Practical Assessment, Research, and Evaluation

Volume 13 *Volume 13, 2008* Article 3

2008

What Does it Mean to Be Average? The Miles per Gallon versus Gallons per Mile Paradox Revisited

Antal Haans

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 13, Number 3, April 2008 ISSN 1531-7714

What Does it Mean to Be Average? The Miles per Gallon versus Gallons per Mile Paradox Revisited

Antal Haans, *Eindhoven University of Technology, The Netherlands*

In the efficiency paradox, which was introduced by Hand (1994; *J R Stat Soc A, 157,* 317-356), two groups of engineers are in disagreement about the average fuel efficiency of a set of cars. One group measured efficiency on a miles per gallon scale, the other on a gallons per mile scale. In the present paper, I argue against an operationalistic explanation of the efficiency paradox, by showing that the paradox is neither the result of an ambiguously defined efficiency concept, nor the result of how fuel efficiency is measured (i.e., whether a miles per gallon, or gallons per mile scale is used). The actual paradox is that the two groups of engineers have asked different statistical questions, while using the same mathematical operation. The paradox results from the fact that fuel efficiency is a derived measure, like density and speed, for which end-to-end concatenation (i.e., addition) is not straightforward.

A group of engineers from a multinational car manufacturer is called together to inform the executive staff about which of two types of car engines is the most efficient. Since no data were available to them, the engineers decided to do their own measurements. For this purpose, they took a representative sample of the two types of car engines, and measured, in the controlled environment of their laboratory, each car's fuel efficiency. The group of engineers consisted of two nationalities: English and French. The engineer who was responsible for the recording of the data happened to be an Englishmen, and because the English are accustomed to describing fuel efficiency in miles per gallon (m/g), he recorded each car's efficiency in m/g (see Table 1). The French engineers, however, are accustomed to describing efficiency in terms of gallons per mile (g/m). Because a m/g scale is not intuitively meaningful to them, the French engineers converted the cars' efficiencies into a g/m scale, by taking the inverse of each datum. Just hours before the engineers had to present their findings to the executive staff, they found themselves to be in disagreement. The English engineers found that, on average, Type I engines were more efficient than Type II engines (with efficiencies of 2.5 and 2.0 m/g respectively). The French, however, came to the opposite conclusion, having calculated that an average Type II engine is more efficient than an average Type I engine (with efficiencies of 0.5 and 0.5125 g/m respectively).

Table 1. Fuel Efficiencies of Two Types of Car Engines for the English and French Engineers.

	English (m/g)		French (g/m)	
	Type I	Type II	Type I	Type II
Car 1	1.0	2.0	1.0	0.5
Car 2	2.5	2.0	0.4	0.5
Car 3	2.5	2.0	0.4	0.5
Car 4	4.0	2.0	0.25	0.5
Average	2.5	2.0	0.5125	0.5

NOTE: m/g stands for miles per gallon, g/m stands for gallons per mile. The average refers to the arithmetic mean. If the reader is troubled by the small efficiencies of the cars, either multiply each value by ten, or, as Hand (2004) suggested, think in terms of military tanks rather than cars.

This fictive story contains one of the paradoxes, further to be referred to as the *efficiency paradox*, that David Hand (1994) discussed in his paper titled "Deconstructing statistical questions". In this paper, Hand argues that many statistical analyses are misdirected as the scientific question of interest (e.g., which of two types of car engines is the most efficient?) is not adequately translated into a statistical question. A statistical question includes, for example, the experimental design, sampling procedure, and statistical model (e.g., ANOVA or t-test). When the statistical question does not match the question of interest, researchers receive the right answer to the wrong question: if you ask for coffee, do not be surprised if you do not get tea! Such "errors of the third kind" (Hand, 1994; p. 317) especially occur when the scientific question is poorly formulated (as for example in Lord's paradox; Lord, 1967; see also Hand, 1994).

Several solutions to the efficiency paradox have been proposed in the literature, none of which, in my opinion, have been successful in explaining why the paradox occurs. In this paper, several of these solutions will be discussed, and an alternative explanation that is more tenable will be provided.

