9 Diagrammatic reasoning about growing samples

I would like to compare this [process of symbolizing] with lignification [transformation into wood]. Where the tree lives and grows, it must be soft and sappy. If, however, the sappiness does not lignify, the tree cannot grow higher. If, on the contrary, all the green of the tree transforms into wood, the growing stops.

Frege in a letter to Hilbert (Frege, 1895/1976, p. 59; translation from German[47])

In Chapter 7, we answer the first research question for grade 7 and conjecture that students may develop a notion of distribution by reasoning about growing samples. Because distribution as an object-like entity proved too ambitious for grade 7, we conducted the next teaching experiment in grade 8. For the HLT of the teaching experiment in grade 8, analyzed in the present chapter, we decided to use the new activity of growing samples as the most important means of supporting reasoning about distribution. In Chapter 8, we answer the second research question for grade 7 and conclude that the notion of diagrammatic reasoning proves useful for analyzing students' symbolizing process when they learn to reason about distribution. It also appears possible to summarize the HLT for grade 8 as progressive diagrammatic reasoning about distribution aspects in relation to growing samples. Due to the HLT's formulation, the first and second research questions become strongly related. In the present chapter we therefore answer the two research questions simultaneously for grade 8 by answering the following integrated research question:

How can students with little statistical background develop a notion of distribution by diagrammatic reasoning about growing samples?

To answer this question we do not describe the whole teaching experiment, as in Chapter 7, but only the activities that were related to growing samples.
After providing information about the eighth-grade teaching experiment (9.1), we describe five episodes in which students reasoned about growing samples or shape (9.2 to 9.6). The results of the final interviews offer an image of students' notions of distribution after the teaching experiment (9.7). The last section summarizes the answer to the integrated research question (9.8).

9.1 Information about the teaching experiment in grade 8

In this section we supply relevant information about the class, the approach, the end goal of the HLT, and the method applied in the retrospective analysis in the eighth-grade teaching experiment.

47. Ich möchte dieses [Symbolisieren] mit dem Verholzungsvorgange vergleichen. Wo der Baum lebt und wächst, muss er weich und saftig sein. Wenn aber das Saftige nicht mit der Zeit verholzte, könnte keine bedeutende Höhe erreicht werden. Wenn dagegen alles Grüne verholzt ist, hört das Wachstum auf.

The class. The teaching experiment that we report on here was in an eighth-grade class with 30 students and lasted ten lessons of 50 minutes. After every second lesson we interviewed around eight students about their work (see also 3.7). The students had hardly learned any statistics except the mean and bar graphs, but during mathematics lessons students had learned about line graphs. The students were being prepared for pre-university (*vwo*) or higher vocational education (*havo*). In the Netherlands, the top 35-40% of the students attend *vwo* or *havo*. The remaining 60-65% are prepared for middle and lower vocational education (*vmbo*). In the practice of Dutch mathematics education, the school textbooks play a central role. Students are expected to work through the tasks by themselves, with the teacher available to help them if necessary. As a consequence, tasks are broken down into very small steps and real problem solving is rare. Students' answers tend to be superficial, partially because they have to deal with about eight contexts per lesson (Van den Boer, 2003). The students in the class reported on here were not accustomed to whole-class discussions, but rather to be "taken by the hand" as the teacher called it. The three assistants characterized the class as "passive but willing to cooperate." One of the challenges, therefore, was to get students engaged in statistical reasoning.

Guided reinvention. As mentioned in Section 2.1, we strove for a learning process of guided reinvention and for striking the balance between reinvention and guidance. On the one hand, we wanted to offer students opportunities to share their own ideas and participate in class discussions; and on the other hand, we wanted to guide their learning process towards culturally accepted statistical methods. This issue can be illustrated with a metaphor that Frege used in a letter to Hilbert (Frege was one of the first modern logicians and philosophers of language, and Hilbert was a formalist mathematician). The topic of the letter was using and making symbols in mathematical discourse.

> I would like to compare this with lignification [transformation into wood]. Where the tree lives and grows, it must be soft and sappy. If, however, the sappiness does not lignify, the tree cannot grow higher. If, on the contrary, all the green of the tree transforms into wood, the growing stops. (Frege, 1895/1976, p. 59)

Applied to our context of teaching statistical data analysis, this might imply the following. On the one hand, if statistical concepts such as mean, median, and mode are defined before students even have an intuitive idea of what these concepts are for, then just as the tree is at risk of transforming into wood, the students' conceptual development may be hindered. It could well be that students are not inclined to share their own ideas if they sense that statistical words have a very precise meaning that they do not yet understand. On the other hand, if teachers and instructional materials do not guide students well in a process of reinvention, the tree stays weak and cannot grow higher. Students may even attach idiosyncratic meanings to statistical terms which are hard to 'unteach' again.

End goal of the HLT. As in the seventh-grade teaching experiment, the end goal of the HLT for grade 8 was that students would come to view distribution or shape as an object that can have different properties and that can be used to model data sets and statistical phenomena. As discussed in 2.2 and 5.2, distribution is an organizing conceptual structure with which students can conceive the data set as a whole instead of just individual data points. However, during the teaching experiments in grade 7, we had become more and more convinced that sampling should play a more central role than we had given it. We had noticed how important the process of talking through the process of data creation was as implicitly addressing the sampling issue. Attempts to make sampling more explicit, for instance with the trial of the Pyx and the balloon activities, had not really been very successful, but the growing samples activity in class 1B had proven promising. The goal of this eighth-grade teaching experiment was to test the conjecture that students could develop reasoning about distribution aspects including shape by reasoning about growing samples.
Compared to the seventh-grade experiments, we decided to do a few things a little different. We would pay more attention to the whole investigative cycle that is an important dimension of statistical thinking (Pfannkuch & Wild, in press): asking a question, design a study, take a sample, analyze data, and communicate the results. In the first lesson we let students first think of how one could test two battery brands. In several other lessons, we used newspaper reports and graphs to let students think about the way the information was acquired, which was also meant to address the sampling issue. For two lessons, students had to collect car color data to test a newspaper claim about percentages of car colors.[48] We also spent more time than in grade 7 on discussing the meaning of notions that had been the topic of previous lessons.

Retrospective analysis. For the retrospective analysis, the transcripts were read, the videotapes watched, and conjectures formulated on students' learning on the basis of the read and watched episodes. The generated conjectures were tested against the other episodes and the rest of the collected data (student work, field observations, and tests) in the next round of analysis (triangulation). Then the whole generating and testing process was repeated. Episodes in the transcripts that we considered examples or counterexamples of these conjectures were coded with these conjectures. The transcripts of lessons 4, 6, and 7 have been judged by three assistants. The amount of agreement was high: of the 35 code assignments that were discussed, only two were not judged unanimously. The conjectures that were confirmed are referred

48. In the Netherlands, in the fall of 2001, blue was the favorite car color with 21.9%, followed by red (20.8%), gray (18.9%), green (13.8%), and black (10.3%).

to as C# (see Appendix A). We give an example of an episode that was assigned with two conjecture codes. In the seventh lesson, two students used the four equal groups option in Minitool 2 for a revised version of the jeans activity (7.12).

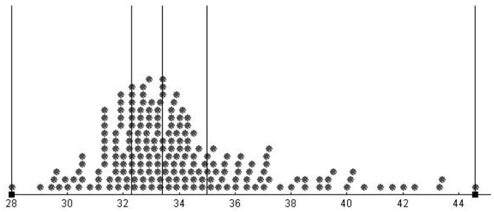

Figure 9.1: Minitool 2 with jeans data set organized into four equal groups

Int.: Why did you choose four equal groups, Sofie?
Sofie.: Because then you can best see the spread, how it is distributed.
Int.: How it is distributed. And how do you see that here [in this graph]? What do you look at then? (...)
Sofie.: Well, you can see that, for example, if you put a [vertical] line here, here a line, and here a line. Then you see here [two lines at the right] that there is a very large spread in that part, so to speak.

Sofie's answers were assigned with C7 and C2. C7 states that "students' notions of spread, distribution, and density are not yet distinguished. When explaining how data are spread out, they often describe the distribution or the density in some area." Related to this idea, C2 states that "students either characterize spread as range or look very locally at spread."
We also give an example of a code assignment that was dismissed in relation to the same diagram.

Int.: What does this tell you? Four equal groups?
Melle: Well, I think that most jeans are between 32 and 34 [inches].

We had originally assigned the code C1 to the this episode (students talk about data sets as consisting of three groups of low, 'average', and high values), because "most jeans are between 32 and 34" implies that below 32 and above 34 the frequencies are relatively low. In the episode, however, this student did not talk about three groups of low, average, and high values (or anything equivalent). We therefore removed the code from this episode.

