The Early History of Average Values and Implications for Education

Arthur Bakker
Freudenthal Institute, Utrecht University

Journal of Statistics Education Volume 11, Number 1 (2003),
ww2.amstat.org/publications/jse/v11n1/bakker.html

Copyright © 2003 by Arthur Bakker, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the author and advance notification of the editor.

Key Words: Design research; Guided reinvention; History of statistics; Measures of central tendency; Middle school.

Abstract

The early history of average values is used as a source of inspiration for instructional design in middle-school classrooms. This historical study helps to define different layers, aspects, and applications of average values and encourages us to look through the eyes of the students, who do not have the same concepts as teachers and instructional designers have. As a result of this study, possible implications for education are considered, such as estimation as a starting point for a statistics course, allowing the midrange as an initial strategy, a visual way of estimating the mean using bar representations in a simple computer tool, and the reinvention of midrange, median, mode, and mean. There turn out to be striking parallels but also important differences between the historical and students’ individual development of statistical understanding.

1. Introduction

Various authors have suggested that studying the history of a topic is good preparation for teaching that topic (Freudenthal 1983b; Radford 2000; Stanton 2001). Obstacles that people in the past grappled with are interesting to teachers because students often encounter similar obstacles. However, students also know things people in the past did not know. In the words of Freudenthal (1983b, p. 1696):

“The young learner recapitulates the learning process of mankind, though in a modified way. He repeats history not as it actually happened but as it would have happened if people in the past would have known something like what we do know now. It is a revised and improved version of the historical learning process that young learners recapitulate. ‘Ought to recapitulate’ - we should say. In fact we have not understood the past well enough to give them this chance to recapitulate it.”
As part of my research in statistics education, I had to design instructional materials for seventh-grade students who had not learned any statistics except the arithmetic mean. To investigate what the "revised and improved version of the historical learning process" could be for these students, I decided to study the early history of statistics, in particular of averages, as well as students' learning.

The study consisted of exploratory interviews with 26 seventh-grade students of different ability and five classroom-based teaching experiments in seventh-grade classes with 12 to 15 lessons per experiment. I audiotaped all interviews and lessons and videotaped the lessons in the last two classes. In five cycles of developing instructional materials, observing classrooms, and revising the materials, the teacher and I were able to develop an instructional unit that helped students learn the concepts of average values through a process of what Freudenthal (1991) called "guided reinvention." This process is consistent with the historical development of the concepts and takes account of the contextual difference between current students and the historical past.

As a framework to study the relation between history and education, I used Freudenthal’s historical and didactical phenomenology, as described in Section 2. The subsequent sections deal with historical and didactical issues related to average values. The last section summarizes the insights about the relation between the historical learning process and the individual learning process for averages.

2. Framework

Freudenthal (1983a) distinguished phenomena that we want to understand or structure and the concepts with which we do so. We can study the relation between phenomena and concepts in different ways. Historical phenomenology is the study of the historical phenomena in which certain mathematical concepts arose in order to understand why these arose. Didactical phenomenology is the study of the relation between the mathematical concepts and the phenomena in which they arise with respect to the process of teaching and learning these concepts and their applications. Studying history can help us see certain phenomena through the eyes of people who did not have the same concepts and techniques. This may help us take a students’ perspective and better understand their learning process.

My approach was as follows. I first collected as many early historical examples with a statistical flavor as possible. The next selection of historical examples was guided by their educational interest: I only selected examples that could be seen as preliminary stages of statistical notions with possible relevance for teaching younger students. For example, if an estimation of a number of years is reached by some method that could be interpreted as an intuitive version of an average, it was included. Estimation as a simple guess would not have been included. Subsequently, the teacher and I conducted the teaching experiments and looked for parallels and differences between the historical and individual learning processes. In the organization of this article, there is an alternation between the two processes to demonstrate their relationship.

3. The Early History of Estimation

The oldest historical examples that were found all had to do with estimation. Four of them are presented here to show preliminary stages of some measures of center.

Example 1
In an ancient Indian story, Rtpurana estimated the number of leaves and fruit on two great branches of a spreading tree (Hacking 1975). He estimated the number on the basis of one single twig, which he multiplied by the estimated number of twigs on the branches and found a number, which after a night of counting turned out to be very close to the real number. It could well be that he chose a typical or an
average twig, since that would indeed give a proper estimation. This then may be seen as an intuitive predecessor of the arithmetic mean, because one average number represents all other twig numbers and this average number is somehow “in the middle” of the others. The choice is presumably made so that what is counted too much on the one hand is counted too little on the other hand. This use of an average, in our modern eyes, has to do with compensation, balance, and representativeness.

**Example 2**
Another example of estimation stems from the Greek historian Herodotus (circa 485-420 BC) on the Egyptians and was found by Rubin (1968, p. 31):

“They declare that three hundred and forty-one generations separate the first king of Egypt from the last mentioned (Hephaestus) - and that there was a king and a high priest corresponding to each generation. Now reckon three generations as a hundred years, three hundred generations make ten thousand years, and the remaining forty-one generations make 1,340 years more; thus one gets a total of 11,340 years ...”

