COPING WITH PATTERNS AND VARIABILITY – RECONSTRUCTING LEARNING PATHWAYS TOWARDS CHANCE

Susanne Schnell

TU Dortmund University, Germany

Dealing with the phenomenon chance in a mathematically sustainable way requires paying attention to predictable patterns in the long run, and observing variability especially in short series of repetitions. This paper exemplifies how students of 11 to 13 years of age cope with expecting patterns on the one hand and experiencing deviations from these patterns on the other in an experimental approach to probability. Empirical findings are presented from design experiments using a specifically-developed experimental learning environment, and are analysed using an interpretative framework.

Probability education is widely reported as quite challenging for students, as pre-instructional conceptions seem to persist through school and beyond, even if they oppose mathematically sustainable ones (overview in Shaughnessy 1992). The impact of these initial conceptions can be explained in terms of a constructivist perspective on learning: individual, active constructions of mental structures always build upon the existing prior mental structures by accommodation to experiences with new phenomena. Taking these initial conceptions into account is crucial for substantial teaching-learning processes as students “can’t check these in at the classroom door” (Konold, 1991, p. 144). In comparison to other mathematical areas, probability seems especially rich for pre-instructional conceptions (Prediger, 2008). The reason for this might be the proximity of the everyday phenomenon of chance and the close — but still different — mathematical one (Prediger, 2008); furthermore, it might be the nature of chance itself: patterns and variability are seemingly counterparts but only together constitute what stochastics is about.

Building on a previous CERME paper (Prediger and Schnell, 2011), this article will focus on the question of which constructs concerning variability and patterns are built in individual learning pathways and how they are used for making predictions.

PATTERNS AND VARIABILITY AS DEFINING ELEMENTS FOR THE PHENOMENON OF CHANCE

Moore (1990) uses the interplay of variability and patterns as a defining element for the phenomenon of chance itself: ”Phenomena having uncertain individual outcomes but a regular pattern of outcomes in many repetitions are called *random*. ‘Random’ is not a synonym for ‘haphazard’ but a description of a kind of order different from the deterministic one” (p. 98, emphasis in original). This definition gives way for different insights that are crucial for stochastics:

Patterns and variability: While especially frequentist approaches to probability are concerned with identifying *patterns* in many repetitions of chance experiments (as estimations for unknown probabilities, see below), it is the omnipresence of *variability,* which is causing the uncertainty that stochastics deals with (cf. Moore 1990, p. 135; Wild and Pfannkuch, 1999). Only when students learn to cope with both, they will get a substantial idea of the phenomenon of chance.

Short-term and long-term distinction: Even though variability is omnipresent, patterns can still be identified, especially in the long run. This is well-known as the empirical law of large numbers: While relative frequencies vary a lot for only a *few repetitions* of an experiment (short-term perspective), the variation of relative frequencies is smaller for a *larger number of repetitions* (long-term perspective) (cf. Fig. 1).

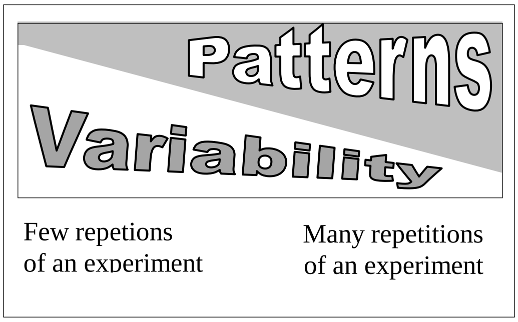

Fig. 1. Interplay of patterns and variability in relation to the number of repetitions

The patterns identified in the long-term perspective are on the one hand an estimation for the theoretical probabilities. On the other hand, theoretical probabilities (determined by a classical approach, for example) serve as prediction of the patterns which will be observable in series of many repetitions. Both the quality of prediction and the quality of the estimation are related to the distinction of the short-term and long-term perspective: While students naturally wish to be able to predict the next single outcome of an experiment (cf. Konold, 1991), probabilities only make statements for long series of repetitions.

