Puzzles and Paradoxes Involving Averages:
An Intuitive Approach

Scott L. Feld[1]

Department of Sociology
Purdue University, West Lafayette, Indiana

Bernard Grofman

Department of Political Science
and
Institute for Mathematical Behavioral Sciences
University of California, Irvine

Abstract   Questions involving (re)aggregating parts into wholes are fundamental to all branches of scientific inquiry (Saari and Sieberg, 2001). After discussing two very general insights we derive from the aggregation literature, we look at how wholes may sometimes be reconstructed from their parts when we weight the parts appropriately, and we show that the notion of weighted averages can be used to make sense of seeming paradoxes: including how family income can go down even though per capita income is going up; how standardized test scores can go down even though the scores of every racial and ethnic group taking the test is going up; why most people think that roads are crowded even though most of the time there may be hardly any cars on the road; why teachers can teach mostly small classes, yet students can take mostly large ones; why George Bush could have won the Electoral College in 2000 even though he loses the popular vote; and why your friends can be expected to have more friends than you do.

Questions involving aggregating parts into wholes are fundamental to all branches of scientific inquiry. (Saari and Sieberg, 2001). For example, there is huge literature on aggregation issues in social choice theory (see reviews in Saari, 1994, 1995; Nurmi, 1999), and another one in statistics with statistical applications ranging from demography to brain imaging (see e.g., King, 1997) to information pooling (see e.g., Grofman and Owen, 1986). After a brief review of some insights from this literature we will focus on the implications of the simple idea of a "weighted" average, and show that understanding the properties of weighted averages allows us to understand some seeming paradoxes and puzzles involving composition effects. We conclude this essay with a few brief observations about other important types of averages most notably the geometric mean and the harmonic mean.

1 Three Insights into Aggregation

Before we proceed to a discussion of specific issues involving weighted averages, we wish to make two broad points derived from our reading of the aggregation literature, and then state the third insight that is at the heart of this essay

Insight 1: if data is broken into different pieces, the properties of the whole may be different from a summation performed on each of its parts, and exactly how we divide something into parts can matter a great deal.

Here we are not making any kind of metaphysical claim about "emergent properties." Rather we are simply observing that how we divide things up into pieces matters for the results, and, thus in particular, aggregating information from some of the possible piecewise divisions need not give us the same result as looking at the whole.

Most of the social choice discussion of this insight involves properties of majority rule and related voting rules such as so-called paradoxes like the *paradox of cyclical majorities*, *Hillinger's paradox, Anscombe's paradox, Ostrogorski's paradox*, the *referendum paradox*, and the *paradox of compound elections* (see e.g., Saari, 1994, 1995; Nurmi 1999; Saari and Sieberg, 2001). But in this essay we will be looking at the majority rule preference aggregation process only in passing. Instead, we will illustrate the insight above with a familiar mathematical operation, the *median*, since this illustration makes many of the same points, but without the "philosophic baggage" that comes with discussing terms like democracy or majority rule.

Consider the sequence of numbers 1 through 9. The median is, of course, 5. If we take these nine numbers and divide them into the three sets of three numbers each: $\{1, 2, 3\}$, $\{4, 5, 6\}$ and $\{7, 8, 9\}$, the three medians are 2, 5, and 8, and thus the "median median" is again 5. But now, divide the set into groups as $\{1, 2, 3\}$, $\{4, 6, 8\}$ and $\{5, 7, 9\}$. Now the medians are 2, 6 and 7, and so the median median is 6. We can also readily create groups whose median median is 4, e.g., $\{1, 2, 5\}$, $\{3, 4, 6\}$ and $\{7, 8, 9\}$. The alert reader will notice, however, that there are constraints on

how far away the median median can be from the overall median in this example, and we can work out the mathematics to define those constraints.

It is useful to consider the limiting case. Imagine a very very large set from 1 to n, n odd, to be divided into k still very large equally sized pieces, with k a divisor of n, such that each piece has s = n/k elements, with s also an odd number.[2] Let us put the s(k-1)/4 smallest numbers into the first (k-1)/2 pieces, and let us put the next s(k-1)/4 smallest numbers into the remaining (k – ((k-1)/2) = (k+1)/2) pieces, with the same number of elements in each piece. Now none of the smallest s(k-1)/2 elements can be a median median, since the median median will be a median in the last set of pieces, and in each of those the median element must be some element with a larger value, since a *minority* of the elements in each such piece come from the smaller numbers. But, with this "trick" in mind we can construct a partition in which the median median is simply the median among the largest (sk – (s(k-1)/2) = s(k +1)/2) numbers. But that means that, for n large, we can construct examples where the median median is ((n+1)/2 + n)/2 = (3n + 1)/4 $\cong$ 3n/4 . In a similar manner, we can work out a partition in which the median median is ((n+1)/2 + 0)/2 = (n + 1)/2 $\cong$ n/4.

