: a long-term archaeological study. *Ecology and Society* 16(1):25.

Oswalt, Wendell H. 1973. *Habitat and technology: the evolution of hunting.* New York: Holt, Rinehart, & Winston.

———. 1976. *An anthropological analysis of food-getting technology.* New York: Wiley.

Powell, Adam, Stephen J. Shennan, and Mark G. Thomas. 2009. Late Pleistocene demography and the appearance of modern human behavior. *Science* 324:1298–1301.

Premo, Luke S., and Steven L. Kuhn. 2010. Modeling effects of local extinctions on culture change and diversity in the Paleolithic. *PLoS ONE* 5:e15582.

Read, Dwight. 2008. An interaction model for resource implement complexity based on risk and number of annual moves. *American Antiquity* 73:599–625.

Riede, Felix. 2008. The Laacher See-eruption (12,920 BP) and material culture change at the end of the Allerød in Northern Europe. *Journal of Archaeological Science* 35:591–599.

Shennan, Stephen J. 2001. Demography and cultural innovation: a model and some implications for the emergence of modern human culture. *Cambridge Archaeological Journal* 11:5–16.

Shott, Michael. 1986. Technological organization and settlement mobility: an ethnographic examination. *Journal of Anthropological Research* 42:15–51.

Torrence, Robin. 1983. Time budgeting and hunter-gatherer technology. In *Hunter-gatherer economy in prehistory: a European perspective.* Geoff Bailey, ed. Pp. 11–22. Cambridge: Cambridge University Press.

———. 1989. Re-tooling: towards a behavioral theory of stone tools. In *Time, energy and stone tools.* Robin Torrence, ed. Pp. 57–66. Cambridge: Cambridge University Press.