Thus, when learning about probabilities in an experiment-based approach, students not only have to make sense of patterns and variability. It is the distinction of a short-term and long-term context, which is crucial to making sense of these different sides of the same phenomenon (Prediger and Schnell, 2011).

PROCESSES OF LEARNING ABOUT VARIABILITY AND PATTERNS

There has been some research undertaken about how students cope with variability in statistics (especially in exploratory data analysis; overview in Shaughnessy, 2007). Results such as Ben-Zvi (2004) show how students gradually develop a sense of how to take variability in statistic samples into account: They use frequency overviews as well as address outliers and compare variation within and between distributions.

For probability education though, research on learning pathways coping with variability and patterns is still limited (cf. Jones et al, 2007, p. 928): For instance, Pratt and Noss (2002) investigate the learning processes of students (age 10 and 11) and describe how they can successfully progress from a short-term context to a long-term context when working in a supporting learning environment. While initially conceptions such as unpredictability of single events are dominant, new conceptions emerge in the learning process: patterns are more stable for large numbers of experiments as well as expected and observed patterns can be linked to theoretical probabilities. This gives insights how students develop a sense for the long-term context and sustainable patterns in the long run, but it leaves questions open on how students encounter and cope with variability.

What becomes clear though is the crucial role of the learning environment to facilitate substantial learning.

LEARNING ENVIRONMENT 'BETTING KING' AND ITS INTENTIONS

The learning arrangement 'Betting King' (published in Prediger & Hußmann, 2013) was developed as part of a textbook, introducing probability in grade 6 (students of age 11 to 12). The main activity consists of predicting single or aggregated outcomes of a chance experiment: The players bet on the outcomes of a race with four coloured animals, being fuelled by throws of a coloured die with asymmetric colour distribution (red ant: 7/20, green frog: 5/20, yellow snail: 5/20, blue hedgehog: 3/20). The game is first played on a board, which is later replaced by a computer simulation (Fig. 2). The differentiation of the short-term and long-term contexts is realised by the possibility to choose the number of throws of the die, after which the results of the game are determined (between 1 and 10,000 in the simulation). Record sheets and written tasks are used to support the learning processes.

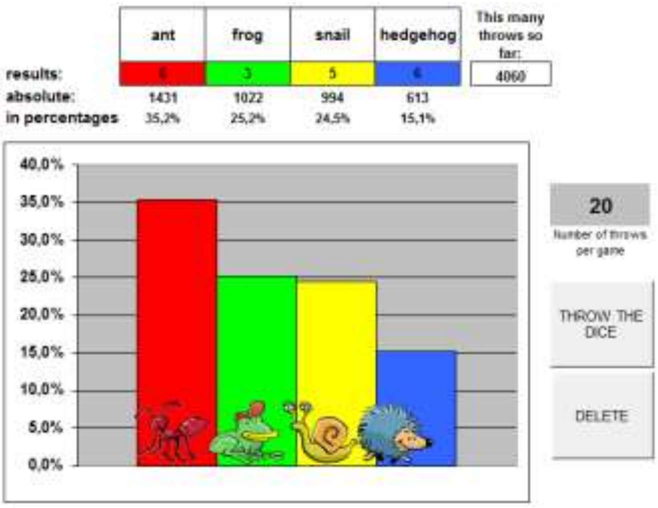

Fig. 2. Layout of the computer simulation when looking at percentages (translated and slightly rearranged for this paper)

There are two consecutive game settings used to motivate explorations of the generated game results: In the first setting, "Betting on Winning", students bet on the animal that will go the farthest. Here, students usually soon discover the pattern that the red ant wins more often and can link it to the unfair colour distribution. Next, the security of this pattern is addressed in relation to the total number of throws: While for short games the pattern is insecure (i.e. the overall winning animal varies a lot), it is rather stable for large total numbers such as 2,000 (i.e. variation of the winning animal of each game is hardly observable).