More generally, we can "force" the median median to be (essentially) any number in the range from the first quartile to the third quartile in the distribution by choosing our partition appropriately. (With small n and k values we cannot quite get to these limits, as shown in the example we looked at earlier.)

Insight 2: the results we get when some portion of the information about the data is not available to us, need not be the same as when we have all the data to examine.

What we have identified as insight 2 seems like such an obvious point that one might wonder why we have listed it as a fundamental insight. But in fact there are examples where we can get unexpected insights from realizing this simple point (though usually only after doing a lot more thinking and sophisticated theoremizing about its implications).

Consider, for example, the work of vdr Hout, de Swart and ter Veer (2006). They axiomatize properties of pure list systems of proportional representation in terms of what they call the *plurality ranking rule*, namely one which "assigns to each combination of individual preference orderings of the parties a social ordering of those parties, where a party higher (receives more seats) when it is the first preference of more voters (receives more first votes)" (2006: 460) They look at a property they call consistency which says, very roughly speaking, that if a party "does better" in two disjoint subsets than does another party, then it should still do better than the other party when the subsets are combined into one, a key question in studying aggregation. They prove some theorems about what kinds of procedures satisfy consistency and some other properties generally thought to be normatively desirable such as neutrality, anonymity, and faithfulness (which looks at what would happen if

---

[2] We restrict ourselves to odd numbers so as to avoid complications caused by finding *the* median for an even number of cases.

there is only one voter, and requires that that voter's preferences be honored). For present purposes, however, their most important result has to do with party fragmentation. Roughly speaking, they show (2006: 466) that only proportional representation rules that make use of only first preference information are *party fragmentation proof*, i.e., such that splitting a party into two separate lists can never improve its summed seat allocation. Thus, they argue that a normative case can be made for rules that "throw away" information.[3]

On the other hand, Saari (1994, 1995) has made a strong case for the Borda rule to be used in choosing a single winner because that rule is the most attractive of all the "scoring rules" and scoring rules, unlike pairwise comparisons, take into account the entire structure of voter preferences and do *not* "throw away" information by only looking at pairwise comparisons. Moreover, by recognizing that pairwise comparisons throw away information, Saari argues that results such as Arrow's Theorem which require the condition of independence of (so-called) irrelevant alternatives, should not be regarded as either paradoxical or problematic for democratic theory.

Insight 3: while the whole is not the simple sum of its parts, sometimes we can reconstruct the whole from the parts by *appropriately* **weighting the parts.**

2 Parts and Wholes

How can it be that most households in the United States are headed by unmarried adults, yet most adults are married? How can family income be going down even though per capita income is going up? How can standardized achievement test scores (e.g. SAT scores) be going down over time even though the scores of every racial and ethnic group taking the test is going up? How can most people think that roads are crowded even though most of the time there may be hardly any cars on the road? How can it be that most classes at a university are small and yet nearly all the students find that most of their classes are large? How can George Bush win the Electoral College even though he got fewer votes than his opponent?

It might appear as if these puzzles have nothing in common. Certainly, they deal with totally different substantive arenas. Yet, as we show below, each of these apparent paradoxes can be resolved by understanding the notion of average in two different ways, (1) as an average over a whole, and (2) as an averaged average, i.e., an average over a set of parts. Since these parts are not necessarily equal to one another,

---

[3] I might also note that, when vdr Hout, de Swart and ter Veer (2006) use the term 'plurality ranking rule' to characterize list PR the implicit linkage they draw between plurality voting rules and list PR voting is at odds with how these systems are commonly treated in political science. In the electoral systems literature, with only a handful of exceptions (see Kurrild-Klitgard, 2008; Grofman, forthcoming) list PR systems and plurality systems are regarded as at opposite ends of a proportionality continuum, and their great similarity in focusing on only first place preferences tends to be overlooked.

we may think of an "averaged average" as a "weighted" average -- where the weights are related to the relative sizes of the parts. The two types of averages, weighted and unweighted, need not coincide and, indeed, can be very far apart.[4] Each type of average gives us a different "perspective" that helps us make sense of the world. [5]

