FROM OBSERVING AND EVALUATING VARIATION TO MEASURING AND COMPARING VARIATION

Erna Lampen
School of Education, University of the Witwatersrand, Johannesburg, South Africa
ernalampen@gmail.com

Teachers are tasked with engaging students' everyday reasoning in order to develop statistical reasoning, but there is little guidance for statistics teachers to understand relevant patterns in everyday reasoning about data contexts. In this paper I report a discursive analysis of a discussion of the data context of prices of used cars by a group of teachers in an introductory statistics course. Everyday reasoning and informal statistical reasoning are proposed as points on a continuum along the development of statistical reasoning, and defined in terms of discursive patterns. I describe patterns in everyday discourse about data contexts and illustrate how such patterns influence reasoning about context in a classroom discussion. Analysis of the participants' word use indicates that everyday terms like "value" and "price" should not be treated as synonyms in discussions of data contexts, even if they refer to the same numerical data.

BACKGROUND

Wild and Pfannkuch (1999) emphasise the driving role of context knowledge at the inception of a question or idea that can be investigated statistically. They say "The earliest stages are driven almost entirely by context knowledge. Statistical knowledge contributes more as the thinking crystallises" (Wild & Pfannkuch, 1999, p. 228). This natural progress in statistical investigation implies that reasoning about the context itself with the purpose of statistical treatment is a learning task. The pedagogical question is: How can teachers know when statistical thinking crystallises from context knowledge in discussions of the data context? Inappropriate idiosyncratic reasoning about data-contexts is based on one's own experience (Jones, Langrall, Mooney, & Thornton, 2004), lacking alienation and abstraction. In particular, in such reasoning measured values are seen as characteristics of the cases and not interpreted against an abstract "background of possible measurement values" of an alienated variable (Bakker, 2004, p. 101). The result of the lack of abstraction and alienation is that students tend to compare data sets by contextually important values, such as a desired maximum or minimum value, rather than by the use of statistical measures of centre and spread. This statistically problematic discourse has roots in everyday reasoning patterns, especially at the start of the statistical investigation cycle. In particular teachers need to be aware of words and their uses that conflate measurable properties with subjective characteristics. Hence statistics teachers must be cognisant of pervasive everyday reasoning patterns that give rise to idiosyncratic statistical reasoning.

PATTERNS OF EVERYDAY REASONING

Reasoning about statistical problems in everyday situations (outside the classroom) is influenced by the perspective people take on the context and their reasoning goals (Gigerenzer, 2002; Nisbett, 1993). Two patterns are relevant for this paper.

Immersion in the Context

According to Zwaan's (2004) *Immersed Experiencer Framework*, language is embodied and text re-produces experiential memories together with their referents as mental representations. Although these mental representations are diffuse and weakly organised at first, they yield more and different information than that presented in the text. Zwaan and Kaschak (2009, p. 368) argue that comprehension of language is dependent on this ability to enact contexts through imagination. Such imagination of action in a context provides a personal situation model without which we cannot answer *why*-questions or distinguish relations like *because* and *therefore* (Kintsch, 1994). However, statistics education research shows that idiosyncratic reasoners answer *why*-questions directly from immersed experience, rather than from (statistically) reasoned experience. For statistical reasoning to emerge, one has to consciously forego the contextual judgements of a personal situation model, but consider the relevant and appropriate experiential or perceptual

information to construct reference classes and units for measurement, and relate variables. Hence, statistical reasoning requires that the Immersed Experiencer gains distance from the context and answer *why-* questions based on objective description and data-informed evaluation.

Subjective-objective Perspectives on the Data-context

In their experiments Nisbett et al. (1993) framed problems as objective or subjective in their study of informal inductive reasoning. Objective problems were designed to suggest distance between the participant and the context, and could be answered by imagining and comparing observable properties such as abilities of sports people, achievements of others, and physical illness in general. Subjective problems dealt with personal preferences among objects, assessment of leadership potential, and judgements about moral behaviour. Not surprisingly they found that objective problems were more likely to evoke informal statistical reasoning (57% of participants provided informal statistical responses to objective problems against 26% to subjective problems), since the sample spaces and units of measurements are relatively clear, and chance variation is relatively transparent. The surprising fact is that almost half of their participants did *not* show informal statistical reasoning for objective problems. Instead these participants transformed the objective questions into subjective questions as they immersed themselves as actors rather than observers in the imagined context. As actors they evaluated the context in a desire to make "best" decisions in which comparison to aggregates had no relevance.

