Manuscript

Roth, W.-M., & Temple, S. (in press). **On understanding variability in data: a study of graph interpretation in an advanced experimental biology laboratory. Educational Studies in Mathematics. DOI: 10.1007/s10649-014-9535-6**

ABSTRACT Data analysis is constitutive of the discovery sciences. Few studies in mathematics education, however, investigate how people deal with (statistical) variability and statistical variance in the data to be interpreted. And even fewer, if any, focus on the uncertainties with which scientists wrestle before they are confident in the data they produce. The purpose of this study is to exhibit the work of coping with variability in one advanced research laboratory, as exemplified in a typical data analysis session. The study shows that when the scientists are confronted with novel data, their understanding of the variability does not arise in straightforward fashion and a lot of normally invisible (interactional) work is required to constitute understanding. Tentative conclusions are provided for the implication to mathematics education.

KEY WORDS: variability; data interpretation; discovery sciences; data analysis; mathematical practices

There is a "huge need" for statistical competencies in the modern workplace and, therefore, a need for statistical education (Scheaffer, 2011). Mathematics educators are interested in the way students interpret data and in their statistical practices and competencies related to data analysis (Garfield & Ben-Zvi, 2007; NCTM, 2000; Roth & McGinn, 1998). Mathematics educators often ask students to collect data that exhibit functional relations (e.g., Confrey, 1995; Gerovsky, 2010; Hardy, 2009; Kaput, 1988) and interpret "well-defined" data and samples (Konold & Pollatsek, 2002); and school science courses and textbooks tend to emphasize function-like relations between variables (Roth, Bowen, & McGinn, 1999; Roth, Pozzer-Ardenghi, & Han, 2005). Unsurprisingly perhaps, when asked to find trends in actual data, even preservice science teachers with BSc and MSc degrees had trouble dealing with variability (Roth, McGinn, &

Bowen, 1998), whereas eighth-grade students following a curricular unit where they designed investigations, transformed data, interpreted data, wrote reports, and argued over and about their findings often used sophisticated statistical techniques to deal with variability (Roth, 1996). Conversely, scientists asked to interpret graphs from introductory college courses pointed out, however, that function-like data and graphs are abnormal in science (Roth, 2003). Allowing students to be able to interpret the variability of data values from empirical inquiry generally and to separate chance variability from systematic variation specifically are important goals of mathematics education not only in primary (English, 2012) and secondary schooling (Cobb & Tzou, 2009; NCTM, 2000) but also at the tertiary level (Garfield, delMas, & Zieffler, 2012; Wild & Pfannkuch, 1999). Comparing models and empirical data, and dealing with variability, are integral parts of learning statistical practice (Garfield & Ben-Zvi, 2008; Pfannkuch, 2005). In the development of statistical literacy, understanding the relationship between models and data is an important dimension (Garfield, 2011; Garfield & Ben-Zvi, 2007). This is so because there might be patterns even though data exhibit considerable variability; extracting such patterns in the face of inevitable variability was, until recently, not a competency that mathematics education would have allowed students to develop (Konold & Pollatsek, 2002). One of the difficulties people of all ages have arises from everyday confusions between randomness and excessive alterations (Falk & Konold, 1994). Similar to other studies investigating mathematics on the job for the purpose of thinking about how mathematics curricula might be organized (Gainsburg, 2007; Hall & Stevens, 1995; Noss, Bakker, Hoyles, & Kent, 2007; Williams, Wake, & Boreham, 2001), this study investigates what do scientists do when they are uncertain about their own graphs and the variability (and statistical variance) displayed therein. The study presents for the interpretation of mathematical culture suitable "thick [ethnographic] description" (Geertz, 1973) of data investigation practices in a situation where research scientists found themselves in a highly uncertain situation about the source of variability in their data.

Research design

This study was designed to investigate what scientists do with graphs when they are in situations of considerable uncertainty and, therefore, are not in a situation to provide some routine explanation to something that they are highly familiar with (as an economist with a supply-demand graph). The sources of the data are the records from a 5-year ethnographic project in one experimental biology laboratory, where we were members and participated in the design of studies and publication of research results. The particular research project from which the present recording stems concerned the absorption of light in the rod-shaped photoreceptors of coho salmon eyes. There are two absorbing chemicals, one based on vitamin A_1 (rhodopsin), the other one based on vitamin A_2 (porphyropsin). Our research attempted to answer the question of how the relative amounts of porphyropsin and rhodopsin in the photoreceptors changed in the course of the coho salmon's life cycle. In this study, we focus on parts of one data interpretation meeting that the team periodically organized.

Five team members were present: the laboratory head (Craig), a research associate with background in physics (Theo), a postdoctoral fellow (Elmar), the second author who was a doctoral student in biology at the time (Shelby), and the lead author (Michael), a physicist and applied cognitive scientist by training. Shelby operated the computer and projected data onto a screen; a chalkboard behind his back was also used for drawing graphs. At the time, Craig already has had more than 30 years of research and publication experience in this field; he now holds an endowed research chair at another university. Shelby, too, has had many years of experience doing research on fishes around the world, had worked as a biologist in various jobs prior to completing an MSc and was working on his PhD. Theo had done undergraduate studies in physics and work experience in developing analytic and visualization software.

The present study treats results "not as specifications of practice but rather as guides for locating the field properties that make the practices recognizable for what they are" (Koschmann & Zemel, 2009, p. 234). It can be read as a micro-genetic study of scientists' methods of dealing with variability while using data and graphs, here in an advanced laboratory on the neuroethology of fish vision. We are concerned with providing an ethnographically *adequate* description, which means that the "ethnographer must articulate the same hesitant and momentary contexts that the natives are displaying to each other and using to organize their concerted behavior" (McDermott, Gospodinoff, & Aron, 1978, p. 246). Thus, we treat the data "as an occasion for construing the work done by participants in interpreting the document as it unfolds before them" (Woolgar, 1990, p. 127) and before they know how it will ultimately be understood. In the same way other research concerning the natural sciences are conceived, as the present is a study *of* rather than *about* scientists' work (Garfinkel et al., 1981). Ethnographic adequacy also implies that the "analyst must be *vulgarly* competent in the local production and reflexively natural accountability of the phenomenon of order* he is 'studying'" (Garfinkel & Wieder, 1992, p. 182). This "vulgar competence" allows "the analyst to recognize, or identify, or follow the development of, or describe phenomena of order* in local production of coherent detail" (p. 182). In our situation, these requirements are met, as we have been active members of the laboratory.

Variability and ways of handling it in a data analysis session of an advanced biology research laboratory

This study was designed to investigate how scientists deal with uncertainty in data and graphs while the latter are produced and interpreted *in the course of* discovery work, that is, prior to the scientists' realization that they have found something of interest. In the following, we analyze an extended, 15-minute episode from a laboratory meeting. The episode was occasioned by two problems that Theo and Shelby identified in preparation for this meeting: (a) there was considerable variability in their data (which expressed itself in large standard deviations the team calculated) and (b) their compiled data differed from those published 8 years earlier by another research group, which was consistent with the current dominant scientific theory. This part of the meeting was designed to understand the first of the two problems: rather than having a lot of A_2 values more sharply distributed around some mean, their A_2 values were spread across the entire spectrum of values (from 0–100%) (Fig. 1a). To get at the levels of porphyropsin (A_2) in the photoreceptors—that is, the real data of interest—the team needed to know the position of the maximum of the absorption curve and the width of the light absorption curve at half-height (half-maximum bandwidth, HBW) (Fig. 1b). Craig had drawn a curve on the chalkboard of what he expected the graph to look like when HBW was plotted against the amount of porphyropsin (%A_2) (Fig. 2a). Shelby plotted HBW against %A_2 from the data that the team had collected thus far (Fig. 2b): as everyone in the room could see, the data apparently did not fit the expected curve. The discussion during the meeting segment analyzed was concerned with understanding the pattern in the data, its difference from the expected curve, and the nature of the variability that at one point in the meeting expressed itself as variance (standard deviations in Fig. 5).