The statistically important point in this quotation is the assumption that three generations was reckoned a hundred years. This assumption was made to estimate the total amount of years between the first Egyptian King and Hephaestus. Of course, three generations were not always exactly a hundred years; sometimes a little less, sometimes a little more, but the errors are roughly evened out. That is why this method may be seen as a preliminary stage of the development of the average. As in the first example we see the aspect of compensation and representativeness (typical number of years for generations).

Rubin also found other old examples of statistical reasoning in the work of one of the first scientific historians, Thucydides (circa 460-400 BC). The quotations in Examples 3 and 4 are from his *History of the Peloponnesian War*. The reader is invited to decide how he or she would translate these two excerpts into modern statistical terms (Rubin 1971, p. 53; translation by R. Warner).

**Example 3**

“(The problem was for the Athenians) ... to force their way over the enemy’s surrounding wall ... Their method was as follows: they constructed ladders to reach the top of the enemy’s wall, and they did this by calculating the height of the wall from the number of layers of bricks at a point which was facing in their direction and had not been plastered. The layers were counted by a lot of people at the same time, and though some were likely to get the figure wrong, the majority would get it right, especially as they counted the layers frequently and were not so far away from the wall that they could not see it well enough for their purpose. Thus, guessing what the thickness of a single brick was, they calculated how long their ladders would have to be ...”

**Example 4**

“Homer gives the number of ships as 1,200 and says that the crew of each Boetian ship numbered 120, and the crews of Philoctetes were fifty men for each ship. By this, I imagine, he means to express the maximum and minimum of the various ships’ companies ... If, therefore, we reckon the number by taking an average of the biggest and smallest ships ...”

In Example 3, we could see an implicit use of the mode, here indicated by “the majority.” Note that “the majority” probably means “the most frequent value” and not necessarily “more than half.” In this situation, the Greeks assumed that the most frequent number would be the correct one. To find the total height of this number of bricks, they supposedly needed another estimation, namely of the expected or
the average thickness of a single brick.

In Example 4, we again see estimation with the help of an average value. Thucydides possibly interpreted the given numbers as the extreme values, so that the total amount of men on the ships can be estimated. He suggested that taking the average of these two extremes would provide an estimate. In fact this is called the midrange, defined as the arithmetic mean of the two extremes. Rubin (1971, p. 53) writes about this:

“This technique of averaging the extreme values of the range to obtain the arithmetic mean or mid-range can be justified if certain assumptions are defensible, i.e., that the underlying distribution is at least approximately symmetrical or rectangular.”

Alternative translations do not give other interpretations on these statistical issues (Rubin 1971; Thucydides 1975). Thomas Hobbes, in the oldest translation from Greek into English (1629), only uses the word "mean" where Rex Warner uses "average."

We encountered certain phenomena that were organized by predecessors of contemporary statistical concepts. In Examples 1, 2, and 3, a kind of average similar to the arithmetic mean was used. In Example 3, we could also recognize the mode. In Example 4, Thucydides described a method that we could call taking the midrange. Note that these notions were not defined or used explicitly, although many mean values were known in those days (see Section 5); the literature on the history of statistics indicates that precursors to the median before 1599 are very unlikely (Eisenhart 1974).

4. Students’ initial views of the average

The historical examples illustrate that it can be rather difficult to make implicit aspects of average values explicit. Did Thucydides really think of the midrange in the second quotation? In exploratory interviews, I encountered a similar difficulty of translating student arguments into statistical terms. From only 26 students I got a rich variety of answers to the question, “Do you know what the average is?” (In Dutch there is only one word for average and mean, "gemiddelde," which has both the informal and the statistical meaning of average.) All of these students had learned the arithmetic mean but no other statistical notion. They therefore have a different background than most American seventh-graders, for example, who learn mean, median, and mode years before. Consider the following sample of student responses to the question "What is the average?"

| What is the average?                  | Possible statistical interpretations                      |
|---------------------------------------|----------------------------------------------------------|
| Jennifer: The half. The whole, and in between the half, that is the mean. | Part of the algorithm: dividing by two.                  |
| Charissa: Everything together.        | Part of the algorithm: adding all values                  |
| Bart: You look between the highest and lowest. | Midrange? Somewhere in between?                           |
| Centina: The most.                    | Mode, typical?                                           |
| Claire: What you think it is roughly. | Estimation or representativeness                         |
| Lisa: The mean is about a bit in balance. | Balance point                                            |
| Kerster: In between.                  | Midrange?                                                |
| Frank: The midpoint.                  | Midrange, median or center of gravity?                   |
| Others: Add and divide by the number. | Algorithm                                                |
If we have a closer look at some of the student answers, it becomes clear that there is often no unique statistical interpretation. If Frank says “the midpoint,” does he refer to the point in the middle of the lowest and highest value, to the middle-most value, to the point with minimal absolute distances to the others, or to a center of gravity? We cannot be sure.

This illustrates that for students there are no clear distinctions between the different aspects of these notions. Early in history, we see the same phenomenon (recall the examples in Section 3). When people organize their world and solve problems, they are urged to become clearer and define their methods more precisely. For instance, with symmetrically distributed data, there is no evident need to distinguish the midrange, median, mean, or balance point. For instructional materials, we must therefore choose contexts that ask for clear distinctions. For example, a skewed distribution can show the limitation of the midrange and require a different measure of center.