The second game setting, "Betting on Positions", motivates to look more closely at the generated patterns: Students play against each other and predict for each animal on which position it will land in relation to a pre-set (large) total number of throws. The bet closer to the result wins one point per animal. Soon, the game itself is left and students are guided to investigate patterns and variations of positions presented in different representations offered by the simulation (numerical and graphical). Here, the focus first lies on comparing results for series of games with the same total number of throws (e.g. patterns and variability within a series of games of 20 throws each compared to a series of games with 2,000 throws) and later introduces a dynamic perspective, where relative frequencies (presented as percentages) can be investigated when the total number of throws is growing gradually.

Observations, patterns and variability play a central role in both game settings: a high variability means the chances to get a point are low, while a stable pattern holds a good chance to win.

RESEARCH QUESTIONS AND DESIGN OF STUDY

The following empirical findings will describe how students address the observations of patterns and variability when working on the learning environment ‘Betting King’. As the empirical data shows, the repertoire of different patterns observed by students is so rich (cf. Schnell, 2013) that this paper will focus on the scenes when students explicitly notice variability in the data, which is contrasted then with the related patterns. The following questions will be addressed:

> *Which constructs of variability are developed over the course of working on the learning environment and how are they linked to an according pattern?*
>
> *How does the overall conception of variability develop over the course of the design experiments?*

Data collection

The foundation for this paper is a study in a laboratory setting, which was conducted to research the individual learning pathways in more detail (cf. Schnell 2013; Schnell and Prediger, 2012 for the broader design research study). Here, the learning environment was used in design experiments (cf. Cobb et al. 2003) with nine pairs of students (age 11 to 13). Each design experiment lasted at least four consecutive sessions of 90 minutes. They were semi-structured by an experiment manual that defined the sequence of tasks as well as interventions for anticipated problems in crucial learning. The data corpus of the study includes videos of all 40 experiment sessions, records of the computer simulations as well as all written products.

Data analysis

The *in-depth data analysis* is guided by an interpretative approach, identifying scenes in which new discoveries are made and integrated into the individual conceptions, here conceptualized as a network of *constructs* (cf. Schnell and Prediger, 2012 for the theoretical background of the analytical model). In these scenes, students' individual constructs are reconstructed with respect to four elements: the emerging insights (summarised by <*proposition*>), the *stochastic context* in which an insight occurs, the *situational context* to which it refers (such as the representation; due to page limitations this will not be addressed in this paper) and which *function* the construct has in the specific situation (e.g. explanations, descriptions of deviations, etc.). Due to the complexity of the game situation, the stochastic context was specified further than short-term and long-term context: Students' constructs can relate to **O**ne or **M**any games (differentiation O-M) and the game(s) can have **S**mall or **L**arge total numbers of throws (differentiation S-L). While patterns in accordance with the colour distribution of the die can be best observed for the context ML (Many & Long), variability is more visible for SM and single deviations (e.g. outliers of positions) in OS but also in OL.

In this article, the learning pathway of one pair of students (Ramona and Sarah) is used to exemplify how learners are coping with variability. For this, constructs with the function "description of deviations" were picked out and complemented with constructs describing the according pattern. To show the development of how variability is experienced, the constructs are contrasted and differences regarding the stochastic context, the function and the influence of predictions are identified.

EMPIRICAL FINDINGS: RAMONA AND SARAH'S PATHWAY OF MAKING SENSE OF VARIABILITY AND PATTERNS

The case of Ramona and Sarah shows how students start with initial conceptions of variability which then get complemented when the focus is shifted from single games (OS-context) to series of games (MS and ML). The girls are one of the focus pairs of the study as their learning pathway is especially rich and well verbalised. Comparisons with other pairs of students will be made in the conclusion.