9.2 Larger samples in the battery context

9.2.1 HLT for lessons 1-3

Sampling was an explicit issue from the first lesson onwards. For the first lesson, we decided to let students invent a method of testing two brands of batteries before any data were provided. We would also stimulate students to think of aggregate features of samples larger than those they would think of: what would a sample[49] of a good brand look like? In the first lesson, the battery activity was used to stimulate discussion on features of both battery distributions (one normal and one skewed). To prevent students from being distracted by computers in the lab during a class discussion, we decided to have them switch off the monitors during discussions.

We did not use the elephant estimation activity as we assumed that these students, with more than one year experience in computing report grades, would already have a reasonably good sense of the mean. Moreover, we wanted to focus on the sampling conjecture in this teaching experiment and avoid the mean distractor effect (7.3.1).

The second lesson would be devoted to a discussion on students' answers on the battery problem. Just as in grade 7, the balloon problem would be used as preparation for the growing samples activity as used in grade 7.

In the third lesson, students would invent data sets according to specific aggregate features such as "brand A has a long life span but is not reliable." As in grade 7, this activity was used to stimulate students to represent aggregate distribution features (mainly average and spread) in graphical representations.

9.2.2 Retrospective analysis

One observation that struck us in the first lesson was that no student wanted to test more than two batteries per brand. Most thought one was enough; some added a second "for reserve." Even when prompted about this small number, many students thought that two batteries would be enough. Some were willing to test more if the machine in which they would be tested required more than two. We formulated the conjecture that students are inclined to think of small samples when first asked about how one could test something (C3). This was confirmed by the different data sources we had (audio, video, student work). A related conjecture that was confirmed (C4) was that students did not expect variation in this (industrial) context. This could explain why they did not think of larger samples.

From these observations we concluded two things. First, a notion of variability is a prerequisite for all statistical investigation (Chapter 5). Without a sense of variability, as in the battery context, there is no need of taking a sample or describing distribution aspects. Second, our approach of paying more attention to the design and sampling issue revealed that students did not expect variation in battery life spans of one type of battery. In grade 7, when we only talked through the process of data creation, this expectation had remained hidden. In hindsight, it would probably have been better to have chosen a context in which students do expect variation.

49. The Dutch term for sample is *steekproef*, which is a technical term most students of this age would not use in everyday language. In English the term 'sample' is informally used for examples of food or other products, but the Dutch language does not use the term *steekproef* for that.

Starting from students' ideas, we prompted them to think beyond the sample size of two. One question[50] that the interviewers asked was, "What if one brand had a good and a bad battery, and the other brand had two good batteries; what would you know about the brands?" From the analysis we concluded that such what-if questions proved useful to let students think about sampling in aggregate terms (C5). We give a few examples of students' answers to this what-if question.

Armin:	[Thinks a long time.] I would still think that the other brand [with two good ones] is better. But it could be coincidence. No, I don't know.
Int.:	What if you had measured 100 batteries and one of that first brand was bad?
Armin:	Then it could be coincidence, for example, but the fewer batteries you take, the more you doubt [*hoe meer je gaat twijfelen*].

Later he realized that it was impossible to test all batteries, and his peer said that it would be a lot of work to test many batteries. This shows that a short discussion can address sampling in a nutshell: sample should be large enough to draw reliable conclusions, but a large sample also has a price.
Another pair of students reasoned with proportions in answer to the same question, "What if one brand had a good and a bad battery, and the other brand had two good batteries; what would you know about the brands?"

Melle:	Then you still don't know very much.
Int.:	Why would you then take two [batteries]?
Sofie:	Yeah, I don't know, in the beginning we thought,... we had not really thought it through.
Int.:	What if you had measured ten and one was not very good. Would you know more?
Sofie:	Yes, you would actually.
Melle:	Yes.
Int.:	Or could it still be that one out of 100.
Melle:	In fact you cannot say anything, because [interrupted].
Sofie:	Actually, you could [*Eigenlijk wel ja*].
Int.:	When could you say something?
Melle:	If all [batteries] of one brand were good, and the other brand has seven bad and three good ones. Then you would be able to say something about it. Then the quality of the other one would be much better.[51]

When working with Minitool 1, students came up with arguments similar to those in grade 7 (such as: D has a higher mean, K has outliers, D is more together). There were also examples of case-oriented views, and certainly not all students had under-

50. This mini-interview question was not discussed before the lesson as we had not anticipated that students would all choose such small sample sizes.
51. On the basis of this episode and three others we conjectured that this kind of what-if questions could be used for making a connection with probability within a statistics unit. We have not yet been able to test this conjecture (cf. Biehler, 1994).

stood what the data stood for. It could be that some students did not think of samples because they did not expect variation in battery life spans (C4). Rick, for instance, wanted to buy the battery with the longest life span, the bottom bar in Figure 9.2.

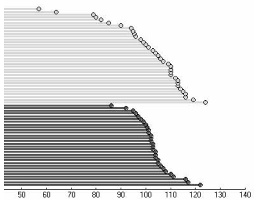

Figure 9.2: Battery data set in Minitool 1 (brand D: dark bars; brand K: light bars)

Int.: Rick, why would you buy the bottom one?
Rick: Because it goes furthest; it goes all the way beyond 120 [hours].

This underlines that understanding what the data stand for and describing them with aggregate terms was not self-evident for students. After a discussion on the reliability of their methods, students started to realize that it makes sense to take larger samples (they thought about ten would be reasonable).

In line with the what-if question of the mini-interviews in the first lesson, the teacher asked students a ‘growing samples question’ in the second lesson.

> Assume your two batteries lasted 85 and 105 hours and you took a larger sample. When is your conclusion ‘the brand is bad’ and when is it ‘it is good’?

There were two types of answers. One type referred to spread only (example 1) and the other incorporated values or proportions of values (example 2). We give one example of each type.

> *Example 1.* The brand is good if all batteries live as long, if you take 10 batteries. The brand is bad if there is a lot of difference between the batteries, also with 10 batteries.
> *Example 2.* Good: if you test 10 batteries, there must be at least seven that last more than 100 hours and have about the same life span. Then there may be a few bad ones. Bad: if you test 10 batteries, and there are four bad ones, it is not a good brand.

Stimulating students with such what-if questions to discuss aggregate features in relation to samples had been reasonably successful if we take into account that this was the second lesson.

In the third lesson, students invented graphs in Minitool 1 of battery brands with specific aggregate features. Their inventions were similar to those in grade 7. We concluded that students were reasonably fluent in interpreting and producing distribution aspects such as average (life span) and spread (reliability) in this battery context. In terms of diagrammatic reasoning, students had diagrammatized specific aggregate features, mentally experimented with larger samples, and reflected on sample size in relation to features of battery brands.

9.3 Growing a sample in the weight context

9.3.1 HLT for lesson 4: towards shape as an object

The strategies for solving the balloon question that had been homework for the second lesson, had been similar to those in grade 7. Students had also made their first predictions of weight graphs. Thus the balloon activity prepared the growing samples activity in the fourth lesson. The overall goal of the growing samples activity in the weight context as formulated in the HLT was to let students develop a notion of distribution in relation to sampling. We conjectured that students could eventually conceive the stability of distribution shapes between samples as well as growing one sample, and that shape could become a topic of discussion. Our conjecture was that this transition from a discrete plurality of data values to a continuous entity of a distribution is important to foster a notion of distribution as an object. During teaching experiments in the seventh grade, in two American sixth-grade classes, and a visit to an American group of ninth graders, we observed that reasoning with continuous shapes turned out difficult to accomplish, even when we explicitly asked for it. It often seemed fruitless to nudge students towards drawing the general, continuous shape of data sets represented in dot plots. Our assumption was that students needed to construct something new, with which they can view the data or the phenomenon differently, for instance a notion of distribution (see the motto of Chapter 8).

Compared to the growing samples activity in grade 7, we would do a few things a little different. First, we decided not to let students measure their own weights because that could be too sensitive, but rather use real data sets from other classes. Second, to make sure that all students would formulate their own ideas (not just the ones that participated in a class discussion), we let them all write their comparisons down. We also tried to strike the balance between engaging students in statistical reasoning and allowing them to use their own terminology on the one hand, and guiding them in using conventional and more precise notions and graphical representations on the other.