The observation that the oldest historical examples had to do with estimation gives rise to the question, “Is estimation a good starting point for statistics education?” The answer from these experiments turned out to be yes. As the historical examples show, estimation involves many qualitative aspects of the average that are neglected if students only learn the algorithm of adding all values and dividing by the number of values. These qualitative aspects include representativeness, somewhere in the middle, balance, and compensation (also see Strauss and Bichler 1988; Mokros and Russell 1995). At first sight, “somewhere in the middle” may seem very vague, but the just mentioned research shows that many students tend to forget even this intermediacy property of the mean when they calculate it blindly.

Another observation from history is that in estimations the average was used to find a total number. I expected that students would reinvent some aspects of the average if they had to estimate total numbers. In many textbooks, the order of teaching is reversed: students first learn to compute the mean and then have to discover in which situations they can apply the mean sensibly.

For education, the historical examples needed revision in two ways. First, most students we worked with already knew the algorithm of calculating the mean, although they often did not understand it well. Second, the historical contexts were not suitable for instruction: how many students would be interested in estimating the number of years between the first Egyptian king and Hephaestus? For the first task of the course I decided to choose a context that would be more appealing to students, namely estimating the number of elephants in a picture (see Figure 1, from Boswinkel et al., 1997).
The students in all classes used four main strategies with some variants:

1. Make groups, guess how many there are in each group and add all numbers.

2. Make a group with a fixed number and estimate how many groups fit into the whole (in Figure 1 part 2 the students estimated groups of 10).

3. Count the number of elephants in the length and width and multiply these (readers who have seen the video "Goodnight Mr. Bean" may recognize his method of counting sheep).

4. Make a grid, choose an “average box” and multiply by the number of boxes in the grid.

Strategy 4 is indeed based on an intuitive sense of the average. When the teacher and I asked the students what they meant with an average box they described it as “a box with not too little and not too many”. A similar description can be found in Aristotle's *Nichomachean Ethics* (*Aristotle 1994*).

### 5. Mean values in Greece

Aristotle (384-322 BC) mentions the arithmetic mean, but also defines a philosophical form of the mean, namely the “mean relative to us.” With this notion he explains what virtue is. About the difference between the arithmetic mean and “the mean relative to us” he writes (*Aristotle 1994*, Book II, Chapter VI, p. 5):

"By the mean of a thing I denote a point equally distant from either extreme, which is one and the same for everybody; by the mean relative to us, that amount which is neither too much nor too little, and this is not one and the same for everybody. For example, let 10 be many and 2 few; then one takes the mean with respect to the thing if one takes 6; since 10-6 = 6-2, and this is the mean according to arithmetical proportion [progression]. But we cannot arrive by this method at the mean relative to us. Suppose that 10 lb. of food is a large ration for anybody and 2 lb. a small one: it does not follow that a trainer will prescribe 6 lb., for perhaps even this will be a large portion, or a small one, for the particular athlete who is to receive it; it is a small portion for Milo, but a large one for a man just beginning to go in for
athletics.” (Italics added)

Later in the same chapter he writes about virtue (Aristotle 1994, Book II, Chapter VI, p. 9):

"Virtue, therefore, is a mean state in the sense that it is able to hit the mean."

For him, the mean relative to us was an ethical ideal. In this way, Aristotle extended the notion of the mathematical means to situations in daily life, though his goal with it was different from what we mostly aim for in statistics.

The description “not too much and not too little” is one that students used in all five teaching experiments. When these students explained their strategies, they defined “an average box” in a grid as a box in which there were “not too many and not too few” elephants. Though it could refer to “in the middle,” balance, or compensation, this expression is vague. Therefore it is crucial to continue to ask questions. We asked, for instance, what would happen if we chose a box with too few elephants? More follow-up questions appear in Section 8.

The arithmetic mean was not the only mean value known to the Greeks. In Pythagoras’ time, around 500 BC, three mean values were known, namely the harmonic, geometric, and arithmetic mean (Heath 1981; Iamblichus 1991). At least eleven different mean values had been defined only some 200 years later (Heath 1981). For a historical phenomenology it is relevant to study the phenomena that gave rise to these concepts. It turns out that the theory of the three mentioned mean values was developed with reference to music theory, geometry, and arithmetic.

Examples of the mean values in music theory and geometry show that we cannot always simply translate history into education. First, consider the musical proportions 6:8:9:12 on a string. The proportion 6:8 = 9:12 as a musical interval is called a fourth, 6:9 = 8:12 is a fifth, and 6:12 an octave. All these proportions form consonant intervals. Moreover, 8 is the harmonic mean of 6 and 12, and 9 is their arithmetic mean. The proportion 8:9 is a second, which is dissonant. This example shows a historical relation between a phenomenon, musical intervals on strings, and the related concepts, namely proportions and means. It also demonstrates that what comes early in history need not come early in education.

