Chapter 11

Analytical Egalitarianism, Anecdotal Evidence and Information Aggregation via Proverbial Wisdom

Can the multitude possibly tolerate or believe in the reality of the beautiful in itself as opposed to the multiplicity of beautiful things, or can they believe in anything conceived in its essence as opposed to the many particulars?

Plato *Republic*

11.1 Analytical Egalitarianism

When Adam Smith attacked the doctrine of innate differences of people, he singled out the "vanity of the philosopher" for his belief that he was somehow superior to the common porter.[1] We have linked the discussion of sameness vs. difference to the attacks on the analytical egalitarianism of classical economists. Those attacks focused on the desirability of self-direction vs. the need for direction by experts, one's betters. Consequently, the debate over equality vs. hierarchy centered on outcomes or policies that emerged from self- vs. expert-direction. Examples of the sorts of policy concerns that were central in the debate over hierarchy vs. egalitarianism include work effort, savings rates, participation in the democratic process, immigration, and sterilization.[2] Each of these concerns requires that we take the facts of the matter, the phenomena under discussion, seriously.

Here, we set aside the specifics of the policy recommendations to consider instead the methodological underpinnings of the debate over analytical egalitarianism vs. hierarchy. At the origin of the methodological debate, in Plato's *Republic,* Socrates says that ordinary people cannot be philosophers because they remain content with many

[1]Recall the passage quoted in full, as an epithet to Chapter 1 above; Smith (1776, I.II.§ 4).

[2]Above, Chapters 3, 4, 5, 6 and 7. The "scientific" breeding of people is an example of the debate over the direction of ordinary people by experts that has been taken up by followers of Adam Smith vs. Plato (Chapter 6).

surface phenomena, whereas the philosopher inquires after the one true explanation.[3] Since Plato's time, "surface phenomena" have acquired a new label, "anecdotal evidence," and Plato's "one explanation" might be called a model. The methodological debate has re-emerged as a debate about coherent model-based optimization versus heuristics. In this chapter, we compare how individuals acquire and process information relative to their scientific counterparts. We suggest that individuals rely on a heuristic, what we call "proverbial wisdom", while experts rely on models. We then examine the properties of "proverbial wisdom" relative to models.

Today, we often think of a model as something precise. And we tend to think of stories and proverbs as imprecise. Consequently, we may believe that models and stories exist in separate and incommensurate realms of discourse. As a preliminary step towards comparing models and proverbs, we therefore propose commensurate idealizations of models and proverbs.

We regard the Platonic principle, that the expert's model is superior to anecdotes of any sort, as a system of analytical hierarchy. In such a view, the expert's knowledge necessarily dominates that of ordinary people.[4] The valuable contribution of rational

---

[3]"Can the multitude possibly tolerate or believe in the reality of the beautiful in itself as opposed to the multiplicity of beautiful things, or can they believe in anything conceived in its essence as opposed to the many particulars?" Plato *Republic* 493e.

[4]Gigerenzer and Todd (1999, p. 28): " ... the heuristics-and-biases approach views heuristics and unreliable aids that the limited human mind too commonly relies upon despite their inferior decison-making performance, and hence researchers in this tradition seek out cases where heuristics can be blamed for poor reasoning."

expectations in public choice has been to show that in important contexts the expert only *weakly* dominates ordinary people. But we propose to go beyond this point of view. We demonstrate that aggregated anecdotal evidence can *improve upon* the expert's model-based estimation if the model is not exactly correct. Thus, neither the expert nor ordinary people dominates the other and so we allow for the possibility of the relationship of exchange.

Our point of view has much in common with the argument by Gigerenzer and Todd regarding "fast and frugal heuristics" that can in fact improve upon optimization procedures. [5] Unlike Gigerenzer and Todd, however, we examine a means by which informal heuristics might work.[6] So, we present a "model" of heuristics. Although we note the irony of our approach, we believe there is merit to considering the expert and ordinary people on the same analytical grounds so that their use of information can be evaluated.

There are five pieces to the argument below. We begin with a series a definitions, in order to proceed with a common understanding of our terminology. Next, to motivate the structure of our approach, we consider Smith's wage theory in which he supposes that

[5]Gigerenzer and Todd (1999, p. 28):" ... we see heuristics as the way the human mind can take advantage of the structure of information in the envirornment to arrive at reasonable decisions, and so we focus on the ways and settings in which simple heuristics lead to accurate and useful inferences."

[6]Gigerenzer and Todd (1999, p. 22): "This book adopts a different, adaptive view of rational behavior. We do not compare human judgment with the laws of logic or probability, but rather how it fares in real-world environments. The function of heuristics is not to be coherent. Rather, their function is to take reasonable, adaptive inferences about the social and physical world given limited time and knowledge. Hence, we should evaluate the performance of heuristics by criteria that reflect this function."

ordinary people *ex post* interpret random events as if they were deterministic. In our fourth section we consider how people might randomly select an observation, and give it a deterministic interpretation. Since they are aware that such a randomly-selected observation is not the only one possible, people make decisions on the basis of a process which centers such selected observations.

The technical burden of the argument is then a demonstration that the expert's model does not dominate the information aggregation possibility of ordinary people. Our technical construct – what we call the median of anecdotal evidence – supports the *possibility* of analytical egalitarianism by providing a *theoretical* counter-example to analytical hierarchy. Our evidence consists of the sort of monte carlo study employed to study estimation procedures in non-ideal circumstances. (Andrews *et al.* 1972). We summarize the argument in section 11.5 below, and present the technical details in Appendix 11.1.

