C. Tolmie

REFIGURING ANTHROPOLOGY

First Principles Of Probability & Statistics

David Hurst Thomas
American Museum of Natural History

Waveland Press, Inc.
Prospect Heights, Illinois

For information about this book, write or call:

Waveland Press, Inc.
P.O. Box 400
Prospect Heights, Illinois 60070
(312) 634-0081

For permission to use copyrighted material, the author is indebted to the following:

FIG. 2.1. By permission of the Trustees of the British Museum (Natural History).
TABLE 2.3. (p. 24) *From Physical Anthropology: An Introduction* by A. J. Kelso. Reprinted by permission of the publisher, J. B. Lippincott Company. Copyright © 1974. (p. 25) Reproduced by permission of the Society for American Archaeology from *Memoirs of the Society for American Archaeology*, Vol. 11, 1956.
FIG. 3.1. From Hulse, Frederick S. *The Human Species: An Introduction to Physical Anthropology*. Copyright © 1963 by Random House, Inc.
FIG. 3.2. From Dozier, Edward P., *The Pueblo Indians of North America*. Copyright © 1970 by Holt, Rinehart and Winston, Inc. Reproduced by permission of Holt, Rinehart and Winston. [This book reissued 1983 by Waveland Press, Inc.]
FIG. 3.4. Reproduced by permission of the Society for American Archaeology from *American Antiquity*, Vol. 35 (4), 1970.
FIG. 3.5. Reproduced by permission of the American Anthropological Association from the *American Anthropologist*, Vol. 73 (3), 1971.
FIG. 13.14. From *Biometry* by Robert R. Sokal and F. James Rohlf. W. H. Freeman and Company. Copyright © 1969.

Second Printing
The 1976 version of this book was entitled *Figuring Anthropology.*

ISBN 0-88133-223-2

Printed in the United States of America.

15 Sampling Problems in Anthropology

● *MORISON'S MUSING: No theory is valid that makes nonsense of what follows.*

15.1 THE NATURE OF ANTHROPOLOGICAL SAMPLES

The time has come to backtrack somewhat and examine an issue glossed over in previous discussions. Raymond Firth's census of Tikopian kinship was considered earlier to illustrate use of the chi-square statistic (Section 11.3). Membership in three Tikopian clans was cross-tabulated against residential village, and the chi-square statistic computed for this 4×3 table resulted in a rejection of H_0; clan membership is significantly associated with residential area in Tikopia. A chi-square statistic was also applied to the Freeds' data from northern India; in this case the age of the family head was significantly associated with occupation. These examples were considered only as illustrations of the specific computing procedure. The logic and rationale for such testing now require close inspection.

The objective of statistical inference is to generalize about true characteristics of a population based upon limited sample information. Chi-square told us that Tikopian clan membership associates with the residential unit. Specifically, the chi-square statistic indicated that the probability of finding so strong an association between two truly independent variables by chance alone is less than 1 in 100. The null hypothesis—that kinship and residence are statistically independent—was rejected on the basis of a seemingly aberrant sample. Although kinship and residence cannot be framed into an immediate causal relationship on the basis of chi-square, these findings seem to suggest at least a functional linkage between the two factors.

The difficulty should be obvious. Statistical inference generalizes from sample to population. The sample consisted of observations on all 218 informants who live in the four primary villages of Tikopia. But what is the population from which this sample was selected? The residents of the Island of Tikopia surely cannot be the population, since Firth completely enumerated all residents of the large villages for his "sample." Perhaps the true population is the Solomon Islands which are politically allied with Tikopia into the British Solomon Islands Protectorate. But Tikopia was not randomly selected from within the Protectorate; moreover, Tikopia is known to have stronger cultural ties with Samoa and Tonga than with the Solomons. Perhaps the true population consists of the islands comprising the "Polynesian fringe" in Melanesia. Or perhaps the population includes all of Polynesia. Could it be that Tikopia represents all of the smaller islands of the Pacific? Would the real population please stand up!

How could this problem have arisen? All statistical methods require random sampling, and a random sample can result only from a random selection procedure. Each variate within a population must have had an *equal and independent* probability of inclusion in the sample. Was Tikopia selected for fieldwork from a table of random digits or by repeated coin flipping? Hardly. Raymond Firth chose to work where he did because Tikopia was one of the few islands remaining with minimal outside contact in 1928. "Primitive Polynesians are rare nowadays," Firth tells us. "Most of the islanders have taken to farming, to cricket, to politics and even a few to anthropology." No random sampling here. Firth selected Tikopia by default.

But I don't intend to demean Raymond Firth; his field methods actually stand him in the best twentieth-century tradition of ethnographic fieldwork. In fact, like Firth's work on Tikopia, most fieldwork in anthropology reflects a curious blend of scientific foresight, logistical planning, and happenstance. Consider another example—how Margaret Mead came to work among the Arapesh of Highland New Guinea:

> *The long climb into the mountains on slippery trails, sometimes up almost perpendicular cliffs and sometimes in riverbeds, was slow and difficult, particularly as I had to be carried, but there was no other way into the interior. When we were part way there, the accident of Reo's success in attracting carriers from farther inland boomeranged. Our carriers left us stranded with all our gear in a village on a mountaintop with no one to move our six months' supplies in either direction—into the interior or back to the coast. So we had no means of reaching the people we had intended to study and no choice but to settle down, build a house, and work with the simple, impoverished Mountain Arapesh, who had little ritual and less art, among whom we now found ourselves. Earlier, when Reo had made his brief trip into the interior to organize carriers, he had found out, as he put it, that "these people haven't any culture worth speaking of—sisters-in-law are friends!" Now that we were stuck in the mountain village, Alitoa, he decided that he would study the language*—(Mead 1972:194–195)

Mead's Arapesh experience is probably more typical than not for ethnological fieldwork, and similar expediencies influence the rest of anthropology's samples as well. Archaeological sites are selected for excavations for a wide variety of reasons: A site will soon be destroyed by a high rise or a highway contractor, or by erosion, or by pothunting vandals; a site has just been discovered and excites the excavator; a handy site is needed for a Saturday "dig" class at nearby Utopian University. The fact is that relatively few (archaeological) sites have been excavated by previous design.

