International Journal of Mathematical Education in Science and Technology, Vol. 00, No. 00, 15 July 2009, 1–13

RESEARCH ARTICLE

Using Mental Imagery Processes for Teaching and Research in Mathematics and Computer Science

Pierre Arnoux[a] and Alain Finkel[b*]

[a] *Institut de Mathématiques de Luminy, CNRS U.M.R. 6206, 163, avenue de Luminy, Case 907, 13288 Marseille cedex 09, France*;
[b] *LSV, ENS Cachan & CNRS 61, av. du Président Wilson, 94230 Cachan, France*

(*Accepted for presentation at ICME-11, the International Congress on Mathematical Education, eleventh edition, Conference TSG 17: Research and Development in the Teaching and Learning of Advanced Mathematical Topics, Monterrey, Mexico, July 6 - 13, 2008*)

The role of mental representations in mathematics and computer science (for teaching or research) is often downplayed or even completely ignored. Using an ongoing work on the subject, we argue for a more systematic study and use of mental representations, to get an intuition of mathematical concepts, and also to understand and build proofs. We give two detailed examples.

Keywords: Mental representations, images, kinesthetic models.

1. Introduction

1.1. *Motivation - From our Experience*

It is a common experience for students or young researchers to realize in retrospect that a large part of understanding consists in building representations of various types. These representations are often both external (e.g., a picture) and mental (feelings). Some of these representations are well-known, e.g. the geometric image of complex numbers as the complex plane, or the kinetic model for elementary calculus. When such a representation is well-chosen, it readily suggests hints of solutions and proofs for related problems and theorems; the difficult task to formalize these solutions and proofs of course remains.

The teaching of mathematics is mainly focused on this formalization task. The creation of appropriate mental representations is largely left to the students; many of them simply fail to build appropriate representations, and acquire a mathematical knowledge that is essentially devoid of meaning. This prevents them from solving non-trivial problems, a common (and painful) experience for many teachers.

In this paper, we present two examples of representations in theoretical computer science and mathematics, and show how these representations lead to an understandable proof of two non-trivial theorems (the pumping lemma and Chebyshev's inequality); without a representation, the usual formal proof of these theorems is

*Corresponding author. Email: finkel@lsv.ens-cachan.fr

ISSN: 1045-1129 print/ISSN 1029-0346 online

DOI: 10.1080/1045112YYxxxxxxxx
http://www.informaworld.com

usually a challenge for students, and the theorems often remain unusable as tools. We argue for a systematic study of mental representations, which should include:

- Building a dictionary of representations;
- Studying which representations are useful for which purpose, and for which students;
- Studying how such representations can be transmitted in teaching in an effective way;
- Creating new representations;
- Understanding how such representations work at the brain and body levels.

1.2. *Mental Imagery and Representations*

The study of mental representations (in particular visual ones, i.e. visual images) has a long history. One may find interest in mental representations in philosophical works since Plato. The philosophers of phenomenology (Husserl, Heidegger and Merleau-Ponty) studied mental activity during thinking. The scientific exploration of mental states, started 150 years ago, used introspection (mainly represented by Binet in France and Tichener in Germany, both between 1850 and 1900) and described the phenomenology of mental states.

Contemporary cognitive scientists (Kosslyn, Dretske, Fodor, Shepard, Pinker, Denis, Mellet [1–3]) explore mental images (Shepard is well-known for a convincing proof of the existence of mental rotations) and representations from different points of view (neuro-physiology, neuro-psychology, philosophy). For instance, many papers are devoted to the study of the different modalities of representations in itinerary strategies, see [4] for a review about mental imagery.

Among the different representation systems, two of them enjoy a crucial role: the visual and the verbal systems. Paivio was the first to prove the role of double coding (verbal and visual) to facilitate memorization.

Contemporary cognitive scientists also recently started to use neuro-imaging to study mental imagery, and it is now becoming possible to check, using technologies from neurosciences, that different areas of the brain are recruited to process verbal processes or visual processes.

The domain of imagery is rich and there is an ongoing debate about the nature of the representational format of visual mental imagery. There are two viewpoints: the propositional side is to consider that all representations are, in fact, propositional, which means that all representations are linguistic or symbolic; the analog side is to say that visual imagery shares some non-symbolic properties with the perception experience. This debate is important because we believe that traces of it may be retrieved in the opinions of those who think that images are important to understand, and those who believe that words are sufficient and better than images.

We strongly believe that the construction of multi-modal representations (visual, symbolic/verbal, haptic,...) is important for understanding mathematics (and for understanding in general).