««««« Insert Fig. 1 about here »»»»»

««««« Insert Fig. 2 about here »»»»»

Possible sources of variability

An important requirement for getting a handle on variability exists in knowing what the exact nature of the data is, what has been included and what has not been included, what the criteria for inclusion and exclusion are, whether what has been done can be defended in terms of such criteria, and what the level of the integrity of those curves is that result after multiple transformations of the original data have been implemented. Because each step in the translation involves crossing an ontological gap, a crossing that is inherently grounded in and legitimized by scientific practice (Latour, 1993), the scientists can guarantee the quality of their work only if they are (relatively) certain that each step in the chain of transformations from “natural object” to the ultimate claim is defensible. It does not then come as a surprise, perhaps, to hear Craig repeatedly ask about the criteria that were applied in accepting or rejecting spectra. The following discussion emerged after Craig had asked whether the plot came from all data (“the whole gamut”) and, therefore, was in turn reflected in the data that the team was interested in (the amount of porphyropsin in the photoreceptors or %A_2). Craig has identified a potential problem. Before he can provide a reading, should this occur, the background necessary for an appropriate reading of the graph that is now projected on the screen (Fig. 2b) needs to be articulated in and for the group. This background is part of the whole research procedure that one needs to know to make sense of any graph that issues from it (e.g., Noss et al., 2007; Roth, 2003). This is what is happening in the current subsection. The problem our team faces is that we do not know what the source of the variability is: it could be within the fish eyes, due to systematic error from the instrumentation, an artifact arising from the curve-fitting and extraction of HBW, or random error.

Articulating the procedure for extracting the data Shelby describes that he "can present" what "is in [his] database." Theo formulates the existence of a problem ("the problem is"), and begins with a description of the curves as being "fairly long." As Craig before, Theo gets up and walks to the chalkboard. Upon arriving there, he produces a curve that we recognize to be an idealized but widened absorption spectrum (Fig. 3a). Theo marks three areas where he "can look." He refers to a rule in the field, which states that the important region is defined by the right part of the curve above the 50% of its height. To determine the total height, Theo states needing to know where the base lies (left and right circle) and what the maximum height of the absorption curve (λ_{max}) is. First for the right side of the curve and then for the right-hand "foot," Theo draws the kinds of features that he observes—and willingly demonstrates and explains to visitors to his office—while analyzing the data that he receives from the wet laboratory (where he participates in some of the data collection sessions operating the software). These features, sudden drop offs in otherwise "clean spectra," make the determination of the baseline points difficult and, therefore, the precise determination of the height of the absorption curve and introduces variability in its location along the wavelength axis. Although an "otherwise clean" graph does contain such problems, Theo explains that he does not reject the data at this point "as long as [he has] enough peak in here, under, above the 50% line." Such a curve still allows him to determine λ_{max}, and, therefore, the amount of porphyropsin present. This is how he sees the situation, and based on it he keeps the data.

«««« Insert Fig. 3 about here »»»»»

In this sequence of turns, the participants (and the analyst) therefore find out that Shelby has included in the plot currently displayed all the data that Theo has provided him with. These, in turn, are based on what Theo has received from Shelby, who collected the curves from the retina in the dark laboratory. Theo explains that there indeed are curves that exhibit sudden drop offs at the left and right end of the otherwise clean absorption spectra. He does not throw these data out, at least he does not as long as he has "enough peak in here," that is, between the half-maximum point and the top of the peak at λ_{max}. Although not specifically articulated as such, what is available *in* and *as* their relation is a possible reason for their HBW data to be noisier than they should be in some ideal case. Theo's presentation actually anticipates possible problems in the interpretation of the data, as he explains possible variability in the data that might wash out the true trends that possibly exist or that make interpretation possible.

Explicating the possible source of the variability When there is considerable variability in the data, the identification of "deep structure" may be difficult. When scientists articulate trouble reading or interpreting data because of variability that may not be understood, observers are provided with opportunities to see what they do when the deep structures visible in idealized graphs—such as those that Craig and Theo have drawn on the chalkboard—are not immediately available in real, experimental data. At this point in the meeting, the team is confronted with a plot of the half-maximum bandwidth of the spectra it collected (i.e., HBW) against porphyropsin levels (%A_2). This plot (Fig. 2b) reflects the real data; it constitutes a stark contrast to the idealized curve Craig has sketched on the adjoining wall. This idealized graph presents the expected ("preconceived") deep structure whereas the graph that has arisen from the exchanges between Shelby and Theo presents the real case, the deep structure of which is at issue in this subsection. What will scientists do to come to grips with the apparent variability in the currently displayed data plot? In the sciences, the exclusion of outliers is a common practice. However, if a "point" represents many measurements rather than a single one, excluding it becomes more problematic (unless, perhaps, if scientists recognize some systematic error in the many measurements underlying the single point). In this part of the episode, the possible exclusion of points is the topic of the talk.

Theo, who is still standing next to the diagram that he has earlier drawn immediately follows Shelby by saying that there is “a little noise on here” while pointing to the left side of his absorption graph. There is, as all laboratory members know, the familiar “ringing”—oscillations overlaying the left branch of the approximately Gaussian absorption curve—the origin of which was not known (Fig. 3b). Because of this variability, the half-maximum bandwidth is measured to the point where the “ring” first crosses half of the maximum height, so that its measurement ends up being smaller than it should be. The algorithm that Theo has designed at that instant in the research does not project where the Gaussian *would* lie but takes the half-maximum bandwidth at the location along the wavelength axis where the absorption curve first falls below half of maximum height.

In this situation, Theo modifies the original graph for the purposes at hand. That is, the original smooth graph was to show one feature and was produced for the purposes of communicating it: the places that are of importance in determining the height of the curve and, with it, the half-maximum bandwidth. Now that the data exhibit a lot of variability, he modifies the *ideal* and idealized graph to introduce one of the aspects that can produce such variability. That is, the variability visible in Shelby’s graph does not arise from the source, the retinal cells, but from the transformations that the measured absorption curves undergo towards producing the porphyropsin content (determined by means of λ_{max}) and the half-maximum bandwidth, which the team hopes to turn into a second reliable measure of A_2 levels. It allows Theo to articulate where variability comes into the determination of the half-maximum bandwidth as existing in the database at the present point in time. Theo proposes something, though the speech volume becomes so low that it is no longer possible to retrieve what he is saying while Craig overlaps uttering the tentative proposal to throw those data points out (“huck those points”). He proposes this elimination because Shelby’s curves did not fit this template (curve), due to large variability, even though each point plotted represents the calculated mean of a series of measurements, and therefore are not simple outliers (singularities).

In this situation, the graphs on the chalkboard have a different function from the graphs produced from the data: these are tools to think, as research on diagrammatic reasoning has shown in school settings where students discuss variability in data (Bakker & Hoffmann, 2005) or in contexts where scientists are asked to interpret unfamiliar data (Hoffmann & Roth, 2005). In such situations, graphs offer opportunities for the "hypostatic abstraction," that is, "the formation of objects (what can be talked about)" (Bakker & Hoffmann, 2005, p. 350). In the present instance, the "objects" to be talked about are possible sources that generate the variability in the data to be interpreted.

Call for disattending to variability

Once scientists know that they can legitimately ignore some data points, for example, because these can be justified to be outliers, then they are in a position to read the data in new ways. They may, as happens here, *provisionally* ignore some data points to see what it yields, much in the way painters do after placing a stroke and then stepping back to evaluate the effect of their move and then to correct or continue with what they have done. That is, it would be typical to articulate a move and thereby to objectify it to be in a better position for evaluating its implications. The advantage of such moves is that new insights thereby arise from an external, epistemic action that reveals its possibilities by inspection and further possibilities for theorizing through reflection.

Interpretive difficulties Following Theo's explanation of a possible source of the variability in the data, Craig provides a first reading—which we may hear as a first attempt in articulating the deep structure in the data. He points to the right-most peak ([i], Fig. 4a) and marks it as something that "he does not know what's going on" and, prior to articulating a trend, he suggests "ignoring it for the time being." He also points to the two center peaks upward ([ii] & [iii], Fig.

4a) and to the two downward peaks ([iv] & [v], Fig. 4a), marks them verbally as "extremes," and suggests "ignoring them." He finally proposes "running a regression on this" (Fig. 4a). At this point, Shelby brings out a contrast by suggesting that "this is not just a couple [of] points," but in fact "this is the whole shebang." That is, we see and hear a proposal for "hucking [throwing out]" or "ignoring" certain "extremes," but this proposal is not accepted but rather confronted with the statement that there are 3,000 data points. Shelby contrasts this with the description of not just being "a couple of points." This large number, as he elaborates, implies that there are *some* extremes. But because this is a large number, any trend would be rather stable with respect to a small number ("some") extreme points.