2.1 Types of households

How can we explain that almost half of the households in the United States are headed by unmarried adults, yet most adults are married? Well, it's not really very complicated once we realize that every marriage has two partners. Consider a simple example: suppose that there are 15 single adults and 10 married couples, composing 25 households in a very small town. Here households either contain single persons or married persons. The majority of households (15/25) are headed by single people, but the majority of the people are married and dwelling in households headed by married people (20 married out of 35 adults). We can think of this in weighted average terms as follows: if we count people, then most are married; if we look at units (households), then to convert back to people we must weight the proportions of household containing married persons by the number of married persons in such households, and then normalize by the ratio of households to people (25/35). In other words, the proportion of adults who are married

---

[4] Because Simpson (1951) was perhaps the first to clearly state the apparent paradox of different outcomes for averages (or other features) of parts and whole, compositional paradoxes of the sort we review here are often known generically as *Simpson's paradox* (see e.g., Wainer and Brown, 2004). However, the statistical intuition goes back at least as far as Yule (1903), and the basic intuition has been well known for hundred of years, showing up for example in Eldbidge Gerry's manipulation of constituency boundaries in early 19th century Massachusetts to yield a majority of the seats with less than a majority of the votes, a now classic instance of what has of what has come to be called in the U.S., in Gerry's (dis)honor, the *gerrymander*. Nurmi (1999) refers to gerrymandering as a special case of what he calls the *referendum paradox*. We prefer to think of both gerrymandering and the *referendum paradox* as special cases of paradoxes involving weighted averages, and thus as special cases of what we might think of as the *generalized Simpson's paradox*. (See also our earlier discussion of compositional effects involving the median.) There are numerous essays that touch on Simpson's paradox. Some look at how Simpson's paradox is related to causal inference (in particular, the problem of confounding variables: see e.g., Pearl, 2000), or at other statistical issues (see e.g., Blyth, 1972; Samuels, 1993), while some look at instances of the paradox in various substantive domains (see e.g., Baker and Kramer, 2001; Wainer, 1986).

[5] Also, depending upon how we weight there can be different types of weighted averages for the same data.

= 20/35
= (2*10 + 0*15)/35
= 2*(10/25)*(25/35) + 0*(15/25)*(25/35).

So to convert back from averages based on units, here the proportions of units headed by a married person, to an average for the society as a whole, here the proportion of married persons in the society, we weight the units containing married persons by 2*(25/35) and we weight the units containing unmarried persons by 0*(25/35).

Note that we are not claiming that one type of average is correct and the other wrong. A sociologist studying marriage and divorce would certainly need to know the proportion of people who are married. For real estate agents, on the other hand, the fact that most households are headed by unmarried adults is what matters, because that tells them an important fact about the mature of the clientele for real estate rentals and sales.[6]

2.2 Family and per capita income

How can family income be going down even though per capita income is going up? Well, once again the key is to understand that units (here families) come in different sizes. Imagine a simple world. In the beginning, there are 15 families of mean size six, giving us a population of size 90, and average per capita income is $5,000. Here, average family income is $30,000. Some years later, there are 25 families of mean size four, giving us a population of 100 persons, and average per capita income is $7,000. In this second time period, average family income is $28,000. Even though society is getting much richer -- in the sense that total GDP has gone up by 56% (from $450,000 to $700,000) and per capita income has increased by 40% -- if we look at families they appear poorer. Why is that? Well, basically families are smaller in period two than they were to start, so when we convert per capita income to family income we would need to take into account family size.

We can again think of this in weighted average terms as follows: to convert from family income to per capita income, we simply weight family income by the ratio of families to persons. In other words, per capita income in time one

= $5,000
= ($30,000)* (15/90).

Similarly, per capita income in time two

---

[6] Also relevant to real estate agents is the average size of household. We would expect average size of household to differ between households head by someone married and those not.

  = $7,000
  = ($28,000)* (25/100).

Once again both per capita income and family income are meaningful numbers. But, if we compare family incomes for time periods when families are very different in average size, we are in effect, comparing apples and oranges. Also, if per capita income is increasing, but family size is going down at an even faster rate, so that the ratio of per capita income to persons per family declines, then we will observe a decline in family income even though the society may well be getting richer (as in the example above)! So we must be very careful to understand why the two indices of income don't need to go up (or down) in sync.

2.3 Standardized test scores

How can SAT (standardized test) scores be going down even though the SAT scores of every racial and ethnic group taking the test is going up? Well, what we have here is a compositional effect that can easily be understood in terms of weighted averages. The average SAT score is given by the scores of the various groups taking the test multiplied by (i.e., weighted by) the proportion of test takers coming from that group.