1.3. *Mental Imagery for Mathematics and Computer Science*

Understanding a mathematical definition, a proof, a concept, . . . is strongly related to building and handling mental objects [5]. One of the oldest and most striking examples is probably the plane representation of complex numbers, discovered more than two centuries after their invention, which made clear many of their properties; it is unthinkable today to teach complex numbers without teaching the complex plane, and this mental representation is invaluable in many respects for understanding concepts, or finding proof strategies (for an example, see Tristan Needham's book [6]). Jacques Hadamard was a pioneer in cognitive psychology of mathematical thinking. His book [7] is still interesting to read. More recently, the Working Group on Representations [8] synthesized four different interpretations of the term "representations". For a bibliography about representation systems, in particular for mathematical education, see [9], pages 160-165. Recently, Carlson, Oehrtman and Thompson studied the mental representations (in terms of mental images and mental actions) associated with functions [10]; Corter and Zahner studied the mental representations associated with probabilities [11]. Thompson studied the relationships between mathematical reasoning and imagery [12]. Mesquita discussed external representations in geometry [13]. However, the usefulness of building representations is often downplayed as an aside, or ignored; one reason is that good representations are not easy to obtain (it took more than 200 years for the complex plane!). Another reason is that many teachers and researchers do not know that it is important to communicate about them.

Let us refer to [9], page 139: "Visualization, spatial and kinesthetic representations, image schemata, and imagistic representation in general, including metaphor and metonymy, came to be recognized as critical to understand how mathematical concepts are meaningfully understood".

Goldin presented a united framework for representation systems. He distinguished between external and internal representations, and gave a classification of internal representational systems, see [9], page 148. Following Goldin, systems of internal representations are composed of five sub-systems:

- verbal/syntactic systems: the language;
- imagistic systems (including visual/spatial, auditory/rhythmic, and tactile/kinesthetic);
- formal notational systems of mathematics;
- a system of planning, monitoring, and executive control;
- a system of affective representation.

We will not discuss this model here, but focus our attention on the imagistic systems (the second category of his classification). We exemplify the imagistic systems with different examples coming from non-trivial mathematics and computer science. To the best of our knowledge, mental images are sometimes presented in primary education, more rarely in secondary schools and occasionally at the university level. The same is true when researchers try to present their scientific results to a large public, and also to other researchers.

1.4. ***Objectives***

We have started a modest team work of study of mental representations, and we would like to argue for a more systematic study, at several levels:

- Classifying different types of representations (visual, kinesthetic,...). There is already some work in this direction, but usually not applied to mathematical teaching. It would be good to have a set of useful representations of different types. For example, it is unclear to us (maybe because of our own limitations) what can be an auditive representation which is non-verbal.
- Studying the efficiency of various representations in a systematic way. We know, because many students have told us so, that some representations are useful to understand concepts and proofs, but this remains very unsystematic at the moment. It is highly likely that no representation can be universal, and that

different people will get help from different representations. A good lecture would be one that can use representations of various types, so that all audiences can find one that is appropriate.

- Building a catalogue of representations. Good representations are difficult to come by; one cannot expect that a teacher, even an outstanding one, can find by himself a good model for all the concepts in his lectures. Furthermore, any person will naturally find a particular type of representation, and will be able to speak efficiently only to a part of the audience. A large catalogue of mental representations, with their type and the way they can be used efficiently, would be a very useful teaching tool.

We give a few examples below.

2. Models for Graphs and Automata

What is a graph G ? It can be *defined and said* as a relation G between elements of a set V of *vertices*, or as a subset G of the cartesian product $V \times V$ (even though mathematicians know that a relation is a subset, this is not the usual meaning of the word relation in natural language); if the student knows what a function is, a relation may be defined as a function which associates a subset $V_x \subseteq V$ with each element $x \in V$. These definitions are mainly verbal in the sense that they do not explicitly induce a visual image or a feeling. A graph G can also be *defined and seen* as a picture with nodes (vertices) $v \in V$ connected by arcs, hence $G = (V, \to)$ where $\to$ is the set of arcs. Concrete examples are maps (nodes are towns and arcs are roads, nodes are metro stations and arcs are railtracks).

A different view is that of the adjacency matrix, a square matrix $M_G = (m_{i,j})$ indexed by the nodes, where $m_{i,j}$ counts the number of arcs joining nodes i and j. Counting the number of paths of length 2 from i to j leads in a very natural way to the classical formula $\sum_k m_{i,k} m_{k,j}$, since any path of length 2 must go through an indeterminate node k. This motivates efficiently the formula for the matrix product; a large chunk of algebra can be illustrated in this way. We believe that a student who has the verbal representation (knowing the formula of product of matrices) *and* the vision of paths in a graph would better understand, memorize and solve problems about matrices and graphs. Extending Paivio, we say that students who mentally represent mathematical objects both verbally and visually will succeed more often.