1. First encounters with variability: Constructs with function explanation (OS-context)

CONSTRUCS EXPLAINING PATTERNS				CONSTRUCTS EXPLAINING VARIABILITY			
Short-title	**Proposition**	**Sto. con.**	**Time**	**Short-title**	**Proposition**	**Sto. con.**	**Time**
ANT-BEST	<Red ant is always best>	MS	1[1]; 21:15				
COLOUR-DISTRIBUTION	<Red ant: 7, green frog: 5, yellow snail: 5, blue hedgehog:3>	All	1; 28:05	HEDGE-HOG-LUCK	<When blue hedgehog is lucky, it will get the blue faces a lot>	OS	1; 30:17

[1]The first number indicates the experiment session in which the construct was first verbalised

Ramona and Sarah identify the red ant as the best animal (ANT-BEST) already at the end of the second game (first game: 25; second game: 38 throws). When roughly seven minutes and two more games later the asymmetric COLOUR-DISTRIBUTION of the die is discovered, this is backing up the ANT-BEST pattern-construct (for more detail cf. Prediger and Schnell, 2011). At this point though, the girls notice that the order of the animals on the board game (with a total number of 28 throws) does not fit the expected order from the colour distribution: blue hedgehog is third and yellow snail is last. Ramona thus constructs HEDGEHOG-LUCK in order to explain the deviation in this one single game from the expected pattern:

Session 1; beginning 30:17;
After having played a game of 28 throws, the colour distribution is discovered

483 Ramona: Blue has good chances, too, because-

484 Sarah: Yes.

485 Ramona: You also have- blue has sometimes a lot of luck and then it gets the three faces sometimes very often.

In this scene, variability is noticed as deviation from the first empirically observed, then theoretically backed-up pattern ANT-BEST. The discrepancy between the order of animals according to the COLOUR-DISTRIBUTION and the empirical result on the board is solved by building the construct HEDGEHOG-LUCK. Its function is to explain why this deviation occurred (i.e. the supposed "luck" of the blue hedgehog, which seems overlay COLOUR-DISTRIBUTION). In several other instances, LUCK is used again to explain *single deviations* of expected patterns, for example when blue hedgehog lands on unexpectedly high position in 'Betting on Positions'.

2. Patterns versus absence of patterns (ML- and MS-context)

CONSTRUCTS ADDRESSING PATTERNS				**CONSTRUCTS ADDRESSING VARIABILITY**			
Short-title	**Proposition**	**Sto. con.**	**Time**	**Short-title**	**Proposition**	**Sto. con.**	**Time**
ANT $_{100,1000}$	<For series of 100 and 1000, red ant always wins >	ML	2; 52:20	DIFFER-ENT$_{1,10}$	<For series of 1 and 10 it is always different>	MS	2; 52:20
RANGES $_{2000}$	<For series of 2000, observed positions vary within certain ranges>	ML	3; 45:53	NO-RANGES$_{20}$	<For series of 20, positions are always messed up and no ranges are observable>	MS	3; 46:16
HIGH-BARS-LESS	<The higher the total number of throws, the less move the bars in the diagrams >	ML	3; 87:39	LOW-BARS-MORE	<For series of low total numbers of throws, bars move a lot>	MS	3; 88:04
PERCENTAGES-LIMITS	<Percentages have limits beyond which they don't change anymore>	ML dyn	4; 35:05	PERCENTAGES-CHANGE	<With steps of 20, percentages change a lot in the beginning>	MS dyn	4; 42:30

At the end of the first game type "Betting on Winning", the students are asked to fill in an overview of all used record sheets, summarizing which animals won at least once for the total numbers of 1, 10, 100 and 1000 throws. After this, the girls are asked to verbalise their observation:

Session 2; beginning 52:15; after having filled in the overview table:

1079	Sarah:	For 100 and 1000, [red] ant always wins. And for 10 and 1, it's always different-

In this statement, Sarah contrasts the stable *pattern* of the winning red ant for high total numbers of throws ($ANT_{100,1000}$; stochastic context ML) with the *absence of a pattern* of a specific winning animal for low numbers of throws ($DIFFERENT_{1,10}$; MS). Here, the function of the variability-construct is not an explanation of a single deviation, but a description of an observation Sarah made for a series of many games with a small total number of throws.

