



Variability: External Data, Internal Apprehension, or a Balance between Them? A Mathematical Argument and Some Upshots

Éric Brian*

Centre Maurice-Halbwachs (CMH), EHESS, ENS, CNRS, PSL*, Research University, 75014 Paris, France

Abstract

This paper addresses how one apprehends variability. Statistics does provide estimators like variance and standard error. After a recall of usual formulae, the already known partial variance is set as a strict axiomatic extension. This setting is the technical originality of the paper. This extended variance is discussed and used to distinguish the usual measure of variability and conditions of some warped apprehensions. It opens a window for various possible operators of probabilization on the space of possible proofs and couple of proofs taken into account or not. Such operators, in the calculus, may absorb part of the usual variability, fitting or not with possible subjective apprehensions. A game is given as an example.

So where does the sense of variability come from? From uniform and systematic comparisons? From partial apprehension of the phenomenon? From rules of computation? And what if these mathematical operators were the analytic *locus* for the habituation to stable environmental variability (environment here being considered in the Darwinian sense)? In addition, the other way around, may we interpret them as projectors of interior sensitiveness onto the environment? In conclusion the proposal made here is quickly compared, on the one and with some foundational elements proper to general sociology after É Durkheim or P Bourdieu and on the other hand with A Berthoz' argument on simplicity in adaptation.

Keywords

Variance, Social coercion, Habitus, Simplicity

Usual Mean and Variance of Probable Events

This paper is in no way an attempt to express a synthesis of the last decades of works on stochastics in biological and social sciences however, something is there to be done. Meanwhile the reader may get a panorama of quite half a century of formation of such a new social morphology inaugurated by Prigogine and Nicolis [1] and developed since then for instance toward Perunov, Marsland and England [2], with the overview of methods offered in Gardiner [3]. The aim here is to address the routinized computing and understanding of variance and variability. The aim is to identify points where the issue of variability may be rephrased, and re-articulated with some foundational problematics in Sociology and Cognitive Science, offering by the way a path between these two domains based on some of their own conceptual keys.

Since the 17th Century, authors usually introduce

probabilities considering a countable number of possible events, let us call them here, e_i , for i going from 1 to N . The geometer Leonhard Euler (1707-1783), commenting in December 1776 a paper given by Daniel Bernoulli (1700-1782) to the Saint Petersburg Academy of sciences, proposed to estimate an unknown value computing known observations in a manner we call today a weighted arithmetic mean, Euler [4]. This was of crucial interest in astronomy. He called his weight the “goodness” of each observation, and the two memoirs published in 1778 have open a European discussion for a few decades about the

***Corresponding author:** Éric Brian, Centre Maurice-Halbwachs, 48 boulevard Jourdan, F-75014, Paris, France, E-mail: eric.brian@ens.fr

Received: January 31, 2017; **Accepted:** July 13, 2017; **Published online:** July 15, 2017

Citation: Brian e (2017) Variability: External Data, Internal Apprehension, or a Balance between Them? A Mathematical Argument and Some Upshots. Ann Cogn Sci 1(2):28-38

distribution of this so called “goodness”. At the formal level, the calculus was analogous to the one of mathematical expectation as shaped by Blaise Pascal (1623-1662), Christiaan Huygens (1629-1695) and Jakob Bernoulli (1654-1705): if one considers a scalar value dependant upon probable events, therefore later called a random variable, the mathematical expectation $E(X)$ for a probabilization p is the average of the possible values attributed to events weighted by the probability of these events, Dodge [5].