Against an Operationalistic Explanation

Hand (1994; also 2004) considers the efficiency paradox to be the result of the concept of fuel efficiency being ambiguously defined. As a result of this ambiguity, several alternative operationalizations of fuel efficiency are possible, each of which might lead to different conclusions. In other words, the paradox occurs because the numerical value that is assigned to a specific car by the English engineers (who use the m/g scale) is different from the numerical value that is assigned, to that same car, by the French engineers (who use the g/m scale). To resolve the paradox, Hand argues, both groups of engineers should have used the same system of assigning numbers. When determining which of the two operationalizations is the most appropriate, Hand (1994) concludes: "in using a car, we are generally interested in how many gallons it will take to cover a given distance (to travel from A to B) rather than how far we can travel on *x* gallons, before we run out of petrol. That being the case, the gallons per mile calculation will be the more appropriate (with the implication that the English are wrong!)" (p. 324). A similar line of reasoning is to be expected from the French engineers, who, perhaps with a dash of nationalism, will indeed persist that efficiency should be measured in gallons per mile. Equally patriotic, the English engineers will, however, persist on the m/g scale, since they have been using that scale for centuries already. Clearly, historical conventions will prevent the engineers from coming to an agreement about which scale is the most appropriate. In such a case, Hand (2004; also 1994) suggests that it might be best to only focus on the ordinal relations between the cars, which is possible since the order of the cars is the same for each scale (i.e., the m/g and g/m scales are monotonically related). Indeed, if one calculates the medians rather than the arithmetic means, the paradox disappears (i.e., The French and English engineers then agree that a Type I engine is the most efficient).

Although the solutions proposed by Hand (1994; 2004) resolve the efficiency paradox, the operationalistic explanation is unsatisfactory: the efficiency paradox is neither the result of an ambiguously defined efficiency concept, nor the result of how fuel efficiency is measured (i.e., in m/g or g/m units). First, the English and French engineers were in perfect agreement about what was meant with fuel efficiency. Both groups understood that the executive staff was interested in the relationship between distances traveled and volumes of fuel consumed, rather than, for example, in how efficiently the consumed fuel is actually used (i.e., the percentage of energy in the fuel that is transferred into shaft rotations). Secondly, there is no need to choose between the m/g and the g/m scale. It does not matter whether a particular car's fuel efficiency is advertised in m/g or in g/m units. A car with an efficiency of 2 m/g is as efficient as a car with an efficiency of 0.5 g/m: the prospective owner of the car will need 5 gallons of fuel to complete a 10-mile trip.

What might be confusing is that the two scales are not linearly related. The m/g scale is linear in respect to mileage: an increase of efficiency by one unit of m/g means that the car can drive one additional mile with each gallon of fuel in the tank. The amount of fuel that is saved by an increase of one unit of m/g depends, however, on where on the scale the car was initially located. Whereas an increase from 1 to 2 m/g saves you half a gallon per mile, an increase from 4 to 5 m/g saves you only 0.05 gallons per mile. The reverse is the case for the g/m scale, which is linear in respect to fuel consumption, but not mileage. Although the two scales are not linearly related, both satisfy the desired relationship between distances and volumes of fuel. As a result, measures taken on the m/g scale can be compared with measures taken on the g/m scale. To do so, all scores simply need to be expressed in the same metric (for example, m/g or euros per mile). If we compare the English average Type I car with the French average Type I car, then it is easily shown that the two must be different. In the Netherlands, the former car would cost you 2.55 euros per mile, whereas the latter would cost you as much as 3.26 euros per mile. Although both groups of engineers determined the average car by calculating the arithmetic mean from the data in Table 1, they must have been asking a different statistical question. The question is: what does it mean for a car to be average?

The Efficiency Paradox Revisited

Calculating the arithmetic mean is, for example, asking the question: if I have a set of *n* rods of different lengths that in an end-to-end concatenation measure *y* meters (i.e., when their lengths are summed), then what would be the length of an average rod, *n* of which together also measure *y* meters? Note that in this example, n arithmetic average rods can replace all the rods in the original set without changing the result of the end-to-end concatenation. Also, note that for the arithmetic mean, each rod contributes to the end-to-end concatenation only once (i.e., regardless of their lengths, all rods are given the same weight).