The arcs of a graph can be labeled by letters belonging to an alphabet Σ (an alphabet is a finite set and we define Σ^* to be the set of all finite words on Σ including the empty word ϵ; formally Σ^* is the free monoid on Σ with the associative operation . defined as concatenation; ϵ is the identity element for concatenation); we thus obtain a *labeled graph*: $G = (V, \Sigma, \to)$ where $\to \subseteq V \times \Sigma \times V$. A finite automaton A (a basic and fundamental object in computer science) can be *defined and said* as a triple $A = (V, \Sigma, \to)$, or *defined and shown* as a labeled graph. Usually, the role of a finite automaton is to recognize a (potentially infinite) set of words (i.e., a language) on the alphabet Σ. Let A be a finite automaton $A = (V, \Sigma, \to)$ where $\to \subseteq V \times \Sigma \times V$. Fix an initial state v_0 and a finite subset $F \subseteq V$ of final states and define the language $L(A)$ of A as the set of finite words $w \in \Sigma^*$ labeling a path (in A seen as a graph) from v_0 to a final state $v \in F$: formally, $L(A) = \{w \in \Sigma^*; v_0 \overset{w}{\to}^* v, v \in F\}$ where $\overset{w}{\to}^*$ denotes the finite sequence $\overset{w_1}{\to}\overset{w_2}{\to} ... \overset{w_n}{\to}$ with $w = w_1 w_2 ... w_n$, $w_i \in \Sigma$ and $v_0 \overset{w_1}{\to} v_1 \overset{w_2}{\to} v_2 ... v_{n-1} \overset{w_n}{\to} v_n = v$.

Now one may consider a finite automaton as a machine for accepting (or recog-

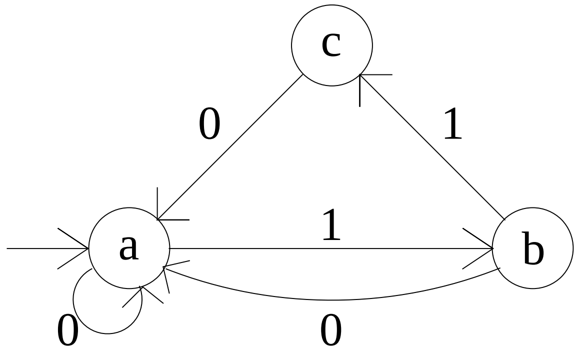

Figure 1. A usual representation of a finite automaton

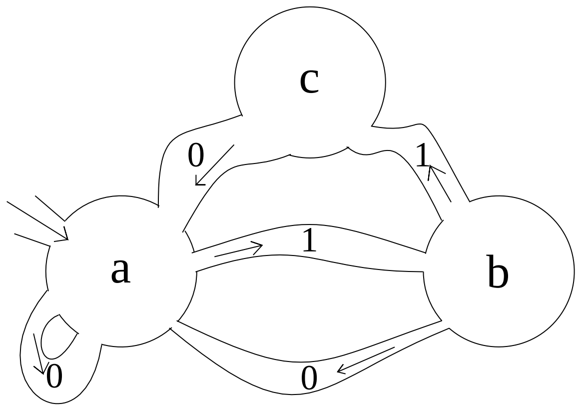

Figure 2. A representation of a finite automaton as a maze

nizing) words put on an input tape or for writing words on an output tape.

- a *recognizer* that reads some letters on an input tape,
- a *producer* that writes some letters on some output tape.

Both the reading and the writing are conditioned by the set $\rightarrow$ of transitions.

These different visions are *not cognitively equivalent*, even though they are mathematically (almost) equivalent; they define the class of regular languages. For instance, the environment is important for the recognizer view (one needs letter on the tape to consume them), but does not exist in the producer view (there is no obstacle against writing on an infinite tape). These representations are mainly visual and partially kinesthetic (in the sense that they use some movement and feeling of movement). Every student would probably have his own preference.

Another view of a finite automaton in terms of graph, which is suitable even for children, is that of a maze with rooms (states) and corridors (transitions); the corridors are one-way only, and they have a name taken in the alphabet Σ (several corridors can have the same name). The maze has one entrance (the initial state), and one or several exits (the final states).

To recognize a word, the player is given this word as a stack of cards, each one showing a letter of the word. He enters the maze by the entrance; at each step, he chooses a corridor named by the letter of the card on the top of the stack, discards this card, and follows the corridor to the next room.

In Fig. 1, we show the conventional representation of a finite automaton $A = (V, \Sigma, \rightarrow)$ with 3 states a, b, c and whose alphabet is $\Sigma = \{0, 1\}$; Fig. 2 gives the representation of the same automaton as a maze. The initial state is $v_0 = a$, represented in the pictures as an incoming arrow; every state is a final state, hence we have not found useful to represent them.

Let us note that the word 111 is not recognized by the associated automaton because, starting from initial state a, the first letter leads to state b, the second letter leads to state c, and from this state there is no letter labeled by 1. The language is in fact the set of words w such that 111 is not a subsequence of w, i.e., $\{w\} \cap \Sigma^*111\Sigma^* = \emptyset$, hence $L(A) = \Sigma^* - \Sigma^*111\Sigma^*$.