Let e_i for $i = 1$ to N be a countable number of events, $X(e_i)$ be a random variable on this set of events, and $p(e_i)$ for $i = 1$ to N their respective probabilities.

$$E(X) = \sum_{i=1}^N X(e_i) \cdot p(e_i) = E_p(X) \quad (1)$$

If all $p(e_i)$ are equal and if their sum is 1 (all possible events together is taken for certainty), then each $p(e_i)$ is equal to $\frac{1}{N}$. The previous formula becomes:

$$E(X) = \frac{1}{N} \sum_{i=1}^N X(e_i) = E_{\frac{1}{N}}(X) \quad (2)$$

The formulae (1) or (2), the usual definition of *mathematical expectation*, are a lot more than a definition, it is a *schema*, criticized by the geometer Jean Le Rond D’Alembert (1717-1783) in *L’Encyclopédie* and rehabilitated by Condorcet (1743-1794), one of his followers at the Paris Academy of sciences. I have studied their positions in Brian [6, 7]. Geometer and metaphysician of calculus as it was said at that time Condorcet was very explicite in his title: «Reflexions on the general rule that prescribes to take as the value for an uncertain event, the probability of this event multiplied by the value of the event itself». Under this rule, (1) means that mathematical expectation is the sum of all unknown possible values for a given random variable. In the following pages, I’ll play with generalizations of schema (1) in order to reach the announced discussion.

Writing like Fréchet (1878-1973) and Halbwachs (1877-1945) in their wonderful introduction to the calculus of probabilities: mathematical expectation (1) is a central value, a centre, one among the possible central estimators mathematicians have built in order to characterize a random variable, as far as the calculus is possible from a mathematical standpoint, such a condition being always satisfied for a finite number of cases.

Usual writing of variance

Following again Fréchet (1878-1973) and Halbwachs [8], the introduction of variance is intuitive. Once a central value established, one other interesting parameter is an estimator of the variability around this centre, something like a diameter or half of it. Here, for a

variable, randomness is analogous to being somewhere in a sphere. With these geometrical notions, necessarily comes a metric. For each event, one can consider the square of the *discrepancy*, the deviation between the value of X and the center $E(X)$. The expectation of this new random variable comes again from schema (1). It is the definition of the *variance*, $Var(X)$. The standard deviation, σ , is the square root of this variance, for reasons of dimensional homogeneity with $E(X)$:

$$Var(X) = \sigma^2 = \sum_{i=1}^N [X(e_i) - E(X)]^2 \times p(e_i) = \frac{\sum_{i=1}^N [X(e_i) - E(X)]^2}{N} \quad (3)$$

Rewriting variance considering couples of events

Let us now consider all couples of possible events (e_i, e_j) and the random variable ΔX defined on these couples $\Delta X(e_i, e_j) = [X(e_i) - X(e_j)]$, the squared deviation between two random values of X is then $\Delta X(e_i, e_j)^2 = [X(e_i) - X(e_j)]^2$. This is a random variable too but based on the product set of events (the set of couples). The number of possible couples could be taken as N^2 , but for each couple (e_i, e_j) , the couple (e_j, e_i) gives the same $\Delta(e_i, e_j)^2$. So, let us consider as well $E_{\frac{1}{2N^2}} \Delta X^2$, the raw moment for all *distinguishable* possible squared deviations without double counting, following schema (1) or (2). It is:

$$E_{\frac{1}{2N^2}}(\Delta^2) = \frac{\sum_{i=1, j=1}^{i=N, j=N} [X(e_i) - X(e_j)]^2}{2N^2}$$

$$E_{\frac{1}{2N^2}}(\Delta^2) = \frac{\sum_{i=1, j=1}^{i=N, j=N} [X(e_i) - \overbrace{E(X) + E(X)}^{=0} - X(e_j)]^2}{2N^2}$$

Hence $E_{\frac{1}{2N^2}}(\Delta^2)$ is the addition of three terms:

$$\frac{\sum_{i=1}^N [X(e_i) - E(X)]^2}{2N}$$

$$\frac{\sum_{j=1}^N [E(X) - X(e_j)]^2}{2N}$$

$$\frac{\sum_{i=1, j=1}^{N, N} [X(e_i) - E(X)][E(X) - X(e_j)]}{N^2}$$

Each of the two first terms is $\frac{1}{2} Var(X)$.