Consider a simplified world where we divide test takers into two groups, Hispanic and non-Hispanics and compare two different points in time. Imagine that, the SAT test scores for both Hispanics and non-Hispanics increased from time one to time two, but that in both periods the SAT scores of Hispanics are lower than those of non-Hispanic test takers. Imagine further that the proportion of Hispanic test takers increased between the two test periods. To make this example concrete, let us imagine that, at time one, Hispanic SAT Verbal scores are, say, 500, while non-Hispanic SAT Verbal scores are 600, and that 5% of the test-takers are Hispanic; while at time two, Hispanic SAT Verbal scores are 520 and non-Hispanic SAT Verbal scores are 610, but now 20% of the test-takers are Hispanic. The mean SAT score in the first period is .95*600+.05*500 = 595; while the mean SAT score in the second period is .8*610+.2*520 =   592. How could average scores in toto go down even though performance within each group of test-takers was going up? Well, quite simply, a higher proportion of the lower test-scoring group was now taking the exam. Overall average scores here are given by the *weighted* average of the scores of each of the groups taking the test, with the weights the proportion of test takers coming from that group.

There are numerous variants of this kind of compositional effect. Imagine, for example, that in a given university, within each graduate department, men and women are equally likely to be accepted relative to their proportions in the applicant pool. Yet overall, it still might be the case that, say, more women than men are accepted to the university if women and men do not apply to all departments in equal proportions, and if the acceptance rate is higher in the units where mostly women are applying. We can even construct hypothetical examples in which the GPA (or GRE)

score of students who are not admitted for graduate study in a given university is higher than the GPA (or GRE) of students who are accepted to that university All we need to do is to find a case in which the departments that admit (substantially) more students have (substantially) lower thresholds of acceptance in terms of under-graduate GPAs or GREs.

2.4 Crowded roads

How crowded are the roads? Well, if we took a snapshot of the roads at various times of day, we would find that a large part of the time the roads are (nearly) empty, and, thus, from this perspective, we would probably conclude that, on average, roads weren't very crowded. But now look at things from the perspective of the drivers. When they are out, it is likely that they will be on the road when it is crowded; thus, most drivers will experience crowded roads.

To see how this works imagine a very simplified example. Imagine that the roads are either empty, which they are 22/24ths of the time, or every driver is on the road (rush hour, coming and going), which occurs 2/24ths of the time. Now the average crowdedness of the roads is 8.3% relative to the total number of cars that might be there ($=1* 2/24 + 0*22/24$), but every driver thinks the roads are 100% crowded because that is their only experience! When the roads are empty there are no drivers on the road to experience that emptiness!

2.5 Class sizes

Assume, for simplicity, that, at a given university, each of its m faculty teaches the same number of courses, say k, and that the total enrollment in all courses taught in some semester is E. Clearly we have n, the number of courses, equal to mk. The average class taught by faculty, $\bar{S}$, thus contains E/n students. It must also be the case that, if the ith class is of size s_i, then

$$\bar{S} = \Sigma \, s_i/n \; . \tag{1}$$

But, how large is the class size experienced by the average student? Well, if the ith class is of size s_i, then exactly s_i students experience a class of that size. Thus, the average class size experienced by students is the student-weighted class average, i.e.,

$$\Sigma \, s_i^2 / \Sigma \, s_i \; . \tag{2}$$

 In general, these two numbers, weighted and unweighted averages (given by Eq. (1) and Eq (2) respectively), will not be the same, and they can be very far apart. Feld and Grofman (1977) used data from the classes at the State University of New York at Stony Brook to analyze this ratio for several different majors and for the university as a whole. For example, for the university as a whole, classes had a mean size of 40.5 with a standard deviation of 65.8. While faculty thought (correctly) that they

were teaching classes with an average of just under 41 students in each, students thought (correctly, as well) that they were taking classes which averaged over 150 students![7] {See text box below for mathematical details.)

The well known Herfindahl-Hirschman (H-H) index of concentration (Hirschman, 1945; Herfindahl, 1950; cf. Taagepera and Grofman, 1981), $1 - \Sigma p_i^2$, is a special case of a weighted average. In effect it is simply a size-weighted average, i.e., if the proportions in each of the various units are given by p_i, then the H-H index gives the average proportion weighted by *itself*, and then subtracted from one. Here, the larger the size of the bigger components in the distribution the smaller will be the H-H index. The paradox of class sizes discovered by Feld and Grofman (1977, 1980) is closely related to the Herfindahl-Hirschman index.