Second, an example from geometry, namely a theorem of Pappus, shows that the Greeks studied the mean values for their beauty (see Figure 2). If in the semicircle ADC with center O one has DB perpendicular to AC and BF perpendicular to DO, then DO is the arithmetic mean, DB the geometric mean, and DF the harmonic mean of the magnitudes AB and BC (Boyer 1991). Clearly, this theorem does not belong in a statistics course at the middle-school levels. These examples also demonstrate that what might look statistical from the term "mean value" need not be statistical at all.

---

**Figure 2.** Theorem of Pappus on arithmetic, geometric, and harmonic mean.
Some other aspects of these mean values, however, do have important didactical implications. Greek mathematics had a different form and aim than modern mathematics, because it was highly geometrical and visual: for instance, lines were used to represent numbers and magnitudes (see Figures 2 and 3). This difference between Greek and modern mathematics can also be illustrated with the difference in definitions of the arithmetic mean. The Greek definition, as we saw in the quotation of Aristotle is as follows: the middle number $b$ of $a$ and $c$ is called the arithmetic mean if and only if $a - b = b - c$. Note that this definition differs from the modern one, $(a + c) / 2$, and that it refers to only two values. The Greek version shows that the mean is in between the two extremes and is difficult to generalize, whereas the modern version highlights the calculation and is easy to generalize. With education in mind, it is important to note that the Greek definition shows other qualitative aspects than the modern quantitative one. For example, we can immediately see from the Greek definition that the mean is halfway between the two other values.

How can we benefit from the Greek representation and the generalization of the modern definition? An answer follows in the next sections.

---

A __________
B ______________
C ______________

Figure 3. B as the arithmetic mean of A and C, in the way that numbers are represented in Euclid’s Elements around 300 BC (Euclid 1956).

---

6. Bar representations

Many students forget this intermediacy aspect if they calculate the mean and sometimes get an answer outside the range, as illustrated with one of the exploratory interviews. The students had a bar graph and a table of average monthly temperatures in front of them.

Interviewer How would you estimate the average annual temperature in the Netherlands from this graph or table?
Jennifer Add everything.
Interviewer And then?
Lisa Divide by 2.
Interviewer Why divide by 2?
Lisa Because that is the average.

They got 55 degrees Celsius and did not realize that this was very hot. When I again asked what the average temperature was, Jennifer again divided by 2 roughly; "25," she said. If students develop a visual estimation method with more than two, this kind of dividing by two would probably happen less often.

Bars proved to be helpful representations when developing understanding of the mean. When asked to
estimate the average annual temperature in the Netherlands from a bar graph, some interviewed students spontaneously came up with a compensation strategy. They said: “I give a bit of July to January, from August to February and so on.” The teacher and I used this “leveling out” idea successfully in the subsequent teaching experiments. A simple computer tool, called *Minitool 1*, designed by Gravemeijer, Cobb, and colleagues ([Cobb 1999](#)) scaffolded this reinvention process. An activity that stems from their teaching experiments is about the life span of batteries. Students had to decide which of two battery brands was the better choice, using *Minitool 1*. Students used different arguments such as “Brand K had more higher values,” “Brand K has outliers,” “Brand D is more reliable,” and “Brand D has a higher mean.” In the Dutch experiments, students indeed developed a strategy of estimating means with the so-called value bar (see [Figure 4](#)). Students use their imaginations to "move" the pieces that stick out at the right of the value bar to the shorter bars left of the value bar.

In answer to the question of how to benefit from the Greek representation and the generalization of the modern definition, we may conclude that we might use the Greek bar representation but with more than two values.

![Figure 4. Compensation strategy in the battery problem (Brand D on the left, Brand K on the right).](#)

### 7. Fair share, midrange, and generalization to $n$ values

The above-mentioned classroom observations on compensation have their counterparts in history. The average has to do with fair share in trade and insurance contexts, and taking the mean of only two extreme values, the midrange, was a predecessor of the arithmetic mean of more than two values in the context of astronomy. These are the topics of this section.

In the first millennium before Christ, there was lively sea trade in the Mediterranean ([Plön and Kreutziger 1965](#)). Small vessels sailed with valuable merchandise from harbor to harbor, carrying for instance grain, oil, wine, condiments, amphorae, cloth, obelisks from Egypt, slaves, and even animals for the circus such as elephants and lions. The captains of these ships encountered many dangers such as sudden thunderstorms, cliffs, piracy, sand banks, and loading merchandise into smaller boats to bring it to the shore. The danger of capsizing sometimes forced them to cut away the mast. During a storm, the captain sometimes decided to throw some cargo overboard to save the rest of the cargo. This throwing act became known as "jettison" of cargo.