««««« Insert Fig. 4 about here »»»»»

The extreme points are not just there. To exist qua orderly object in the hand of the researchers and therefore at hand, any "point" has to be demonstrably "extreme." Here, this is achieved in the sequentially produced contrast Craig produces between the point made salient by the indexical gesture and the subsequent hand movement that not only bypasses the point but also does so to such an extent that it becomes an instructible instance of "extreme." Although much of the work on gestures does not treat the two gestures as distinctive—other than denoting them as "deictic" and "iconic"—there is an ontological difference between them. This difference constitutes the heart of the dynamical understandings in physics and mathematics (Châtelet, 1993). This is so because the hand movement enacts a curve, represented in mathematics by a relation, such as $G(x,y) = 0$, whereas the individual points on such a curve, denoted by indexical gestures, represents a parametric or function of the kind $y = f(x)$ where the separation of data and result is already accomplished. It is precisely in the gesture that a virtual relation becomes actual and material, in and through the body of the scientist (Châtelet, 1993). That is, "the gesture envelops before grasping and sketches its deployment before denoting or exemplifying" (p. 33, my translation).

Hypothesizing a pattern Craig has suggested "huck[ing] those points" that vary a lot from the apparent trend (i.e., the outliers). This outlier aspect exists in a demonstrable manner such that when required, the spiky nature of the “downward spikes” can be exhibited for anyone not yet convinced. He again suggests ignoring some of the outliers, but Shelby has responded saying that the dataset is large and that in this case it is likely to have some that reflect large variations. Craig now suggests running a “polynomial on it” to “see what sort of correlation we can come up with,” but adds that they first have to “*recursively* go back through the data” to see whether “those points [outliers]” are “due to errant spectra.”

Craig further proposes that these could “have provided considerably more variance to that point in the transition,” thereby referring to data reflecting—according to the “dogma” that framed their work at the time—a changeover from high porphyropsin levels thought to be typical of salmon in freshwater to lower porphyropsin levels thought to be typical of salmon in saltwater. As he continues, Craig first acknowledges what Shelby has said (“yea”) and then provides an elaboration of the data by showing the trend that they display (Fig. 4b). Placing his pen against the projection screen, Craig gesturally “draws” a curve (Fig. 4b) that looks like the one he has earlier drawn on the chalkboard (Fig. 2a)—even though he has just prior to this instant exhibited the opposite relation. There is a high degree of similarity, even in the relative position of the HBW, which is higher for 100% A_2 than for 0% A_2. (Shelby has drawn the abscissa from left to right, and, therefore, in the reverse of what Craig has done on the chalkboard.) In this situation, therefore, Craig explicitly notes that the trend he has just exhibited in and through his hand/arm movement, “this,” “looks like that,” the earlier graph. The same movement in the reverse that earlier had produced the line on the chalkboard now, ephemerally, exhibits the trend in the data. In this way, what we might alternatively see as a parabolic trend, visible if a regression were to be performed on the data displayed, becomes an idealized inverted parabolic trend (deep structure) represented in the trajectory of the pen over the data.

In the course of his talk, Craig has actually changed what he brought out as the trend underlying the data. Initially, he gestures—indexically/iconically in his movement, iconically with the hand inclined to show the linear relation from lower left to upper right at the right end of the plot. He then turns and points to the graph on the chalkboard, which in fact runs differently from what he has shown just now. He articulates a curvilinear relationship with a maximum in the center (Fig. 4b). That is, the hand movement both exhibits what is required to demonstrate the curvilinear relationship and serves as an instruction for seeing (again) this very relationship. It therefore exists twice, in the form of “two technologies”: as orderly phenomenon that everyone present can see once he has learned how to look and as embodied practice, which, when enacted, *gives* immediate access to the so articulated phenomenon of the curvilinear relationship.

In this situation, the HBW versus %A_2 graph provides an indication of the variability in the data, perhaps more sensitive than other indicators because of the reasons that Theo earlier elaborated. Craig insists on checking the data that they all meet the criteria, especially those “low points.” He is concerned with the fact that data such as those in the “downward spikes” would introduce unwanted variability in the determination of the actual proportion of A_2. Craig later brings the conversation back to the issue of what has been included as data in the plot, to which Shelby responds that he did something quick. Craig insists, “does that mean *every*thing?” and Shelby affirms. But again, he insists that there are “a lot of data points,” as if he wanted to warn the participant to “huck [throw out] those points” all too quickly.

Articulating further uncertainty Craig begins making a statement about being “pretty . . .” but he is interrupted by Michael’s question about how many data points there are in each of the spikes. This question can be heard as co-articulating uncertainty about what exactly each data point reflects: individual or collated measurements. Shelby ascertains having heard my utterance as a “good question” but says that he does “not know how to tease this apart.” He articulates being surprised about the “spiky” nature of the graph and that he may have “chosen the wrong type of graph.” He elaborates by reiterating that there are 3,000 data points, which would mean that it “takes a lot” for the curve “to go down that far.” Then he states again to be “kind of surprised.” We can hear him express surprise about the variability and, therefore, co-express his anticipation to have a much smoother graph. Craig then articulates the hypothesis that Shelby has "mix[ed] quite a broad spectrum of fish" and "that could have something to do with it." In fact, the fish that have contributed to the data were from different hatcheries and from different river systems. This "could have something to do with [the variance]." Here, only the familiarity with the background to the data collection allows Craig to articulate this hypothesis.

Reducing complexity

Reducing the complexity in and of the data by limiting the source was one of the strategies that was repeatedly observed in the course of this study. In the present instance, this occurred immediately after Craig's articulation of the hypothesis that the variability might be caused by pooling all the fish in their study. Shelby suggests that he "could just choose one set." He proposes using fish that had come from the Kispiox River in northern British Columbia. Craig accepts by restating that possibility. The selection of a subset of the data, which was introduced as a way of decreasing variability actually appears to have increased it. However, a direct comparison between the former and the new plot shows that there are in fact new features that appeared and old features that had disappeared (Fig. 4c). Thus, the one peak that Craig had repeatedly pointed to and marked as a candidate for elimination has disappeared. But on the right-hand side of the plot, there not only is a new peak but the existing peaks are as high as those in the middle earlier designated for elimination. What is important to this meeting, however, is what arises in the relation between Craig and Shelby: the two lower peaks that are still present. Shelby characterizes them, therefore, as particular to the Kispiox fish. After and thereby breaking the long pause, Craig tentatively ("maybe") proposes a plot with error bars.

Trying to get a grip on variability

Replotting the data together with a statistical measure of variability The preceding move has not led to the intended reduction in variability of the data. In such a situation of uncertainty, when scientists are in the dark and without the knowledge that later comes with hindsight—that is, before scientists are familiar with the domain—any action appears as good as any other (Roth, 2004). It is only after the fact that a value “good” or “poor” can be attributed to action and its ascribed intention or, in other words, that the quality of fit can be determined between situated actions and the plans that are said to underlie them. It may not come as a surprise then that Craig proposes trying another avenue: plotting the data with error bars. The uncertainty of what such a move may bring about is expressed in the adverb “maybe,” which modalizes the verb “to try” as a possible but not necessary next move. The verb “to try” implies decision between alternatives and finding out about something that is doubtful or obscure. At first, Shelby expresses uncertainty about whether “that” is “going to work with that” and, speaking increasingly faster, states that he will give it a try.

When the scatter plot appears on the screen (Fig. 5a), Shelby notes that the two lowest data points also are those without error bars ([i] & [ii], Fig. 5a), and that they therefore represent unique points. Their salience is increased by the cursor, which he moves directly over one then the other point ([i] & [ii], Fig. 5a), moving back and forth three times. *In and as the relation* with Craig, this statement is confirmed and marked as salient, as if this fact provided a reason: “okay, so there you go.” That is, in this instance, the visibility of the two points exists in and as their relation. It is not just there, merely pointed out by Shelby. Rather, the distinction of these data points and their relation to the interpretation is a result of the relation. “They *are* just two unique points,” that can be pointed to (by means of the cursor) and that can be seen and confirmed on the part of the listener (“yea,” and “here you go”).