Compared to the length of rods, fuel efficiency is expressed in ratios of distances and volumes of fuel. Fuel efficiency is, in other words, what Campbell (1920) called a derived measure (like density or speed). The concatenation of derived measures is not straightforward, as is illustrated in the following classic problem. On a Sunday morning, Beryl sets out to visit her parents. On the trip to her parents' house, she drives with a speed of 60 km/h. On the trip back home, she drives with a speed of 100 km/h. What was Beryl's average speed on this round trip? Most people will give the arithmetic average of 80 km/h as the answer to this question (see, e.g., Lann & Falk, 2006). However, one cannot simply concatenate speeds, only the distances and durations from which they are derived. Assume that the distance to her parents' house is 60 km. The distance of the round trip (i.e., 120 km) should be divided by the total duration of the roundtrip (i.e., 1 hour in one direction and 0.6 hours in the other). Beryl's average speed, thus, is 75 km/h. In other words, two average speeds of 75 km/h, can replace all the speeds in the original set without changing the result of the end-to-end concatenation (driving the two trips one after the other). By calculating the arithmetic mean, one wrongly assumes that each trip is of the same duration. Instead, the trips should be weighted proportional to their contribution to the total duration of the round trip (which is similar to weighting each trip by the inverse of its speed). In this case, the correct average is the harmonic mean, not the arithmetic mean. The efficiency paradox is this classic problem in disguise.

By calculating the arithmetic means from the data in Table 1, the French and English engineers weighted all cars equally, regardless of their efficiencies. By doing so, the English engineers assumed that each car had an equal volume of fuel in the tank. In other words, the English engineers asked the following question. Take a set of *n* cars which, when each of the cars is given *x* gallons of fuel, can together travel a distance of *y* miles. What would be the efficiency of an average car, *n* of which can replace the original set of cars? In contrast, the French engineers assumed that, regardless of fuel efficiency, each car traveled an equal distance. In other words, the French engineers asked the following question. Take a set of *n* cars which, when each of the cars travels *y* meters, together consume *x* gallons of fuel. What would be the efficiency of an average car, *n* of which can replace the original set of cars? To answer the same question as the English engineers, the French should not have weighted each car equally. Instead, each car should have been weighted proportional to its contribution to the total volume of fuel that is consumed (i.e., should have been weighted by the inverse of its efficiency). In other words, to answer the same question as the English engineers, the French should have calculated the harmonic mean.

If the cars are assumed to have equal amounts of fuel in the tank, then the most efficient car contributes more to the total distance that the cars can travel, than when the cars are assumed to drive equal distances. Therefore, the English arithmetic average Type I car is more efficient than the French arithmetic average Type I car. Although both groups of engineers calculated the arithmetic mean, they have asked different statistical questions. At least one of two groups should have calculated the harmonic mean to resolve the paradox.

Discussion

I have argued against an operationalistic explanation of the efficiency paradox, by showing that the paradox is neither the result of an ambiguously defined efficiency concept, nor the result of how fuel efficiency is measured (i.e., in m/g or g/m units). Of course, other solutions have been proposed as well. Hand (2004), for example, suggests using a logarithmic transformation of the data, as ratios of positive values often show skewed distributions. The French engineers in our example, who do find such a skewed distribution (see Table 1), might have made a similar suggestion. Although a logarithmic transformation makes the paradox disappear (i.e., The French and English engineers will then agree that a Type I engine is the most efficient), the English engineers will no doubt object. They might argue that their data do not appear to be skewed, and that the French engineers themselves transformed the data into a heavily skewed distribution by taking the inverse of each datum (a transformation that, ironically, is commonly used for the normalization of data; see, e.g., Osborne, 2002).