9.3.2 Retrospective analysis

In retrospect we have come to see students' reasoning about growing samples as diagrammatic reasoning. As in Chapter 8, the process of hypostatic abstraction of shape appeared to consist of multiple steps. In this section, we analyze the lesson in three cycles, each consisting of making diagrams of a hypothetical situation (diagrammatization and a thought experiment) and comparing those sketches with diagrams displaying real data sets (reflection). In this section, we take all figures and written explanations from three students because their work gives an impression of the whole class's work in the following sense: their diagrams cover all types of diagrams made in this class and their learning abilities varied considerably. Ruud and Chris's report grades on the total of all subjects were in the bottom third of the class and Sandra had the best report grade of the class.

First cycle

The text of the activity sheet started as follows:

> Last week you made graphs for a balloon driver with data that you had invented yourselves. During this lesson you will get to see weight data of students from another school. We are going to investigate the influence of the sample size on the shape of the graph.
>
> a. Predict first a graph of ten data values, for example with the dots of Minitool 2.

The sample size of ten was chosen because the students found that size reasonable in the battery and balloon contexts. Figure 9.3 shows the different diagrams students made to show their predictions: there were three value-bar graphs (such as in Minitool 1), eight with only the endpoints (such as with 'hide bars'), and the remaining nineteen plots were dot plots (as in Minitool 2) (Fig. 9.4).

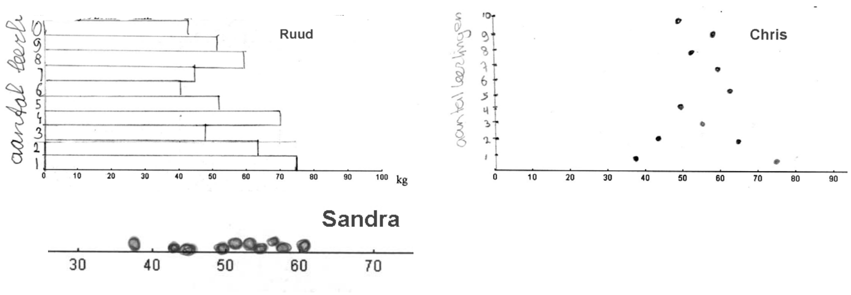

Figure 9.3: Student predictions for ten data points (weight in kg)

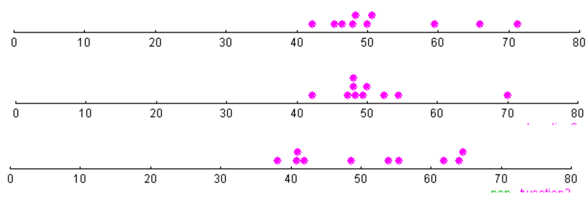

Figure 9.4: Three real samples in Minitool 2.

To stimulate the reflection on the diagrams (the last step of diagrammatic reasoning) the teacher showed three samples of ten data points on the blackboard and students had to compare their own diagrams with the diagrams of the real samples.

> b. You get to see three different samples of size 10. Are they different from your own prediction? Describe the differences.

The reason for the three small samples was to show the variation between samples.[52] We have found no clear indications, though, that students conceived this variation as a sign of sample sizes being too small to draw reliable conclusions, but they generally agreed that larger samples were more reliable.

There was a short class discussion on the diagrams with real data before students worked for themselves again. During this reflection activity, students were stimulated to describe features of the data sets, some of which were aggregate (together, further apart).

Jacob:	In the middle [graph] there are more together.
Teacher:	Here are many more together, clumped or so. Who can mention other differences?
Jacob:	Well, uh, the lowest, that is the furthest apart.
Teacher	Those are all furthest apart. Here they are in one clump. Are there any other things that you notice, Gigi?
Gigi:	Yes, the middle one has just one at 70.

The latter answer is a sign of a case-oriented view. The written answers of the three example students were the following.

Ruud:	Mine looks very much like what is on the blackboard.
Chris:	The middle-most [diagram on the blackboard] best resembles mine because the weights are close together and that is also the case in my graph. It lies between 35 and 75 [kg].
Sandra:	The other [real data] are more weights together and mine are further apart.

Ruud's answer is not very specific, as most of the written answers in the first cycle of growing samples were. Chris used the predicate 'close together' and added numbers that indicate the range, probably as an indication of spread.[53] Sandra used such terms as 'together' and 'further apart', which address spread. Many other students in this class also used daily-life words such as 'together', 'spread out', and 'further apart' to describe something that is both a property of the dots in the diagram and of the data.

This process of using predicates is also called predication[54] and can be considered a prerequisite to hypostatic abstraction: before 'spread' can be taken as a topic of common attention, a set of dots needs to be predicated with 'spread out.' For the analysis of the process of hypostatic abstraction it is important to note that the students used predicates (together, apart) and no nouns (spread) in this first cycle of growing sam-

52. See the research literature on resampling (e.g. Konold, 1994; Simon & Bruce, 1991).
53. Range was also historically the first sample measure of variability (David, 1998a).
54. Van Oers (2000) uses 'predication' in a slightly more specific sense: "Predication is the process of attaching extra quality to an object of common attention (such as a situation, topic or theme) and, by doing so, making it distinct from others" (p. 150).

ples. This changed in the second cycle of producing a diagram and comparing it with a real sample.

Second cycle

With the feedback of the samples of ten data points in dot plots, students had to make predictions for a whole class of 27 students and also for three classes with a total of 67 students (27 and 67 were the sample sizes of real data sets we had). During this cycle, all[55] students made dot plots, probably because the teacher had shown dot plots on the blackboard and because drawing so many value bars is laborious (Figures 9.4 and 9.5). In the seventh-grade experiment, we had left a lot of space for reinvention during the growing samples activity (7.9). In this case we wanted to guide the process a bit more, for instance by stimulating students to use statistical words. For research purposes, we also wanted to know what these terms meant to them.

c. We will now have a look how the graph changes with larger samples. Predict a sample of 27 students (one class) and of 67 students (three classes).

d. You now get to see real samples of those sizes. Describe the differences. You can use words such as majority, outliers, spread, average.

When the teacher showed the two real data sets (Figure 9.6), a short class discussion recurred in which the teacher capitalized on the question of why most of students' predictions now looked like what was on the blackboard, whereas earlier predictions varied more. No student had a reasonable explanation, which indicates that this was an advanced question.

The written answers to question (*d)* of the same three students were the following (Figure 9.5).

Ruud: My spread is different.

Chris: Mine resembles the sample, but I have more people around a certain weight and I do not really have outliers because I have 10 of about 70 and 80 and the real sample has only 6 around the 70 and 80.

Sandra: With the 27 there are outliers and there is spread; with the 67 there are more together and more around the average.

In his written answer, Ruud addressed the issue of spread, although we cannot infer from this short answer what he meant by it. Chris was explicit about a particular area in her graph, the category of high values. Sandra used the term 'outliers' in this stage, by which students meant high or low values as we have seen in other classes. She also seemed to understand that many students are about average.

55. Only one student made a value-bar graph in the sample of 27, but she switched to a dot plot for the sample of 67.

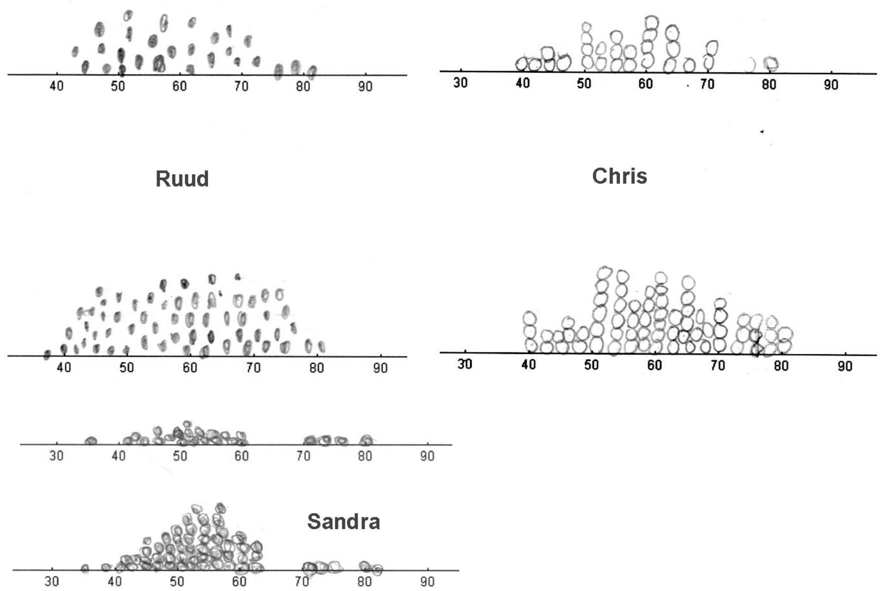

Figure 9.5: Predicted graphs for one and for three classes.