From about 700 BC, merchants and shippers agreed that damage to the cargo and ship should be shared
equally among themselves. What a merchant had to pay was called his contribution. The set of rules became customary law, now known as the Rhodian Sea-Law (Ashburner 1909). By the seventh or eighth century AD, there was a Greek text with many rules on what should be done in different situations. For example, rule 9 out of 47 in part III reads as follows:

“9. If the captain is deliberating about jettison, let him ask the passengers who have goods on board; and let them take a vote what is to be done. Let there be brought into contribution the goods ...” (Ashburner 1909, p. 87)

The Roman emperor Justinian I (483 - 565) became famous by his order to collect all available laws, the codification of Roman Law in 534, now known as the Digest of Justinian. One part became known as the "lex Rhodia de iactu," the Rhodian law on jettison. The basic principle in the Digest XIV.2.1 is:

“The Rhodian law decrees that if in order to lighten the ship merchandise has been thrown overboard, that which has been given for all should be replaced by the contribution of all.” (Lowndes and Rudolf 1975, p. 3)

The rest of the text explains what should be done in specific situations and raises questions such as, “In which proportion should be paid?” In Digest XIV.2.2.4, it is written that the equalized portion should take into account what the value of the saved and the lost cargo was. For the price of the lost cargo, the purchase price should be reckoned, but for the saved cargo, the selling price should be estimated. This implies that calculations of the fair share must have been complicated, if the contributions had not just been estimated.

The number examples in the Latin texts are extremely simple and not very explicit. For example, in Digest XIV.2.4.2 we read:

“If therefore, for instance, two persons each had merchandise valued at 20,000 sesterces and one lost 10,000 due to water damage, the one with the saved merchandise should contribute according to his 20,000, but the other on the basis of the 10,000.” (Spruit 1996; translation AB)

We can find more realistic examples in books of the nineteenth century that tell how to calculate averages (see, for example, van der Hoeven 1854; Hopkins 1859). These averages were calculated by a so-called "average-adjuster," which was a kind of accountant. It must have been a serious profession, because there was even an Association of Average Adjusters in the nineteenth and early twentieth century (Lowndes and Rudolf 1975). From the examples in this section we see that the average in this sense originally had to do with fair share and insurance, but how did the term "average" also come to signify the arithmetic mean?

The Oxford English Dictionary (Simpson and Weiner 1989) writes that one of the meanings of "average" in maritime law is “the equitable distribution of expense or loss, when of general incidence, among all the parties interested, in proportion to their several interests”. In its transferred use it became to signify the mean: “The distribution of the aggregate inequalities (in quantity, quality, intensity, etc.) of a series of things among all the members of the series, so as to equalize them, and ascertain their common or mean quantity, etc.” and “the arithmetical so obtained." The exact etymological origin of average is uncertain (Simpson and Weiner 1989). Some authors think that average ultimately stems from the Arabic "awariyah" - damaged goods (see Schwartzman 1994), but Heck (1889) argues that this is not very likely.

This origin of average in combination with children’s intuition on fairness implies that fair share might be a suitable instructional context too (see, for example Boswinkel, et al. 1997).
Another possible precursor to the arithmetic mean is the *midrange*, which is the mean of the two extreme values, used for example in Arabian astronomy of the ninth to eleventh centuries, but also in metallurgy and navigation (*Eisenhart 1974*). Nowadays we know that many observations and errors in those contexts follow the normal distribution. Therefore, the midrange probably was a sensible value to take in those situations.

Not until the sixteenth century was it recognized that the arithmetic mean could be extended to $n$ cases: $\left( \frac{a_1 + a_2 + ... + a_n}{n} \right)$, Székely (*1997*) supposes that the invention of the decimal system by Stevin in 1585 facilitated such division calculations. Astronomers wanted to know a real value, such as the position of a planet or the diameter of the moon, but there are always measurement errors. Using the mean of several measured values, scientists assumed that the errors add up to a relatively small number when compared to the total of all measured values. The method of taking the mean for reducing observation errors was indeed mainly developed in astronomy (*Plackett 1970*). From the late sixteenth century onwards, it gradually became a common method to use the arithmetic mean to reduce errors in other areas too (*Plackett 1970; Eisenhart 1974*). This implies that repeated measurement might also be a useful instructional activity for developing understanding of the mean and distribution.

In **Section 5**, the question appeared of how to benefit from the Greek representation and the modern general definition. Until now we have seen ingredients for an answer. The interview in which students divide by 2 instead of 12 and similar observations indicate that it may be better to start with more than two values. This is possible because students nowadays already know the decimal system and calculational techniques to find the mean of more than two value bars. We should learn from history, like with the bars representations, but not follow history literally, because students now know things that people in the past did not know.

**8. Defining mean, median, mode, and midrange**

In **Section 5** it became clear that clarifying an "average box" in the elephant task as one with "not too few and not too many elephants in it" is a starting point that requires elaboration. In this section, we will see how students became more explicit and precise and what this has to do with the historical development. When prompted to further explain what they meant, a few students suggested to count the emptiest and the fullest box and calculate the mean of these two, or the midrange. This is a little more precise than "somewhere in the middle," which some students said. But with counterexamples other students pointed out that this method is not reliable. One girl said: “But if you have one that is 100 and the rest [of the numbers] are 1, then you wouldn’t take 50, would you?” A boy stressed that one has to look how the rest of the numbers lie in between the lowest and highest number. They started, in other words, to look at how the data were distributed. This last step can be seen as one step away from the midrange and one step towards the insight that the mean accounts for all data in the data set.