««««« Insert Fig. 5 about here »»»»»

A first interpretive attempt Shelby first draws a comparison to what has been: “yea, it’s quite different.” He then states an implication for the trend of the data: “you do get a general increase from zero to a hundred, very, very subtle.” As he makes the statement, one can see the cursor move from the cluster of points at the left margin of the graph to the aligned cluster of points on the right ([iii], Fig. 5a). Although the relationship is different from the one Craig has articulated before, nobody confronts the two different hypothetical deep structures in the data and the trends that they express. At this point, there are no comments to the fact that Shelby proposes a linear correlation whereas Craig has made visible a curved relationship: one of these is still visible in the line on the chalkboard, the ephemeral gesture perhaps present in perceptual memory.

A second, alternative interpretation At first, Theo asks Shelby to change the scale on the "y-axis." Together they establish the new limits on the plot. Craig affirms and names what is happening "a higher resolution there." There is a 3.90-s pause, then the new graph appears (Fig. 5b), and then another 6-second pause unfolds. Craig is the first to speak. He gets up, walks to the screen, and then moves his hand in front of it to exhibit three trends in three parts of the plot. Craig points out—not unlike in his first analysis—that there is a linear trend on the right-hand side of the plot ([i], Fig. 5b). He then makes salient a "component," which, as the shape of the hand trajectory shows, is approximately inverse parabolic ([ii], Fig. 5b). He repeats the gesture first backward and then forward again. In the 0.68-second pause that follows, his index finger first makes a slight upward movement following what might be seen as the trend of the first 11 data points ([iii], Fig. 5b), then moves back toward the point where the parabolic movement ended and produces a second movement now slightly downward ([iv], Fig. 5b) while uttering "down again."

When we compare the hand movements produced in this situation (Fig. 5b), we note the similarities and differences with earlier presentations. These differences, however, are not explicitly articulated as changes in the idea or assessment of what they are looking at. The first part of the gesture, on the right end, is similar to what Craig produced in the context of the Kispiox fish, but different from the inverse parabola that he produced even earlier. The central part does have the inverse parabolic shape. Finally, on the right end, Craig first makes a movement *upward* from right to left and then, almost in the same breath, changes the description and hand movement to one that goes *downward* from right to left. The hand gestures here are forms of epistemic movement that constitutes thinking itself rather than being a *re*presentation thereof (Radford, 2003; Roth, 2012). Again, the trends depicted in the movement are different from the “general increase,” a “very, very subtle” one that Shelby has articulated together with a nearly linear trajectory of the cursor from the left to the right. We therefore observe contradictions both within the articulations and movements that Craig produces—even within a fraction of a second—and with respect to those that Shelby previously articulated. However, these contradictions again are not made the topic of discussion. It is as if they did not exist or were salient.

Another attempt at understanding variability as it expressed itself as variance (error bars)

Shelby then evaluates Craig’s presentation saying that he finds it “not surprising” and continues to articulate a reason. He talks about the data representing not individual fish but rods that are grouped. He ends by saying that he is skeptical of those data near 0% and 100% (A_2). Craig follows the comment about being skeptical by querying the nature of the error bar—whether they represent ± 1 standard deviation. By asking the question in this way, the normative character of this value over other possible values comes to the fore. Shelby responds by saying that he had plotted using 95% confidence intervals. There is a long pause and then begins some elaboration about wanting to use 1.96 standard deviations and just as he was beginning to suggest what this would lead to “then you would get,” Theo interrupts him making salient the points “up there.” (The two numbers, *SD* = 1.96 and 95% confidence intervals are equivalent.) Shelby continues, “one [1 SD]” “would be a bit cleaner.” In this, he articulates a reference to Craig’s question about the error bars corresponding to *SD* = 1. Theo suggests that the 95% confidence intervals are a bit much, thereby adding to the preceding question about the standard deviation being 1 and the assertion that “one would be a bit cleaner.” Michael is asking about the number of fish corresponding to the right side of the plot: Is there a smaller number of fish? A reason for the query follows immediately, “because the overall variability increases there, as you go to the right?” Shelby ascertains that it makes sense to have greater variability when there is a smaller number of fish; and he adds that most of their fish have been exhibiting near 0% A_2.

A sudden end of the discussion of HBW and its treatment in the final publication

At this point, this discussion of the half-maximum bandwidth ends suddenly. There is no clear conclusion any one member offers but a clear proposal and acceptance to move to some other topic. As in many other instances during this research project, rather than making the problem the core topic of the discussion, the scientists go on to some other point. They do not wonder why something does not work. As soon as it does, they go on with their business never wondering what it was that did not allow them to make sense or do their work (Roth, 2004). There is perhaps an underlying and unstated assumption that the problems will go away once the data set is only large enough; alternatively, any remaining problems can be dealt with should they remain to the end. In the journal publication of this work, the team noted the high within-fish variability in percent A_2 and described different ways in which it controlled this variability (Temple et al., 2006). At the experimental level, only photoreceptors from the dorsal area of the eye were taken. The measurements actually taken were subject to inclusion criteria—presence of a baseline in the absorption curve, peak absorption within the commonly accepted range, minimal absorption by other materials on the slide, and a signal-to-noise ratio greater than 5 to 1. To further control variability in the determination of A_2, the team smoothed the curves using an averaging function and determined the location of the peak on the wavelength axis. Although half-maximum bandwidth was mentioned to broaden in the life history of fishes and although was shown in one of the figures, it was not described as having been used in the determination of A_2.

Discussion: graphs and graphing in discovery work

This study was designed to understand how scientists deal with variability that shows up in their graphs (e.g., in the size of the error bars they plot), and the associated uncertainties that arise for them in making sense of their data. It draws on data collected during a 5-year ethnographic effort in one advanced scientific research laboratory. The results show that although there is some understanding how variability results from the production of the data and although the members of the research team knew that there was trouble, they did not know what to make of the variability (and statistical variance) in their data and what patterns these expressed. In an attempt to make interpretation easier, they only considered part of their data. The session concerning these data ended without a conclusion and despite the inconsistency of the data with the theory, and despite very different patterns proposed not only between scientists but also within a scientist. In focusing on the local practices of scientists to bring order to their data through sequentially ordered turn-taking that made for an orderly meeting, our study is oriented very differently from those that attempt to identify mental processes that affect perception of graphs or that operate in the identification of non-obvious properties (Friel, Curcio, & Bright, 2001). It also differs from studies that "heavily rel[y] on [participants'] *explanations* of their work and thinking" (Gainsburg, 2007, p. 482) when this is done *after the fact* rather than during and as integral part of the work.

In the episode as a whole, one can see that because the team members did not have a ready answer, they did not and could not know whether some action will produce the anticipated changes in the data displayed. But any orderliness that after the fact was accounted for arose from the practical aspects of the work, which led to "transcendental orderliness" (Garfinkel, Lynch, & Livingston, 1981, p. 141) of the natural objects in the team's hands. In the preceding section, it is the "transcendental" disorderliness, the properties of *this* display that the work of the scientists comes to provide. In fact, orderliness and disorderliness reflexively make each other possible as ground against which the other emerged as figure. It has been noted that inscriptions may in fact hide certain aspects of mathematics in the workplace—it becomes a black box (Williams & Wake, 2007). Work has to be done to unpack that which is hidden and invisible—such as the number of data underlying each point plotted invisible to Craig but pointed to by Shelby. In the present instance, the particular representational form Shelby had chosen exhibited and hid variability. It exhibited variability that the team wanted and needed to get under control or at least understand; it was variability that muddied the anticipated or real trend in the phenomenon. But it also hid the variability of each data point, itself the mean of many measurements, which total over 3,000.

In the end, the team members present abandoned what they themselves had identified as problematic. It may be just a further example of abandonment that is typical of everyday problem solving, where it constitutes one possible way of going about a problem as coming up with some definite answer (Lave, 1988). It stands, of course, in striking contrast to school situations, where students tend to be required not only to have but one single answer in contrast to multiple answers, but where they are not generally allowed to abandon what the teacher has designed as the problem at hand (Roth & McGinn, 1997). In the present situation, the lead scientist works out some "preconceived notions" as to the relationship between two salient variables, %A_2 and half-maximum bandwidth of the absorption spectrum. This relationship is important, because the scientists were hoping to use it for the determination of the relative amounts of rhodopsin (A_1) and porphyropsin (A_2) in the pigment of the fish eyes, which are the relevant variables for determining—so the Nobel prize winning canon at the time—the readiness for the saltwater journey. According to Theo, he has not yet "checked what the bandwidth gives [them]." This is what they pursue in the episode analyzed here.