DISCOURSE AS A THEORETICAL FRAMEWORK

My research focuses on classroom discussions rather than experimental situations. Classroom discussions are not as carefully controlled as experiments since there are many more possible discursive moves among a diverse group of students and a teacher. I adopt a discursive orientation which holds that discourses are patterned, that learning can be observed as shifting discourses, and that everyday discourse patterns are inevitable when novices are learning to reason statistically (Sfard, 2008). Informed by experimental results as described above I distinguish between everyday reasoning and informal statistical reasoning as follows: Everyday reasoning is aimed at personal decision making and action, and is observed as *evaluation discourse*. Informal statistical reasoning is distinguished by *exploration discourse* aimed at informal comparison within or between suitable contextual aggregates (reference classes) and hence basing informal inferences on a large number of observations. Since formal discourses evolve from their informal root discourses (Sfard, 2008), these discourses can theoretically be posed as two points on a continuum between everyday reasoning and formal statistical reasoning.

I report on word uses that enable and constrain discursive shifts from everyday reasoning to toward informal statistical reasoning in a discussion of prices of used cars by a group of high school teachers in an introductory statistics course. The results provide information about patterns in so-called idiosyncratic responses at the start of the cycle of statistical investigation. Knowledge to identify patterns in novice statistical reasoning will inform a teacher's planned instructional discourse.

THE RESEARCH QUESTIONS

What discursive patterns emerge in a classroom discussion of a data context at the start of an investigative cycle? How do the discursive patterns constrain or enable statistical reasoning?

RESEARCH CONTEXT

The episode I report on was part of my PhD research, a case study of a group of twelve high school teachers who attended a semester course in introductory statistics for teachers. Throughout the course the students and I engaged in rich discussions about ill-structured data contexts as a start to cycles of statistical investigation. The particular discussion (during one three hour session in the third week of the course) was about prices of used cars in order to statistically define *a reasonable price.* The students had a data table of 85 cases and various variables as shown in Table 1. In addition they had a depreciation table of official "book values" of cars, and excerpts from a blog about buying and selling used cars. The classroom discussions were video-recorded and transcribed. The language of learning and teaching is English.

Table 1: Excerpt from a data table of used cars.

www.cars4sale.co.za 15/7/2008					
Model	Year	Km	City	Colour	Price (R)
Toyota RunX 140i RS	2007	3000	Pretoria	GOLD	129950
Toyota RunX 180i RX	2007	5000	Cape Town	SILVER BLUE	165000

I analysed the discussions as discourses, looking for discursive patterns and shifts from everyday reasoning to informal statistical reasoning, as defined above.

RESULTS

Evaluation discourse is indicative of immersion in the experience of the context and practical reasoning aimed at decision making. Comparisons are made to personal factors rather than a reference class of other prices. In this discourse cars have "value" and "worth" based on their "condition" and "age". Condition and age can be directly observed rather than measured, and value and worth is determined by prudent judgement. *Exploration discourse* is proposed to be informal discourse about abstracted properties of cars, such as "price", "kilometre reading", and "age" seemingly "as if they occur without human participation" (Sfard, 2008, p. 295). I will illustrate discursive positions within evaluation discourse (or everyday reasoning) by excerpts from the discussion and show how the lack of clarification of the relationship between value, worth and price constrained informal statistical reasoning about the context.

Example 1: Worth and Value in Extreme Evaluation Discourse

29	SM	It is not worth to buy a car.
30	KH	Mmm [tentative].
31	GK	What do you mean it's not worth to buy a car? You mean it's not worth to buy a new car or an old car?
32	SM	It does not matter, if it is a new car or an old car. As long as it is a car.
33	GK	Mm?
34	SM	Ja.
35	GK	The reason it is (inaudible) the petrol price?
36	SM	The interest rate, service…cars are very expensive, OK? And then, when you buy it, it depreciates.

In Example 1 the worth of a car is tightly related to the perceived goal of SM's investigation, namely buying a car for himself. Price is not a variable – cars are only expensive, and depreciation is an undesirable property of a car. The narrative in Example 1 can be described as a value-for-me narrative and belongs to evaluation discourse.

Example 2: Appropriate Price in Evaluation Discourse

81	SDS	OK, if I was trying to buy a car, and I was getting this data, I would look at specific criteria. Like if I was looking for a specific colour, I would choose all the ones with the colour I want and look at what is an appropriate price. Not too cheap so that there is probably something wrong with it, also not something too expensive, because then I can't afford it. So I'm looking for, I wanna say an appropriate price for each criteria – colour, or model, or area…

In this example SDS changes the task from finding a reasonable price to finding an appropriate price. SDS's narrative suggests everyday practical reasoning aimed at decision making,

but her narrative contains "seeds" of comparison within an aggregate of other cars-with-prices. She refers to a data set with multiple variables issued as course material; she considers an "appropriate price" among a range of prices in relation to specific "criteria" (sources of variation) and suggests that the appropriate price is not extreme. This suggests she is apparently imagining an average price, but average in terms of value-for-her rather than a more objective value-of-cars. A closer look at her word use reveals jarring comparison scales: SDS uses the term "appropriate price" in relation to cheap and expensive. But she does not judge cheap and expensive relative to the same measure. "Too cheap" indicates "something wrong with the car" and "too expensive" indicates "…I can't afford it". "Too cheap" is referenced to a judgement of the condition and value of a car, and "too expensive" is related to a judgement of its cost to her as an individual. Hence the "price" which should be "not extreme" or "in the middle" cannot be abstracted to "average" or "mean" on an objective scale of measurement.