And in the last term, for a countable number of events, one can carry out the two sums one after the other. But by definition $\sum_{i=1, N} [X(e_i) - E(X)] = 0$. Therefore, this last term is null and:

$$Var_{\frac{1}{N}}(X) = E_{\frac{1}{2N^2}}(\Delta^2) \quad (4)$$

We have seen that the classic definition of variance (3) fits the average schema (1) or (2) for squared deviation between the probable values $X(e_i)$ and the mathematical

expectation $E(X)$ and that the definition of mutual deviation Δ on the set of couples (e_i, e_j) brings us, following schema (1) or (2) again, towards $E_{1/2N^2}(\Delta^2)$ the raw moment (the average) of all *distinguishable* squared deviations between two possible probable values. The previous proof, and its conclusion, formula (4), may be phrased as a theorem: for a countable number of cases the second central moment of X , its variance $Var(X)$, equal the raw moment of all distinguishable squared deviations.

A known application: partial variance: In other words (4) is a rewriting of variance and a second schema. Its advantage it to cut with routinized measures and uses of variances, habits that squeeze the room where new things may be caught. So, its interest lies at the conceptual level, where one can catch that this is only a basic average on a product set. The rewriting is not new: Lebart and Banet [9] have proposed to consider this property in order to define notions of partial variance and covariance, mentioning earlier similar proposals. Later Lebart, Morineau, *et al.* [10] have integrated partial variance defined with matrix among the options offered in the software SPAD™ today used for big data analysis. More recently Baba, Shibata and Sibuya [11] proposed to use partial correlation for measuring conditional independence. And even more recently Larson, Sonenberg and Nadon [12] have shown that analysis of partial variance improves genomic explorations.

Around thirty years ago, the mathematical context was computer multivariate statistics and its algebraic background. Lebart and Banet considered the probabilization of couples of possible events or statistical individuals or cases [9]. In this context there is a $N \times N$ matrix of weights attributed to each couple of events to be taken or not into account. For instance, for geographical data, this matrix of weights may record the contiguity of cases ei and ej : a 0 may mark the absence of contiguity between two cases and the corresponding deviation Δ will not be taken into account. The empirical advantage of the technique in computer statistics is clear: computing the variance for all (i, j) couples provide a general variance, based on a systematic and uniform comparison of all couples of possible events. Employing one another probabilization matrix of couples will give a measure of the part of variability due to couples actually taken into account, etc.

Some light on a known alteration of a spontaneous calculus: Usually the *standard deviation* $\sigma = \sqrt{Var(X)}$ is taken as the second parameter after the central value. Square rooted to the same dimension. σ is not a deterministic number but a random variable. The first prompting idea for estimating σ is to use:

$$S = \frac{\sqrt[2]{\sum_{i=1}^N [X(e_i) - E(X)]^2}}{(N)}$$

In his statistical encyclopedia Dodge [6] recalls, p. 157, that finding for σ a fine estimator has not been an evidence. At the beginning of the 20th Century, two British statisticians K Pearson (1857-1936) and WS Gosset (1876-1937) (known as STUDENT) have observed that S was biased. Namely: $E(S^2)$ is not $Var(X)$. The statisticians arrived later to a non biased estimator S' , so that $(E(S'^2) = Var(X))$:

$$S' = \frac{\sqrt[2]{\sum_{i=1}^N [X(e_i) - E(X)]^2}}{(N-1)}$$

Justifications of this opportune $(N-1)$ are running in statistical textbooks. Generally they consist in observing that $E(S'^2)$ is fine. But with the above rewriting, there is a simple analytic argument: there are N couples (i, i) where $X(e_i) = X(e_i)$ makes $\Delta(e_i, e_i) = 0$, and they count for nothing in the sum in the expectation of Δ . Then the probabilization on the product space of couples should not be $1/N \times N$ but $1/N \times (N-1)$. In the previous proof, the first sums are not inducing $N/N^2 = 1/N$ but $N/N \times (N-1) = 1/(N-1)$ (QED).