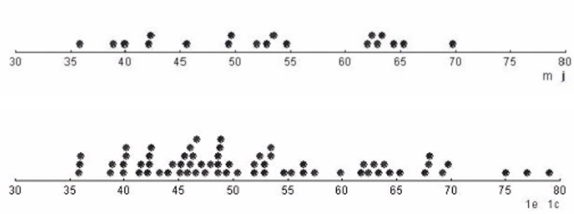

Figure 9.6: Real data sets of one and three classes in Minitool 2.

Sandra used the term ‘outliers’ in this stage, by which students meant high or low values as we have seen in other classes. She also seemed to understand that many students are about average.

These examples illustrate that students used statistical words to describe properties of the data and diagrams. From a statistical point of view, these terms were not very precise. By ‘mean’, students generally meant ‘about average’ or the ‘majority’; by ‘spread’, they meant “how far the data lie apart” (as in grade 7). By ‘sample’ they

seemed to mean just a bunch of people, not necessarily the data as being representative of a population (cf. Schwartz et al., 1998).
In contrast to the first cycle, students used nouns instead of just predicates to compare the diagrams. Ruud (like others) used the noun 'spread', whereas students earlier used only predicates such as 'spread out.' From a semiotic perspective this is an important step since it can be the sign of a hypostatic abstraction. This might seem a trivial linguistic trick, but statistically it makes a difference whether we say "the dots are spread out" or "the spread is large." In the latter case, spread is something that can have characteristics that can be predicated (large, small) or even measured (for instance, by the range or the interquartile range). Other notions, such as sample and average, were also used as nouns: that is, as objects that can be talked about. Recall that Peirce defined objects as things that could be talked or thought about. As the example of 'outliers' shows, the objects that students form during the process of hypostatic abstractions need not be the ones that we aimed for. From the context and from students' reasoning with notions as tools, we have to determine what they refer to when they use these notions.

Third cycle

So far, students had not talked about the shape of their graphs. In this last cycle of growing the sample, we asked for a graph that would show data of all students in the city, not necessarily with dots (Figure 9.7), and asked students to describe the shape of their graphs. The aim of asking this was to stimulate students to use continuous shapes and dynamically relate samples to populations without yet making that distinction explicit.

e. Make a weight graph of a sample of all eighth graders in Utrecht. You need not draw dots. It is the shape of the graph that is important.

f. Describe the shape of your graph and explain why you have drawn that shape.

The written answers to question f of the same three students were the following.

Ruud:	Because the average [values are] roughly between 50 and 60 kg.
Chris:	I think it is a pyramid shape. I have drawn my graph such because I found it easy to make and easy to read.
Sandra:	Because most are around the average and there are outliers at 30 and 80 [kg].

Ruud's answer resembles that of students in seventh grade who indicated a range of the average values or the majority. His answer focuses on the average group, or 'modal clump' as Konold and colleagues (2002) call such groups in the center. During an interview, Ruud literally called his graph a 'bell shape' though he had probably not encountered that term in a school situation before.

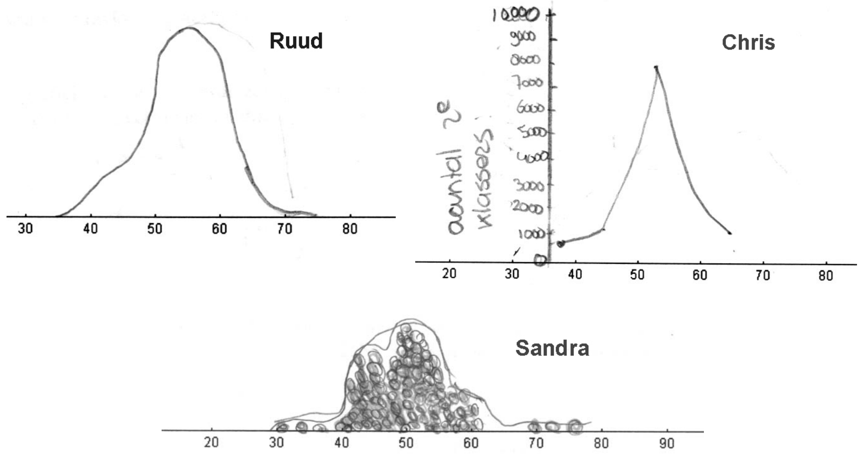

Figure 9.7: Predicted diagrams for all students in the city.

Chris's diagram was probably influenced by line graphs that they made during mathematics lessons (compare this with Mike's graph in Section 8.6). One of the aims in the HLT was indeed that students would draw continuous shapes of an informal notion of a probability distribution. However, this graph of Chris's shows the problem of using line graphs for that: line graphs cannot represent frequencies in the way Chris's diagram shows it because it is not clear what the class width is. Shape sketches without a vertical axis informally describe the course of the density, not the frequencies.

Sandra's diagram shows both dots and a continuous shape. It could well be that she started using dots and then drew the continuous shape. We had not anticipated such combined representations in the HLT. Her answer is an example of an episode that we coded with C1, on describing data sets as consisting of low, average, and high values.

In this third cycle of growing samples, 23 students drew a bump shape (mostly continuous). The words they used for the shapes were pyramid (three students), semicircle (one), and bell shape (four). Of course, we did not exactly know what these shapes meant to them. Therefore, students' reasoning with these shapes was taken up in the sixth lesson.

About the activity

This activity of growing samples involved short cycles of constructing diagrams of new hypothetical situations, comparing these with other diagrams of a real sample of the same size. We analyzed students' reasoning as an instance of diagrammatic reasoning, which typically involves constructing diagrams, experimenting with them, and reflecting on the results of the thought experiments. Students' diagrams were strongly influenced by the two Minitools they had used (Minitool 1) and seen (Minitool 2), but they also used line graphs taught in mathematics lessons.

How did the process of hypostatic abstraction of spread evolve? Instead of just writing that the data were more spread out, students wrote or said that the spread was large. From the terms used in this fourth lesson, we conclude that many issues from Table 5.7, such as center (average, majority), spread (range and range of groups), and density (shape) had become topics of discussions (hypostatic abstractions) during the growing samples activity. Some of these words were used in a rather unconventional way, which implies that students need more guidance at this point. Shape became a topic of discussion as students predicted that the shape of the graph would be a semicircle, a pyramid, or a bell shape.

The growing samples activity combines different heuristics formulated in Chapter 6. First, it often stays away from data so as to avoid students from adopting a case-oriented view. Second, by asking students to compare their own diagrams with those representing real data, we invited them to "compare forests instead of trees", as the metaphor of another heuristic goes (Chapter 6). Third, by letting students predict a situation, we create the need to use conceptual tools for predicting that situation. The quick alternation between prediction and reflection during diagrammatic reasoning probably created ample opportunities for hypostatic abstraction.

In earlier lessons we had noticed that these students found it hard to concentrate during class discussions for longer than about ten minutes. A cycle of producing a diagram for a sample of a specific size, comparing it with a real sample requires short periods of concentration. Providing real data in between their inventions demanded short periods of reflection and feedback. We found it striking how well students knew the context of weight; their predictions resembled the actual samples in many respects. The delicacy of this subject might explain part of their engagement during class discussions.

9.4 Reasoning about shapes in the weight context

9.4.1 HLT for lesson 6: skewness as a topic of discussion

From the mini-interviews in previous lessons, we had concluded that students had a notion of distribution as consisting of three groups of low, average, and high values (C1). For example, Tula said, "A few low ones and a few high ones and more around average."

In the fourth lesson, almost all diagrams looked roughly symmetrical, which supports the hypothesis H21, which is based on the history of distributions, that students initially assume implicitly that distributions are symmetrical. In real life, however,

the phenomenon of weight shows distributions that are skewed to the right because of a "left wall effect" (two students had in fact drawn a left wall in the fourth lesson). By a left wall we mean that the lower limit (say about 30 kg) is relatively close to the average (53 kg) and the upper limit (sumo wrestlers can weigh up to 350 kg) is relatively far away from the average. This left wall in combination with no clear right wall causes the distribution to be skewed to the right. We therefore wanted skewness to become a topic of discussion as well.
In collaboration with the teacher, we invented the following activity to focus the students' attention to shape and skewness. We would draw the three student shapes on the blackboard and add two skewed shapes, which resulted in a pyramid, a semicircle, a bell shape, a unimodal distribution that was skewed to the right, and one that was skewed to the left (Figure 9.8). Students had to explain which shapes could *not* match the context of weight. We expected that it would be easier for students to engage in the discussion if they could argue which shapes were not correct instead of defending the shape they had chosen. Moreover, we anticipated a wider variety of reasoning than if all students defended one shape.