Consider the reasoning which led me to select the Reese River Valley of central Nevada for some of my own archaeological fieldwork. The Reese River area was very carefully chosen for a number of reasons:

> *(a) it is accessible, yet only by dirt road, so it is relatively free of "pothunting" activity; (b) good maps (15′ and Reese River Survey, both U.S.G.S.) are available for the entire valley; (c) most land is federally owned, thus obviating the problem of uncooperative landowners; (d) the slopes have abundant piñon trees, which have yet to be successfully related to archaeological activities, in spite of their ethnographic importance; and (e) the area has not been subjected to archaeological research to date, so its resources are completely at our disposal*—(Thomas 1969:94).

Yet this published account only tells part of the story. There were additional compelling factors about the Reese River Valley which were (quite properly) omitted from the scholarly publication. For one thing, the Reese River is a beautiful place—I liked the land and it seemed like a nice place to spend my summers. Because much of the actual archaeology was to be performed by students enrolled in my University of California summer courses, I also made *physical comfort* another major concern. One sterling lesson shone through from my own undergraduate archaeology days: uncomfortable students are grumpy students, and grumpy students are inefficient fieldworkers. The work schedule and the budget were both too rigid to allow for student inefficiency due to sunburn, heat stroke, dehydration, fatigue, or mutiny. So physical living conditions and outright personal preference were prime factors in my deciding to spend three field seasons in the Reese River Valley. I doubt that my reasons are much different from many other (candid) archaeologists'.

Dozens of comparable examples could be trotted forth to illustrate the vagaries and vicissitudes of anthropological fieldwork. But these cases should illustrate my point: Anthropologists select the sites of their fieldwork for a number of diverse—and sometimes colorful—reasons, and *random selection is rarely one of these reasons*.

Random sampling is itself a rare event in anthropology, an assertion which should startle nobody. I don't mean to imply that random sampling is not a worthwhile procedure. It is. But a realistic approach to quantitative methods demands that we see just what violence is wrought by the failures and deficiencies of sampling in anthropology.

15.2 THE HYPOTHETICAL SAMPLING UNIVERSE

> ● *OCCAM'S RAZOR: Entia non sunt multiplicanda praeter necessitatem.* (*translation:* In explaining obscure matters, imaginary things should never be postulated as existing.)

If misery enjoys company, then anthropologists are indeed a comfortable lot. Anthropology's sampling difficulties are shared among a wide range of other sciences. Many of the techniques we know as "inferential statistics" evolved from practiced statistical applications to agricultural field studies. These experiments were designed to answer some pragmatic questions regarding levels and rates of fertilization, efficacy of hybrid grains, dietary requirements of livestock and so forth. Agricultural experiments are commonly repeated year after year under a wide variety of soil and climatic conditions, to insure that diverse experimental conditions are represented. But the suitability of these experimental conditions varies a great deal, in direct proportion to the doggedness and insight of the investigators. The difficulty here, as with anthropological fieldwork, is that statistical inference allows only generalization from a sample to the population from which that sample was randomly selected. How can an agricultural experiment qualify as a "random sample from specified population?"

The sampling problem is thus not a new one. W. S. Gosset began his classic 1908 paper (this paper defined the basics of the Student's *t*-test) with the statement:

> *Any experiment can be regarded as forming an individual from a "population of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population*—(Student 1908:1).

Gosset's programmatic statement defines a new sort of population. The most important characteristic of Student's population is that it is *hypothetical.* This is a population of all possible experiments which *could* have taken place. Experimental findings are thus said to represent a *random sample* generated from the total population of all hypothetically possible samples. Why a *random* sample? Why not—it's only a hypothetical population, so we can "sample" it in any way we please. This hypothetical population is infinite for the "hard" sciences, as long as experimental replications can theoretically continue indefinitely. But for disciplines with a historical dimension, such as prehistoric archaeology and human paleontology, experimentation is restricted to a large, yet finite, universe of possibilities. A finite number of Pueblo IV sites exist today, there will never be any more. We can never hope to create more specimens of *Australopithecus robustus.* New finds will still be made, of course, but our study will forever remain restricted to those materials now buried.

The key phrase in Student's statement is "under the same conditions." The hypothetical universe consists of all experiments which could have taken place *under similar conditions.* The physical scientist has relatively little trouble specifying these standard conditions. Whenever a chemist compresses two gasses, he always specifies STP, which means *standard temperature of 0°C and standard pressure at 760 mm of mercury.* But what constitutes "similar conditions" for cultural events?

I mentioned my motives in selecting the Reese River Valley for three seasons of archaeological fieldwork, motives neither statistical nor random. I was more concerned with the practicalities of performing archaeological field investigation. The specific objective of this fieldwork was to reconstruct the prehistoric settlement patterns of this area. But the Reese River Valley is over 100 miles long, and the financial resources were not sufficient to examine so large an area. A 15-mile-long cross section of the valley was thus selected for intensive settlement pattern survey. A 500 meter grid system was imposed over this restricted test area, and a 10 percent random sample of 1400 of the 500 m^2 squares was generated. I knew beforehand that this area had been occupied since about 2500 B.C., so the specific objective was to reconstruct the settlement patterns for this 4500 span of prehistoric time within this specific area. Fieldwork took three summers with a crew ranging in size from 23 to 45 people, and the settlement patterns were ultimately reconstructed (see Thomas 1973). Because I was very careful to collect a 10 percent *random sample* from *within* the study area, I felt quite safe about generalizing the results to the entire 15 by 20-mile area originally sampled. I had statistics on my side when inferring from a randomly generated sample to the population from whence the sample was selected.