This representation is visual (the maze is a graph!) and the dynamics of the automaton is illustrated with the help of imagination of mental movements on the graph. This representation is then both visual and more kinesthetic than the previous ones.

Recognition fails if at some moment no corridor in the present room is named by the letter on the top card, or if the last room is not an exit; moreover, exiting the maze is allowed only when the stack of cards becomes empty.

There are possible differences between a maze and a finite automaton. In a *real* maze, we may suppose that it is possible to come back and try another path but here, we must suppose that this is not possible. There are other differences, finding them and understanding them may help understand finite automata better.

This model can be understood at a very basic level, and allows one to attack and solve a number of elementary (and not elementary) problems. It is not complete: for example, what happens if two corridors from the same room bear the same name? But this incompleteness is in fact convenient, because it allows one to introduce interesting concepts (deterministic/non-deterministic) in a natural way, often following questions from the students.

This definition/representation can be presented as a game and involves more deeply the imaginary movement of the body in an imaginary space. This would recruit different zones of the brain than a purely verbal definition.

If one views word recognition as a walk on a finite graph (random or deterministic, depending on the type of the graph), then there are interesting consequences. For example, such a walk, if long enough, must contain loops, and loops can obviously be iterated: one gets a very simple and intuitive proof of the classical pumping lemma, usually stated in this way: If a finite automaton has k states, any word w recognized by the automaton, of length at least k, can be factorized $w = xuy$, where $x, u, y \in \Sigma^*$, $u \neq \epsilon$, so that all the words xu^ny, for all integers n, are recognized by the automaton.

Let us now recall that a language is *regular* if it is the empty set or if it can be obtained from the singleton language $\{a\}$, $a \in \Sigma$, and from the empty word language $\{\epsilon\}$ by using a finite number of the following three operations: union ($\cup$), concatenation (.) and the star operation ($*$) defined by: $L^* = L^0 \cup L^1 \cup L^2 \cup ... \cup L^n \cup ...$ where L^n is the language $L.L...L$, where L is concatenated n times. For instance, the language recognized by the automaton above can be denoted by the regular expression $\{0, 10, 110\}^*.\{e, 1, 11\}$. If now, one considers regular languages, then the pumping lemma becomes:

Pumping lemma: *Let L be a regular language, then there exists an integer k_L such that for every word $w \in L$ such that $|w| \geq k_L$ there exist three words $x, u, y \in \Sigma^*$ such that $u \neq \epsilon$, $|xu| \leq k_L$, $w = xuy$ and for all $n \geq 0, xu^ny \in L$.*

This lemma is not so simple to prove from the regular expression of L. But there is a deep connection between the class of languages recognized by finite automata and the class of regular languages. The Kleene Theorem states that these two classes of languages are the same. Due to this theorem, one may use either the finite automaton view (with visual and kinesthetic graph-representations) or the regular expression view (more oriented towards verbal representations).

In theoretical computer science, and more generally in all sciences, some viewpoints produce some results more easily than others. For instance, knowing that finite automata generate exactly the class of regular languages (which may be also characterized by formulae of logics and/or algebraic characterizations by finite quotients of monoids), we may easily prove the classical pumping lemma for regular languages. Hence, a student who would have understood and memorized these two representations (graph-theoretic and algebraic representations) would probably succeed more often in solving problems.

There is another version of the pumping lemma for stack automata and context-free grammars. Still, a visual reasoning on the derivation tree associated with a context-free grammar helps a lot in guessing how the (double) iteration occurs; and this iteration lemma is very difficult to guess and prove when only considering stack automata. We are not saying that visual reasoning is better than verbal/algebraic reasoning! But to be able to have representations is important and to be able to adapt our representations allows us to better understand and make statements and proofs.

3. Mechanical Model in Probability

The first concepts of probability theory (expected value, variance) are often mysterious for the students, and basic results, such as Chebyshev's inequality, are usually experienced as pure algebraic computations void of meaning. The usual definition of the basic object in probability, a random variable, is itself very abstract, and does not allow for an easy intuition. But there is a mechanical model for a random variable which, although not complete, can give a much better grasp on these concepts.

3.1. *The Mechanical Model*

A random variable is usually defined as a real function, denoted by X, on a space endowed with a measure μ of total mass 1; in most elementary computations, one only uses the image measure $X_*\mu$ on the real line.

Hence we view a random variable as a distribution of measure, or probability, on the real line; and we represent this probability distribution as a metal bar of variable width and finite weight. One should take some time to play with this representation for many cases: the uniform distribution on an interval is just a metal bar of constant width and finite length, like a usual metal ruler; a Bernoulli distribution, associated with a game of heads and tails, is given by two point masses (atoms) linked with a rigid bar without weight, a limit model of a dumbbell ; a bell curve distribution is an infinite bar which becomes extremely thin for large values, and which can be shown explicitly by the graph of the bell curve. We can consider this bar as being made from small atoms; the random variable consists in choosing randomly an atom in the bar, and it is given by the location of the atom. A random choice will go more often to the places where the bar is thicker, as there are more atoms there. We will always suppose that the total mass of the bar is finite, otherwise there is no simple way to make a random choice, and we can normalize the mass to 1, as is usual in probability.