The actual paradox, however, results from the fact that fuel efficiency is a derived measure, like density and speed, for which concatenation (i.e., addition) is not straightforward. Although formulated in a different way, the efficiency paradox is similar to the classic problem of averaging speeds (see, e.g., Falk, Lann & Zamir, 2005; Lann & Falk, 2006). By calculating the arithmetic average cars, the French and English engineers have asked different statistical questions, and at least one of two groups should have calculated the harmonic means to resolve the paradox. The question that remains is: who were right? Although this question falls outside the scope of the present paper, let us consider the tacit assumptions that the engineers made. The English engineers assumed that car owners purchase equal volumes of fuel, regardless of the efficiency of their cars. In contrast, the French engineers assumed that car owners travel equal distances, regardless of the efficiency of their cars. It appears, that the question should not be about which group of engineers is right, but about which assumption is the least wrong. It is to be expected that car owners adjust the distances they travel to the efficiency of their cars, i.e., that people with inefficient cars generally drive less miles (with the implication that the English engineers are the least wrong).

Jones (1994) argues in favor of calculating the geometric mean, since the geometric means of the m/g and g/m data are reciprocals of one another. Note that, calculating the geometric mean is asking the same question as taking the arithmetic mean of the logarithmically transformed data (which can be easily shown by expressing the two averages in the same metric). For the geometric mean, each car is weighted by the inverse of the square root of its efficiency (see, e.g., Hoehn, 1984). It remains, however, unclear why this specific weighting should be the most appropriate. As Hand (1994) stated: "... this average ... merely corresponds to yet another question that the researchers might really want to answer. It is not clear to me that this particular question is the one that is 'needed'." (p. 352). Perhaps, the most appropriate weighting is based on the actual relationship between fuel efficiency and car usage. Alternatively, one could, for example, calculate the arithmetic mean of the efficiencies (in m/g) of all cars of a certain type that stop for fuel at a certain petrol station, during a certain interval of time (i.e., a self-weighted average; e.g., Falk et al., 2005; Lann & Falk, 2006).

Derived measures, such as fuel efficiency and density, are frequently used, not only in physics and engineering, but in psychology as well: for example, body mass index, skin conductance response (also Hand, 2004), digit ratio, or sleep efficiency. When averaging derived measures, it is the specific weighting (and thus the assumptions on which this weighting is based) that determines the statistical question that is posed (for an overview of several means and their weightings, see, e.g., Hoehn, 1984). I hope that the present paper is a helpful complement to Hand's (1994) "Deconstructing statistical questions", so that more readers will take Hand's advice to stop and reconsider whether they did prefer the coffee, or were actually interested in the tea.

References

Campbell, N. R. (1920). *Physics: The elements.* Cambridge, UK: Cambridge University Press.

Falk, R., Lann, A., & Zamir, S. (2005). Average speed bumps: Four perspectives on averaging speeds. *Chance, 18,* 25-32.

Jones, M. C. (1994). Comment on D. J. Hand, Deconstructing statistical questions. *Journal of the Royal Statistical Society A, 157,* 344-345.

Hand, D. J. (1994). Deconstructing statistical questions (with discussion). *Journal of the Royal Statistical Society A, 157,* 317-356.

Hand, D. J. (2004). *Measurement theory and practice: The world through quantification.* London: Arnold.

Hoehn, L. (1984). A geometrical interpretation of the weighted mean. *The College Mathematics Journal, 15,* 135-139.

Lann, A., & Falk, R. (2006). Tell me the method, I'll give you the mean. *The American Statistician, 60,* 322-327.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68,* 304-305.

Osborne, J. W. (2002). Notes on the use of data transformations. *Practical Assessment, Research and Evaluation, 8*(6). Retrieved August 14, 2007 from http://PAREonline.net/getvn.asp?v=8&n=6.

Note

This research was supported by the JF Schouten School for User-System Interaction Research at Eindhoven University of Technology, Eindhoven, The Netherlands. The author thanks Steven Ralston for his language support.

Citation

Haans, Antal (2008). What Does it Mean to Be Average? The Miles per Gallon versus Gallons per Mile Paradox Revisited. *Practical Assessment Research & Evaluation*, 13(3). Available online: http://pareonline.net/getvn.asp?v=13&n=3

Practical Assessment, Research & Evaluation, Vol 13, No 3
Haans, The Efficiency Paradox Revisited

Author

Antal Haans
Eindhoven University of Technology (IPO 1.26)
P.O. Box 513
NL-5600 MB Eindhoven
The Netherlands

A.Haans [at] tue.nl