While the Feld and Grofman (1977) calculations are in terms of raw numbers, it is easy to convert their formulae to percentages. In particular, if we divide through by E, the total enrollment, then we get the average class size as a proportion of total class enrollment, $\bar{p}$, as being given by

$$\bar{p} = \Sigma p_i/n = 1/n \ . \tag{1'}$$

Similarly, the class size proportion experienced by the average student is given by

$$\Sigma p_i^2 / \Sigma p_i = \Sigma p_i^2 \tag{2'}$$

Feld and Grofman (1980) show that R, which is the ratio of Eq. (2) to Eq (1), or of Eq. (2)' to Eq (1)', is given by

$$R = 1 + \sigma^2 / \mu^2 \ . \tag{3}$$

Thus, the class size paradox (and also, we should note, the H-H index) can be directly linked to familiar ideas in statistics; namely the mean and the variance of a distribution (cf. Feld and Grofman, 2007).

A professor with a six course load who finds herself teaching, say, six classes with 75 students each may think it better to teach one 400 person class to allow her to offer five seminars to 10 students each. However, in her "improved" situation, almost all her former students (400 of the 450) are enrolled in a 400-person class, and the average class size from the perspective of the students is a whopping 356, even though the professor sees herself with the same average class size she had before, namely 75 students per class! Many universities, in part at the urging of their faculty, have "changed" their course offerings in this way, apparently without recognizing the very large consequences for their students, who see themselves at a multiversity despite the seemingly large number of small seminars the faculty (correctly) insist are being offered the student body!

---

[7] See also Feld and Grofman (1980).

2.6 Bush v. Gore: The 2000 U.S. presidential election

In addition to the claim that George W. Bush (and/or the U.S. Supreme Court) stole the election from Al Gore because of what happened in the Florida recount process, a frequent claim about the 2000 U.S. presidential election has been that Bush won the Electoral College vote despite Gore having won a plurality of the popular vote because of the biases in the Electoral College introduced by overweighting the small states. The Electoral College is a form of weighted voting, with states as units. Each state gets a number of Electoral College votes equal to the combined size of its House and Senate delegations. Of course, while House seats are allocated proportional to a state's population,[8] each state gets two Senate seats regardless of population.

It is true that, in the 2000 presidential contest, Bush did better than Gore in the smallest states, and Gore did better than Bush in the largest states, [9] which makes blaming the discrepancy between Electoral College and popular vote outcome plausible. [10] But even if (a) states were equipopulous and (b) each state's share of Electoral College votes had been perfectly proportional to its population, it is still possible for Bush to have won the Electoral College while losing the popular vote. Such an outcome could have happened under a winner take all rule for each state's electoral college vote if, Gore's *average* margin of victory in the minority of states that Gore won was much larger than Bush's average margin of victory in the majority of states that Bush won. Had this occurred, Gore would have wasted more votes than did Bush.[11]

To see how this could work consider the simplest possible example. Imagine that we have just three equipopulous districts, with equal turnout in each. Republicans win two with 60% of the vote, Democrats win one with 85% of the vote. The Democrats won more votes, yet capture a minority of the seats.[12] The knowledge-

---

Young, 1982).

[9] The correlation between state size (measured by the size of its congressional delegation in 2000) and the Bush share of the two party presidential vote in 2000 was only -.16.

[10] For the record, we should note that Bush's 2000 presidential victory was overdetermined, i.e., can be "blamed" on many factors, from failure by Gore to hold his home state of Tennessee, to excessive wasted votes by Gore in the Electoral College, to felon disfranchisement that disproportionately froze out potential Democratic-leaning voters (especially minority voters) from political participation in 2000, to Gore's legal team's mismanagement of the legal issues in Florida, to the "spoiler" candidacy of Ralph Nader.

[11] We can make the story more complicated by allowing for apportionment and turnout effects (Grofman, Koetzle, and Brunell 1997; cf. Grofman, Brunell, and Campagna,1997) .

[12] In actuality, however, Bush won his 31 states with an average of 58% of the two-party vote; while Gore won his 19 states (plus the District of Columbia) with an average of 57% of the vote.

able reader will immediately recognize that what we have here is a form of gerry-mandering (Grofman, 1990).