This contrasting of small and large total numbers of throws becomes a conscious strategy in the following, where Sarah first verbalises the construct RANGES: <Positions stay in certain ranges> and Ramona immediately projects this on the differentiation between small and large total numbers of throws:

Session 3; beginning 45:3; Predicting positions for games of 2,000 throws

769	Sarah:	I was throwing the dice just like that, then I saw that 400 and 500- *points to screen)* mostly this gets 500 and sometimes also 400, this *(points to screen)* gets 700 and 600 and there 200 and 300.
		(…)
772	Ramona:	But that's only like that for <u>2,000</u>. Ah yes, of course, right? Yes. Type in 20.
		(…)
790	Ramona:	*(after having looked at a series of games of 20 throws)* Ah, no, for 20, it's always mixed up.

Again, variability is constructed as the absence of stability ($NO\text{-}RANGES_{20}$). The idea of contrasting long-term and short-term contexts in the search for the existence of patterns is then also used for the bar-charts (constructs HIGH-BARS-LESS and LOW-BARS-MORE).

Furthermore, the influence of variability and patterns for finding a secure bet is made explicit in relation to the stochastic context:

Session 4, 73:30, at the end of playing 'Betting on Positions'

1057f.	Sarah:	[For 10,000] nothing moves, I think. No, not much. So it's easier to bet when you have these high numbers than the smaller ones.

Variability now seems to be a phenomenon, that is actively addressed and influencing the security of bets. It is also complementing pattern-constructs while being restricted to the short-term context.

This restriction of an absence of a pattern to games with small total numbers of throws becomes a challenge when the girls encounter variability also for large total numbers of throws: When the game switches to percentages, a dynamic view of games' results is introduced and the results of every game are cumulated (which is discussed in detail with the interviewer at 33:15). While Ramona verbalises the construct <Percentages have limits beyond which they don't change anymore> (Session 4; 35:05) after a few games v, the relation to the high total number of throws is not yet clear. Thus, the girls are asked to keep experimenting with the games.

Session 4; beginning 38:30; (Sarah keeps generating games of 20, so far 14 games of 20, i.e. a total number of 280 throws)

567	Ramona:	The first digit [of the percentage, which are shown with three decimals, e.g. 35.0; cf. Fig. 2] always stays the same. And the last digit doesn't – never.
568	Interviewer:	Hm.
569	Ramona:	And in the middle, it stays the same, too. *(Sarah is generating more games of 20 throws; now roughly 3000 throws in total)* Yes, sometimes. But not always. For [blue] hedgehog it does- no it does not. Oh man!

While Ramona's initial reaction to percentages was to expect a stable pattern (PERCENTAGES-LIMITS), she is now struggling as the percentages keep changing. Maybe as a solution for this, she splits the percentages into the individual digits, for which she can find a stable pattern for the first decimal, the absence of a pattern for the last decimals, but then also encounters problems with the digit in the middle. Aside from the individual approach to percentages, this shows that for Ramona, the absence of patterns (and thus variability) means it *is* changing. A notion of changes that occur *more or less frequently* seems not yet established.

It takes four more minutes and generating several series of growing numbers before the girls can construct the variability observation as counterpart for PERCENTAGE-LIMITS and relate both to their stochastic context.

CONCLUSION AND OUTLOOK ON OTHER STUDENTS

When Ramona and Sarah start working in the learning environment, they perceive variability as the *deviation of a single game* from an expected pattern. This leads to building a construct which provides an explanation as to why this occurred — which is later used for outliers of all kinds. This initial conception of variability gets complemented when variability also occurs in series of games with small total numbers of throws. Here, the girls not only notice the *absence of a pattern* but keep actively seeking for it when confronted with new representations. They are not giving explanations for these constructs but make use of the interplay of patterns and variability to evaluate the security of predictions in relation to a long-term and short-term context. Ramona's struggle with percentages shows, however, that the division of the two contexts might be too rigid when it comes to a dynamic perspective. Since, up to this point, series of games with the same total number of throws were addressed, the contrast of patterns and no-patterns was sustainable. But for the dynamic perspective — and percentages specifically — the occurrence of occasional variations is posing a new challenge for Ramona.