The present analysis exhibits the work that the scientists produce for finding "what the bandwidth gives them." Craig makes an argument for eliminating some of the data points, but Shelby resists, for, as he argues, there are about 3,000 data points. This means that there will be some large variations, on the one hand, and relative stability of the data, on the other hand. Theo does provide an explanation of why some individual HBW measurements might be much smaller than they ought to be: His algorithm cuts off one part of HBW when the "ringing" that all of them are familiar with crosses the HBW line before the best-fit curve does. Craig provides a first reading of the trend, which, in his gestural movement, takes on a curvilinear shape that matches his "preconceived notion." The scientists here attempt to control the variability by looking at a subset of the data. But the variability that is the salient aspect in their work remains when they look at the data from one of the seven sources. The next move is to establish a better understanding of the variability within each data point by "trying to plot with error bars." Here it turns out that two of the “spikes” do not have error bars and, according to the team, are “unique points.” Once they have changed the scale of the plot, we observe the scientists enact two readings of the data.

Shelby, by means of the cursor, traced out a more-or-less linear relation, “a general increase from zero to a hundred” that is “very, very subtle.” In this way, the trend was precisely of the kind that a dominant researcher in the field (Hárosi) and others previously published. It had, as predicted, a lower bandwidth for rhodopsin (A_1) on the left end of the graph and a higher bandwidth of porphyropsin (A_2) on the right end of the graph (in sharp contrast to the “preconceived notion” that Craig had articulated earlier). This is also the way one would expect it to occur if one knew that the HBW values plotted here are not measured in nanometers, that is, on the same scale that the absorption spectra are plotted. Rather, HBW, as plotted here according to the values on the abscissa, would vary linearly with $\%A_2$ if the widths of the spectra, measured in nanometers, vary as a negative square function. This is not salient in the situation, however, and therefore does not play a role in its internal dynamic of the data analysis meeting.

Craig also provides a reading; in fact, he provides two different readings of the data. There is a linear increase on the right-hand side of the plot similar to what Shelby has shown. In the center, however, Craig has a parabolic relationship, as exhibited in his “preconceived notion” and as gestured in an earlier plot. Finally, Craig first gestures a linear relation with a negative slope at the left end of the plot; he then changes to a linear relation with a positive slope. In the course of the data analysis, there are different trends proposed. There are also trends that have been published in the research for nearly two decades. However, in this meeting, the question is never posed or answered why the measured curve does not match the anticipated curve, as articulated at the beginning. This even though this very issue is a stated goal for the laboratory. There is also no discussion about the different proposals concerning the trends.

The graphs not only are the topic of talk but also ground for articulating relations that would be much more difficult to present in verbal form. Verbal indices (“here,” “there,” “this”) are used to “tie” the indexical, indexical/iconic, and iconic (movement) expressions to the graph (e.g., Ochs, Gonzales, & Jacoby, 1996). In the deictic gestures, the graph as a whole or individual points are the topic, whereas the graph is the ground against the *dynamic* hand/arm movement that expresses something not itself in the display. It is layered over and above the display but represents something "underneath" or underlying the data, the deep structure. The erratic data points are but concrete expressions of extraneous variability of some true signal that is made visible in and by Craig's hand movements. Gestures play an integral role in the search for trends (Roth, 2012), which are found because gestures are generative (de Freitas & Sinclair, 2012), semiotic means of objectification (Radford, 2003). These are a way of orienting visual perception through the maze of the data that Craig did "not know what [was] going on here." Of course, the gesture itself was motivated by the perceptual gestalt that the data themselves offered; the gesture, as words in other situations, is but a means that "lets something be seen, namely what is talked about, and *for* the speaker (the medium) or for the interlocutors" (Heidegger, 1977, p. 32). The trends existed in their relation with each other, where they take on the *public* visibility of their characteristics; and the trends revealed themselves such that others could perceive them too. Even if there are multiple accounts of possible trends, these are in any case articulated and therefore rationally accountable trends. When Craig moved his hand/index finger across the display, he actually directed and taught others and himself how to look and make visible a possible trend. Similarly, in moving the cursor across the data display, Shelby exhibited in his movement *how* others in the meeting had to look so that they "[got] a general increase from zero to a hundred, very, very subtle." The movement of the cursor exhibited the material that justified the description of the general but subtle increase from the very left to the very right of the display. The movements constituted the natural accountability that grounded the observation of the trend.

Some preliminary thoughts on graphing and uncertainty in schools

Studies of how scientists deal with variability, while telling us a lot about knowing and learning applied mathematics, should not be taken as indicative of strategies for school mathematics curricula without careful prior discussion and consideration. The following are but some preliminary thoughts about the relevance of this study of graphs and graphing in the discovery sciences to mathematics education.

In this study, research team members apparently talked about and wrestled with variability, a topic of interest to mathematics and statistics educators (e.g., Cobb & Tzou, 2009; English, 2012; Pfannkuch, 2005). Random variations that obscure true relationships underlying data are potential trouble spots in university students' modeling activities (Doerr, 2000; Falk & Konold, 1994). This is so because random errors may exhibit a linear relationship when the true relationship is exponential. This has to be particularly the case when the size of the error that comes with each data point is unknown but subsequently understood to be systematic. Attempting to understand the source of variance or limiting the statistical variance by controlling the nature of (variability in) the data, therefore, may be an important step to understanding the true relations. Understanding and getting a grip on variability may be even more important than finding patterns. There are suggestions that students of science may learn more and better understand a phenomenon if they understand when, where, and how it can get lost (Garfinkel, 2002; Roth, 2013). It may be more important for students to discuss sources of variability, changes in the phenomenon when different types of systematic sources of statistical variance modifies the signal, and the role of stochastic error (e.g., Cobb & Tzou, 2009).

Some scholars have observed that we "can expect *learners* to generate multiple interpretations of inscriptions, use multiple meanings of words, and focus on different aspects of inscriptions" (Moschkovich, 2008, p. 578, emphasis added). As the present investigation shows, this state is not only characteristic of students—who are in the process of learning something they do not know and therefore cannot evaluate—but also of highly experienced and successful research scientists confronted with some data first time through when they are in the process of establishing possible deep structures of the data. The present analysis shows that there may be alternative and contradictory interpretations articulated in the same meeting *without* there being a contrasting evaluation that might lead to further learning. Yet research has shown that learning arises precisely from the opposition and confrontation of different propositions or while reasoning about alternative hypotheses and, in the process, build consensus and sets of shared rules about data interpretations and variation (Cobb & Tzou, 2009). In school mathematics, this might be the place where teachers—in the process of observing and overhearing students—decide to enter their discussion and assist students in making the differences explicit and in making the resulting contradictions the topic of discussion. It is precisely in the ensuing dialogical relations that learning results from inquiry. Do we have to conclude that scientists, too, may need such support? Or do we have to conclude that students do not need such support because scientists themselves do not use opposition? Concerning the discovery sciences, at least, the very fact that scientists work at the cutting edge of their field and the fact that these meetings focus on topics in real time that prevents a reflective approach may mitigate such attempts.

In this study, the variability in the data arose from the scientific phenomenon itself and from the ways in which measurements were gathered, inclusion criteria were defined and applied, and ways in which graphs were transformed to produce the A_2 values that were presented plotted against time of year. This may be used as part of an argument for interdisciplinary approaches to science, technology, engineering and mathematics education (e.g., Artigue, 2012) because it provides students with opportunities to understand variability and its effect on comparisons even at the university level (e.g., Tra & Evans, 2010). As an above-mentioned study had shown, eighth-grade students familiar with the analysis of data that they collected themselves, and therefore familiar with variability and how to control it, outperformed university graduates on data interpretation tasks (Roth et al., 1998). On the other hand, students in mathematics classes presented with data sets were asking questions extensively to understand the provenance of the data and what might possibly have affected them (Cobb & Tzou, 2009).