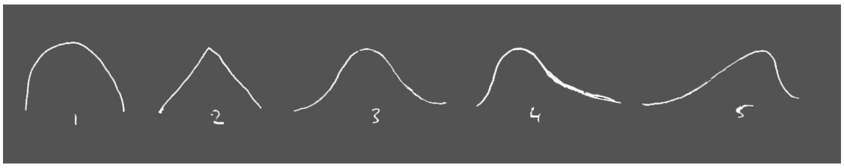

Figure 9.8: Five shapes as on the blackboard (1) a semicircle, (2) pyramid, (3) normal distribution, (4) distribution skewed to the right, (5) distribution skewed to the left. There were no axes with numbers.

9.4.2 Retrospective analysis

The teacher asked the class which shapes on the blackboard could not be right. For all shapes except the third, many students raised their hands. Apparently, most students expected a 'normal' shape (3). The teacher pointed out which students were to explain which shapes could not be right. Just as in class 1B we tried to involve all students in classroom discussions to avoid a situation in which just a small group revealed their thinking.

1 First, Gigi explained why the semicircle (1) could not be the right shape. One of his arguments was that there were too many people with low or high weights (the graph was relatively high at the endpoints of the shape).

Gigi: Well, I thought that it was a strange shape (…) I thought that the average was about here [a little more to the right than the top] and I found this one [peak of the hill] was a little too high. It has to be lower. And I thought that it was about 80, 90 [kg] here and I don't think that so many people weigh so much [points at the height of the graph at the part of the graph with higher values].

Teacher: (…) Does everybody agree with what Gigi says?

Tom: Yes, but I also had something else. That there are no outliers. That it is straight [vertical] and not that [he makes a horizontal gesture with two hands that looks like the tails of a normal distribution]. I would expect that it to slope more if it goes more to the outside [makes the same gesture].

Tom apparently understood that very low and high values do not occur very often, which means that the tails of the shape should be low and horizontal. Just as in grade 7, students used the term 'outliers' for low and high values that occur infrequently or for any values that deviated from the majority. As mentioned before, we concluded that we should have taken more effort to reserve that term for exceptional or suspect values and introduce another term for the tails of the shapes (such as low or high values; or the tails of the shapes).

2 Based on the students' reactions the teacher judged that they agreed that the semicircle was not the right shape, so she wiped it off the blackboard and turned to the pyramid shape (2). This discussion involved 'outliers' and the mean in relation to shape. It was Mourad's turn. So far students had only reacted to the teacher's questions–a very common type of classroom interaction (Van den Boer, 2003). In this lesson, however, students started to react to each other.

Mourad: Well, I didn't think this was the one, because, yeah, I don't think that a graph can be that rectangular.

Teacher: The graph is not so rectangular? [inviting him to say more]

Mourad: No, there are no outliers or anything.

Alex: It does have outliers; right at the end of both. It does have outliers.

Wim: That's just the bottom [of the graph].

Alex: At the end of the sloping line, there is an outlier, isn't there? (…)

Anna: But the middle is the mean and everything else is an outlier. [Other students disagree, e.g. Fleur:]

Fleur: Who says that the middle is the mean?

Anna: Yes, yes, roughly then.

Teacher: Tom, you want to respond.

Tom: Look, if you have an outlier, then it has to go straight a bit [makes the same horizontal gesture as before]; otherwise it would not be an outlier (…) but that is not what I wanted to say. I wanted to react, that it [this graph] could not be the right one, because the peak is too sharp and then the mean would be too many of exactly the same.

Mike: He just means that for exactly one weight all these kids weigh the same, so if the tip is at I-don't-know-how-many kilos, maybe 60 kilos, that all these kids are exactly 60 kilos.

In short, there were two aspects of the pyramid shape that students found inappropriate: the sharp peak (center) and the straight tails. In our interpretation,

these students understood that the frequencies of the weights around the mean would be similar; only if there were more of the exact same value as the mean, would the shape exhibit such a peak. We consider this kind of reasoning a form of mental experimentation in which students use statistical notions to reason what the cause of the sharp peak could have been. It is likely that their experience with dot plots in Minitool 2 and in lesson 4 enabled them to make this connection between the mean and the continuous shape. In retrospect, we often wished that we had asked students for more explanation: what do you mean by outliers? What do you mean by mean? However, it took quite some effort to have a class discussion and we decided to focus on the five shapes in the first place.[56]
Because the students agreed that this was not the right shape, the teacher also wiped this shape off the blackboard. As the interaction shows, students started to participate; their passive attitude started to change. Hearing the confusion between mean and mode, we decided to return to this issue at some point in the discussion.

3 Next, Sofie was to explain why the bell shape (3) could not be the right shape. Before the discussion almost all of the students thought this was the right shape (one girl admitted she did not know).

Sofie: I didn't have this as the one, because there are also overweight kids. Therefore, I thought that it should go a bit like this [draws the right part a little more to the right, thus indicating a distribution skewed to the right, Figure 9.5, shape 4].

The other students were not convinced. For instance:

Rick: That means that there are more heavier kids, but there are also kids who are underweight, so the other side should also go like that [this would imply a symmetrical graph].
Tom: Guys, this is the right graph!

Because several students still thought it was the right shape, the teacher did not wipe the normal shape off the blackboard.

56. Students were not always motivated to express their thoughts again and again: during interviews between lessons, a girl said, "In the beginning, I liked it, the first activities were pretty interesting, but then I thought that it was too much about one subject. Then it became a bit boring. Because, we had already discussed it, but you all kept asking about what we thought, though we had already explained it all. I found that a bit boring."

4 Next, Mike had to explain why the fourth graph could not be right.

Mike: I didn't think this was it because... if the average is, maybe, if this is the highest point, then this [part on the left] would be a little longer; then it would have a curve like there [left half of shape 3]. I don't think that this can be right at all, and I also find it strange that there are so many high outliers. Then you would maybe come to 120 kilos or so.

Some students argued that the mean need not be the value in the middle. Since students at this point argued about the mean versus the value that occurred the most, we decided to introduce a name for the mode, which these students had not learned before. Another reason we wanted to mention the mode (and in the ninth lesson the median) was that the instructional unit had to replace a schoolbook chapter on statistics in which the mean, median, and mode were addressed, which other teachers of the school were to teach. The teacher therefore needed to 'cover' these notions in this replacement unit. After the fourth shape had been discussed, we introduced a definition for the mode.

Researcher: The value that occurs the most often also has a name; it is called the mode [pointing at the value where the distribution has its peak]. (…) Who can explain in this graph [skewed to the right] whether the mean is higher or lower than the mode? (…)

Most students expected the mean to be higher (one raised her finger for lower). They found it hard to explain.

Tony: Most of what comes after it [the mode] is more than left of it, at the low side, so to speak. So there are more people with a high weight and few with low weight. [The students do not understand his explanation]
Res.: Who can say it again in his own words?
Rick: There are just more heavy people than light people, and therefore the mean is higher [in reference to shape 4].

Their remarks make sense if we interpret 'heavy' as heavier than the mode and 'light' as left of the mode and take into account that the right tail of shape 4 is further away from the mode than the left tail.

This is an example of how we used opportunities to introduce statistical terms, when students already talked about the corresponding concepts or informal precursors to them. Traditionally, the mode is introduced as the value that occurs the most in reference to data values (the upward perspective in Table 5.7), but we introduced it as a characteristic of a distribution (a downward perspective). Compared to the rather academic discussion of median and mean in class 1B, the discussion here of middle, mode, and mean was a heated debate in which many students became really engaged.

5 Last, Ellen said about the fifth graph:

Ellen: Well, I think this one is also wrong because there are more heavy people than light people. And I think that eighth graders are more around 50 kilos. That's it. [Note that there were no numbers in the sketch.]

Tom then objected that "nowhere does it say 50" and a lively discussion between the two evolved. We then asked Ellen to add numbers to her shape. She put 50 in the middle of the range, which would explain her saying that in this sketch "there are more heavy people than light people." Thus, as anticipated in the HLT, skewness became a topic of discussion in terms of heavy and light, even in relation to the mode and the mean, and the tails of the shapes, but not literally in terms of 'skewed.'

Students then wanted to know what the right shape was. Using a symmetry argument, we explained that it had to be something in between shapes 3 and 4.

Res.: Assume the shape were symmetrical, that left and right were exactly the same. Are there students in the Netherlands who weigh 90 kilos, do you think?
Students: Yes!
Res.: [Draws a shape with 50 as the mode and the right tail up to 90]. If it were symmetrical, then 10 kilos could also occur?
Students: No!
Res.: So it cannot really be symmetrical. It starts pretty low, 40 occurs. Perhaps even 30. And [there is] a tail to the right but not as long as in this shape [4].