The here presented scenes are snapshots from the learning pathways of two girls. Even though these pathways are highly individual, some common features of the pathways towards variability can be identified that appeared with all nine pairs of students: at the beginning of all design experiment series, constructs with the *function of explaining single deviations* could be reconstructed. These conceptions are used less often over time when the *focus shifts to series of games* with a large total number of throws and are almost exclusively used for deviations of single games with small or large total numbers of throws.

While Ramona and Sarah find a meaningful way to deal with patterns and the absence of them by relating them to different stochastic contexts, the omnipresence of variability can also pose challenges of which patterns can be observed: Diana and Julia for example focus on the still visible changing of the bars' heights for games of 10,000 throws, but they relate this only to the green frog and blue hedgehog. When then asked to compare this with games of 30 throws, they first address that more animals are changing before they describe the larger and smaller differences in the movements for the different contexts.

Lastly, Hannah and Nelly are over-generalising variability per se: Even after having played the game 'Betting on Winning' for long series of games with 10,000 throws in which only the red ant won, they refer to this as "chance" and claim it could be different every time. They deny the existence of (meaningful) patterns at all, before and even after discovering the asymmetric colour distribution.

Overall, most students from the empirical study managed to evaluate the security of predictions correctly in relation to short-term and long-term contexts. However, how they cope with encountering variability is highly individual. To understand these complex processes in more detail, further research focuses on a broader overview of the different conceptions of variability (cf. Schnell, 2013) and tries to identify types of processes of how conceptions develop when variation is encountered (cf. Schnell and Prediger, 2012; Schnell, 2013).

LITERATURE

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal*, 3(2), 42-63.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R. & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, 32(1), 9-13.

Jones, G.A., Langrall, C.W., & Mooney, E.S. (2007). Research in probability: Responding to classroom realities. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909-956). Charlotte, USA: Information Age Publishing.

Konold, C. (1991). Understanding students' beliefs about probability. In E. von Glasersfeld (Ed.), *Radical Constructivism in Mathematics Education* (pp.13-156). Dordrecht: Kluwer.

Moore, David S. (1990). Uncertainty. In L. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academy Press.

Pratt, D. & Noss, R. (2002). The Micro-Evolution of Mathematical Knowledge: The Case of Randomness. *Journal of the Learning Sciences*, 11(4), 453-488.

Prediger, S. & Hußmann, S. (2013, in press). Spielen – Wetten – Voraussagen. Den Zufall im Griff? To be published in S. Prediger, B. Barzel, S. Hußmann & T. Leuders (Eds.), *Mathewerkstatt 6*. Berlin: Cornelsen. (Textbook for German middle schools)

Prediger, S. & Schnell, S. (2011): Individual pathways in the development of students' conceptions of patters of chance. In: M. Pytlak, T. Rowland & E. Swoboda (Ed.), *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education* (pp.885-895). University of Rzeszow, Rzeszow.

Prediger, S. (2008). Do you want me to do it with probability or with my normal thinking? Horizontal and vertical views on the formation of stochastic conceptions. *International Electronic Journal of Mathematics Education*, 3(3), 126-154.

Schnell, S. & Prediger, S. (2012). From "everything chances" to "for high numbers, it changes just a bit". Theoretical notions for a microanalysis of conceptual change processes in stochastic contexts. To appear in: *ZDM - The International Journal on Mathematics Education*.

Schnell, S. (in preparation for 2013). *Entwicklung von individuellen Vorstellungen zum Phänomen Zufall* (working title). PhD Thesis. TU Dortmund University (Supervisor S. Prediger).

Shaughnessy, J. M. (1992). Research in probability and statistics. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: MacMillan.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 957-1009). Greenwich, CT: Information Age Publishing, Inc. and NCTM.

Wild, C. J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (with discussion). *International Statistical Review*, 67(3), 223-265.