But the ultimate objective was an *anthropological* statement rather than a *statistical* statement, and the problem arose as to *how far can this pattern be generalized beyond the 1400 square tracts in the initial population*? After all, three years is a considerable investment in fieldwork, and merely describing 300 square miles hardly seems worth the effort. Could I legitimately extend the observed settlement pattern to the entire 100 miles of the Reese River Valley? Or what about generalizing to the entire central Nevada region? Or perhaps across the entire Great Basin? Have I found *the* settlement pattern for western North America? Could I crow that my three years' fieldwork laid bare the settlement patterns of all hunter-gatherers, regardless of time or space? What is the population against which I can project my sample results?

In a purely statistical perspective, the only valid generalization is the first statement: A 10 percent random sample indicates that patterns *X*, *Y*, and *Z* hold for the entire 100 percent of the population. These inferences emanate from a sample and extend to the exact population from which that sample was randomly selected. This is as far as I can go on *statistical* grounds. All statements beyond this rock-bottom level deal with a *hypothetical population* such as that suggested by Student. As an anthropologist, I am entitled to extend my findings as far as I wish, provided Student's caveat *under similar conditions* can be defined. The Reese River sample comes from a very large hypothetical population of *similar samples which exist under similar conditions*. The boundaries of this hypothetical population are defined on nonstatistical grounds and must be justified as such. Suppose I decide to make the following case: Since the critical resources of the survey area are virtually constant for the entire 100-mile length of the Reese River Valley, the aboriginal settlement pattern was probably identical throughout the entire valley. That is, I could assert that all prehistoric occupants of the Reese River Valley lived *under similar conditions*. If you wish to assume the role of skeptic, you are free to challenge my assertion of *similar conditions*. You might point out that the Reese River actually flows underground for a distance, and this lack of water must have influenced the prehistoric settlement pattern. Or you might observe that local drainage patterns create a large *playa* in the middle of the valley; the fish and waterfowl associated with this playa must surely have affected the settlement pattern. The anthropological literature is littered with concrete (and heated) examples of this sort of debate regarding similar conditions or the lack of similar conditions.

The point is that two distinct levels of inference are operating. As long as my inferences at Reese River dealt with a randomly generated sample from an explicitly defined population, I was on firm statistical grounds. The results were grounded in probability theory and as such were relatively robust to challenge. But once I shifted reference, from the *concrete population* to the *hypothetical population*, my arguments lost steam because of the difficulties in defining "similar conditions." Resorting to the hypothetical population opens the door to challenges based upon substantive anthropological and demographic evidence. Statistical inference based upon the true population was free from such challenges.

And yet, when applied properly, the hypothetical population construct remains a valid tool. Strictly speaking, Firth's Tikopian census is a statistical population, not a statistical sample. Applying a chi-square statistic to a statistical population is meaningless. But if Firth's census is viewed as a random sample from the hypothetical population of all such samples which exist *under similar conditions*, then the situation is framed in a probabilistic light. Here it does not matter one whit that Firth's sample is not random. Nobody said it was. The sample statistic of $\chi^2 = 62.786$ tells us that *if* Firth's sample had been random, such findings would occur fewer than one time in one thousand trials. Firth's census differs a great deal from that expected by chance alone. The critical difference between *actual* sampled populations and *hypothetical* populations existing under similar conditions enables the anthropologist to do a great deal of meaningful inference, *even though the sample is not random.*

There are, of course, difficulties in using the concept of the hypothetical universe, and one of the most important difficulties involves the potential abuse of the concept of randomness. Statistical results emerge as "significant" because of three kinds of factors: (1) There could be a real discrepancy between the values posited under H_0 and those observed in the sample; (2) extraneous variables, uncontrolled in the experimental design, could have intervened to inflate the sample statistic; or (3) random factors could have occurred. This third vector, the influence of chance events, is relied upon to explain a good many deviations from the expected values. Random events are generally considered to be small, independent conditions which conspire to upset expectation. We know that coins will almost never produce exactly 50 percent heads because minor events influence each toss of the coin: small breezes, variability in the way an individual tosses the coin, coins dropping different distances at different rates. If all these random factors could simultaneously be controlled—perhaps by some marvelous error-free coin-flipping machine—then the results should be predictable with total certainty. Only the laws of physics would control each outcome.

Similar random errors operate within anthropological data to confound the results. Three basic sources of random error seem to be involved: errors of sampling, errors of content, and errors of analysis. A *sampling error* occurs when a sample poorly represents the population from which it was drawn. All but the most uniform populations have a few variates falling on the extremes of the distribution; whenever these extreme variates are included in the sample, that sample becomes slightly less representative of its population. I am reminded of the time I demonstrated the "laws of chance" to some skeptical anthropology students in an introductory statistics course. Mysterious forces operate in the real world, I told them, which make chance phenomena operate in a highly predictable manner. I pointed out how well the unknown can be predicted from probabilities of past performances, citing baseball averages and human sex ratios as supportive evidence. I showed them how radiocarbon dating is based upon random decay of the C^{14} molecule (and we all know how scientific radiocarbon dates are!). Then I launched my gambling examples, which I figured should clinch the argument. I expansively asked one skeptical student to flip a quarter 20 times. We would let that quarter demonstrate the simple case of $p = q = 1/2$. When the experiment resulted in 19 tails and but a single head, I could do little but sheepishly plead sampling error.