The following figures are images of this representation for the case of the uniform distribution (Fig. 3), the Bernoulli and binomial distributions (Fig. 4) and the Gaussian and Cauchy distribution (Fig. 5). It would be better to give material

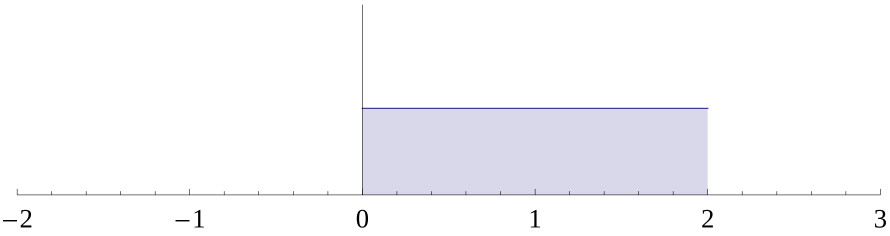

Figure 3. A representation of the uniform distribution on the interval [0, 2].

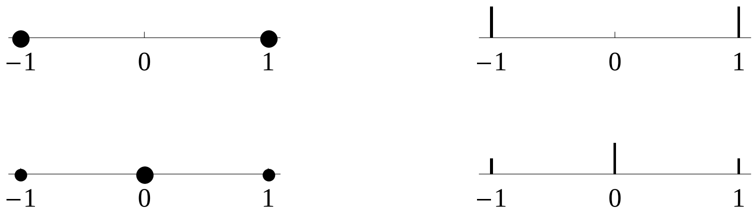

Figure 4. Two representations of the balanced Bernoulli distribution taking values -1 and 1(top) and of the balanced binomial distribution taking values -1, 0, 1 (bottom)

Figure 5. A representation of the Gaussian distribution (left) and the Cauchy distribution (right)

models, and not only figures, during a real course; in this paper, we can only give a hint of the representation.

It takes time to explain this analogy, which can be made rather precise. It is not complete, but its deficiencies illustrate in concrete ways fundamental problems of measure theory and distribution theory, related for example to countability; depending on the audience, one can choose to overlook them, or explain the fine details.

Several concepts follow immediately from this representation.

3.2. *Expected Value and Centre of Mass: the Non-Random Part of a Random Variable*

The probability of a set A, $P(X \in A)$, is the mass of the corresponding part of the bar. Once the representation is understood, the computation to perform is usually clear (although it often requires some analytic skills).

The expected value is just the centre of mass of the bar, and gives in some sense the "nonrandom part" of the random variable. It always exists for a bounded bar, but its existence for a non-bounded bar is not guaranteed: a bar that is infinite in one direction, of width $\frac{1}{1+x^2}$ for $x \geq 0$, has a centre of mass at infinity (there is too much mass very far from the origin, or, in probabilistic terminology, the tail of the distribution is too large), and a bar that is infinite in both directions, of width $\frac{1}{1+x^2}$ for all x, has no centre of mass at all! A consequence is that, if one repeats the experience of randomly choosing a point, and takes the average of the successive experiments, these averages will go to infinity in the first case, and will vary in an erratic and non-bounded way in the second case; this experiment is easy (and very enlightening) to realize on a computer.

By playing with this representation, it is possible to acquire some intuition: it is physically clear that, if the bar is bounded, it has a centre of mass; if the bar is

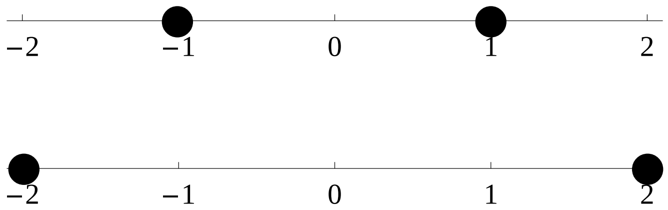

Figure 6. Two balanced Bernoulli distribution with expected value 0; the top one has variance 1, the bottom one has variance 4

infinite, but with a width that decreases very rapidly (for example exponentially), it is natural to consider that the small part of the bar that is far from the origin will play almost no role (an example is the Gaussian law); but if the mass is spread very far along the line, with a large tail for the distribution, it can almost be felt physically that the centre of mass will be ill-defined. Precise statements, of course, can only be made in the framework of calculus; but the model gives meaning to the question of convergence.

As further developments of the theory show, the expected value is the value around which the random variable fluctuates: under suitable hypothesis, the averages of successive independent trials converge in a precise sense to this expected value (laws of large numbers).