About the activity

The aim of the HLT for this lesson was that students would learn to reason about skewed shapes, and they did so in terms of heavy and light. The satisfactory thing about this activity was that they came to reason with notions in a way they had not demonstrated before and that they were more engaged in the discussion than ever before, including the students with low grades for mathematics. We conjecture that the lack of formal rules, such as for manipulating algebraic formulas, makes it easier for low-achieving students to participate in the discussion. We furthermore conjecture that the lack of data, the game-like character and students' knowledge about the context were important factors, but also the fact that they had to argue against certain shapes. Such reasoning is safer than choosing the shape they think is right and defending that one. As we envisioned in Chapter 5, students started to develop a downward perspective on data: mode, average, and other statistical notions were discussed in relation to shape and were not just operations on data values.

With reference to the lignification metaphor (motto of this chapter), we could say that we had been successful in getting students to participate in reasoning about these shapes. However, they often used terms (in particular 'outliers') in unconventional or vague ways. On the one hand, if statistical concepts are defined before students even have an intuitive idea of what these concepts are for, then students' conceptual development could be hindered (as discussed in Section 2.2). On the other hand, if teachers and instructional materials do not guide students well in a process of reinvention, the tree remains weak and cannot grow higher. It is evident that the notions of average, outliers, distribution, and sample of students in the present research needed to be developed into more precise notions, but at least students developed a language that was meaningful to them. They developed an image that could be sharpened later on; the sappy part of the tree that could be lignified in the future.

In terms of diagrammatic reasoning, this lesson was mainly devoted to reflection on shapes, but there were also examples of mental experimentation (what would the shape look like if...). Skewness was addressed within the weight context, but had not been predicated yet with terms such as 'left-skewed' or 'right-skewed.' Students mainly used two distribution aspects in their reasoning, average and the tails ('outliers'). These notions are hypostatic abstractions which have become reasoning tools. However, as we have mentioned before, students mostly had somewhat idiosyncratic but understandable interpretations of these terms. From the analysis we concluded that students probably had the following understanding of distribution: there are many values around average (high rounded part in the sketch) and few low and high values, which is evidenced by the horizontal tails of the shape.

In terms of emergent models (2.1), the shapes had become models of data sets such as weight. We envisioned that students would next use these shapes as models for a more mathematical reasoning about distribution, and would recognize shapes in other situations.

9.5 Growing the jeans data set in Minitool 2

At the end of the sixth lesson, we prepared the jeans activity (cf. 7.12) by asking students to design an experiment to find out about the waist sizes of men for a factory of jeans.

In grade 7, the jeans activity had turned out to be too complicated, partially due to the sampling issues involved. Nevertheless we expected eighth graders with more sampling experience to do better. Students were told they could earn 1500 guilders with a good report (this was just before the Euro was introduced). We made sample size an issue by offering students the opportunity to 'buy' samples. Small samples were cheaper than large samples. In the seventh lesson, students compared the waist data sets of different sample sizes in Minitool 2 (Figure 9.9). We cite from a mini-interview to illustrate how high-achieving students reasoned about the distribution aspects including shape.

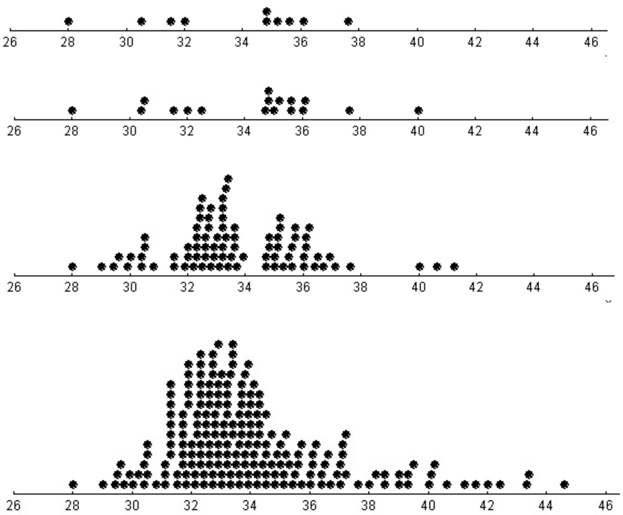

Figure 9.9: Jeans data sets of different sample sizes (10, 17, 82, 200); waist size is in inches. Students could scroll through them with the scroll wheel of the mouse.

Int.: You had first opened the data set with 10 values. Can you tell me how it changed when you got more data [17]?
Tula: Then you get a larger difference.
Int.: What exactly do you mean by that?
Tula: Well look. You can see that a bigger hump already merges here [around 32-34 inches].
Ellen: And it is more spread out with more jeans because you have [inaudible].

Like most other students, these two mainly focused on two distribution aspects: spread (difference, spread out) and center (hump, majority). Despite the previous lesson, skewness (the position of the hill in relation to the other values) was not an issue that students addressed themselves.[57]

Int.: What about the next one. What changed here [third graph with 82 dots]?
Tula: Then you see that a real hill begins to take shape [*dat hier echt al een heuveltje begint te vormen*]. (...)
Int.: And what happened to the spread? Is it larger or smaller?
Tula: Well, there are more dots. (...)
Int.: If you see this, what shape might you expect for a larger sample?
Tula: Well, that there would be even more of a hill here.
Ellen: And that something would come in between here as well.
Tula: And here little less, so that’s quite spread out.
Int.: Ellen, why do you expect something to come in between here?

57. In one of the homework tasks students had to predict the shape of train delays, which is a very skewed distribution.

Ellen: If you have more people, then you'll also have people with that size.
Int.: Let's have a look. [They open the data set with 200 values.]
Ellen: Yeah, you see, I was right.

The mini-interview shows that Tula expected a hill to emerge from the growing sample and that Ellen expected the holes in the dot plots to be filled in. This example shows how two high-achieving students started to see shape as a pattern in the variability of the data. In hindsight, however, their reasoning did not appear to be representative of the whole class.

9.6 Growing samples from lists of numbers

9.6.1 HLT for lesson 8: recognizing shapes in dot plots

The HLT for the eighth lesson aimed at using shape as a tool in reasoning about distributions and sample size. The intention of the HLT for this eighth lesson was that students would learn that sample size is important and that the shape of a sample in a diagram will stabilize if the random sample is big enough. In the activity, students had to draw a growing sample from a set of 250 numbers, plot their data, and stop if they thought they knew the shape of the distribution (we had made uniform, normal, skewed distributions, but students did not know these terms yet). We expected students to recognize certain shapes. After this activity we showed the students dot plots of the 250 numbers of each distribution, and taught the terms 'uniform', 'normal', 'skewed to the left', and 'skewed to the right' in an informal way. We knew in advance that this activity without a context would not be easy for students, but we did not want to underestimate their abilities.

9.6.2 Retrospective analysis

This activity was not a success: students did not see the shapes in their samples and they mostly did not tell more than where the mode was. There were too many problems to ascertain what the core problem actually was. We mention a few obstacles.
First, the target shapes were often not visible from students' samples; they were not even apparent to us and we were privy to the source from which the samples were taken. Sometimes, the sample size was just too small, and sometimes students just had bad luck. For instance, Melle had taken 100 values from the first list and had gotten Figure 9.10, which looks far from uniform.
Second, there were also occasions in which we saw a right-skewed shape, whereas a student could not see it. For instance, Melle did not regard his fourth shape as skewed, although his peer Sofie did. Third, some students took an unsuitable scaling which resulted in no particular shape whatsoever. Fourth, there was no every-day context such as weight or height that could help students attribute meaning to the numbers. Fifth, in the sixth lesson students had reasoned about continuous shapes of population distributions, not about shapes of sample represented in dot plots. These

two situations were too different. In the latter case, there is a lot of noise around the signal of the shape.