This case involved a sample which was probably a poor indicator of the population. One wonders how many "randomly generated" samples have been rejected—tossed out, ignored, recomputed—because they seemed too unbelievable. Cross-cultural ethnographic samples constantly run the risk of unrepresentative random samples, as do archaeological surveys in which the randomly generated samples poorly reflect the actual population from whence they sprang.

Errors of content arise whenever the anthropologist records faulty data. The first chapter of Naroll's *Data Quality Control* (1962b) discusses many cases in which errors can creep into the ethnographer's notes. Archaeologists seem especially vulnerable to errors of this sort, perhaps because of the large crews necessary on most excavations: level bags are mislabelled or misplaced; radiocarbon or pollen samples are easily contaminated; housefloors are destroyed before they have been recognized; vertical stratigraphy is overlooked by hasty or incompetent excavators. The archaeological laboratory is a chamber of horrors where random errors arise as if by orthogenesis. Artifacts are miscatalogued and catalogue numbers disappear or become illegible. The detailed measurements of physical anthropology and archaeology are in error due to the faulty calibration of balances, inexperienced personnel reading the vernier scales of calipers, and a host of allied difficulties. There is also the serious danger of systematic error whenever measurements must be estimated from incomplete or damaged specimens. But sometimes one must make strategic guesses or else diminish one's sample beyond practical utility.

Errors of this sort are the nemesis of anthropology. Fortunately, we know from the celebrated Heisenberg uncertainty principle that there is an absolute limit to the precision in any set of data. If a physicist were to measure the velocity of an electron to, say, within 10 percent, that measurement would still be greatly in error. Even though the error in position is miniscule, a 10 percent error is as great in this case as an error of one mile in a human stature estimate. The Heisenberg principle assures us that errors of content will *always* be present. There is always an absolute limit to precision of measurement, although anthropologists are a long way from approaching this limit. At least Heisenberg has spared us the goose-chasing search for the ultimate anthropometric device which will forever eliminate errors and lead us to the land of blissful numeration.

Finally, there are *errors of analysis*, a category including phenomena as diverse as keypunch errors, arithmetic mistakes in finding descriptive statistics, misidentification due to illegible field notes, computer malfunctions, and so on.

I am reminded of the time as a graduate student when I personally (and laboriously) keypunched over 8000 cards of artifact measurements. The cards were carefully machine-verified and reverified to eliminate all possible error. When I was finally satisfied, the bulky boxes of cards were temporarily stored in a closet. Once I was finally ready to actually analyze these data, I discovered that a steam heater had so badly warped and deformed the cards that they produced "garbage" when processed by the computer. Fortunately, the magnitude of error was so great that I quickly discovered the problem and threw away the entire batch of cards. Presumably, errors of this sort plague the unfortunate processor of quantitative data, and there are times when computers qualify under Marvin Harris' appellation "labor saving devices which increase work."

The point is that significance testing techniques cannot distinguish these random errors from (1) actual differences and (2) uncontrolled systematic factors. Inferential statistics merely compute the probability that random influences in general could have conspired to produce a given result. A problem thus arises with the hypothetical universe of possibilities. What are the potential sources of random error in Firth's Tikopian kinship data? Sample error as such cannot exist because the sample consists of *all* available informants. Sampling occurs only in a hypothetical sense, so errors cannot possibly arise. It is impossible to draw a nonrepresentative sample from our hypothetical population. Errors of analysis also seem rather remote. Only 218 informants were queried and there were no keypunching or laboratory measurements to introduce error. It is possible, though I think unlikely, that errors were made in the hand tabulation of the census data. Possibly some of the Tikopians were misclassified as to clan or village, or possibly the same informant was counted twice.

The point is that when entire populations are treated *as if* they were samples, a good deal of potential "sampling error" is lost as explanation for deviant results due to "chance factors." Furthermore, the meticulous transcription, labelling, and processing of field data greatly minimizes the likelihood of errors of content and analysis. So, it is possible when dealing with hypothetical populations to "dry up" the potential sources for random error. If such errors cannot possibly have operated, isn't it meaningless to measure the probability that they *could have* occurred?

One alternative to this awkward situation is simply to perform sloppy fieldwork and shoddy analysis. We could thus go forth confidently computing our probabilities, secure in the knowledge that we have introduced plenty of error into our data. Perhaps a wiser course of action is simply to recognize the problem of diminishing sources of random error when dealing with the hypothetical universe: The fewer errors we can possibly have, the less heavily we should rely upon chance to explain the observed results. In reality, data are probably never totally free from error. Even if such were the case, how would we know it?

Regardless of the difficulties in application, it is clear that many anthropologists, whether or not they are aware of it, rely upon the hypothetical universe of possible samples. This elusive population is admittedly an artificial construct, designed in part to ease the social scientist over a difficult methodological hurdle. Some mathematical statisticians reject this construct out of hand. "The onus lies on the exponent of statistical theory," argues Hogben (1968), "to furnish irresistible reasons for adopting procedures which have still to prove their worth against a background of three centuries of progress in scientific discovery accomplished without their aid." In fact, even aggressive advocates of the hypothetical universe concept advise caution; Hagood (1941: 304, 425), for instance, admits that the construct is "relatively new and not too well defined . . . it is therefore only an imagined possibility and whether or not one wishes to utilize the concept is still at the discretion of the individual research worker." More to the point, anthropologists *do* use the construct, whether or not they wish to defend it.

I personally think that anthropologists are justified in using many of the "softer" statistical concepts (such as the hypothetical universe of possibilities) on the grounds that we remain a primitive science. All legitimate attempts to truly advance the state of our art must be encouraged. In addition, anthropologists do not seem quite so readily enamored with statistical elegance per se as many of our more quantitatively sophisticated colleagues in social science. That is, we don't have as much to lose from bending the rules of statistics, since most anthropologists probably only half-believe their statistics anyway. The true efficacy of concepts such as the hypothetical universe of possibilities for anthropology should probably be judged not strictly upon abstruse theoretical grounds, but rather upon the quality of the solid results which spring therefrom. At least we must judge success, or lack of it, upon anthropological rather than purely statistical criteria.