Our next step in the instructional sequence was to create situations in which the intuitive variants of center would create cognitive conflicts and ask for clearer definitions. The teacher asked, “Assume we had estimated not elephants but something else, what would have been an “average box” here in Figure 5?” This might sound strange since the numbers are already there, but this question was meant to let students explain what they meant by an average box. The students did not have trouble with this hypothetical question; the game-like activity drew their attention and evoked statistical reasoning.
This matrix with a skewed distribution of numbers helped more students see that taking the midrange can be a poor method for estimating the total number. Some students proposed to look at a number with six numbers below and six above it; they reinvented the median. A few students said they looked “where the most were.” This could be seen as a modal class (Konold, et al. 2002). Others used an estimated mean, because they felt the need to account for the deviations; the deviations on either side should even out. In this way, a once "fluid" field of similar concepts started to get sharper borders. The teacher wrote down the different methods on the blackboard and taught the modern statistical names after the different student methods had been discussed. Freudenthal (1991) would call this “guided reinvention.” That is, students reinvent existing concepts under the guidance of the teacher and with help of certain instructional activities.

In this respect, it is interesting to discuss the origin of the word "definition." The Latin "finis" means "end, border, boundary." Schwartzman (1994, p. 68): “When you define something you ‘put boundaries around’ what it can mean. A good definition puts an end to confusion about what a term means.” This implies that students should first explore the subject area before they can appreciate and understand clear definitions. Most textbooks take the opposite direction. They define mean, median, and mode, and then let students practice the procedures and applications.

Additionally, in most school textbooks, the midrange is avoided because it is not a robust measure of central tendency. In our case though, the discussion on the midrange formed an intermediate step towards the meaningful understanding of other statistical notions such as mean, median, and distribution. Without the historical study, I would probably not have thought of the midrange as a precursor to the mean or of allowing the midrange as an initial strategy.

The approach of guided reinvention is in line with the historical development of statistical concepts. For example, the median and mode were used implicitly long before they got their present names and definitions in the nineteenth century (Walker 1931; David 1995, 1998). It is striking that the median only gained importance when skewed distributions became topic of study in the nineteenth century. In that light it is surprising that most textbooks introduce mean, median, and mode as a trinity. As we saw in earlier sections, the mean has a long history with many applications, the mode appears implicitly in some situations, but the median is a recent concept. I am not claiming that because it appeared late in history, the median is more difficult to grasp than the mean. I only want to stress that the median has difficulties that are often overlooked, namely its close relation with distribution and outliers. Most students have not yet developed a sense of a skewed distribution and outliers, but they need this for deciding between mean and median (Zawojewski and Shaughnessy 2000).

9. Faces of the mean

One of the instructional difficulties with the mean is that it has so many faces. The historical study helps us to tease out some subtle aspects and define differences in the aspect of representativeness for
instructional design. The historical examples until about the nineteenth century always had to do with finding a real value, for example the number of leaves on a branch or the diameter of the moon. In all the older examples, the mean was used as a means to an end. It took a long time before the mean was used as a representative or substitute value as an entity on its own. The Belgian statistician Quetelet (1796-1874), famous as the inventor of *l'homme moyen*, the average man, was one of the first scientists to use the mean as the representative value for an aspect of a population. This transition from the real value in astronomy to the representative value of Quetelet, which is a statistical construct in the social sciences, was an important conceptual change. Therefore, there are several layers of understanding the mean as a representative value. The following example from the interviews may illustrate this.

Students have little problem in seeing an average Dutchman as a typical Dutchman, but have difficulties with artificial constructs like the average size of a family. When asked to explain that families have an average of 2.5 persons, several students thought that this referred to two adults and one child. This is an example of where the historical learning process needs revision. Students already know the word "average" in its common usage, meaning "typical," but they do not yet see it as a representative construct in the technical sense.

Moreover, the aspect of representativeness was already present implicitly in the estimation tasks, because when finding a total with an average, this average could be seen as representative. In the case of the number of leaves, the average was the number of leaves on a typical branch. In the case of the elephants the average box was representative for a box with a typical number of elephants. In these examples, the average was used as a multiplicand to find a total.

The average can also have a different role, namely to find a number instead of a total. To clarify this I mention three components of the mean calculation: the number of values \( n \), the sum or total \( \Sigma \) and the mean \( \mu \). These components can have different roles and it is useful for instructional design to categorize the three possibilities \( n \times \mu = \Sigma \), \( \Sigma / n = \mu \), and \( \Sigma / \mu = n \).

1. Estimation often has to do with finding the total number: \( n \times \mu = \Sigma \). The fact that a kind of average value is used often stays implicit because the focus is on the total. In this way, students can develop an understanding of many aspects of the mean without using it explicitly. In Section 4 we saw an example of this: the activity of estimating the number of elephants.

2. Fair share has to do with finding the mean: \( \Sigma / n = \mu \). This calculation answers the question of how much everyone would get after fair redistribution (Section 7). The mean is also useful as a measure for fair comparison, for instance if we need to compensate for the number of values in different groups. We then use parts per million, a percentage, gross national product per head, et cetera. Cortina, Saldanha, and Thompson (1999) call this the mean as a measure. Also see Stigler’s historical examples of coin testing from the twelfth century and measuring with a 16 feet rod from 1535 (Stigler 1999).