The 2000 House of Representatives was apportioned based on the 1990 census. Had the apportionment used in 2002 (based on the 2000 Census, and thus closer to the actual population figures in the states ca. 2000) been available for use instead, i.e., had the Electoral College allocation that will be used in 2004 been used in 2000, then Bush would have increased his Electoral College share from 50.5% to 51.8%. Indeed, even had the 2000 Electoral College allocated votes based just on the 2002 House seats (with no bonus for Senate seats), but again using a statewide winner take all system, then (putting Florida into the Bush column, as before), Bush would still have won the 2000 election with 50.1% of the votes. Thus, if all other things were equal, Bush might have been predicted (ca. 2002) to do better in 2004 than he did in 2000! And, of course, in fact he did.[13]

2.7 Friendship networks

Most of us find that we have fewer friends than some of our friends do. In fact, Feld (1991) shows why we ought to expect that the average person's average friend has a friendship network that is larger than her own. That is because the relatively few people with many friends include many of us in their large friendship nets. Conse-quently, we number among our own friends a disproportionately high proportion of people with many friends. When we "average in" the size of their very large friend-ship networks with those of our other friends with smaller rolladexes, we can still expect to find that, on average, friends of ours have a large friendship network than we do.

3 Different Types of Averages

Besides the distinction between weighted and unweighted averages we have empha-sized here, there are other types of averages than every social scientist should know about.

3.1 Median

Consider the *median*, the value such that half or more of the items in a distribution are at or above that value while half or more of the items in that same distribution

---

[13] Indeed, in 2004, even were there no Senate two seat "bonus", Bush could still carry a weighted majority of the states.

are at or below it as well. In addition to playing a central role in models of party competition over a single policy or ideological dimension (see e.g., Downs, 1957), there are useful things to say about the median versus the mean highly relevant to social science theory.

For example, Seymour Martin Lipset's famous thesis (Lipset, 1959) that rich countries are more likely to be democratic than poor countries is almost always tested with data on *mean* income (Diamond, 1992). But that's nonsense. The appropriate test is *median* income; otherwise you fall prey to the "Abu Dhabi fallacy" of thinking that a county with a few billionaires and lots of poor people is a rich country. In fact, unpublished work by one of the present authors has shown that essentially all of the current counterexamples to rich countries being democratic are countries where there are huge differences between mean and median income.[14]

3.2 *Geometric mean*

Another neglected type of average is the *geometric mean*, the square root of the product of the values to be averaged**.** Statisticians regard the geometric mean as appropriate when losses or gains can best be expressed in percentage terms; when rapid growth is involved in the development of a bacterial or viral population, or when the data span several orders of magnitude (Good and Hardin, 2003:96). Consider competing claims about the value of some quantitative variable where one side has an incentive to lie on the high side and the other side to lie on the low side**,** and the two claims are orders of magnitude apart**.** The physicist turned electoral systems specialist Rein Taagepera (personal communication, September 5, 2003) has proposed that. rather than taking the arithmetic average of the two values, we are more likely to come closer to the truth if we take the geometric mean instead.

Taagepera tells an amusing story to illustrate this point. While still a high school student in Morocco, Professor Taagepera heard two quite different estimates of the number of people killed in an incident involving French troops and Moroccan protesters. The protestors claimed that roughly 4,000 were killed; the French officials that only 40 were. Taagepera did not believe either estimate but he did believe that they provided bounds on the feasible range. If we simply took the arithmetic average of the two estimates to get a "best" estimate, we would get 2,020, far closer to the "official" than to the "local" estimates. However, we might more plausibly assume

---

[14] Some scholars have tried to "save" the Lipset thesis by restating it as one where countries that are rich and which *become* democratic are likely *to stay* so. But if we look at median income rather than mean income there is no need to "rescue" the thesis in this way. Moreover, the principal underlying mechanism implied by Lipset, namely the greater likelihood of there being a substantial middle class in wealthier countries than in poorer countries, makes far more sense if operationalized in terms of median income (or of other related features of the income distribution) than when it is operationalized by per capita income.

that each estimate is off by the same proportion, x, i.e., $4000/x = x/40$. Thus, $x^2 = 4000*40$ and the "best estimate" of the number of protesters killed is simply the square root of 16,000, i.e., 400. Hence, assuming each estimate is off by the same *proportion* gives us the rule that the "best estimate" is the square root of the product of the estimates, a.k.a. the *geometric mean*. As it turns out, the young Taagepera had a friend in the police force who was in a position to provide a more reliable estimate of deaths than the public estimate, and this estimate proved to be very close to the geometric mean calculated above.[15]