15.3 SIR FRANCIS AND HIS PROBLEM

● *MURPHY'S LAW:*
(1) Nothing is as easy as it looks.
(2) Everything takes longer than you think.
(3) If anything can go wrong, it will.

The origins of cross-cultural ethnology are generally traced back to the work of E. B. Tylor. Tylor, a professor at Oxford, pioneered anthropological thinking, particularly in the field of cultural evolution. In 1888, Tylor presented a paper entitled "On a Method of Investigating the Development of Institutions," in which he considered the functional relationship between patterns of postmarital residence and ritual avoidance of one's in-laws. Tylor assembled a massive worldwide corpus of data in the hope of demonstrating that "the development of institutions may be investigated on the basis of tabulation and classification." Although his data consisted largely of hearsay reports from adventurers, missionaries, and soldiers of fortune, Tylor was convinced that anthropology could be systematized. Using what he called "social arithmetic," Tylor commenced with a rather modest beginning to an ambitious project. "The point I chose was a quaint and somewhat comic custom as to the barbaric etiquette between husbands and their wives' relatives," explained Tylor, "They may not look at one another, much less speak, and even then, avoid mentioning one another's names." Tylor tabulated data on 282 societies to determine just how the "comic custom" of kin avoidance related to postmarital residence. If Tylor were alive today, he would doubtless judge the significance of association with a chi-square statistic, hopefully coupled with a coefficient such as ϕ to assess the strength of association. But the mathematical statistics of Tylor's day had no measures of correlation or significance, so Tylor did his best to compute the "probable closeness of causal connection." He figured that such extreme associations "would show their concurrence only once or twice by chance in so large a sample of societies." Tylor presented his findings orally to the Royal Anthropological Institute of London in 1888. The minutes of that meeting record that Tylor's work was generally well received, especially his efforts to compile comprehensive files of primitive customs throughout the world. In fact, the 1889 Tylor paper is now considered a classic of anthropological research, and was reprinted by Moore (1961).

As was the custom, the president of the Royal Anthropological Institute commented at length on the paper at hand. The president of the Institute at the time happened to be Sir Francis Galton, master of topics as diverse as genetics, human evolution, and biostatistics; in fact, this is the same Galton who coined the term *regression*, discussed in Section 13.2. Galton expressed deep concern over Tylor's "social arithmetic." In today's idiom, Galton questioned Tylor's contingency tables. Galton feared that the trait tallies might be inflated because the same trait could be counted several times—once for each closely related society. The islands of the Malay region were a case in point. They were tallied as individual cases, even though the common cultural heritage of all these islands is indisputable. Is each island an independent case, or should the entire Malay region as a whole be taken as one case? When related societies are taken as independent, these "duplicate copies" of the same phenomenon will ruin any probability calculation, and Galton wondered just how many of Tylor's 282 societies could be considered truly independent. In other words, Galton was chiding Tylor for not using independent events. Historically related cultures may not be independent because of diffusion, and the probability of a given trait can be seriously inflated by such duplication.

Interestingly enough, the difficulty of assessing the independence of cases in cross-cultural samples has become known as *Galton's problem*, a misnomer. In fact, Galton simply made an insightful observation; the real problem was Tylor's. The implications of Galton's problem have had a strange impact upon the trajectory of anthropology. The influence of statistical methods upon Franz Boas, acknowledged dean of early twentieth-century anthropology, is a case in point. Boas was himself an accomplished statistician who ingeniously applied some advanced statistical methods to the study of living racial types, to genetics, and also to ethnological data. In fact, Boas later credited himself with being the first anthropologist after Tylor's 1888 discussion to apply statistics to the field of mythology in his *Indianische Sagen*. In retrospect, Boas commented (Boas 1940:309) that, "I might have established some nice coefficients of correlation for elements of mythology" [had such coefficients been available at the time]. Shortly after his study of mythology, Boas travelled to the interior of British Columbia, where he discovered clanless tribes with family organization and patrilineal trend clearly adopted from their coastal neighbors. Boas quickly became convinced that the complexities of cultural diffusion will negate effective statistical handling of ethnological data such as Tylor advocated (and Boas defended in the 1890s). Boas later confided to Robert Lowie, a former student, that the overall gravity of Galton's problem will always prevent the useful statistical analysis of ethnological data (Lowie 1946: 229). To what degree the objections of Galton ultimately influenced Boas is uncertain, but once Boas became disillusioned with Tylor's statistical methods, he flatly denied the validity of statistical analysis of any ethnological data. As a direct result, the statistical methods lay dormant in ethnology for decades until their revival by A. L. Kroeber and his students at Berkeley during the mid-1930s. Even with today's more sophisticated statistical methods at our disposal, ethnologists must still wrestle with the problem Galton first recognized over 85 years ago.

15.3.1 Solving Galton's Problem

Two general strategies have been followed in the attempt to compensate for the lack of statistical independence among societies of the world.[1] Naroll (1970b:977) terms the first approach the *sifting method*. Quite simply, the ethnologist attempts mechanically to rid the sample of duplicates. Beatrice Whiting's monograph on Paiute sorcery is a good example of this sifting procedure in practice (Whiting 1950). After a lengthy period of fieldwork among the Northern Paiute of southern Oregon, Whiting determined that social control was exerted largely through ties within the extended family. Wrongs were often adjudicated by social relationships, frequently taking the form of physical violence and sorcery. When one person wronged another, he feared not only assault but also sickness induced by sorcery. Based upon her experience among the Northern Paiute, Whiting hypothesized that sorcery and coordinate social action must be functionally linked, and she tested her hypothesis using cross-cultural methods. A sample of 50 societies was selected from the Cross-Cultural Survey files (now renamed the Human Relations Area Files) and each sample society was rated on (1) importance of sorcery and (2) methods of social control.