Example 3: Value, Worth and Price in Evaluation Discourse

      The third example indicates a shift from personal decision making to discourse about properties of cars. The discussion is about the relationship between the variables "mileage" and "price". In this example the variable "price" is still problematic. The students are confused about the relationship between price, value and worth.

63	Lecturer	Are you saying, if you buy a car in 2005, and you buy it in December 2005…
64	KH, GK	Will we sell it at the same price?
66	GK	Maybe, can we say are the cars worth the same price? Are the values of the cars the same?
68	GG	But, if a person bought it in January, and didn't do much, say, ja, there is no mileage on the car, then it is probably better to take the one from January.
69	Lecturer	Why?
70	GG	Well, if…the one in January…doesn't do much driving…and the one in December has decided to do a road trip for example, so it's mileage, ja…and I suppose it's also the condition of the car as well.

      In Example 3 worth, value and price are expressed in relation to each other and not as different turns of phrase of the concept price. The narrative is about value-of-cars and the implication is that the amount of rand that is the price of the car cannot be taken as measurement of the worth and condition of a hypothetical used car. Price is therefore not yet abstracted as a statistical variable. That the narrative belongs to evaluation discourse is evident from how the students frame themselves as "selling" cars and making prudent decisions like "better to take the one from January."

      Just how strong the hold of contextual complexity and observed variation is on the discourse in the class is illustrated in Example 4. The discussion in Example 4 follows on from Example 3, which was about the influence of the condition of used cars on prices.

Example 4: Evaluation Discourse Prevents Comparison to a Standard

70	GG	Well, if…the one in January…doesn't do much driving…and the one in December has decided to do a road trip for example, so it's mileage, ja…and I suppose it's also the condition of the car as well.
71	KH	They [used car dealers] don't look at that. They take a look at the car, take out the [official depreciation] table and say this is the trade value of the car.
72	RK	What if the cars are used on different terrains, you know like…the person uses the car on a good surface, and then a person uses the car for a short time…driving you know, not on the road…
73	RK	I am saying if they are using mileage to to to assess the value of…then they might be misleading this customer.

KH (Turn 71) tries to shift the discussion to the use of the industry standard, the depreciation table, in order to compare prices of used cars. But the official depreciation table which is compiled from mean trade prices is discounted as a tool to standardise used car prices. Within the evaluation discourse that surrounds KH's reference, the statistical tool is discarded as misleading.

DISCUSSION

Contrary to the statistical meaning of value-as-price in the depreciation table, the concept of value was consistently appropriated on a personal scale by the students during the classroom discussion. Through the discussion, "value" became a more relative judgement in the meaning of "relative worth, a good value at the price" ("Value", n.d.) but the variable under discussion remained value and not price. An extreme evaluation narrative (such as Example 1) constrained comparison of prices of used cars to judge relative value. Instead, the worth of a used car remained a subjective judgement. Narratives in such a *value-for-me* discourse cannot be endorsed or rejected by anyone else, and hence prevents shift toward statistical discourse where consensus is based on the use of statistical measures. Prices-as-numbers are high or low, while price as a characteristic of used cars are cheap or expensive. In literate statistical discourse values of the variable price are alienated from personal judgements in terms of expensive or cheap. Without the alienation, statistical summaries like the mean price may remain contextual judgements on a personal scale of value-for-me, rather than an objective description of a data-set.

GG's judgement in Turns 68 and 70 (Example 3) is an example of placing cars on a relative scale of *value-for-money.* This evaluation scale is based on intuitions and experience about what one can get for a given price. Value-for-money discourse is more alienated than value-for-me discourse, although its narratives remain evaluative. The evaluation is not completely subjective anymore (there is no indication of "too expensive for me"), but related to other measurable properties of cars, in this case "mileage".