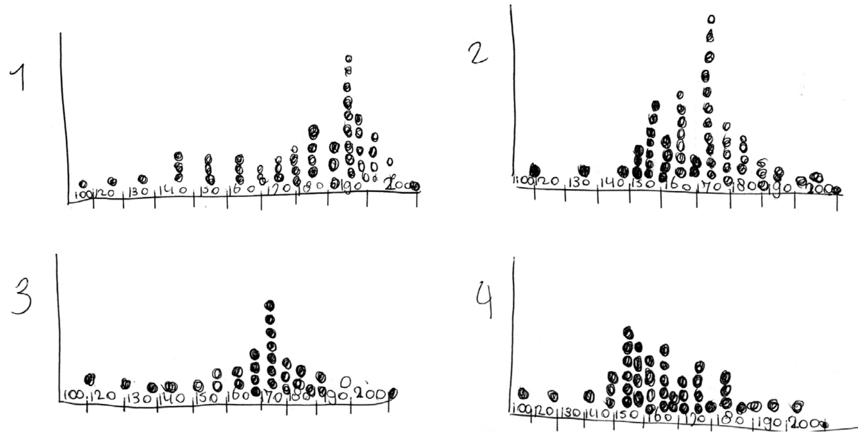

Figure 9.10: Melle's samples from the four lists (uniform, normal, left-skewed, and right-skewed).

In general, students did not interpret the diagrams the way we did, which indicates that they had not made the hypostatic abstraction steps we had expected them to make. Apparently you need to already have the shape in your mind to be able to see it in a diagram. Experts are inclined to separate signal and noise. The signal is the continuous shape with which they model the data, and the noise is the variation around the signal. Students had not yet learned to use the shapes they knew to model the samples they made, let alone to distinguish between signal and noise.

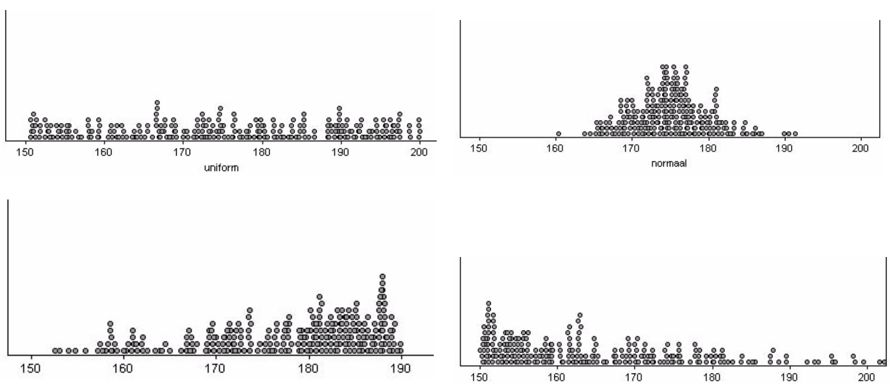

Figure 9.11: The populations of the four data sets from which students grew their samples (created in Fathom)

We mention these experiences as an illustration of a step in the HLT that was too big. The reasoning that students had demonstrated in the previous lessons appeared insufficient to recognize shapes in an abstract context. As in Section 8.4, the analyses of lessons 4, 6, and 8 also show that concept development is a gradual process with many instances of hypostatic abstraction. Additionally, the semiotic notion of diagrammatic reasoning, with experimenting as the second step, helped us to analyze what had gone wrong, as shown above. And from the fact students generally did not see the shapes that we saw in these diagrams, we conclude that they still had to perform certain steps of hypostatic abstraction. In semiotic terms, the same sign had different interpretants for the students than for us, because they had not constructed the same objects as we had in mind.

In Chapter 4 we mentioned the problematic relationship between phenomena and concepts as thought objects. We stated that people with different conceptual understandings can perceive different things. This section provided an example of that.

The analyses on diagrammatic reasoning about growing samples also indicate what the software could be useful for: experimenting with diagrams, the second step of diagrammatic reasoning. As shown in previous sections, students clearly made use of the two Minitool diagrams that they had used. In the present section it turned out that students had not had enough experience with an option that the software offers: scaling. If we had drawn more attention to this option and had highlighted it more in the HLT, the results may have been better. In terms of research, an advantage of this too large a step is that we were immediately able to see that these students certainly could not deal with much more advanced statistical problems.

9.7 Final interviews

In the final semi-structured interviews, we interviewed five pairs of students for about ten minutes per pair. In the fourth lesson, different shapes had been discussed and in the eighth lesson four types of distributions were given a name (uniform, normal, left-skewed, and right-skewed). Through these interviews, we wanted to find out what 'distribution' meant to students and whether they regarded measures of center as characteristics of a distribution. We decided to check whether students could correctly characterize a continuous sketch of a skewed distribution, and to ask them to indicate the position of mean, median, and mode. We discuss the results per question.

Table 9.1: Interview format

1	What is a distribution?
2	What kind of distributions are there?
3	Can you indicate mean, median, and mode in this distribution? (Figure 9.12) If students mistake the median for the midrange we also ask: Do you remember how you found the median in Minitool 2?

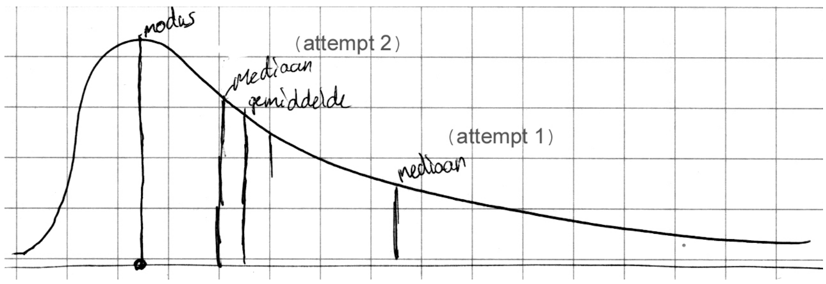

Figure 9.12: Sketch of a right-skewed distribution used for question 3. *Modus* = mode, *mediaan* = median, *gemiddelde* = mean. In the first attempt, the median was mistaken for the midrange.

What is a distribution?
Based on previous lessons we had expected that students would say something along the lines of "how the data are distributed in the graph." We present all the actual answers.
Pair 1. Steven thought that distribution was the "step size of the axis." Lynn added, "Then I think of the distribution of the data across the graph." Steven's answer probably refers to how the axis is labeled or how the data are structured, for instance in equal intervals. It could well be that he interpreted the Dutch term *verdeling*, which also has the connotation of division, as standing for how the axis was divided. Lynn's answer is an example of what we had expected.
Pair 2. Other students, such as Sanne, seemed to think of spread, or how the data are spread out. Her peer, Fleur, reacted with surprise.

Sanne:	If you have, for example, a graph with a lot of dots, then they can all be on one stack, so to speak. Then it is a very small distribution, because they all lie close to each other. But if they lie at the end and the beginning, then it is a very large distribution because they lie spread out across the line [the axis].
Fleur:	Is this the spread? Or is that the same? (...) I think the distribution is for example how you, how you for example uh, on this line, how you divided it, by 5, and then you get 10, 15, and then 20. How you divided it.

Where Sanne referred to spread, Fleur seemed to think of how the axis was divided (like Steven).
Pair 3. What Natasha and Sandra said came closer to what we had anticipated:

Natasha:	That is, uh, how do you say that. How, let's say everything (...). How everything, not the spread, but how the dots are in the graph.
Sandra:	How you put it then in the graph, the distribution of the uh, the dots then in the graph, where they are.

Pair 4. John described the process of getting a distribution, and it could be that he was referring to the transition from Minitool 1 to Minitool 2:

John:	Uh, … [laughs]. Well, you have a graph, with dots that have fallen down. And then at the bottom you have numbers, and the distribution is, how it is, how do you say that?
Int.:	Well, you already point it out, how it is...
John:	How it is put down, uh, for example, 20, 30, 40, 50, 60 and then the dots at uh, then you have for example 40 dots and there are for example fifteen of 20 and uh, fifteen of 60, and so on. But then you distribute it over it. That is the distribution.
Int.:	And what would you say, Ron?
Ron:	A distribution can also be straight, uh, be skewed, and uh, normal, that it is distributed evenly.

John's answer is interesting with respect to the process-product issue we address in Chapter 5. In that chapter we asked what the process aspect of distribution could be. Students' descriptions, such as John's might provide us with a clue: imagine that dots are distributed one-by-one over a variable; the result is the distribution. In fact, this is what students do if they grow a sample. We conjecture that the focus on growing samples in this teaching experiment has influenced student views on this process aspect of distribution or has provided them with a language to express their thoughts about distributions.

Ron was the only one who mentioned names of distributions learned the week before. Because students often used 'spread out' and 'distributed' interchangeably, we asked John:

Int.:	And what is spread again?
John:	You mean how the data is spread out or, in fact [laughs] or distributed over the graph.

This is another example of conjecture C7 that students' notions of spread, distribution, and density are intertwined. This supports the conjecture we make in Chapter 7: to develop a notion of distribution we might as well focus on spread and ask students to describe how data values are spread out, for example in dot plots.

Pair 5. Erno's answer also demonstrates this close connection of spread and distribution.