Whiting scored the importance of sorcery by determining the degree to which people believed that sorcery was responsible for sickness. Social control was assessed by determining the society's treatment of murder. If murder was generally settled by retaliation, then the social control was classified as *coordinate*; but when murder cases were solved by a specifically delegated authority, the society was considered to have *superordinate* control. Whiting's hypothesis predicted that sorcery should be most highly developed in societies with coordinate control and will be virtually absent where superordinate control is present. Table 15.1 shows the contingency table for Whiting's 50 sample societies. These results show a rather strong relationship between coordinate control and sorcery—$\phi = +0.57$ and $p < 0.01$ by chi-square—thus supporting the Whiting hypothesis of sorcery.

But Whiting admitted that the results were open to criticism from two major sources. There is the initial question of whether or not she had accurately coded

[1]This section on Galton's problem relies heavily upon the work of Raoul Naroll, particularly Naroll (1970b).

TABLE 15.1 The 50 randomly selected societies used to test Whiting's hypothesis on sorcery (adapted from Whiting 1950: table I).

Status of Sorcery	No Superordinate Justice (coordinate control)			Superordinate Justice	
Important	Arunta	Apache	Barama Caribs	Ashanti	Hill Maria Gonds
	Buka	Chuckchee	Delaware	Azande	Kwakiutl
	Dieri	Copper Eskimo	Jivaro	Chagga	Lamba
	Dobu	Ifugao	Witoto	Fiji	Sanpoil
	Kiwai	Kutchin		Kamilaroi	
	Kwoma	Mala		Tiv	
	Lesu	Maori		Venda	
	Murngin	Paiute			
	Orokaiva	Yurok			
	Trobriands	Zuni			
Unimportant	Lango			Bali	Cayapa
				Japan	Cheyenne
				Kazak	Crow
				Lepcha	Samoa
				Masai	Tikopia
				Ontong Java	Tonga
				Riff	
				Tanala	

the importance of sorcery and social control. Under ideal conditions, such coding should be performed by at least two independent judges who are unaware of the hypothesis being tested. She was unable to arrange such coding, but attempted to verify her evaluations against other independent studies.

Whiting noted further that a more critical point could be made "on the grounds that in selecting the tribes used in the correlation no allowance was made for historical connections and diffusions" (1950:88). Perhaps the Australian tribes (the Dieri, Arunta, Murngin, and Kamilaroi) share so many traits in common that they are really duplicate cases, and hence lack statistical independence. The New Guinea groups (Kwoma, Orokaiva, and Kiwai) could produce similar biasing effects. Although Whiting did not use the term, she was concerned about the magnitude of Galton's problem. The difficulty was controlled by filtering (or sifting) these interdependent societies. Whiting operationally defined *interrelated* as sharing "an area which is generally recognized as having cultural unity." Only a single society from each cultural area was retained, and the sample was reduced from the original 50 societies to only 26 cases, each of which belongs to a distinct cultural area. This filtered sample (Table 15.2) still supports the research hypothesis ($\phi = +0.73$ and $p = 0.00019$ by Fisher's Exact method).

Interestingly enough, the relationship appears to be stronger in the filtered sample than among the raw data. If Galton's problem were truly operative in the initial sample, then one should expect a *decrease* in association once the redundant cases are removed. Either Galton's problem was not such a problem after all, or else Whiting's sifting techniques failed to solve the difficulties of intercorrelation. Naroll (1970b: 978) has pointed out that Whiting's solution to Galton's problem is vulnerable because diffusion cannot legitimately be assumed to stop at culture areas or even at continental boundaries. The triad

TABLE 15.2 The 26 remaining societies after sample has been *sifted* for historical contact and common origin (adapted from Whiting 1950: table III).

Status of Sorcery	No Superordinate Justice		Superordinate Justice	
Important	Arunta	Barama Caribs	Ashanti	
	Chuckchee	Kwoma	Azandi	
	Copper Eskimo	Trobriands	Hill Maria Gonds	
	Delaware	Tupinamba	Kwakiutl	
	Ifugao	Witoto		
	Paiute			
	Zuni			
Unimportant			Cheyenne	Bali
			Japan	Cayapa
			Lepcha	Kazak
			Masai	Samoa
			Riff	
			Tanala	

of domesticated maize-beans-squash is known, for example, to have diffused throughout most of the Western hemisphere.

A second, more successful strategy of attacking Galton's problem is known collectively as the *propinquity method*. Although there are at least seven different propinquity solutions, all the methods begin with the assumption that diffusion can be measured as a function of geographical proximity (propinquity). Using various probabilistic models, similarities between neighbors are taken to reflect diffusion, whereas similarity between more distant peoples more probably reflects a functional relationship rather than a simple historical or diffusionary contact. The propinquity methods are relatively simple to understand and inexpensive to operate, but lack of space precludes detailed summary of the methodology. See Naroll (1970b) for a cogent review of both mechanics and logic of the propinquity solutions to Galton's problem.

15.3.2 Expanding Galton's Problem

Most anthropologists think of Galton's problem only in the context of cross-cultural research, and the attempts to vitiate this difficulty are concerned only with the contaminated samples of worldwide cultural groups. But in the broader perspective, Galton's problem extends far beyond the focus of cross-cultural ethnology. Galton cited a rather specific instance of the more general difficulty known as *spatial autocorrelation* (Loflin 1972). Whenever observations are known to be interdependent, such results will conspire to produce inconsistent and unreliable statistical estimators of the true population characteristics. When the usual statistical tests are computed for such redundant data, serious overestimation of reliability will often result.