There are many presuppositions about what students of science *ought* to do when analyzing data. For example, one study provides a list of questions students are to draw on in their argumentation about graphs (Chin & Osborne, 2010). The questions that the authors provide are categorized according to some of the basic science process skills—for example, observing, comparing, analyzing, predicting—and include: "'Is there anything unusual or unexpected?' 'How are A and B similar?' 'How are A and B different?' 'What is the significance of these similarities and differences?' . . . 'Is there anything I am puzzled about?' . . . 'Which is the better graph? Why?'" (p. 242). As the analysis of the episode shows, the research scientists did not ask such questions, at least not in my presence and in the sessions we recorded. They were able to live with apparent contradictions in the descriptions various members provided for the possible trend in the data and with the apparent contradiction between the theory or a priori conceptions and the actual data. The kinds of processes educators often postulate—including disagreement, challenge, cognitive conflict, and puzzlement—are not (as) explicit in the data analysis sessions that we report on here. But this does not mean that these processes *ought not* be fostered in schools, where they might function as a means to develop competencies that are currently not generally observed in most working place contexts, including scientific discovery work.

Acknowledgments

Several grants from the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada supported this study. All opinions are those of the authors.

References

Artigue, M. (2012). Reflections around interdisciplinary issues in mathematics education. In W. Blum, R. B. Ferri, & K. Maaß (Eds.), *Mathematikunterricht im Kontext von Realität, Kultur, und Lehrerprofessionalität: Festschrift für Gabriele Kaiser* [Mathematics courses in the context of reality, culture, and teacher professionality: Festschrift for Gabriele Kaiser] (pp. 24–33). Wiesbaden: Vieweg+Teubner Verlag | Springer.

Bakker, A., & Hoffmann, M. H. G. (2005). Diagrammatic reasoning as the basis for developing concepts: A semiotic analysis of students' learning about statistical distribution. *Educational Studies in Mathematics, 60,* 333–358.

Châtelet, G. (1993). *Les enjeux du mobile: Mathématique, physique, philosophie* [The stakes of movement: Mathematics, physics, philosophy]. Paris: Éditions du Seuil.

Chin, C., & Osborne, J. (2010). Supporting argumentation through students' questions: Case studies in science classrooms. *Journal of the Learning Sciences, 19*, 230–284.

Cobb, P., & Tzou, C. (2009). Supporting students' learning about data creation. In W.-M. Roth (Ed.), *Mathematical representation at the interface of body and culture* (pp. 135–171). Charlotte, NC: Information Age Publishing.

Confrey, J. (1995). Splitting, covariation, and their role in the development of exponential functions. *Journal for Research in Mathematics Education, 26,* 66–86.

de Freitas, E., & Sinclair, N. (2012). Diagram, gesture, agency: Theorizing embodiment in the mathematics classroom. *Educational Studies in Mathematics*, *80*, 133–152.

Doerr, H. M. (2000). How can I find a pattern in this random data? The convergence of multiplicative and probabilistic reasoning. *Journal of Mathematical Behavior, 18*, 431–454.

English, L. (2012). Data modelling with first-grade students. *Educational Studies in Mathematics, 81,* 15–30.

Falk, R., & Konold, C. (1994). Random means hard to digest. *For the Learning of Mathematics, 16,* 2–12.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*, 124–158.

Gainsburg, J. (2007). The mathematical disposition of structural engineers. *Journal for Research in Mathematics Education, 38*, 477–506.

Garfield, J. (2011). Statistical literacy, reasoning, and thinking. In M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 1439–1442). Berlin: Springer.

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, *75*, 372–396.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting* *research and teaching practice*. Berlin: Springer.

Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory tertiary level statistics course. *ZDM Mathematics Education, 44,* 883–898.

Garfinkel, H. (2002). *Ethnomethodology's program: Working out Durkheim's aphorism*. Lanham, NY: Rowman & Littlefield.

Garfinkel, H., Lynch, M., & Livingston, E. (1981). The work of a discovering science construed with materials from the optically discovered pulsar. *Philosophy of the Social Sciences*, *11*, 131–158.

Garfinkel, H., & Wieder, D. L. (1992). Two incommensurable, asymmetrically alternate technologies of social analysis. In G. Watson & R. M. Seiler (Eds.), *Text in context: Contributions to ethnomethodology* (pp. 175–206). Newbury Park, CA: Sage.

Geertz, C. (1973). *The interpretation of cultures: Selected essays.* New York, NY: Basic Books.

Gerovsky, S. (2010). Mathematical learning and gesture: Character viewpoint and observer viewpoint in students' gestured graphs of functions. *Gesture*, *10,* 321–343.

Hall, R., & Stevens, R. (1995). Making space: A comparison of mathematical work in school and professional design practices. In S. L. Star (Ed.), *The cultures of computing* (pp. 118–145). London: Basil Blackwell.

Hardy, N. (2009). Students' perceptions of institutional practices: The case of limits of functions in college level calculus courses. *Educational Studies in Mathematics, 72,* 341–358.

Heidegger, M. (1977). *Sein und Zeit* [Being and time]. Tübingen: Max Niemeyer.

Hoffmann, M. H. G., & Roth, W.-M. (2005). What you should know to survive in knowledge societies. On a semiotic understanding of "knowledge." *Semiotica, 156*, 101–138.

Kaput, J. J. (1988, November). *Truth and meaning in representation situations: Comments on the Greeno contribution*. Paper presented at the annual meeting of the North American Chapter of the International Group for Psychology of Mathematics Education, DeKalb, IL.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33,* 259–289.

Koschmann, T., & Zemel, A. (2009). Optical pulsars and black arrows: Discoveries as occasioned productions. *Journal of the Learning Sciences, 18*, 200–246.

Latour, B. (1993). *La clef de Berlin et d'autres leçons d'un amateur de sciences* [The key to Berlin and other lessons of a science lover]. Paris: Éditions de la Découverte.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge: Cambridge University Press.

McDermott, R. P., Gospodinoff, K., & Aron, J. (1978). Criteria for an ethnographically adequate description of concerted activities and their contexts. *Semiotica, 24*, 245–275.

Moschkovich, J. (2008). "I went by twos, he went by one": Multiple interpretations of inscriptions as resources for mathematical discussions. *Journal of the Learning Sciences, 17*, 551–587.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Noss, R., Bakker, A., Hoyles, C., & Kent, P. (2007). Situating graphs as workplace knowledge. *Educational Studies in Mathematics, 65*, 367–384.

Ochs, E., Gonzales, P., & Jacoby, S. (1996). "When I come down I'm in the domain state": Grammar and graphic representation in the interpretive activity of physicists. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 328–369). Cambridge: Cambridge University Press.

Pfannkuch, M. (2005). Thinking tools and variation. *Statistics Education Research Journal, 14*(2), 5–22.

Radford, L. (2003). Gestures, speech, and the sprouting of signs: A semiotic-cultural approach to students' types of generalizations. *Mathematical Thinking and Learning, 5,* 37–70.

Roth, W.-M. (1996). Where is the context in contextual word problems?: Mathematical practices and products in Grade 8 students' answers to story problems. *Cognition and Instruction, 14*, 487–527.

Roth, W.-M. (2003). *Toward an anthropology of graphing: Activity-theoretic and semiotic* *perspectives*. Dordrecht: Kluwer Academic Publishers.

Roth, W.-M. (2004). “Tappen Im Dunkeln”. Der Umgang mit Unsicherheiten und Unwägbarkeiten während des Forschungsprozesses [“Groping in the dark”: Dealing with uncertainties and vagaries during the research process]. *Zeitschrift für Qualitative Bildungs-, Beratungs-, und Sozialforschung, 5*, 155–178.

Roth, W.-M. (2012). Tracking the origins of signs in mathematical activity: A material phenomenological approach. In M. Bockarova, M. Danesi, & R. Núñez (Eds.), *Cognitive science and interdisciplinary approaches to mathematical cognition* (pp. 182–215). Munich: LINCOM EUROPA.

Roth, W.-M. (2013). *What more? in/for science education: An ethnomethodological perspective*. Rotterdam, The Netherlands: Sense Publishers.

Roth, W.-M., Bowen, G. M., & McGinn, M. K. (1999). Differences in graph-related practices between high school biology textbooks and scientific ecology journals. *Journal of Research in Science Teaching, 36*, 977–1019.

Roth, W.-M., & McGinn, M. K. (1997). Toward a new perspective on problem solving. *Canadian Journal of Education, 22*, 18–32.

Roth, W.-M., & McGinn, M. K. (1998). Inscriptions: a social practice approach to “representations.” *Review of Educational Research, 68*, 35–59.

Roth, W.-M., McGinn, M. K., & Bowen, G. M. (1998). How prepared are preservice teachers to teach scientific inquiry? Levels of performance in scientific representation practices. *Journal of Science Teacher Education, 9*, 25–48.