3.3. *Variance and Moment of Inertia: Measuring the Randomness of a Random Variable*

If the expected value, or centre of mass, is well-defined, how can we measure the way the random variable is spread around this expected value? How can we measure, by a simple number, the way in which a mass distribution is spread around the centre of gravity? Think of two atoms of mass $\frac{1}{2}$, separated by a distance d, with the centre of mass in the middle. This is the "dumbbell" model of the Bernoulli distribution, see Fig. 6. Try to rotate this dumbbell: the larger the value of d, the more difficult it will be, because the moment of inertia will increase. A little reflection shows that the moment of inertia is indeed a good measure of the repartition, which increases when the mass distribution is spread outward; it is 0 if, and only if, all the mass is concentrated at the centre of mass, in which case there is no randomness.

The variance of a random variable can be defined as the moment of inertia of the bar with respect to the gravity centre, if it exists, and this gives a way to measure the randomness of the distribution: if the moment of inertia is zero, the variable, taking only one value, is not random. One can get a feeling of this moment of inertia, by imagining the metal bar pivoting on an axis through the centre of mass, and the effort needed to make this bar move around the axis; is is obviously 0 if all the mass is in the axis, and large if most of the mass is very far; it is easy to compute for a finite example, like the dumbbell, where it is clearly proportional to the square of the distance separating the two masses (this last sentence needs some knowledge of physics); here, a picture can only convey a small part of the representation, and it would be better in a lecture to have real physical models.

Simple examples, such as the dumbbell (Bernoulli), show that, if the variable is expressed in some unit (for example meter), the variance is expressed in the square of this unit; hence it is natural to consider the square root of the variance, which is called the standard deviation. It is proportional to the variable, and expressed in the same unit.

Playing with different models show that the variance, or preferably the standard deviation, is a first crude measure of the randomness of the variable.

3.4. *An Application: Chebyshev's Inequality*

This inequality is usually stated in the following way:

Theorem (Chebyshev's inequality): *Let X be a random variable with expected value E and variance V. We have: $P(|X - E| \geq d) \leq \frac{V}{d^2}$.*

The proof is generally given as a sequence of inequalities, which are rather obvious, but seem meaningless. It is rarely understood by the students. Let us make this last statement more precise. We may distinguish at least two different modes of understanding: the first one consists in being able to check that a sequence of formulae makes a correct proof. This can be done locally without a global vision of the proof and without any feeling of understanding (this feeling often indicates something important about the deep understanding of the human being; while feeling is not sufficient, we think that it is necessary). To draw a metaphor with computer science and logics, we observe that a program can verify that a given finite sequence of formulas is a correct proof more easily than it can find such a proof, or know whether such a proof exists. The second understanding occurs when there exists a mental representation of the complete proof in the mind of the human; this often comes with a positive feeling. We make a connection between this feeling and a somatic-marker (an hypothesis due to Antonio Damasio) which is a mechanism by which emotional processes can guide decision-making (in this case to decide whether we have understood).

Using the previous model, we can give a sketch of proof for this theorem, which is both intuitive and mathematically sound, as follows.

It is clear that, in a system of masses, if we slide the mass away from the centre of mass, the moment of inertia increases. This is immediately felt by most students, and it is easy to prove from the formula; conversely, if we slide the masses toward the centre of mass, the moment of inertia decreases, an effect well-known by ice-skaters, and easy to experiment physically. Suppose that we have a system of masses such that there is a mass m at distance at least d from the centre of mass, and the rest of the mass at a smaller distance. The minimum of the moment of inertia for such systems will be realized if we slide the mass m at distance d exactly, and slide the rest of the mass to the centre of mass (one can make an exact picture: in this case, we will have two atoms of mass $\frac{m}{2}$ on both sides of the centre of mass, at distance d, and one atom of mass $1 - m$ at the centre of mass). In this case, the moment of inertia is exactly md^2; in any other case, it is greater (see Fig. 7).

We hope that Fig. 5 will help the reader build a visual and kinetic mental representation of the system of masses. For the kinetic part, it would of course be much better to have a physical model; this unfortunately cannot be really done in a paper. If after some time, the reader does not succeed to build a satisfying mental representation, he has the choice to continue his reading without real understanding or to stop his reading. This is a strong argument for providing physical representations to students (and also to our colleagues).

We have "proved" that, as soon as there is at least mass m at distance at least d from the centre of mass, the variance, or moment of inertia, V satisfies $V \geq md^2$, or $m \leq \frac{V}{d^2}$; this is exactly Chebyshev's inequality; the formalization of this "proof" is the usual proof, which appears natural in this setting.