If devices such as our median of anecdotal evidence are used to make decisions, they are certainly neither so-called nor so-conceived in ordinary language. In section 11.6, we return to Adam Smith and consider his discussion of the role of "proverbs" as exemplars of experience.[7] In our reading of Smith, proverbs summarize experience. As

[7]"The general maxims of morality are formed, like all other general maxims, from experience and induction. We observe in a great variety of particular cases what pleases or displeases our moral faculties, what these approve or disapprove of, and, by induction from this experience, we establish those general rules." (Smith 1759, VII.III §13).

such, they are information-aggregating stories which become rules of morality.[8] Thus, we suggest that what we call "median anecdotal evidence" for technical reasons goes by the name "proverb" or "maxim" or "parable" in the real world.[9] Consequently, we are not surprised that when Gigerenzer and Todd point to real-world heuristics, their instances are proverbs.[10] We provide some statistical justification for their conjecture about the role of proverbs in information aggregation.[11]

11.2 "Models" and "Evidence", Defined

We start with two concepts, models and evidence. By a model, we understand a K dimension causal specification which tells us how one or more variable(s) effect(s) another (Hoover 2001). By evidence, we understand N observations or elements of a K dimension empirical distribution. Recent discussions in economic methodology have

[8]The history of proverbial wisdom has yet to be written. Smith's contemporary, Gotthold Ephraim Lessing, analyzed fables and is still discussed in the philosophy of science. "For Lessing, the fable is a genre intended neither for entertainment nor for the communication and direction of emotions but rather for elucidation. ... The moral of the fable is a general claim, and hence is symbolic. For us to understand it clearly (and also for it to motivate us to act), it must be made visualizable; it must be given a concrete form." Cartwright (1993, p. 268).

[9]Discussions with Dan Houser have clarified the possibility of employing experimental procedures to recover what kind of a centering procedure, if any, is used in the formation of proverbs.

[10]Gigerenzer and Todd (1999, p. 31): "Social norms and social imitation can also help us make decisions with limited time and knowledge. Following heuristics such as 'eat what older and experienced conspecifics eat' or 'prefer mates picked by others' can speed up decision making by reducing the need for direct epxerience and information gathering."

[11]Gigerenzer and Todd (1999, p. 363): "Simple heuristics can be advantageous for navigating the complexities of social domains, and can be learned in a social manner, through imitation, word of mouth, or cultural heritage. We suspect that social norms, cultural strictures, historical proverbs, and the like can enable fast and frugal social reasoning by obviating cost-benefit calculations and extensive information search.

drawn attention to the possibility that anecdotes, conceived of as stories based on observations chosen in some unspecified way, and models influence one another in complicated ways. The stories that economists use to explicate the models (McCloskey 1990) may be *required,* to make the models operational (Morgan 2001).[12] We fully agree that stories are used in this way, and we would, in fact, expand upon the argument to suggest that *images* are also used to influence the model. However, we put aside these issues initially, in order to compare how an expert and a well-informed ordinary person might come to understand the world. For our purposes, then, it is not necessary to give priority either to a model or to the evidence.[13]

We proceed from our definition of model and evidence, to define an expert as someone who uses *all* the evidence in a transparent manner to estimate a model. Transparency is central to statistical ethics (ASA 2002) because it allows the requisite movement between the private information of the expert model builder and the common

[12]Cartwright (1993, pp. 270-1): "Models make the abstract concepts of physics more concrete. They also help to connect theory with the real world. How does this work? Typically, we design our experiments to look as much as possible like the models we have available. Then we know what specific forms our general laws should take. Obviously, one single model will not serve; we expand it by piecing in others. That usually doesn't work with fables, and that is in part why Lessing keeps the characters in his fables so thin and featureless. If he were to fill in extra details, the characters would fall under new, different abstract concepts, which may suggest different behavior from the first and perhaps even contradictory behavior."

[13]Thus, we leave aside the important question of whether we collect evidence because a model tells us that these variables are important, or whether the evidence leads us to believe that some variable affects another and should be included in the model.

knowledge of the model users.[14] Thus, transparency allows discarding of "outliers" asymmetrically as long as the reader is appropriately notified.

In some respects, the anecdotal evidence used by ordinary people may be the polar extreme of modeling by the expert. Whereas the expert employs all the evidence available, anecdotal evidence is the minimum required to "connect the dots," to tell a simple story. Whereas the modeler's practice is transparent, the ordinary person may select anecdotal evidence in a unknown fashion.

Supposing that we wish to compare the expert and the ordinary person, it would seem that the case for the wisdom of ordinary people is already lost. As a matter of *definition*, we have just suggested that the acquisition of information by ordinary people does not satisfy the transparency requirement. Put differently, transparency is a scientific requirement and ordinary people are not scientists.

However, it does not follow that the expert's employment of a causal model results in a superior *processing* of information relative to that of ordinary people because there are many ordinary people who process the information. A similar confusion is characteristic of a common criticism of democracy. Ordinary people are said to bring nothing to the political process other than their many anecdotes, each evidently inferior to the expert's

[14]The expert may prefer one outcome over another (Feigenbaum-Levy 1996), but as long as the bias is transparent, the expert does not exploit information asymmetrically.

model.[15] The difficulty with this argument has been pointed out by workers in rational expectations politics. While experts have a deeper understanding of the process, ordinary people may use heuristics to act *as if* they understood the model employed by the expert. Although no individual understands the process as well as the expert, if there are many people and their decisions are made through a political process that centers their informed preferences, then their collective decision will mirror that of the expert (Wittman 1995).

The problem we address here, that of *information* aggregation, is similar to the *preference* aggregation problem raised by Buchanan and Tullock in *The Calculus of Consent* (1962). Of all possible voting mechanisms, why it is that majority rule is so widely employed? We suggest that, just as the political process of ordinary people may be idealized using a median voter idea, information may also be aggregated by ordinary people using centralized anecdotal evidence. As a result of this similarity to the voting context, we call our information aggregation idealization a "regression by voting." We then argue that what Smith and many since have understood as "proverbs" or "proverbial wisdom" summarize experience. As such, proverbs are observed examples of aggregated information.

---

[15] Similar arguments were made in the nineteenth century to immunize "science" from criticism by ordinary people: for, if ordinary people cannot abstract from surface phenomena, they have no basis (except anecdotes) upon which to criticize the scientist. See the remarks on Jevons in the *Economist*: "Jevons's best work is so original and abstract in character, as to go out of the way, and in some respects out of the reach, of the ordinary student." (20 October 1883, p. 122). The argument was used to suggest that women were unable to participate in scientific discussions and as we have noted (Chapter 4) James Hunt left the ethological society to protest the admission of women.