Roth, W.-M., Pozzer-Ardenghi, L., & Han, J. (2005). *Critical graphicacy: Understanding visual representation practices in school science*. Dordrecht, The Netherlands: Springer-Kluwer.

Scheaffer, R. L. (2011). Statistics education. In M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 1482–1484). Berlin: Springer.

Temple, S., Plate, E. M., Ramsden, S., Haimberger, T. J., Roth, W.-M., and Hawryshyn, C. W. (2006). Seasonal cycle in vitamin A_1/A_2-based visual pigment composition during the life history of coho salmon (*Oncorhynchus kisutch*). *Journal of Comparative Physiology A* 192, 301–313.

Tra, Y. V., & Evans, I. M. (2010). Enhancing interdisciplinary mathematics and biology education: A microarray data analysis course bridging these disciplines. *CBE Life Sciences Education, 9*, 217–226.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical inquiry. *International Statistical Review, 67.* 223–265.

Williams, J., & Wake, G. (2007). Black boxes in workplace mathematics. *Educational Studies in Mathematics, 64*, 317–343.

Williams, J. S., Wake, G. D., & Boreham, N. C. (2001). School or college mathematics and workplace practice: An activity theory perspective. *Research in Mathematics Education, 3,* 69–83.

Woolgar, S. (1990). Time and documents in researcher interaction: Some ways of making out what is happening in experimental science. In M. Lynch & S. Woolgar (Eds.), *Representation in scientific practice* (pp. 123–152). Cambridge, MA: MIT Press.

Captions

Fig. 1. a. The A_2 values from 3 sites are spread across the entire spectrum of A_2 (0–100%) rather than clustering more sharply around some mean value. b. A raw absorption spectrum, as it comes from the instrumentation, is fitted with a 7th-order polynomial derived from the literature. The wavelength where the polynomial has its maximum is called “lambda-max” (λ_{max}) and the width of the polynomial at half its height with respect to “baseline” is referred to as “half-maximum bandwidth (HBW).”

Fig. 2. a. Craig draws a graph representing the expected relationship between half-maximum bandwidth (HBW) and amount of porphyropsin (%A_2). b. Shelby plots actual HBW against %A_2.

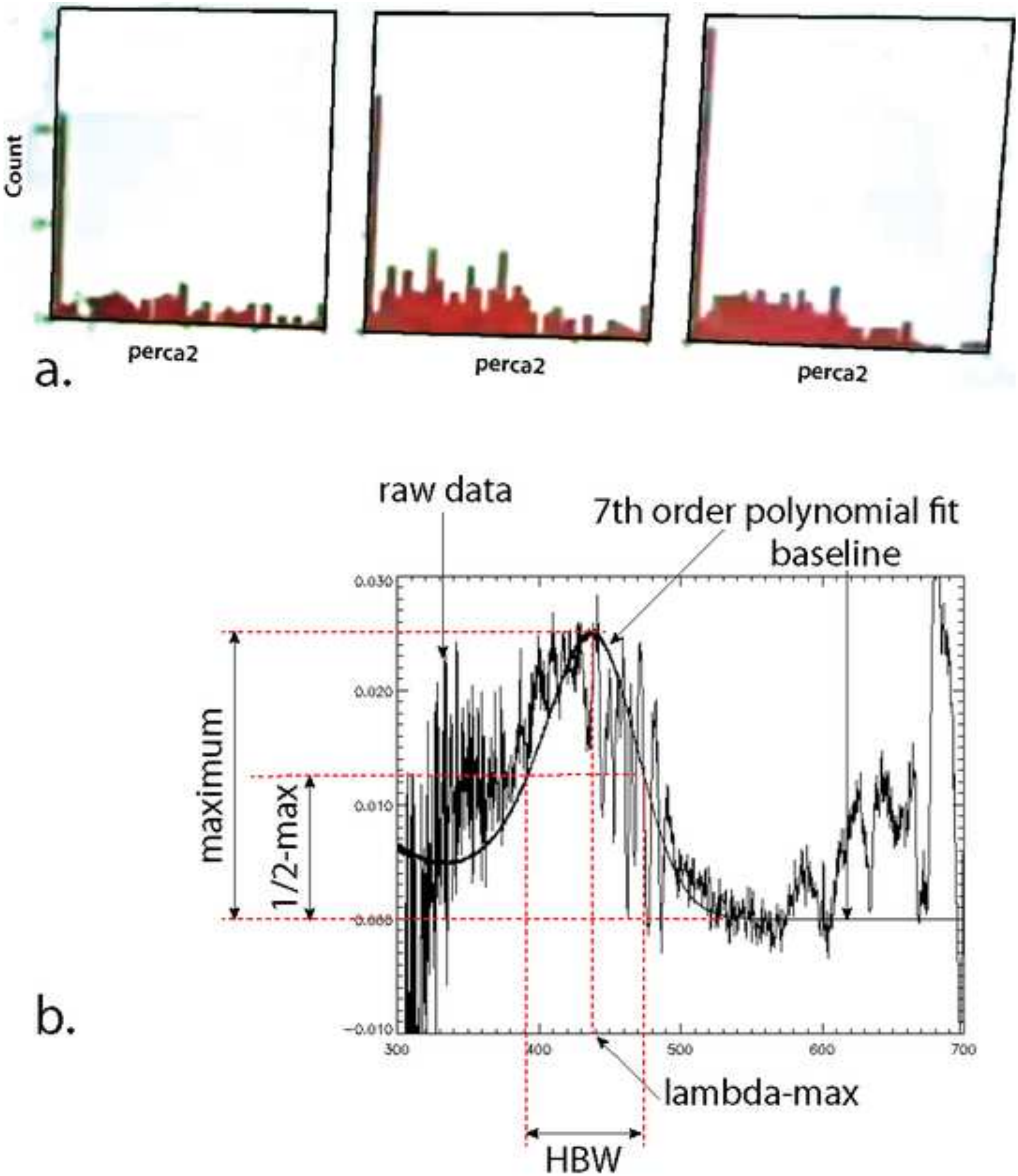
Fig. 3. a. Theo shows how lambda-max and HBW are extracted. b. He identifies the “ringing” as

a problem in the raw data, which introduces variation in the determination of HBW.

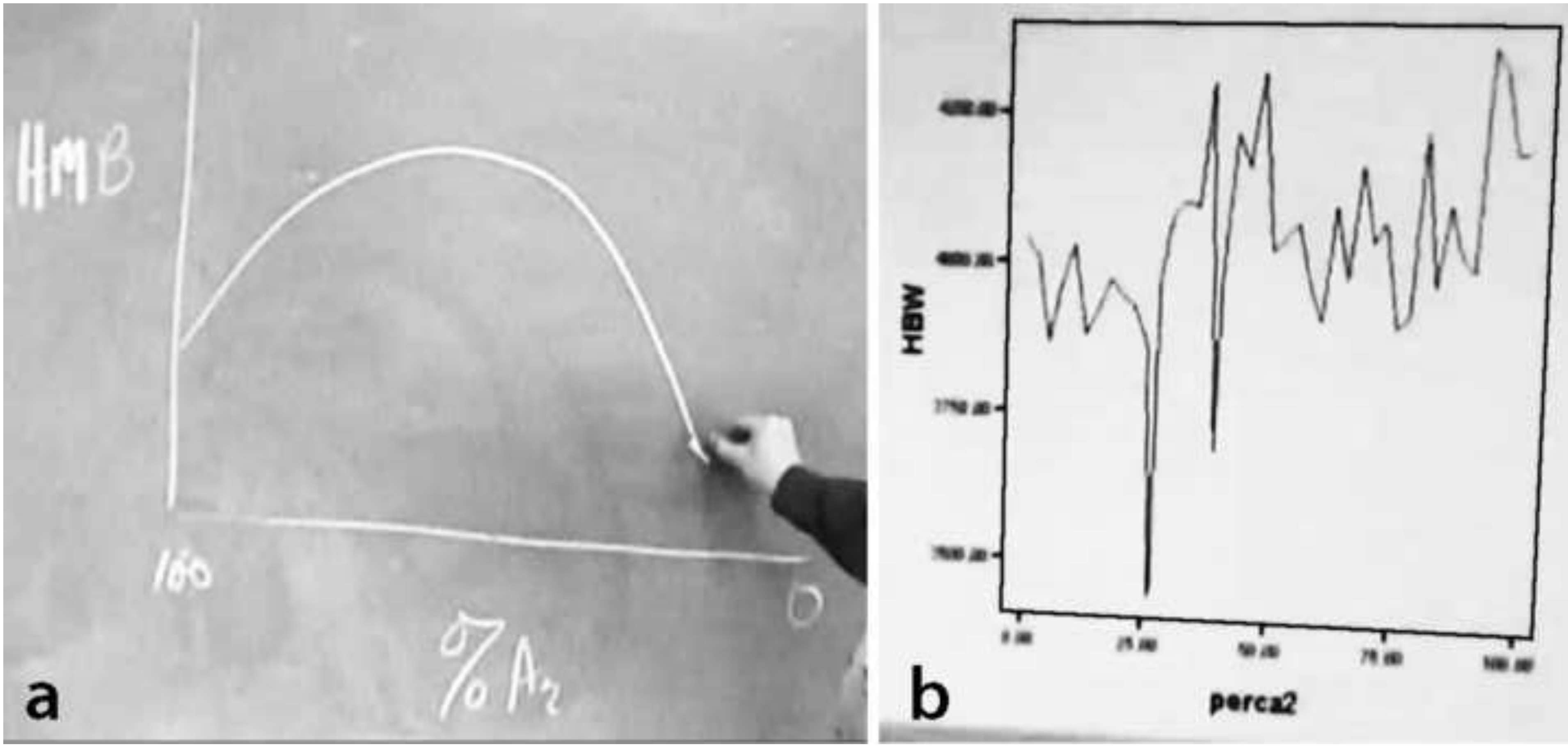
Fig. 4. a. Craig identifies 4 points as outliers ([ii]–[v] and gestures ([i]) where a regression should be conducted. b. Craig gestures a hypothesized best fit. c. The suggestion to plot only a subset of the data does not lead to a reduction in the variation.