Increasingly objective discourse about the *relative-value-of-a-car* emerged as price became consciously related to other properties. SDS (Example 2) as well as GG and RK (Example 4) relate the value of a car to "condition" and in this discourse SDS suggests a not-extreme price as appropriate. I interpret RK's reference to "different terrains" (Turn 72, Example 4), on which a car was used as a proposal of a causal relationship between condition of the road, condition of the car driven on the road, and the value of the car. Less explicitly, the value of a car is also related to overuse and underuse through measurement of mileage (Turn 68, Example 3), hence another precursor of comparison of an actual measurement (mileage) to an average. Yet, GK (Turn 62, Example 3) voices her unease with the practice of assigning the same book value (average price) to cars that were *new* in January and December of the same year, respectively. Although a judgement of "condition" is likely to be on a subjective, evaluative scale between bad or something wrong and excellent, the discourse has shifted to properties of cars that are perceived to influence its value. In this discourse "reasonable price" surfaced briefly as a not-extreme price among other prices, but the concept remained caught between evaluative and descriptive reasoning.

The rejection of the official depreciation table of used car prices has particular pedagogical implications. The "book value" is widely used as a tool for comparison of the prices of used cars in practical situations. During the discussion the students had access to the relevant depreciation table and to a data set of prices of used cars. Yet, neither the car prices in the data table nor the "book prices" in the depreciation table were understood as objective measurements – parsed from value as a property of a specific car. In fact, GK explicitly questions the conflation of trade value and value of a car (Turn 66, Example 3). In Turn 71 (Example 4), KH endeavours to shift the discussion towards accepting the system and comparing values of aggregates of cars, relative to each other, regardless of specific differences. But KH's narrative is rejected. In Turns 72 to 75 (Example 4) RK expresses his doubt that the official book price is a reliable indication of the value of a car. He seems to suggest that the book value is used in practice to negate differences in value of individual cars. This example raises awareness that in everyday discourse applications of statistical tools such as mean prices of cars may themselves be judged on a personal scale and subsumed in evaluation narratives.

CONCLUSION

These examples from an extended classroom discussion at the start of a cycle of data handling tell a story of everyday discourses that are incommensurate with statistical discourse about the data-context. Yet, subtle enabling shifts in the discourse were made towards informal statistical reasoning. These discursive shifts can be described as from *value-for-me* to *value-for-money* to *relative-value-of-cars*.

If I had not allowed this extended classroom discussion of the context I would not have suspected that we were not talking of the same thing, namely objective prices of cars. I treated words like value, worth and price as synonyms (as they often are in everyday discourse) without understanding that the average price which I thought would emerge as a statistical measure of reasonable price would most probably not carry statistical meaning for the students. Discursive analysis of our word use revealed patterns that confirm experimental psychological research about novice statistical reasoning, and informs statistics education research about reasoning at low levels.

At the start of a cycle of statistical investigation the learning task is to abstract and alienate variables as measurements, so that contextual judgements at the "end" of the cycle can be based on applied statistical measures. The students' struggle to frame the discussion as an *investigation* of used car prices, rather than evaluation for prudent personal decision, confirms the hold of the context on the emergence of statistical reasoning. If contextual immersion and subjectification of context is accepted as a psychologically inevitable starting point in discussions of statistical contexts, teachers of statistics must learn to identify jarring discourses, without severing the complexity of the context from the statistical measures to be applied. With descriptive statistics in mind as the target discourse, evaluation narratives must be replaced by exploration narratives about the data-context, and observation and evaluation of variation must consciously change to objective measurement of variation. Drawing students' attention to such discursive properties of a classroom discussion may support a meaningful shift to statistical discourse throughout the investigation cycle.

REFERENCES

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Unpublished Doctoral Dissertation, Utrecht University.

Gigerenzer, G. (2002). *Reckoning with risk. Learning to live with uncertainty*. London: Penguin Books.

Jones, G., Langrall, C., Mooney, E., & Thornton, C. (2004). Models of development in statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 97 - 118). Dordrecht, The Netherlands: Kluwer Academic Press.

Kintsch, W. (1994). Text comprehension, memory and learning. *American Psychologist*, *49*(4), 294-303.

Nisbett, R. E. (1993). *Rules for reasoning*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Nisbett, R. E., Krantz, D. H., & Kunda, C. J. Z. (1993). The use of statistical heuristics in everyday inductive reasoning. In R. E. Nisbett (Ed.), *Rules for reasoning* (pp. 15 - 54). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Sfard, A. (2008). *Thinking as communication*. New York: Cambridge University Press.

Value [Def.3]. (n.d.). In *Merriam-Webster Online*. Retrieved 12 October 12, 2013, from http://www.merriam-webster.com.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35 - 62). New York: Academic Press.

Zwaan, R. A., & Kaschak, M. P. (2009). Language comprehension as a means of "re-situating" oneself. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 368 - 381). Cambridge: Cambridge University Press.