We should say a word about how such proverbial wisdom bears upon the substantive debates we have studied above. Smith cites one of the books of Hebrew scriptures as a collection of proverbial wisdom. One piece of proverbial wisdom – the Biblical-based question which the British Evangelicals asked on behalf of enslaved Africans: "Am I not a man and a brother?" – was critical to the anti-slavery coalition of Classical economists and Evangelicals. This proverb was questioned in the mid-19th century by the newly arisen "science" of anthropology under the influence of James Hunt, for whom Negroes were a species apart, outside the rule of law. We point to this episode[16] and the argument below to suggest that *if* proverbial wisdom and "scientific knowledge" give us opposing directions, *then* we ought to take seriously the possibility that the "science" is at fault.

Just as we idealize the information aggregation possibilities available to ordinary people, we idealize the expert as someone who connects model and evidence transparently. Our idealized expert has no interests other than truth and so estimates a model with all the data in the most efficient manner. Real-world experts sometimes fall remarkably short of this idealization. Perhaps the most grotesque example of this is the expert who attempts to persuade people to disregard the evidence of their senses and/or

[16] We fully recognize that we have deliberately chosen a piece of anecdotal evidence here. While this goes beyond our purview in this chapter, the images and texts throughout the book, suggest that in fact there was much "evidence" of this sort in the public domain during the latter half of the century.

their information.[17] By this we do not mean that the expert explains to people that there are pieces of information which they need to consider. Instead, the expert insists that what people perceive does not “really” exist because it is not consistent with the expert’s model.[18] One episode will stand for many. “Scientific” anthropologists assured their readers that the intelligent, dark-skinned people they might observe could not be “real Negroes” because in their “science” all Negroes were childlike, simpletons. (Chapter 4).

For convenience, we collect our definitions in Table 11-1. The words in quotation marks are supposed to apply to the real, or observed, world.

---

[17]The argument was sometimes gendered: women being supposed less competent to make inferences than men. Henderson (1994) describes an episode in which the male theorists (Francis Galton) attempted to remove the participation of female onlookers. The influential owner of the *Anthropological Review* (James Hunt) left the Ethnological Society over the issue of the admission of women.

[18]It is therefore no coincidence that the great challenge to platonic realism applied to society comes from George Grote and J. S. Mill (Chapter 6 above).

Table 11-1: Definitions	
Terms	**Definition**
Model	Equation specifying an unknown but supposed fixed K-dimension casual relationship subject to a random disturbance
Evidence	N elements of the K-dimension empirical distribution
"Expert"	Person who estimates a model employing all the evidence transparently
Anecdotal evidence	K elements of the K-dimension empirical distribution which exactly fit the model
"Ordinary person"	Person who selects anecdotal evidence in an unknown fashion
Regression by voting	An idealized process that aggregates information by a procedure akin to majority rule
Median of Anecdotal Evidence	A competitive equilibrium in a regression by voting procedure
"Proverb"	Observed Median of Anecdotal Evidence

11.3 Observing Luck, Imputing Genius

Smith's account of gambles in his chapter on the distribution of wages in the *Wealth of Nations* offers insight into how ordinary people interpret random events . In contrast with standard neo-classical theory, Smith argues that ordinary people interpret a winning gamble as one that deserves an additional reward in terms of approbation. As a consequence, it is important not to read neo-classical treatments back into Smith.[19] In a nutshell, Smith tells us that those who succeed are presumed to be deserving of success. To use modern terms, he sketches an updating procedure by which random events *ex ante* become deterministic events *ex post.* Since this updating by rendering events deterministic is central to our model of information aggregation, we pay careful attention to Smith's discussion of gambling.

He begins with a simple statement of the problem of the risky choice of occupation:

> The probability that any particular person shall ever be qualified for the employment to which he is educated, is very different in different occupations. In the greater part of mechanic trades, success is almost certain; but very uncertain in the liberal professions. Put your son apprentice to a shoemaker, there is little doubt of his learning to make a pair of shoes: But send him to study the law, it is at least twenty to one if ever he makes such proficiency as will enable him to live by the business. (1776, I.10 ¶ 25)

Not only does the probability of success differ across occupations, but people also over-estimate the expected value of the gambles:

> In a perfectly fair lottery, those who draw the prizes ought to gain all that is lost by those who draw the blanks. In a profession where twenty fail for one that succeeds, that one ought to gain all that should have been gained by the unsuccessful twenty. The counsellor at law who, perhaps, at near forty years of age, begins to make something by his profession, ought to receive the retribution, not only of his own so tedious and expensive education, but of that of more than twenty others who are never likely to make any thing by it. How extravagant soever the fees of

[19]Milton Friedman and L. J. Savage (1948) offered economists an alternative to Smith's illusions-based account in which i) everyone knows the probability associated with gambles; and ii) no one's perception of the importance of an individual changes after the gamble. Levy (1999) suggests that the second aspect of Smith's approach requires more attention than it has received.

> counsellors at law may sometimes appear, their real retribution is never equal to this. Compute in any particular place, what is likely to be annually gained, and what is likely to be annually spent, by all the different workmen in any common trade, such as that of shoemakers or weavers, and you will find that the former sum will generally exceed the latter. But make the same computation with regard to all the counsellors and students of law, in all the different inns of court, and you will find that their annual gains bear but a very small proportion to their annual expence, even though you rate the former as high, and the latter as low, as can well be done. The lottery of the law, therefore, is very far from being a perfectly fair lottery; and that, as well as many other liberal and honourable professions, is, in point of pecuniary gain, evidently under-recompenced. (1776, I.10 ¶ 25)

Two reasons are offered for this. The first is a bias in estimating one's chances:

> Those professions keep their level, however, with other occupations, and, notwithstanding these discouragements, all the most generous and liberal spirits are eager to crowd into them. Two different causes contribute to recommend them. First, the desire of the reputation which attends upon superior excellence in any of them; and, secondly, the natural confidence which every man has more or less, not only in his own abilities, but in his own good fortune. (1776, I.10 ¶ 26)

The second reason is a possibly correct understanding of the approbation from winning:

> To excel in any profession, in which but few arrive at mediocrity, is the most decisive mark of what is called genius or superior talents. The public admiration which attends upon such distinguished abilities, makes always a part of their reward; a greater or smaller in proportion as it is higher or lower in degree. It makes a considerable part of that reward in the profession of physic; a still greater perhaps in that of law; in poetry and philosophy it makes almost the whole. (1776, I.10 ¶ 27)

Behind the fact of approbation which Classical economists supposed for their wage theory (Peart-Levy 2003a), there is a generating process which moves from an actor's good luck to the spectator's imputation of superior ability: those who succeed are presumed to be deserving of success. Thus, Smith, and the Classical economists who followed in his footsteps, hold that ordinary people do not use models with explicit random components to understand the world. If this is so, then how might ordinary people aggregate information? We offer a simple model of this next.