3. The third combination of the three components is \( \Sigma / \mu = n \). This is a variant of the first possibility and could also appear in estimation tasks. For instance, “How many 12-year-old students could go into the basket of a hot air balloon if normally eight adults are allowed?” The students first estimated adults’ weights for a total weight \( \Sigma \) and then their own average weight \( \mu \), and calculated the number \( n \) by doing \( \Sigma / \mu = n \).

The balloon activity implicitly asks for an average, namely the estimated weight of students and adults. Some students asked a “normal looking child” how much he weighed; others asked a few students and took a value somewhere in the middle. One girl even passed around a sheet of paper to collect others’ weights. In the case of a “normal looking child” the intuitive average is connected with the qualitative
aspect of representativeness.

This example also raises the issue of sampling in a natural way, as students already took small samples by asking students from their class. Put even stronger, I noticed from the interviews that students should develop some sense of sampling from the very start, because sampling is also related to representativeness.

Mokros and Russell (1995) found that the aspect of representativeness is hard to develop for students. They advise to postpone the calculational aspects of the mean until late in the middle grades, “well after students have developed a strong foundation of the idea of representativeness” (p. 38). The findings in this article support their view and supply ways to teach average values in a more qualitative way. This section also showed that even one aspect of the mean, such as representativeness, could have different layers of easier and more difficult uses.

10. Conclusions

This article deals with the relation between the historical and individual learning process for average values. The resulting insights were used to develop a revised and improved version of the historical learning process that the young learner could recapitulate. The examples in this paper show that there are many parallels but also important differences between historical and individual learning processes.

The earliest historical examples of statistical reasoning had to do with estimation. Parallel to this, it turned out useful to start with estimations in teaching experiments as well. During estimation activities, students reinvented measures of center and then learned the corresponding statistical names.

In history, the midrange may be seen as a predecessor of the mean. Parallel to this, students used a method that was taking the midrange when estimating total numbers. The historically late definition of the mean of more than two values and its historically late application, in combination with didactical arguments, support the view that students should only learn the algorithm of the mean in the later middle grades. If students already know the calculation of the mean, the designed activities can also be used to connect students’ daily-life sense of average to the algorithmic procedure.

The Greek way of defining mean values was visual and geometrical. The representation with a computer tool mentioned in Section 6, comes close to the Greek representation. This bar representation helped the students to reinvent the method of compensating and finding or representing the mean visually without calculations. They saw, presumably better than with calculations, that the mean is somewhere in the middle of the data and that it is strongly influenced by outliers. This compensation strategy is related to the word "average," which is has its origin in fair share and insurance in maritime law.

In history, we saw that the mean was used to find a total number and to approximate a real value; not until the nineteenth century was it used as a construct on its own, representing a specific aspect of a population. Likewise, there are also instructional layers in the aspect of representativeness. The historical analysis helped to detect such layers.

A major difference between historical phenomena and useful instructional contexts is that historical questions are generally not very interesting for students. Most historical contexts, therefore, need a modern translation if the designer wants to use them with young students without knowledge of those historical contexts.

Another difference between the historical and individual learning process is that students nowadays know
things that people in the past did not know. For instance, most seventh-grade students know what average is in its daily sense. It would be a waste to follow history too strictly and not to use their cultural knowledge.

A historical phenomenology as meant by Freudenthal (1983a, p. 32) should yield many phenomena that “beg to be organized” by certain concepts, plus an analysis of how these phenomena gave rise to these concepts. The essential point of didactical phenomenology is to translate these phenomena into problems that are meaningful for students and still have the potential power of asking for organization by a particular statistical method. Knowing the historical development of certain concepts can help to anticipate such learning in a process of guided reinvention.

It is a major problem for designers that they know so much and find it hard to forget their knowledge. What seems a minor step for them might have taken centuries to develop in history and might also be difficult for students. A historical study can help to distinguish more aspects, problems, related notions and intermediate stages of the development of certain notions. In other words, it can help us to look through the eyes of the students.

Acknowledgments

This work was supported by the Netherlands Organization for Scientific Research, under grant number 575-36-003B. The opinions expressed do not necessarily reflect the views of the Foundation. The author thanks Koeno Gravemeijer, Cliff Konold, Jan van Maanen, Rob Kooijman, and Viola Heutger for helpful discussions.

Photographs used by permission of Encyclopaedia Britannica.

References

Aristotle (1994), *Nichomachean Ethics*, Cambridge, MA: Harvard University Press.

Ashburner, W. (1909), *Nomos Rhodioon Nautikos: The Rhodian Sea-Law*, Oxford: Clarendon Press.

Boswinkel, N., Niehaus, J., Gravemeijer, K. P. E., Middleton, J. A., Spence, M. S., Burrill, G., and Milinkovic, J. (1997), *Picturing Numbers*, Chicago: Encyclopaedia Britannica Educational Corporation.

Boyer, C. B. (1991), *A History of Mathematics* (revised edition by U.C. Merzbach), New York: John Wiley and Sons, Inc.