3.3 *Harmonic mean*

Lastly, we consider a useful use of the harmonic mean, the simplest case of which, for two values, s_1 and s_2, is defined as $1/(1s_1 + 1/s_2)$ perhaps the most obscure mean of all. We illustrate the uses of the harmonic mean with a Mother Goose tale. As is well known, Jack and Jill went up the hill to fetch a pail of water. Jack fell down and broke his crown and Jill came tumbling after. Now, imagine the hill is one mile high and Jack and Jill went up the hill at only two miles per hour, but they came tumbling down the hill at 6 miles per hour. What was their average speed on the hill? Let d be distance and s be speed. The answer is based on the harmonic mean of 2 and 6 (see figure below), and is given by the formula

$$d/ (1/s_1 + 1/s_2) = 2/(1/2 + 1/6) = 3 \text{ miles per hour}$$

Thus 3 miles per hour is the correct answer, not, the simple average -- four miles per hour.[16]

---

[15] For other uses of the geometric mean see Taagepera (2001).

[16] The figure below was drawn by a student, Gaelan Lloyd, in Professor Grofman's three quarter undergraduate statistics course to illustrate Lloyd's answer to a "Jack and Hill homework assignment" in the section of the course dealing with different types of averages. We are indebted to Mr. Lloyd for permission to reproduce the figure.
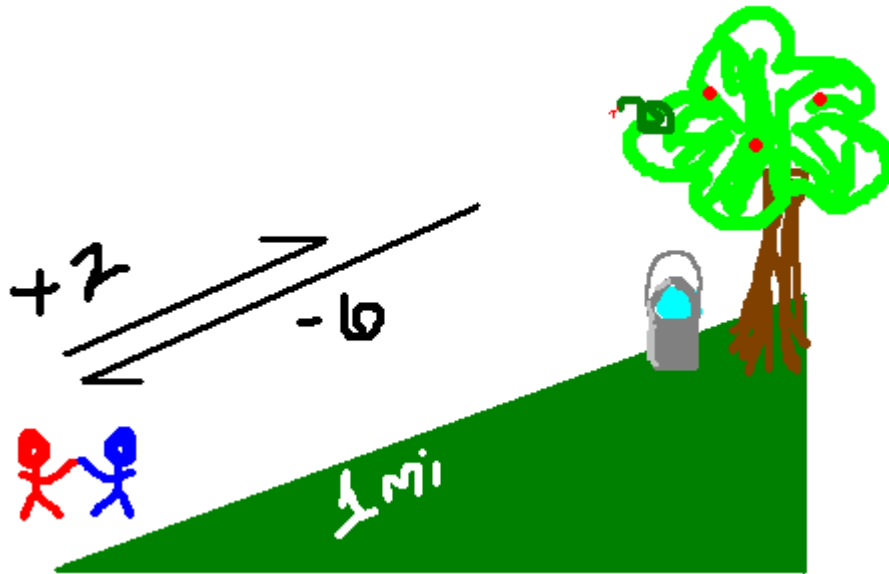
4 Discussion

Understanding the idea of weighted average helps us make sense of many sociological and social psychological puzzles of the sort discussed above, as well as puzzles in other social science disciplines-- such as how a majority can be outvoted on a majority of issues if we vote, not on policies one at a time, but instead on policy platforms combining multiple issues (Anscombe, 1976; Nurmi, 1999; Saari and Sieberg, 2001); or how a minority of the voters can control a majority of the seats in a legislature (see Grofman, 1990 and discussion of the Electoral College above). Moreover, the notion of weighted averages shows up in a number of places where you might not expect it, e.g., in the Hirschman-Herfindahl measure of concentration, which can be thought of as a "self-weighted" average. In sum, issues of aggregation and so-called aggregation paradoxes and the many varieties of average are all topics that should be part of the common wisdom of the educated social scientist.

References

Anscombe, G. E. M. 1976. On Frustration of the Majority by Fulfillment of the Majority's Will. *Analysis* 36: 161-168.

Baker, Stuart G. and Barnett S. Kramer. 2001. Good for Women, Good for Men, Bad for People: Simpson's Paradox and the Importance of Sex-Specific Analysis in Observational Studies. *Journal of Women's Health and Gender-Based Medicine* 19 (9): 867-872.

Balinski, M. L. and H. P. Young. 1982. *Fair Representation: Meeting the Idea of One Man, One Vote*. New Haven: Yale University Press.

Blyth, Colin. 1972. On Simpson's Paradox and the Sure-Thing Principle. *Journal of the American Statistical Association* 67: 364-366.

Diamond, Larry. 1992. Economic Development and Democracy Reconsidered. *American Behavioral Scientist* 35: 450-99.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper & Row.