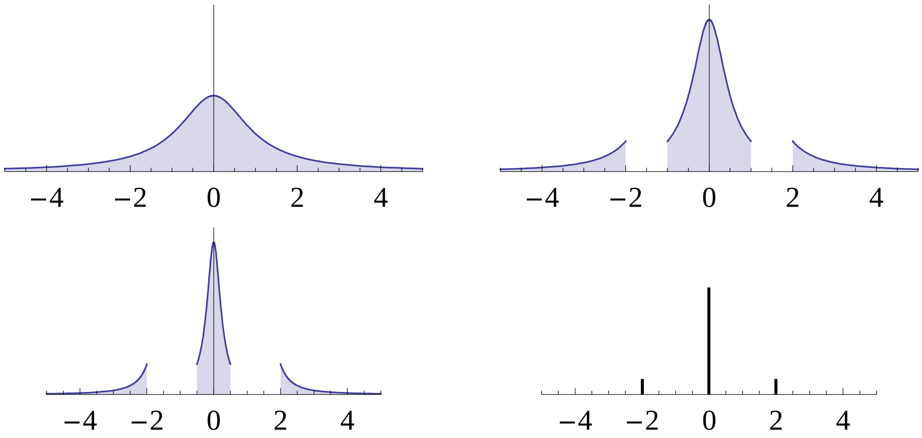

Figure 7. "Proof" of the Chebyshev's inequality: Sliding the mass distribution toward the centre of mass, or the distance $d = 2$, and decreasing the variance (or moment of inertia), with 2 intermediate steps.

3.5. *Further Developments*

This model can be extended to random variables in several dimensions; one must of course start with two dimensions, since most difficulties are already present in this case, yet one can still give a good visual representation, which is difficult in three dimensions, and impossible in higher dimensions.

In this case, the centre of mass is easy to define in the same way as before, but the moment of inertia is no longer a number, it is a tensor. The analogy can only be followed with rather advanced students, but one can then show very interesting phenomena.

A two-dimensional random variable can also be seen as a pair of random variables X, Y on the same underlying space, by taking coordinates, and it defines a distribution of mass in the plane. The tensor of inertia is then defined by a matrix, whose diagonal elements are the variances of the two variables, and the anti-diagonal element (which are equal, since the matrix of inertia is symmetric) are the covariance of the two variables; if this covariance is zero, the variables can be considered as orthogonal, in a sense which can be made precise, in the framework of spaces of square-integrable function (L^2-space).

A special case is that of two independent variables: this means that the distribution of masses is a product distribution, and this easily implies that the matrix of inertia is diagonal and covariance is 0: independence implies orthogonality. It is useful at this point to study a few simple cases: two independent 0-1 variables (head or tail variables), or two independent uniform variables on the unit interval; one can then prove either geometrically or algebraically that the variance of the sum of two independent variables is the sum of the variance.

This is a form of the Pythagoras Theorem: consider the standard deviation as the size of randomness of the random variable, or as a kind of "length" of the random variable, and the variance as the square of the length. Let X and Y be two random variables which are orthogonal, and let Z be their sum $Z = X + Y$. If we consider X, Y, Z as vectors, they form the three sides of a right triangle which admits Z as hypothenuse. We can then expect that the square of the length of Z is the sum of the squares of the lengths of X and Y, and this is indeed the case: it is not difficult to prove that, for two variables with covariance 0 (and in particular, for two independent variables), the variance of the sum is the sum of the variances. Note that in this case, it is not the length (the standard deviation) but the square of the length (the variance) which must be added, and this is far from intuitive if we do not have a geometric picture at hand.

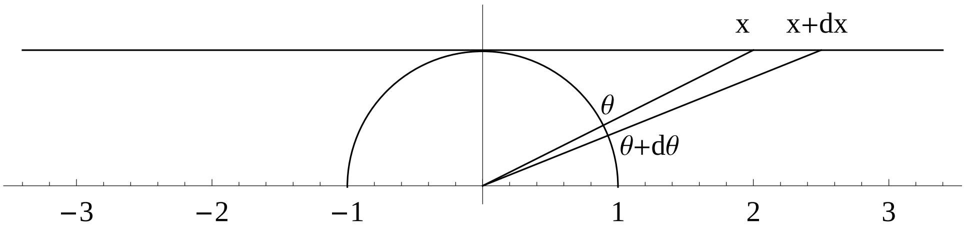

Figure 8. A geometric model for Cauchy density

Students often ask why we define the variance by taking the square of the distance to the expected value, and not the absolute value, which might also give an interesting quantity (it is the L^1 norm for integrable functions), and is often easier to compute; this property of additivity is probably the best answer that can be made: no other quantity would satisfy the Pythagoras theorem, and this leads to deep consequences (the first one being an easy proof of the weak law of large numbers, which justifies the initial definition of the expected value).

3.6. *A Physical Model for Cauchy Density*

Here is a nice physical model for the classical Cauchy density, which is often used as a counter-example.