11.4 Anecdotes and Experts

Suppose we accept the common disparagement modelers offer for model-free evidence, "anecdotal." Anecdotal evidence is regarded as promiscuous empiricism offered without recognition of randomness. The "welfare queen" is not a representative guide to policy; she is a fragment of a story, not a model. "Anecdotes," as the dictionary tells us, are narratives of episodes.[20] We suppose these narratives offer evidence without the beneficent constraint of a model. If Smith is correct, ordinary people interpret such realizations of a random process in deterministic terms.

Here, we inquire into the informational properties of anecdotes. We venture out of well-explored ground because we believe that models and anecdotes, when organized coherently, are competing sources of information and so we propose to model the information aggregation of anecdotal evidence. We suppose that ordinary people aggregate information as if by voting, where the outcome is determined by the median. Then we consider how such information aggregation compares with expert decision making.

The idea we exploit, that voting itself can be viewed as robust estimation (Bassett

[20] "The narrative of a detached incident, or of a single event, told as being in itself interesting or striking." OED (1992) entry "anecdote."

and Perksy 1999), is a continuation of Galton's proposal in two long-forgotten papers published in *Nature* in 1907(Appendix 1) that called for decision-making by majority rule as an estimation procedure which recovered the *sample* median. Galton considers one dimension. We will consider many.

To motivate our conceptualization of the difference between a causal model and anecdotal evidence, we use a picture which relates variable X and Y. This picture also can help make intuitively clear why at least one piece of anecdotal evidence can be thought of faring as well as a causal model. In Figure 1, the empirical distribution, the observations of the model comprise points $[x_a, y_a], [x_b, y_b], [x_c, y_c]$ which we have labeled **a**, **b**, **c**. These three points are, then, three bits of anecdotal evidence, three stories about the relationship between X and Y. The economists' typical causal model relating X to Y is a regression equation which is marked as **OLS**. The regression equation is an abstracting device, passing among the data and encountering none.

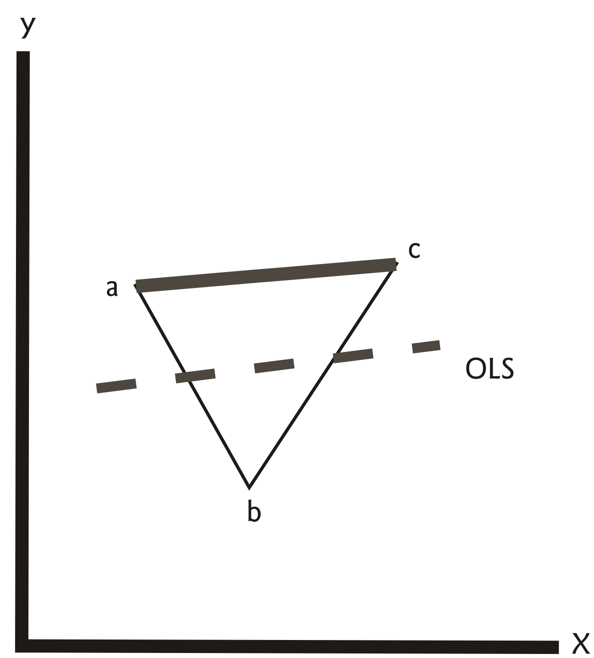

Figure 1.Fitting a Regression

Suppose there are three individuals in society, each of whom has encountered in some unknown manner one and only one of the three pieces of anecdotal evidence. If we ask each of the three what is the relationship between X and Y, we will obtain three different answers. The slope of **ab** is negative and relatively large in absolute value. The slope of **bc** is positive and relative large in absolute value. The slope is **ac** is positive and relative small in absolute value.

Suppose a competitive political process occurs, in which the majority is decisive. Then we can think about the median answer as akin to an equilibrium in "regression by voting." The median of the three slopes here is **ac.**

The picture we have drawn suggests that the slope of **ac** is not all that different than the slope of the expert's regression line. The question that concerns us is whether individuals' anecdotes – aggregated using the majority rule principle – can ever be a better representation of the relationship between X and Y than the expert's regression equation. It turns out that the expert's regression equation will do better under the assumption that the model is always and everywhere true. (Appendix 11.1) In that case, the expert's model is infallible. Under these circumstances it is efficient to process information subject to the (true) model.

What if the expert's model holds with a high probability, but it occasionally fails? Here, imposing the model on the data, might hurt the performance of the expert relative to the aggregated information of regular people. (Appendix 11.1) Looking at data through a sometimes-broken glass can grossly distort the interpretation of the "evidence".

These are strong, perhaps even counter-intuitive, claims. What we have just sketched is the reason why statistical researchers employ "elementary set methods" of regression. The power of model infallibility is bought at a high cost. It is worthy of remark that statistical workers who attempt to find a model which fits the majority of the data might employ such elementary set methods to locate a dependable starting point for their more sensitive procedures.[21]

11.5. "Universal Experience" – Proverbs in Adam Smith

It is in this context of using elementary set methods as the starting point for traditional estimation methods that we wish to reflect upon what specialists tell us about Adam Smith's epistemology: he has none (Harman 1986). Smith does, however, explain how philosophy begins with proverbial wisdom and he offers his teaching on public finance only after giving the reader four maxims on taxation.[22] To demonstrate the importance that Smith attaches to proverbs, we reproduce every occurrence of "proverb" in his two books.[23]

In his *Moral Sentiments* we find that proverbial wisdom – as founded on "universal experience" – is said to come as close to the truth as possible. Here is Smith's account of the nice properties of proverbial wisdom:

---

[21]When Rousseeuw and Leroy (1987) revived elementary set methods with the least median of squares, they also proposed a trimmed least squares that begins iteration from this robust solution. Various M-class estimation procedures introduced in Andrews, *et al.* (1972), are sensitive to starting points and so the sample median is often recommended, e.g., Mosteller and Tukey (1977).