Fig. 5. a. After plotting the error bars for each data point, Shelby points to two critical points ([i], [ii]) and then gestures a best fit ([iii]). b. After Shelby rescaled the plot, Craig uses gestures to show what the trends in the data are.

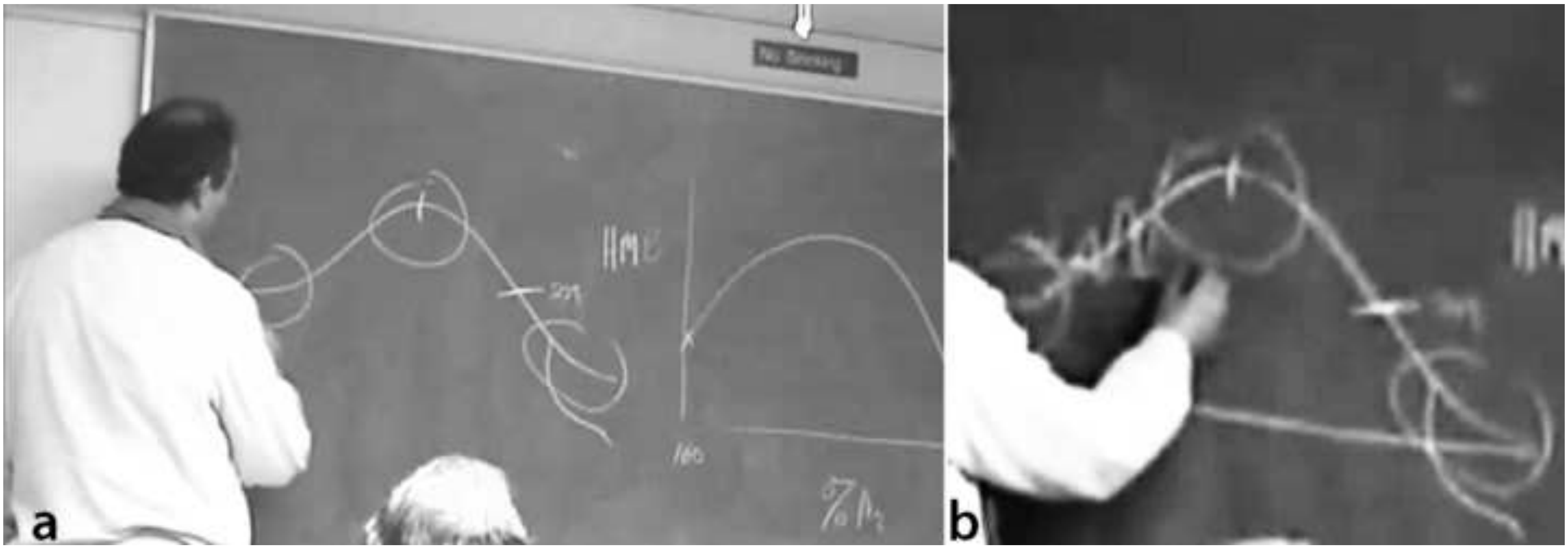
colour figure



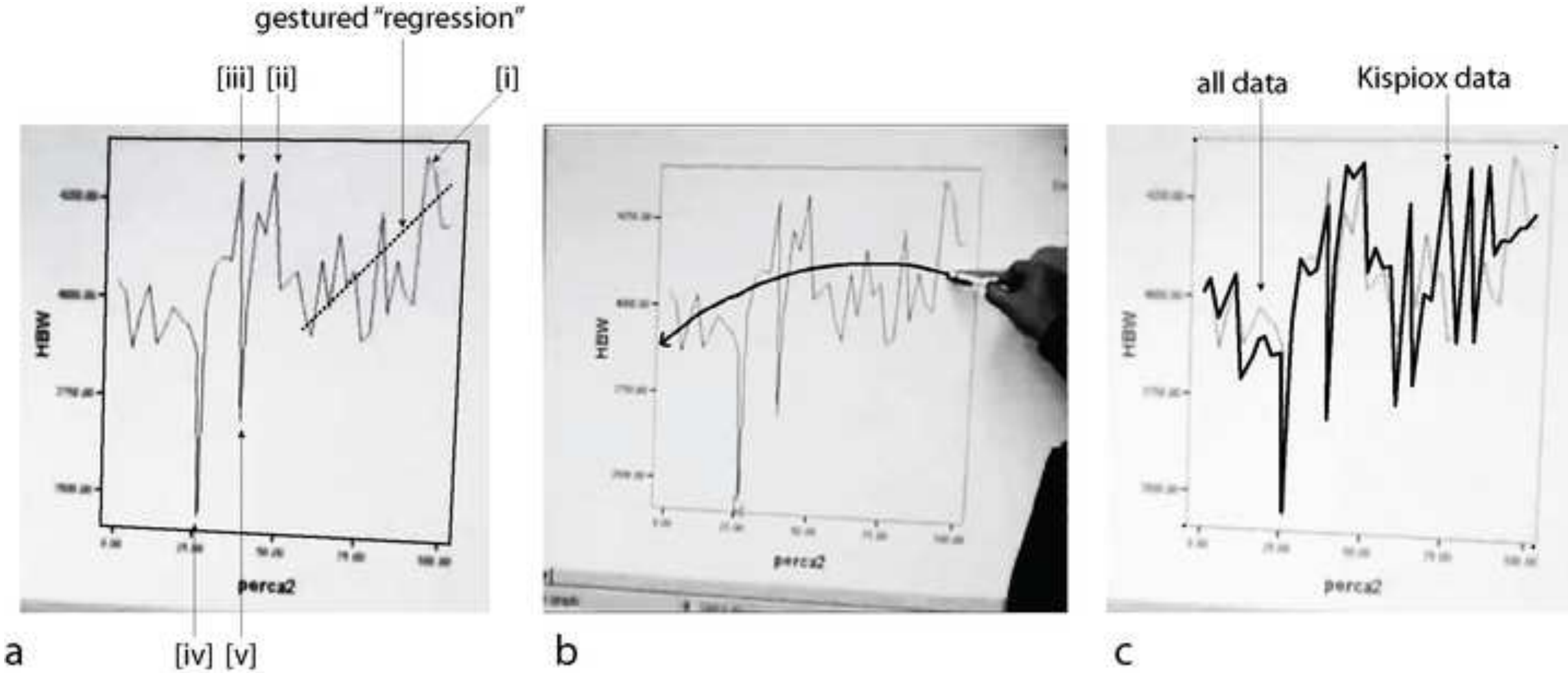
colour figure



colour figure



colour figure



colour figure