Cobb, P. (1999), "Individual and Collective Mathematical Development: The Case of Statistical Data Analysis," *Mathematical Thinking and Learning*, 1(1), 5-43.

Cortina, J. L., Saldanha, L., and Thompson, P. W. (1999), "Multiplicative conceptions of the arithmetic mean," in *Proceedings of the Twenty First Meeting of the North American Chapter of the International Group of the Psychology of Mathematics Education*, eds. F. Hitt and M. Santos, Cuernacava, Morelos, Mexico: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.
David, H. A. (1995), "First (?) Occurences of Common Terms in Mathematical Statistics," *The American Statistician*, 49(2), 121-133.

----- (1998), "First (?) Occurences of Common Terms in Probability and Statistics - A Second List, with Corrections," *The American Statistician*, 52(1), 36-40.

Eisenhart, C. (1974), "The development of the concept of the best mean of a set of measurements from antiquity to the present day," *1971 American Statistical Association Presidential Address*, unpublished manuscript.

Euclid (1956), *The Thirteen Books of the Elements*, tr. T. H. Heath, New York: Dover.

Freudenthal, H. (1983a), *Didactical Phenomenology of Mathematical Structures*, Dordrecht: Reidel.

---- (1983b), "The Implicit Philosophy of Mathematics: History and Education," *Proceedings of the International Congress of Mathematicians*, pp. 1695-1709, Warsaw and Amsterdam: Polish Scientific Publishers and Elsevier Science Publishers.

---- (1991), *Revisiting Mathematics Education: China Lectures*. Dordrecht: Kluwer Academic Publishers.

Hacking, I. (1975), *The Emergence of Probability. A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*, Cambridge: Cambridge University Press.

Heath, T. H. (1981), *A History of Greek Mathematics*, New York: Dover.

Heck, P. (1889), *Das Recht der grossen Haverei [The Law of General Average]*, Berlin: Verlag von Müller.

Hopkins, M. (1859), *A Handbook of Average* (2nd ed.), London: Smith, Elder, and Company.

Iamblichus (1991), *Greek Mathematics* (Vol. 1), Cambridge, MA: Harvard University Press.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A. D., Wing, R., and Mayr, S. (2002), "Students' Use of Modal Clumps to Summarize Data," in *Developing a Statistically Literate Society: Proceedings of the International Conference on Teaching Statistics* [CD-ROM], ed. B. Phillips, Voorburg, The Netherlands: International Statistical Institute.

Lowndes, R., and Rudolf, G. R. (Eds.), (1975), *General Average and York Antwerp Rules* (10th ed.), London: Stevens and Sons.

Mokros, J., and Russell, S. J. (1995), "Children's Concepts of Average and Representativeness," *Journal for Research in Mathematics Education*, 26(1), 20-39.

Plackett, R. L. (1970), "The Principle of the Arithmetic Mean," in *Studies in the History of Statistics and Probability* (Vol. 1), eds. E. Pearson and M. G. Kendall, London: Griffin.

Plön, O., and Kreutziger, G. (1965), *Das Recht der grossen Haverei [The Law of General Average]* (Vol. 1), Hamburg: Otto Meissner Verlag.

Radford, L. (2000), "Historical Formation and Student Understanding of Mathematics," in *History in
Mathematics Education: the ICMI Study, eds. J. Fauvel and J. van Maanen, Dordrecht: Kluwer.

Rubin, E. (1968), "The Statistical World of Herodotus," The American Statistician, 22(1), 31-33.

----- (1971), "Quantitative Commentary on Thucydides," The American Statistician, 25(4), 52-54.

Schwartzman, S. (1994), The Words of Mathematics: An Etymological Dictionary of Mathematical Terms Used in English, Washington, DC: Mathematical Association of America.

Simpson, J. A., and Weiner, E. S. C. (Eds.), (1989), The Oxford English Dictionary (2nd ed.), Oxford: Clarendon Press.

Spruit, J. E. e. a. (Ed.), (1996), Corpus Iuris Civilis; tekst en vertaling [Civil Law: Text and Translation] (Vol. III), Zutphen: Walburg Pers.

Stigler, S. M. (1999), Statistics on the Table. The History of Statistical Concepts and Methods, Cambridge, MA: Harvard University Press.

Thucydides (1954), History of the Peloponnesian War, tr. R. Warner, Baltimore, Maryland: Penguin Books.

----- (1975), Peri tou Peloponnesiakou Polemou [On the Peloponnesian War], tr. T. Hobbes, ed. R. Schlatter, New Brunswick, NJ: Rutgers University Press.

van der Hoeven, P. (1854), Handleiding tot het opmaken van de avarijen [Guide to calculate averages], Dordrecht: P.K. Braat.

Walker, H. M. (1931), Studies in the History of Statistical Methods. With Special Reference to Certain Educational Problems, Baltimore: The Williams and Wilkins Company.

Zawojewski, J. S., and Shaughnessy, J. M. (2000), "Mean and Median: Are They Really so Easy?" Mathematics Teaching in the Middle School, 5(7), 436-440.

Arthur Bakker
Freudenthal Institute, Utrecht University
3506 GK Utrecht
The Netherlands
arthur@fi.uu.nl