[22]Thus, Smith offers *his* thoughts on taxation as maxims: "Before I enter upon the examination of particular taxes, it is necessary to premise the four following maxims with regard to taxes in general." (1776, V.2 § 25).

[23]"Proverbs" seem to be invariably positives; maxims in Smith's account can be "vile."

> The general rules of almost all the virtues, the general rules which determine what are the offices of prudence, of charity, of generosity, of gratitude, of friendship, are in many respects loose and inaccurate, admit of many exceptions, and require so many modifications, that it is scarce possible to regulate our conduct entirely by a regard to them. The common proverbial maxims of prudence, being founded in universal experience, are perhaps the best general rules which can be given about it. To affect, however, a very strict and literal adherence to them would evidently be the most absurd and ridiculous pedantry. (1759, III.I ¶ 121)

Smith holds there is a robustness to proverbs, as they discard troublesome outliers:

> Men in the inferior and middling stations of life, besides, can never be great enough to be above the law, which must generally overawe them into some sort of respect for, at least, the more important rules of justice. The success of such people, too, almost always depends upon the favour and good opinion of their neighbours and equals; and without a tolerably regular conduct these can very seldom be obtained. The good old proverb, therefore, That honesty is the best policy, holds, in such situations, almost always perfectly true. In such situations, therefore, we may generally expect a considerable degree of virtue; and, fortunately for the good morals of society, these are the situations of by far the greater part of mankind. (1759, I.III ¶ 32)

Smith's contemporaries would know that "vulgar" is Latin for popular which he uses to describe the illusions of the rich:

> The homely and vulgar proverb, that the eye is larger than the belly, never was more fully verified than with regard to him. The capacity of his stomach bears no proportion to the immensity of his desires, and will receive no more than that of the meanest peasant. The rest he is obliged to distribute among those, who prepare, in the nicest manner, that little which he himself makes use of, among those who fit up the palace in which this little is to be consumed, among those who provide and keep in order all the different baubles and trinkets, which are employed in the oeconomy of greatness; all of whom thus derive from his luxury and caprice, that share of the necessaries of life, which they would in vain have expected from his humanity or his justice. (1759, IV.I ¶ 10)

Second, Smith explicates the basis of proverbial wisdom in *Wealth of Nations.* There, he explains how common understanding is formulated as rules/maxims and then written as fables or wise sayings

> In every age and country of the world men must have attended to the characters, designs, and actions of one another, and many reputable rules and maxims for the conduct of human life must have been laid down and approved of by common consent. As soon as writing came into fashion, wise men, or those who fancied themselves such, would naturally endeavour to increase the number of those established and respected maxims, and to express their own sense of what was either proper or improper conduct, sometimes in the more artificial form of apologues, like what are called the fables of AEsop; and sometimes in the more simple one of apophthegms, or wise sayings, like the Proverbs of Solomon, the verses of Theognis and Phocyllides, and some part of the works of Hesiod.

Then moral philosophy induces systematic treatment:

> They might continue in this manner for a long time merely to multiply the number of those maxims of prudence and morality, without even attempting to arrange them in any very distinct or methodical order, much less to connect them together by one or more general principles from which they were all deducible, like effects from their natural causes. The beauty of a systematical arrangement of different observations connected by a few common principles was first seen in the rude essays of those ancient times towards a system of natural philosophy. Something of the same kind was afterwards attempted in morals. The maxims of common life were arranged in some methodical order, and connected together by a few common principles, in the same manner as they had attempted to arrange and connect the phenomena of nature. The science which pretends to investigate and explain those connecting principles is what is properly called moral philosophy. (*WN*, V.1.¶ 153)

Modern economists who revel in the restatement of proverbial wisdom – "There is no such thing as a free lunch" or "Don't put all your eggs in the same basket" will find that here too, Smith precedes us:

> Money, says the proverb, makes money. When you have got a little, it is often easy to get more. (*WN* I.9 ¶ 11)

> Our ancestors were idle for want of a sufficient encouragement to industry. It is better, says the proverb, to play for nothing than to work for nothing. (*WN*, II.3. ¶ 12)

> Jack of all trades will never be rich, says the proverb. (*WN*, IV.5. ¶ 55)
>
> Light come, light go, says the proverb; and the ordinary tone of expence seems every-where to be regulated, not so much according to the real ability of spending, as to the supposed facility of getting money to spend. (*WN*, IV.7 ¶ 147)
>
> If the rod be bent too much one way, says the proverb, in order to make it straight you must bend it as much the other. The French philosophers, who have proposed the system which represents agriculture as the sole source of the revenue and wealth of every country, seem to have adopted this proverbial maxim (WN, IV.9¶ 4

11.6 Culture as Information Aggregation

Smith defines "prudential" behavior in terms of proverbs – "The common proverbial maxims of prudence, being founded in universal experience, are perhaps the best general rules which can be given about it." Standard neo-classical economics does not specify where people begin optimization problems; rather, it concentrates on how individuals adjust optimally from where they are. Suppose that what is called prudential behavior is, in fact, how people begin to optimize. This suggests that Smith's insight into prudential behavior as based on proverbs might be an important addition to neo-classical economics as a way into the vexing problem of multiple equilibria. Since we know how to find proverbs – they are still called "proverbs" in ordinary language – we have some hope of coming to an understanding of the starting point problem in consumer choice.[24]

Some of Smith's proverbs confirm the received wisdom of neo-classical economics.

[24]A problem with "conventions" is that they are not so called by the ordinary people who presumably use them.

"Better to play for nothing than work for nothing" is one. But he also offers proverbs which conflict with the results of neo-classical economic theory. This offers a way of testing Smith's insight via experimental economics.[25]

The question of high dimensional narration suggests that complex narration may be especially important for the efficient dissemination of information. This thought suggests that stories/narration might be considered to exist in the same plane of information-aggregation existence that prices and models do. When prices, models and narratives give us the same advice, then we know the answer and proceed to other concerns. When the advice they provide points in different directions, then we might need to ask which one is at fault. What Adam Smith knew, but modern economists have often forgotten, is that proverbial wisdom is a method by which one can test the model.

[25]These last paragraphs have grown out of discussions with Dan Houser.