Feld, Scott L. 1991. Why Your Friends Have More Friends Than You Do. *American Journal of Sociology* 96 (May): 1464-1477.

Feld, Scott L. and Bernard Grofman. 1977. Variation in class size, the class size paradox, and some consequences for students. *Research in Higher Education*, 6(3): 2l5-222.

Feld, Scott L. and Bernard N. Grofman. 1980. Conflict of interest between faculty, students and administrators: Consequences of the class size paradox. In Gordon Tullock (Ed.), *Frontiers of Economics* 3: 111-116.

Feld, Scott L. and Bernard Grofman. 2007. The Laakso-Taagepera Index in a Means and Variance Framework. *Journal of Theoretical Politics* 19 (1): 101-106.

Good, Philip I. and and James W. Hardin. 2003. *Common Errors in Statistics (and How to Avoid Them)*. New York: Wiley-Interscience.

Grofman, Bernard (Ed.). 1990. *Political Gerrymandering and the Courts*. New York: Agathon Press.

Grofman, Bernard. Forthcoming. Electoral Rules and Ethnic Representation and Accommodation: Combining Social Choice and Electoral System Perspectives. In

Brendan O'Leary (Ed.) *Power Sharing in Deeply Divided Societies: Progress and Regression*. Philadelphia: University of Pennsylvania Press.

Grofman, Bernard, Thomas Brunell, Janet Campagna. 1997. Distinguishing between the effects of swing ratio and bias on outcomes in the U.S. Electoral College, 1990-1992. *Electoral Studies* 16(4): 471-487.

Grofman, Bernard, William Koetzle, Thomas Brunell. 1997. An integrated perspective on the three potential sources of partisan bias: Malapportionment, turnout differences, and the geographic distribution of party vote shares. *Electoral Studies* 16(4): 457-470.

Grofman, Bernard N. and Guillermo Owen (Eds.). 1986. *Information Pooling and Group Decision Making*. Greenwich, CT: JAI Press.

Herfindahl, O. C. 1950. Concentration in the Steel Industry. Unpublished Ph.D. dissertation, Columbia University.

Hirschman, A. O. 1945. *National Power and Structure of Foreign Trade*. Berkeley: University of California Press.

King, Gary. 1997. *A Solution to the Ecological Inference Problem*. Princeton, NJ: Princeton University Press.

Kurrild-Klitgard, Peter. 2008. Voting Paradoxes under Proportional Representation: Evidence from Eight Danish Elections. *Scandinavian Political Studies* 31(3): 242-267.

Lipset, Seymour Martin. 1959. Some Social Requisites of Democracy: Economic Development and Political Legitimacy. *American Political Science Review* 53: 69-105.

Nurmi, Hannu. 1999. *Voting Paradoxes and How to Deal with Them*. Berlin: Springer Verlag.

Ostrogorski, M. 1913. *La Démocratie et l'Organisation des Partis Politiques*. Paris: Calmann-Lévy.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Saari, Donald G. 1994. *Geometry of Voting*. Berlin and New York: Springer Verlag.

Saari, Donald G. 1995. *Basic Geometry of Voting*. Berlin and New York: Springer Verlag.

Saari, Donald G. and Katri K. Sieberg. 2001. The Sum Of The Parts Can Violate The Whole. *American Political Science Review* 95 (June): 415-433.

Samuels, Myra. 1993. Simpson's Paradox and Related Phenomena. *Journal of the American Statistical Association* 88 (March): 81-88.

Simpson, Edward H. 1951. The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society, Ser. B* 13: 238–241.

Taagepera, Rein. 2001. Party Size Baselines Imposed by Institutional Constraints: Theory for Simple Electoral Systems. *Journal of Theoretical Politics* 13 (4): 331-354.

Taagepera Rein and Bernard Grofman. 1981. Effective size and number of components. *Sociological Methods & Research* 10(1): 63-81.

Van der Hout, E., H. de Swart, A. ter Veer. 2006. Characteristic Properties of List PR Systems, *Social Choice and Welfare* 27: 459-475.

Wainer, Howard. 1986. Minority Contributions to the SAT Score Turnaround: An Example of Simpson's Paradox. *Journal of Educational and Behavioral Statistics* 11 (4): 239-244.

Wainer, Howard and Lisa Brown. 2004. Statistical Paradoxes in the Interpretation of Group Differences. *American Statistician* 58 (2): 117-123.

Yule, G. U. 1903. Notes on the Theory of Association of Attributes in Statistics. *Biometrika* 2: 121–134