Consider a laser beam which is pivoting on an axis at the origin of the plane, and shoots in a random direction. Take a horizontal wall at unit distance, of equation $y = 1$. The laser beams marks a random point on this wall; it is a small computation (which makes an interesting geometry exercise) to prove that the random variable so obtained has density $\frac{1}{\pi}\frac{1}{1+x^2}$. The basic idea is to prove, using the notations of Fig. 8, that for given x, θ, when dx tends to 0, we have $d\theta \approx \frac{dx}{1+x^2}$. As we remarked above, this random variable has no expected value; it is easy to make computer experiments, by choosing random numbers in the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ and taking their tangent, for example to test the behaviour of the average.

3.7. *Limits of the Model*

This mechanical model has some limitations. First of all, it becomes quickly necessary to give explicit analytic formulas (and this must be one of the aims of a probability course), and to go beyond the figures.

In this case, the simple model with atoms becomes insufficient, as it can only model finite probabilities; but this can be a good opportunity to introduce a number of questions on the foundations of probability and analysis, up to and including measure theory and distributions, if the audience allows it.

It is unclear to us how well this model motivates the important questions of convergence (which could be left aside in a first presentation); however, the model for the Cauchy density given above could motivate the question (and some models without variance for the income distribution could be used to show its practical interest).

More importantly, this presentation, at the basic level, can make sense for a large public; but it will be much easier to grasp for students who have some knowledge of mechanics (centre of mass, moment of inertia, and their computation in some simple situations), and who are able to recognize some formulae they already know.

This is part of a more general fact: intuition can only be built on the foundation of some previous knowledge, informal or formal, and in this sense, it is wrong to oppose learning and understanding: intuition of sophisticated phenomena generally needs some technical culture.

4. Conclusion

We think that understanding a mathematical notion means building an internal (mental) representation which must be sufficiently adapted to the object. We build and internally manipulate these representations (and also the representations from the representations and so on). We believe that the first effective representations are, in general, concrete (visual, kinetic, auditive,...) and they allow one to build other more sophisticated symbolic representations from them. Concrete mental representations (inspired by the real physical world) are interesting for all sciences, including pure and theoretical mathematics and physics. If the thoughts of children are based on a simple and naive physical world (even though we know now that this world does not really exist !), the thoughts of students and colleagues are also based on the internalization of the physical world enriched by more and more symbolically sophisticated pieces of knowledge. Our opinion is that the more we are able to build mental representations, the more we are able to think effectively as student or researcher.

It is however non-trivial to build an effective mental representation. It would be valuable to gather many different mental representations, and to test their effectiveness, the different parameters that play a role in the way students receive them, and the conditions under which they can be useful; this is the effort we started at a small scale.

Acknowlegments: We thank the anonymous referees and our colleagues Stéphane Demri and Jean Goubault-Larrecq for their careful reading of the paper and for their suggestions which allowed us to improve the quality of the paper.

References

[1] M. Denis, E. Mellet, and S. Kosslyn, *Neuroimaging of Mental Imagery*, Psychology Press (2004).
[2] ———, *Neuroimaging of Mental Imagery: An introduction*, European Journal of Cognitive Psychology 16 (2004), pp. 625–630.
[3] M.F. Ehrlich, H. Tardieu, and M. Cavazza, *Les modèles mentaux. Approche cognitive des représentations*, Masson (1993).
[4] N.J. Thomas, *The Stanford Encyclopedia of Philosophy, Mental imagery*, (2008).
[5] J.P. Changeux, *L'homme neuronal*, Fayard (1983).
[6] T. Needham, *Visual complex analysis*, Oxford University Press (1997).
[7] J. Hadamard, *Psychology of Invention in the Mathematical Field*, Princeton Press (1945).
[8] G.A. Goldin and C. Janvier, *Representations and the Psychology of Mathematics Education*, Journal of Mathematical Behavior 17 (1998).
[9] G.A. Goldin, *Representational Systems, Learning, and Problem Solving in Mathematics*, Journal of Mathematical Behavior 17 (1998), pp. 137–165.
[10] M. Carlson, M. Oehrtman, and P. Thompson, *Foundational Reasoning Abilities that Promote Coherence in Students Understanding of Function*, in *Making the Connection: Research and Teaching in Undergraduate Mathematics* MAA, 2007, pp. 150–171.
[11] J.E. Corter and D.C. Zahner, *Use of External Visual Representations in Probability Problem Solving*, Statistics Education Research Journal, International Association for Statistical Education (IASE/ISI) 6 (2007), pp. 22–50.
[12] P. Thompson, *Imagery and the Development of Mathematical Reasoning*, in *Theories of learning mathematics* Hillsdale, NJ: Erlbaum, 1996, pp. 267–283.
[13] A.L. Mesquita, *On Conceptual Obstacles Linked with External Representation in Geometry*, Journal of Mathematical Behavior 17 (1998), pp. 183–195.