Appendix 11.1: Models and Anecdotes

1. Theory. We formulate the problem of comparing how ordinary people and experts make decisions in terms of competing methods of estimating regression equations. Both methods are idealizations. In the case of experts, we thereby avoid the consideration of the expert's private interest. Second, we represent the procedure of "ordinary people" as akin to the equilibrium condition of preference aggregation in a competitive, majority-rule democracy in which candidates locate their position in policy space so as to maximize their appeal (Downs 1957).[26] In this account, democracy is said to enable voters to aggregate their *desires* through voting. We examine the consequences of aggregating *information* in a similar fashion, although we cannot point to a competitive political process in which such an equilibrium is effected.

The most general question to ask in this comparative spirit is whether the decision function of ordinary people is admissible in this formulation. "Admissible" was first defined for a decision function by Wald (1950) in terms of a unknown distribution F, in the class Ω, and a known risk function $r(,)$. Here is Wald's definition:

> A decision function δ will be said to be admissible if there exists no other decision function δ^* which is uniformly better than δ, i.e., if there exists no decision function δ^* satisfying the following two conditions:
> $r(F,\delta^*) \leq r(F,\delta)$ for all F in Ω, and
> $r(F,\delta^*) < r(F,\delta)$ for at least one element of F of Ω. (Wald 1950, 15).

[26]The difficulty for median-voter equilibrium in high dimension preference aggregation (Plott 1967) will re-appear in our information aggregation account below.

Admissibility allows us to take into account costs which are consequences of the different rules by noting that *r(,)* is net of computational costs and the cost of sampling. It requires that we consider all possible alternatives to the one we propose to make sure that one does not dominate it. If we specify the specific alternatives we believe plausible then we can employ the tradition comparison of statistical methods in terms of relative efficiency, i.e., by comparing the number of observations, N_1, one procedure requires to match the precision of another procedure at a given N_2.[27] The efficiency question is the same in statistics as it is in economics: given the same input, what gives the highest output? The efficiency approach assumes, as is plausible in many important contexts, that the cost of observation and computation is invariant to the decision function itself.

Focusing on efficiency issues while keeping that of admissibility in the background, we will abstract from one advantage which goes to ordinary people. There are a lot of us. To conduct the argument in efficiency terms, we allow the experts' regression to have the same sample size (N) as the ordinary people regression. Since the standard errors of the regression techniques, in ideal cases, obey the $\sqrt{N}$ law, the reader can recompute these standard errors. Thus, in our experiments below there are N=5000 observations seen by experts and 2or 3 observations seen by each of 5000 ordinary people. But what if the expert only sees 50 while our 5000 ordinary people still have their limited ability?

---

[27]Here is a textbook description the efficiency a sample median relative to a mean at normality: "it means that in the normal case the median would require about n=157 observations to to achieve the accuracy that the mean achieves with n=100." Lehmann (1983, p. 360).

Obviously the standard errors of the expert's regression (in ideal cases) will fall by a factor of $\sqrt{100}$. In Table 11-2, column 1, the ideal condition for the expert finds the standard error of the voters is twice the expert's. But what would this matter for the question of admissibility of ordinary people's decisions making if, by taking account of the superior numbers of the ordinary people, we could inflate the standard error of the expert's model by a factor of 10?[28]

We propose to make one mathematical concept the foundation of our account, the device which is central to recent nonparametric thinking in statistics: an empirical probability distribution $\hat{F}$ which puts mass *1/n* at each point $x_1, x_2 \ldots x_n$. (Efron 1979 & 1981). Each point in the distribution may then be considered either as an observation or as an anecdote in a story. A model of these observations that satisfies the obligation of transparency is an estimate of location applied to all N (Andrews, *et al.* 1972). The plural of anecdote – so runs the joke – is data.

A transparent model might or might not be an anecdote. The mean of the empirical distribution is generally not an observation. The most obvious example where a model is an anecdote is a median of the empirical distribution when N is odd.[29] The anecdotality of other order statistics is obvious. There is of course more to models than

[28]Robin Hanson is thanked for this paragraph.

[29]When N is even, selecting one of Stigler's (1999) "co-medians" at random would allow the median of the empirical distribution to satisfy the anecdotal property.

estimates of location. Consequently, we turn to consider regression estimation.

We define [**ab**], [**bc**] and [**ac**] in Figure 11.1 as "anecdotal evidence," elements of a multi-dimensional empirical distribution. So defined, anecdotal evidence is collapsed (degenerate) regression. The problem with anecdotal evidence is that compared to the causal model, **OLS**, which gives a unique description of the relationship between X and Y, one can pick any of the combinations of anecdotes in a non-transparent fashion to obtain a wide range of relationships. Although OLS will not in general satisfy the anecdotal property, "elementary set methods" will (Farebrother 1997).

We consider the properties of centered anecdotal evidence – here [**ac**] – which we define as the median of anecdotal evidence [MAE]. Such estimators as what we are calling the MAE have i) desirable properties when the estimating model is mis-specified ii) and serve as a dependable starting point for an iterative procedure. One intuitively appealing method of determining this relationship is simply to find the slopes connecting [**a**, **b**], [**b**, **c**] and [**a**, **c**] and take an average of the three. Such was the first method of computing a regression in the 1750s by Roger Boscovich (Farebrother 1997).

We propose that [**a**, **b**], [**b**, **c**] and [**a**, **c**] are from the set of possible anecdotes from which ordinary people may draw. Each of these is a little story about X and Y; none of these involves an explicit recognition of randomness and so each might appeal to those who view the world as a deterministic narrative rather than a causal model with a random component.[30] The only randomness will be that different people will have different stories to tell. Our method of information aggregation will be to select the median.

Representing experts is straightforward. Instead of taking phenomena at surface value the expert posits a causal model, an ordering relationship with fixed α and β – $y_i = \alpha+\beta x_i+\varepsilon_i$ – for which all i=1..N of the empirical observations are subject. This too is illustrated in Figure 1 as the most popular of the standard regression techniques, ordinary least squares. Only when the reliability of the expert's ability to intuit an ordering relationship has been called into question have elementary set methods been revived. (Rousseeuw and Leroy 1987).