These problems are by no means restricted to conventional cross-cultural studies. Consider the difficulties arising in obtaining samples of fossil primates. Obtaining paleontological samples is a rather different matter from taking samples from among living primate populations, and most paleontological field collections cannot meet the rigid specifications of random-sampling theory. Fossil samples are more heterogeneous than samples from living populations because many specifications such as age, sex, and diet cannot be adequately controlled; prehistoric biotic communities (*facies*) influence such samples, so the paleontologist can never assume that he has animals from a single community, or whether the fossils represent several different biota; the agencies of burial and fossilization make some critters more likely candidates for survival into modern times; natural agencies of exposure make some fossils more readily available than others to the modern scientist; differential weathering introduces bias; the collector himself introduces skewing effects because entire depositions of fossils are rarely retrieved—more commonly, expeditions return only with "exhibition quality" specimens, or selected "target species," or the more portable bones, leaving the unwieldy specimens in the field; some invertebrate fossils such as the brachiopods, gastropods, and trilobites are collected simply because of their intrinsic appeal to the human eye. Literally dozens of factors are known to influence the sampling bias of paleontology.

But I wish to point out that *Galton's problem* also operates to obfuscate the samples of human paleontologists. Between 1929 and 1934, for instance, Tabun and Skhūl caves were excavated on Mt. Carmel (in what is now Israel). Between the two sites, an extraordinary series of fossils was unearthed, representing a dozen or so individuals. The archaeological contexts are virtually identical, and it appears that the two caves were occupied at roughly the same time period, about 36,000 years ago. The Mt. Carmel fossils are extraordinary because their physical type differs from both the Classic Neanderthal form of Europe and from later *Homo sapiens sapiens* morphology. The Mt. Carmel brow ridges are large, but not as large as Classic Neanderthal; the faces are somewhat reduced in size and lack the "inflated" look of Classic Neanderthal; the limb bones seem slender and more moderate of proportion than the Neanderthals. These important skeletons have convinced some modern paleontologists that the Mt. Carmel fossils represent a true *hybridization* between the earlier Classic Neanderthals and contemporary populations of the modern *sapiens* forms who were living to the northeast on the Russian steppes. Other paleoanthropologists interpret the Mt. Carmel material as simply an intermediate stage in the rather orderly evolution from Classic Neanderthal to modern man. Depending upon one's perspective, the Mt. Carmel finds can support either the theory that Neanderthal was in the direct line of human evolution or that the Neanderthals represent a side branch, a dead-end which became extinct during the Pleistocene.

The present point does not concern whether or not Neanderthal stands in our direct family tree, but rather to what extent our interpretation of the fossil evidence has been colored by a version of Galton's problem. How does spatial autocorrelation influence paleontological samples? Nobody knows to what extent the Mt. Carmel finds could have been related. The skeletons might possibly represent a single group of very close relatives. If so, then the "atypical" cranial form could represent nothing more than a family resemblance (restricted gene pool) rather than the far-reaching evolutionary implications that paleoanthropologists have assigned to them. That is, if the Mt. Carmel people were closely related, then they represent only *one* independent case, rather than *one dozen* cases. How far would modern genetic inferences take us if our sample were restricted to a single family from the Little Italy section of New York City? Or Hell's Kitchen? Or Harlem? Or Taos Pueblo? Or Hyannis Port?

Although it seems rather unlikely that the dozen Mt. Carmel skeletons form a single close-knit family, can we deny that they might represent members of a large extended family, who perhaps buried their dead at a shrine sacred to the family? And if this were true, then the skeletons represent only a minuscule fraction of the gene pool operating in the circum-Mediterranean area about 36,000 years ago.

The fact is that a dangerously large proportion of the human fossil record is based upon a few significant clusters of finds. Consider this: Our total sample of pithecanthropine (*H. erectus*) fossils number fewer than a hundred well-documented specimens (Buettner–Janusch 1973: table 8.3). But over 40 of these individuals are "Peking Men" from the single cave of Chou Kou Tien in northeastern China, and ten additional *Homo erectus* specimens are "Solo Men" from the Ngangdon site on Java. In other words, *over half of the known Homo erectus specimens in the world come from only two sites.* Nobody knows to what extent the Peking specimens are interrelated, but who can deny the possibility? Galton's problem of autocorrelation between individual cases might even prove more devastating in the paleontological context than in the cross-cultural matrix. Whether or not specific inferential statistics are computed for *Homo erectus* or the Mt. Carmel finds is irrelevant. The evidence of paleontology comes from samples, and these samples consist of intercorrelated observations. Do we consider the Peking men to be 40 percent of the *Homo erectus* material because 40 skeletons were found, or should we consider them as only a single case because the skeletons are likely to be closely related? There have been no attempts to compensate for Galton's problem in the study of human evolution; none of the cross-cultural solutions seem relevant to this case, and the magnitude of error remains undetermined.

Nor do archaeologists escape the legacy of Sir Francis. A number of prehistorians have recently pointed out the fact that archaeological inference is directly conditioned by the methods of sampling (for example, Thomas 1973), and archaeologists face an unusual variant of the general Galton's problem. To date, few archaeologists analyze their prehistoric data using the methods and logic of cross-cultural analysis. How would an archaeological cross-cultural analysis operate? Although the following scenario is admittedly conjectural, recent studies by McNett (1967, 1970) suggest that methods of this nature might not be far away.