Erno:	Uh, distribution, that is uhm, pff, uh, how, yes, the data are spread out, are distributed.
Int.:	How would you say it, Alex?
Alex:	A distribution is the number of dots from 50 kilo to 100 kilo and then distributed [or divided], and that in the middle there is more for example.

Alex's answer hints at the image of distribution that we described in earlier lessons: a distribution as consisting of low and high values with low frequencies and a middle group with higher frequencies (C1).

What kind of distributions are there?

Most of the students did not remember precisely what the notions of uniform, normal, skewed to the left and right meant. All students except two called a distribution that is skewed to the right 'skewed to the left.' Most of the students who initially said 'skewed to the left' corrected themselves. Instead of uniform or normal they often used terms such as 'even' or 'equal' (*gelijk* in Dutch). Most of the time it was unclear whether they meant uniform or symmetrical. What is interesting from a semiotic point of view is that students generally do not say "the distribution is skewed to the right," but use demonstrative words such as 'that' or 'it' (which are indexical) as in "that one is skewed." In Section 8.3 we present an example of a student saying "that one is normal." Our impression was that students did not feel comfortable using the term 'shape' or 'distribution' itself. We consider this a pre-stage of a proper hypostatic abstraction.

Though it was interesting to see what students' definitions of distribution were, we think in retrospect that we should have stayed closer to what students had done in previous lessons. For instance, we could have asked students to describe the shapes in plots such as in Figure 9.11, to find out what they saw in them and how well they could describe aggregate features of those distributions. We could also have asked students to grow a sample from a large data set and to describe and predict aggregate features of the data set.

Measures of center in the skewed distribution

Students found it easy to find the mode, which is just where the distribution has its peak. However, they were not able to formulate it this way. For example:

Int.:	What was the mode again?
Fleur:	There where it is the highest. Isn't it?
Int.:	What is the highest?
Fleur:	Av... no not the average. How is it called?
Int.:	Do you know? [to her peer]
Sanne:	That line, where it has its highest point.
Int.:	Where the graph, the distribution has its peak.
Fleur:	That's what I meant.

Most could give a rough estimate of where the mean is. Their strategy seemed to be to start in the middle and look for high or low values that would influence the mean in one direction. For the median, all students except one pointed at the midrange, halfway between the minimum and maximum value. This happened often in the seventh grades (C8, see also Section 6.9). This is not really surprising because the median was only introduced on paper during the ninth lesson in a computer task, and was not further discussed by the teacher.

When students indicated the midrange as the median, we asked, "Do you remember how you got the median in Minitool 2?" After such questions, several students concluded that there should be the same amount of dots left and right of the median.

Fleur: So assume there are 100 dots in total, there must be 50 over there and then you put a line roughly there and then there are also 50 over here. [her estimate was quite accurate]

This could be called mental experimentation: Fleur made a mental connection between a dot plot and the continuous shape (see also Sandra's graph in Figure 9.7). Doing so she could accurately estimate the position of the median in the distribution. It is likely that students' experience with the two equal group option in Minitool 2 helped them to answer the question about the median. In our view, being able to locate mean, median, and mode in such a continuous sketch adds extra value to being able to calculate or determine these measures of center from a set of data. Located in a shape sketch, there is a meaningful relationship between these measures and the shape as a whole.

In short, students' notions of distribution could be characterized as follows. They had a sense of distribution as consisting of low and high values that occur infrequently and an average group with a higher density (we also discussed bimodal distributions in one of the lessons). They understand what the consequences of the differences in density are for the shape of the diagrams (low is infrequent and high is frequent). Students imagine a distribution as coming into existence through growing a sample (or collecting and plotting data): the dots are distributed over the axis with a specific pattern ("more in the middle," for instance). In this teaching experiment we tested the conjecture that students could develop a notion of distribution by diagrammatic reasoning about growing samples. In conclusion, we would say that this is indeed the case. By growing samples, students developed a sense of how distributions come into being, where hills emerge. There are indications that they see that holes get filled in at some point, and that the hill might become a little wider due to new extreme values.

9.8 Answer to the integrated research question

In the teaching experiment in grade 8, we tested the conjecture that students could develop a notion of distribution by reasoning about growing samples. From the results we concluded that this was indeed the case. This section summarizes an answer to the question of *how* eighth graders can develop a notion of distribution by diagrammatic reasoning about growing samples.

Distribution is a multifaceted concept that is difficult to learn, but learning to reason about distribution can be stimulated by the activity of growing samples as described in this chapter. We analyzed students' reasoning as instances of diagrammatic reasoning. What played a crucial role in this process is hypostatic abstraction as the formation of objects. From the analyses presented above, it is clear that the development of the multifaceted concept of distribution included several steps of hypostatic abstraction.

We give a few examples of hypostatic abstraction from the previous sections. In the first and third lesson, students described what a larger sample of a good and a bad battery brand would look like. In this way, we stimulated predication and mental experimentation (a what-if attitude). In the fourth lesson, students used predicates such as ‘spread out’ and ‘together’ to say something about the dots (representing the data). Then they came to use nouns such as ‘spread’, ‘average’, and ‘outliers’, which indicated steps of hypostatic abstraction of distribution aspects, and they used these nouns as tools in their reasoning about shapes of distributions. Students started to see shape as something that they could talk about, which indicates another instance of hypostatic abstraction. Students guessed about the shape of the weight distribution and came up with semicircle, pyramid, and bell shape. It must be due to their experimenting with value-bar graphs (in Minitool 1) and dot plots (Minitool 2) that students were able to construct the diagrams of Figures 9.3 to 9.7 and reflect on them. In the sixth lesson, students discussed these shapes and used statistical notions such as mean and outlier to explain why certain shapes could not represent a weight distribution. During the diagrammatic reasoning about this, the statistical notions of mean and ‘outlier’ were used as reasoning tools. Students also implicitly reasoned about frequency and density in this phase. To make skewness a topic of discussion as well, we introduced two skewed shapes.

In the eighth lesson we discovered that we had made too big a step in the HLT. Although students had learned to reason about continuous shapes in lesson 6, they found it difficult to recognize shapes in growing samples represented in dot plots without a meaningful context. If we semiotically compare how students interpret the shapes in lessons 6 and 8, it becomes clear that students had not yet made certain hypostatic abstraction steps necessary to recognize shapes in dot plots with a lot of variation instead of the continuous shapes used in lesson 6. Additionally, what had been missing was a specific type of experimentation with diagrams: scaling. Without a proper scale it is hard to see the shapes of the distributions. Yet students learned to describe shapes and distributions as being uniform, normal, skewed to the left or to the right. This means that shape had become an object they could reason about.

One of the end goals of the HLT was that distribution would become an object-like entity. In Chapter 5, we argue that a distribution is more like a composite unit than an object with a procedural and structural side and wonder what the procedural side of a distribution would be. From the way students talked about distribution, in particular during the final interviews, we inferred that they imagined the process of distributing dots over the variable as if growing a sample in a dot plot. We conjecture that such a process view of distribution could well be the procedural side of the concept, but we realize that our focus on growing samples has also fostered this view of distribution.

The analysis shows that the reification process of distribution is a complex process that involves many steps of hypostatic abstractions. Understanding distribution requires understanding key aspects such as center, spread, density, and skewness. There even seems to be a reflexive relationship between the development of such characteristics of a distribution and the notion of distribution as an object or a shape: by reasoning about the occurrence of low, average, and high values, students expect a particular shape, and by reasoning about shape, students develop the meaning of distribution aspects such as mean, spread, density, and skewness.

We cannot answer the question of whether distribution had indeed become an object for the majority of the students without specifying what we mean by distribution and by object. In Section 10.1.3 we specify different levels.

As a final remark we would like to stress that, for instructional design to be successful, it is not enough that the instructional materials are well designed. The fact that the growing samples activities turned out successful was, in our view, due to a balance between students' background in mathematics, the Minitools, the teacher's ability to orchestrate discussions, and the timing of the activity. In line with findings of Kanselaar, Van Galen, Beemer, Erkens, and Gravemeijer (1999), we contend that these variables cannot and should not be investigated as separate factors. The methodology of design research offers a way to investigate these issues coherently. The metaphor that came to mind is the sound of a symphony orchestra. One oboe player playing too high, a trumpet player playing too loud, or a timpanist playing too early can ruin the chord as a whole. A chord only sounds good if all musical aspects are in tune with each other. The advantage of design research is that this tuning process can be accomplished in different cycles of anticipation and adjustment. As a consequence, we recommend to invest in using computer tools only if other factors such as teacher support, instructional activities, end goals, and assessment are adjusted to using these computer tools.