In terms of Figure 1 we allow the ordinary person access to a randomly selected pair of points – [**a**, **b**], [**b**, **c**] or [**a**, **c**] – from which a slope can be deduced. The resulting slope will be the opinion of the voter on the informational issue.[31] When we generalize beyond the simple regression we allow the voter access to a randomly selected k-tuple from which the appropriate slopes can be determined. The political decision will be made

---

[30]Though there were some exceptions which became increasingly important throughout the century (Peart 1995), the prevailing view among nineteenth century political economists and anthropologists was deterministic; see Hacking 1975.

[31]What if – as will happen – the randomly selected pair of points is [**a**, **a**], [**b**, **b**] or [**c**, **c**]? In the first version of the paper, we do not allow this to happen by checking the rank condition before we did the computations. The results reported below allow the singular or near singular results to be counted in the computations for the median. This results in slightly higher standard errors. Our current procedure raised a technical detail. Our implementation of *Shazam* (Whistler *et al.* 2001) computes the sample median by linear programming. Consequently, one must take care to skip over the occasional "Not a Number" returned by the compiler, an unnecessary step in a computation of the median by sorting. The same caution would be needed if the computations of the median were by a reweighted mean. Mosteller and Tukey (1977) discuss LAD computations in terms of a reweighted OLS regression.

on the basis of the median of the voter's opinion. In terms of the statistical literature the resulting technique is known as the median of pairwise slopes as extended beyond the context of simple regression with a random-selection algorithm.[32] This differs from Boscovich's 1750s approach mainly by the use of median instead of a subjectively-trimmed mean and a random selection process instead of an exhaustive combinatorial approach.

What properties might we expect of this procedure? Bias? Suppose that the expert's causal model is true – subject to a random disturbance, there is some simple linear relationship connecting the pairs (k-tuples). In that case the randomly-selected pair (k-tuple) is an ordinary least squares equation with sample size K. Making the assumptions that i) the median of the true errors is 0 and ii) these errors are independent of the right hand side variables, least squares is a median-unbiased estimator (Levy 1992b). If the estimated slope is b_i then the median expectation $M(b_i) = \beta$, so the mean-expectation is $E(M(b_i)) = \beta$. Thus the process of information aggregation which we model as an election procedure with the decision made by the median voter will be unbiased.

Under what conditions is an elementary set method more efficient than the standard regression? Here we propose to think about the class of elementary set methods as an estimation procedure without the benefit of the standard regression causal model. The "causal model" has a particularly interesting status in standard regression theory: it is

[32] Rousseeuw and Leroy (1987) systematically discuss the median of pairwise slopes in terms of the contribution of Henri Theil and latter workers.

supposed with hold with probability of 1 – the finite which we actually observe and infinite which we might potentially observe. This allows us to think about the model as a constraint added to a maximum likelihood method. Since it is a constraint with probability 1, it can be imposed *without* reducing the joint probability; indeed in a world of finite precision a probability 1 constraint can *increase* the likelihood (Levy 1988).

Thus, *when the model holds*, we cannot reduce the joint probability by imposing it even had we a maximum likelihood method of elementary sets. As we do not have such a method at our disposal, imposing the constraint and employing standard regression methods which are maximum likelihood, will generally result in a more efficient procedure.[33] Even if we had a method at our disposal – such a technique would presumably induce the model from the data instead of having the model imposed upon the data – it must be bounded by finite computational capability which is less good than the probability 1 insight. Thus, we have reason to believe that a maximum likelihood elementary set method will be less precise than a standard maximum likelihood approach when the causal model has probability 1 status. In this case we have reason to believe that an elementary set method would be inefficient, since adding a model constraint cannot reduce the joint probability.

---

[33]The passage from "maximum likelihood" to "asymptotic efficient" is more troublesome than one might like to believe, Lehmann (1983, pp. 403-27). Here is his review of the historical confusions: "Fisher's work was followed by a euphoric belief in the universal consistency and asymptotic efficiency of maximum likelihood estimation, at least in the iid case. The true situation was sorted out only gradually." (Lehmann 1983, p. 482).

Although Figure 1 supposes as a familiar illustration that the expert's method of estimating the casual model is ordinary least squares [OLS], the statement in the previous paragraph does not. There is no good reason to believe that least squares and its linear weighting of all elements of the empirical distribution would be the technique which would dominate an elementary set method. Indeed, least absolute deviations [LAD] at a heavy-tailed distribution might be arbitrarily more efficient than least squares.[34] Indeed, even if we suppose the causal model has probability 1 but allow the tail mass to get thicker we know that the efficiency of OLS falls relative to LAD and MAE. The simulations below consider the error distribution at the extremes of the t-distribution, normal to cauchy. When the model holds, MAE will never become as efficient as LAD; LAD is an admissible alternative to maximum likelihood at a cauchy which takes into account the probability 1 status of the model.[35]

But what if the probability 1 claim is simple vanity and the causal model holds only frequently? The constraint which helped with the causal model was probability 1 will hurt when it falls below this. We test this conjecture below.

2 *Monte Carlo Evidence*

Here, as we noted above, we employ Galton's idea that we can move between

---

[34] Steve Stigler is thanked for emphasizing this point.

[35] Andrews, *et al.* (1972) compare the maximum likelihood estimate of location of a Cauchy with the median and find very little gain in precision to repay the computational complexity. The text presumes we can move from location to regression context.

estimation and an idealized election, and so we present the argument as if we held a election to fit a regression. We consider a polity with 5,000 voters who form policy opinions by picking data points at random from 5,000 observations and computing pairwise slopes. Thus, each of the 5,000 voters in the first case is allowed to observe precisely 2 points. The decision is made by the median. This "regression by voting" is compared with two standard regression techniques instantiating causal models which our idealized experts might employ. The experts are allowed to observe all N=5,000. The experts have at their disposal two techniques. First is the maximum likelihood procedure where the errors are independent normal, OLS. The second is LAD, not quite maximum likelihood when the errors are independent Cauchy. Each experiment is replicated 10,000 times. All computations are carried out in *Shazam* 9.0 (Whistler, *et al.,* 2001). The basis of the comparison between "expert as estimator" and "ordinary people as estimators" is the mean of the estimating procedure and the standard error of the estimates. We have turned the contest alluded to in section 1 above, between friends of Plato and friends of Adam Smith, into a standard monte carlo study. We will have something to say about the bias of this procedure in the paragraph that concludes this section.