A massive compilation of archaeological data would first have to be synthesized into a central data bank; present archaeologists are probably operating at the level of Tylor in 1888. Perhaps some dogged archaeologist will pattern these files after HRAF (Human Relations Area Files). In fact, archaeologists could even dream up a catchy acronym such as the ARF (Archaeological Relations File), or perhaps WARP (World Archaeological Recording Project), or maybe even SHARD (Sample of Holocultural Archaeological Research Data). The major prehistoric cultures of the world could then be coded into objective categories for future statistical analysis. Consider the Cochise culture, for example. *Cochise*, a prehistoric manifestation of the generalized Desert culture, flourished in southern Arizona from about 7000 B.C. to about 1 A.D. The Cochise peoples could be treated as any society of the *Ethnographic Atlas*, and coded upon many of the *Atlas'* variables. The Cochise culture, for example, practiced little or no agriculture (col. 12), hunted a great deal (col. 8), lived in seminomadic bands (col. 30), lived in communities of less than 50 people (col. 31), lacked metallurgy (col. 42), lacked pottery manufacture (col. 48), and lacked class stratification (col. 68). Although some of the *Atlas* codings could not be applied to archaeological cultures—especially categories of kinship terminology, religion, sexual practices, and linguistic affiliation—these variables are more than compensated by relevant ecological and material cultural categories. Many problems will arise with specific codings, of course, but ethnographic cross-cultural coding is at present still not without such difficulty.

Once the data files had been established, investigators could follow the lead of their current cross-cultural brethren and test a variety of hypotheses upon randomly selected prehistoric societies. Specific aspects of interest could be the role of ecological factors in shaping cultural practices, the relationship between population size and productivity, the reasons for migration and warfare, the evolution of the state. In fact, many of the objectives of modern cross-cultural analysis could fruitfully be explored upon the archaeological samples, since only archaeological data possess suitable time depth to test many of the underlying mechanisms of cultural change and social process.

But the introduction of time depth brings us up short. *Time* is at once archaeology's greatest asset and its most significant handicap. Here we must confront Galton's problem. In conventional cross-cultural studies, Galton's problem is just an instance of spatial autocorrelation, in which neighboring or historically related societies interact. Obviously, the Cochise culture did not exist in a vacuum, and Cochise is quite similar to the neighboring, contemporary cultures such as "Lovelock" in western Nevada, the "Oak Grove" and "Topanga" cultures of coastal southern California, and the "Amargosa" culture of the Mohave desert. Is this basic similarity due to similar adaptations to similar environments (parallel cultural process) or due to diffusion of traits through direct and indirect cultural contact (diffusion)? Galton's problem strikes again! Probably the diffusionary and adaptational processes of the past operated similarly to those of the ethnographic present, and Galton's problem is just as large a factor in archaeological as in ethnographic sampling. Perhaps some version of the "sifting" or "propinquity" solutions will be adapted for archaeological sampling. But one obvious difference is that the ethnographic culture area concept remains a constant for historical groups, while archaeological culture areas (area co-traditions) are known to fluctuate through time, making the propinquity methods more tricky to apply.

Even assuming that the problems of prehistoric areal covariation can be adequately solved, we also encounter a more nettlesome version of intercorrelation. Galton objected that Tylor failed to consider the *spatial autocorrelation* between societies. But when archaeologists (inevitably) turn to a hologeistic consideration of their data, they will also be plagued with *temporal autocorrelation*. Societies are interrelated not only through space but also through time. Galton's problem, as stated in 1889, refers only to the contemporary (or horizontal) cultural contact. But archaeological interactions can also involve genetic (or vertical) contact. The Cochise culture illustrates this difficulty quite clearly. Although I discussed Cochise as a single prehistoric society, Cochise has actually been divided into three chronological *phases*:

Sulphur Spring Cochise (ca. 7000 B.C.–5000 B.C.)
Chiricahua Cochise (5000 B.C.–1900 B.C.)
San Pedro Cochise (1900 B.C.–1 A.D.)

All three phases of Cochise show a number of overriding similarities: sparse population, nonsedentary seasonal round, use of sandals (with moccasins rare), ground stone implements, use of *atlatl*, no pottery, and so forth. Yet each phase has some quite diagnostic attributes. The Sulphur Spring phase contains a sparse artifact inventory, including thin, flat milling stones and percussion-flaked, plano-convex tools for chopping and scraping (and little else). The Chiricahua phase introduces a number of new stone-tool types and basketry, and the San Pedro Cochise adds the use of pressure flaking, a new type of steep-beveled end scraper, and the occasional use of the bow and arrow. In addition, the San Pedro peoples began harvesting primitive corn as early as 3000 B.C., with squash and beans available by at least 1000 B.C. Although all three Cochise phases maintain a tradition which lasts for 7000 years, each phase reflects in situ adaptation and evolution.

For purposes of coding and cross-cultural analysis, do we have *one case or three cases*? If the Cochise culture is a single unit, how can the internal time phasing be handled? Or if the three phases must be considered as individual cases, then an alarming degree of temporal autocorrelation will exist. This is a problem not generally encountered with ethnographic surveys, where internal change is a relatively minor proposition. There is presently no name for the bias due to temporal and/or genetic relationships. Presumably some day in the future, an enterprising archaeologist (X) will present a pioneering paper analogous to the 1889 Tylor paper, illustrating the holocultural analysis of archaeological samples. Then an equally insightful critic will arise from the audience to point out the hazards of temporal-genetic autocorrelation. Following tradition, this unfortunate critic's name will inextricably be wed to the problem of temporal interrelationships (X's problem). Until the day when genetic interdependence is reified into X's problem, this brief scenario must stand as warning.

It seems that no subdiscipline of anthropology will be totally immune from the wrath of Galton, although, to date, only the cross-cultural specialists have come to grips with the intercorrelation problem. The remaining anthropological subfields will doubtless someday follow suit.

● *There once was a little girl who wrote a thank-you note to her grandmother for giving her a book about penguins for Christmas: "Dear Grandmother, Thank you very much for the nice book you sent me for Christmas. This book gives me more information about penguins than I care to have."*—H. Smith