There are three conditions which we study. First, the experts are exactly right: there is a causal model and the errors are normally distributed. Second, the experts are perhaps right: there is a causal model and the errors are distributed Cauchy.[36] Third, the

---

[36] The normality assumption in the context considered is inconsistent with the possibility of extending the model by exploratory data analysis (Levy 1999/2000). Other error distributions allow EDA.

experts are very frequently right. The causal model holds 95% of the time but 5% of the time the signs of the regression coefficients switch. And by happenstance when this 5% occurs, the population from which the right hand side is drawn changes from a standard normal to a normal with a variance of 100. This third condition is that characterized by influential observations, the favorite drilling ground for those who would revive elementary set methods. The voting results in Table 11-2 are taken by the median of pairwise slopes. The results are in accord with predictions when the causal model is true: voting is unbiased and less efficient than the expert techniques. (Columns 1-1 and 1-2). When the model fails 5% of the time and these observations are influential, voting is the least biased and most efficient technique (Column 1-3). This is the appeal of elementary set methods.[37]

[37]Contrary to the suggestions in Rousseeuw and Leroy (1987). we find that LAD does not break down nearly as badly as OLS.

Table 11-2. Expert and Ordinary People, Simple Regression N=5000; Replications=10000			
Truth	p=1: y=1+2x+ ε ε ~ Normal	p=1: y=1+2x+ ε ε ~ Cauchy	p=.95: y=1+2x+ ε p=.05: y=-1-2x+ ε ε ~ Normal
Expert Belief	y=α+βX + ε	y=α+βX + ε	y=α+βX + ε
OLS estimate β	Mean = 2.00 Std. Dev. = 0.014	Mean = 1.35 Std. Dev. = 72.2	Mean = -1.36 Std. Dev. = 0.063
LAD estimate β	Mean = 2.00 Std. Dev. = 0.018	Mean = 2.00 Std. Dev. = 0.022	Mean = 1.38 Std. Dev. = 0.075
Voter Belief	y=α+βX	y=α+βX	y=α+βX
Voting on β MAE	Mean = 2.00 Std. Dev. = 0.027	Mean = 2.00 Std. Dev. = 0.047	Mean = 1.84 Std. Dev. = 0.031
Experiment	1-1	1-2	1-3

The case of simple regression is just that, simple. The causal model, such as it is, is hardly something requiring deep reflection so perhaps is it plausible that proverbial wisdom might encompass the idea that two variables related. What about high dimension problems where multiple regression techniques are required?

We now leave the simple regression context and consider multiple regression estimation. In Table 11-3 we present the crux of the matter: a multivariable situation where both X and Z impact Y. We allow the ordinary people insight into either X or Z but not both. In the first situation (column 2-1), X and Z are independent; in (column 2-2), the variables are correlated, in particular Z = .5X + .5 Q where both X and Q are standard normal. ε is a standard normal in both cases.

Table 11-3. Simple Regression in a Multivariate Context N=5000; Replications=10000		
Truth	$Y=1+2X+3Z+\varepsilon$ X, Z independent	$Y=1+2X+3Z+\varepsilon$ X, Z corrleated
Voter Belief	$y=\alpha+\beta X$	$y=\alpha+\beta X$
Voting on β MAE	Mean = 2.00 Std. Dev. = 0.08	Mean = 3.50 Std. Dev. = 0.047
Voter Belief	$y=\alpha+\gamma z$	$y=\alpha+\gamma z$
Voting on γ MAE	Mean = 3.00 Std. Dev. = 0.06	Mean = 5.00 Std. Dev. = 0.06
Experiment	2-1	2-3

What kind of proverbs might deal with higher dimensionality? Perhaps proverbial wisdom is embodied in parables, in stories, and what the causal modeler would call a dimension is handled by a character in a narration. We leave this for further study.

Returning to our competition (Table 11-4), we allow ordinary people to pick a random 3-tuple and solve the equation. To give the expert more of a chance, we cut the probability of contamination in half. (Column 3-3). The results are, from Table 11-2, easily predictable.

Table 11-4. Expert and Ordinary People, Multiple Regression N=5000; Replications=10000			
Truth	p=1: y=1+2x+3z+ ε ε ~ Normal	p=1: y=1+2x+3z+ ε ε ~ Cauchy	p=.975: y=1+2x+3z + ε p=.025: y=-1-2x-3z+ ε ε ~ Normal
Expert Belief	y=α+βX +γZ+ ε	y=α+βX +γZ+ ε	y=α+βX +γZ+ ε
OLS estimate β	Mean = 2.00 Std. Dev. = 0.014	Mean = 7.93 Std. Dev. = 492.6	Mean = -0.86 Std. Dev. = 0.169
OLS estimate γ	Mean = 3.00 Std. Dev. = 0.014	Mean = 2.41 Std. Dev. = 292.6	Mean = -1.29 Std. Dev. = 0.210
LAD estimate β	Mean = 2.00 Std. Dev. = 0.018	Mean = 2.00 Std. Dev. = 0.022	Mean = 1.85 Std. Dev. = 0.035
LAD estimate γ	Mean = 3.00 Std. Dev. = 0.018	Mean = 3.00 Std. Dev. = 0.022	Mean = 2.78 Std. Dev. = 0.036
Voter Belief	y=α+βX +γZ	y=α+βX +γZ	y=α+βX +γZ
Voting on β MAE	Mean = 2.00 Std. Dev. = 0.027	Mean = 2.00 Std. Dev. = 0.052	Mean = 1.96 Std. Dev. = 0.029
Voting on γ MAE	Mean = 3.00 Std. Dev. = 0.027	Mean = 3.00 Std. Dev. = 0.052	Mean = 2.94 Std. Dev. = 0.029
Experiment	3-1	3-2	3-3

Additional monte carlo simulations comparing combinations of cases – ignorant voters and arrogant modelers – give predictably mixed results.

It is perhaps necessary to emphasize that our construction assumes that anecdotal evidence is randomly selected. It is easy see to what sort of bias can occur if the polity were, for one reason or another, to select among a few anecdotes as, for example, when a large number of people obtain their information about other races from the same visualization. Propaganda is presumably the selection, if not fabrication, of such anecdotes.