For a countable number of events, the rewritten variance is not a discovery. We have spent some paragraphs to detail schema (1) and (3) and what we have expressed as theorem (4). In addition we have mentioned some already known applications of this way of understanding variance. These preliminaries will be necessary for strengthening our intuition when moving to the next steps that will transgress routinized notions about variability.

For General Sets of Events and Proofs

Real life is not made of countable sets of events analogous to a game, but of apprehended proofed events. Thinking outside the classic analogy with games must be revised. In addition, if working on a probabilistic mode on general sets of events was only pushing to infinite limits results known for countable games, it would have been acknowledged for a while: on the one hand it was already a point for mathematical and metaphysical disagreements between Condorcet and Laplace (1749-1827), Brian [6], and on the other hand tracking erroneous reasoning based on this abuse has been a motor for mathematical theoreticians of probabilities since then, see for instance Hald, Stigler, Porter ^a [13-15].

Definitions

We must then built a probabilization where X , $E(X)$ and $Var(X)$ will be well established. To do so I will fol-

^a Between Laplace and Condorcet, there was too a divergence in the understanding of the so inspiring memoir and reasoning attributed to Thomas Bayes - this divergence is analysed in Brian [6]. The extension of variance proposed here does not affect bayesian calculus.

low the nowadays widely accepted axiomatic recapitulated in Neveu [16]. This synthesis of 20th foundations of probabilities fits Fréchet's [17] epistemological realism. We consider the non empty set Ω of proofs that realize possible events. Ω , \emptyset , the operators \cap (intersection) and \cup (union) are defining a boolean algebra of parts of Ω . It will be noted A . Each part of Ω - a subset of proofs, is noted here A or B is included in the certain whole Ω is an element of the set of events A .

Concretely Ω is the set of all possible proofs. As such it is certain. If there is no proof at all this is void. A as a subset of proofs is a event that may happen or not, among all possible parts of all possibilities. We are interested in the probability of objects like A or B . For the first centuries of the calculus of probabilities, mathematicians were obsessed by cutting unity in partial probabilities attributed to each possible event. This has been the principle of the first pages here. With the axiomatic formulation, the set of possible proofs is partitioned in events and each part - each event - will get its probability. There is here a fundamental switch from decomposing the measure of certainty to decomposing the set of proofs. For operating this change, a theoretical approach of sets was necessary. This is why; it is typical of the 20th Century. A being a part of Ω it is one element of A . From an epistemological standpoint today, or from the one of metaphysics of calculus to use this ancient expression, the distinction between proofs and events is crucial. One apprehends proofs but not immediately events. Events are apprehended through proofs that encompass them.

A *probabilization* dP of the set of events A is a function from A to $[0, 1]$, so that $\int_A dP(A) = 1$ (this is a normalisation so that the probability of all possible events taken together is certainty), and $\int_{\emptyset} dP(A) = 0$ (here the probability of nothing is null). With some other properties making dP compatible with the algebraic structure of A . They are given by Neveu [16] as two centuries earlier they were sketched by Condorcet [18] using a less powerful language.

Here comes now the general definition of Neveu [16]. A *probabilized space* (Ω, A, dP) is made of a given non empty set Ω (the set of proofs), a boolean algebra of parts of Ω (A the set of events) and a probability dP defined from A to $[0, 1]$. The reader may guess the reasons why I am now introducing the probabilized product space $(\Omega \times \Omega, A \times A, dQ)$, leaving open for a while its probabilization Dq ^b.

For X , a random variable defined from $A \times A$ to \mathbb{R} , one

^bThe mathematician M. Barbut, after reading Brian [19] told the author that he considered the definition of Gini's index (1912) fits this generalization.

can define its deviation ΔX , the random variable from the product space $A \times A$ to \mathbb{R} as $\Delta X(A, B) = [X(A) - X(B)]$. Looking for the homologous of formulae (1) and (3), the definitions of mathematical expectation and variance, we can consider the raw moment of X as far as its integral may be established, a property expressed by $X \in L^1(\Omega, A, dP)$ in Hilbertian terms, then :

$$E(X) = E_{dP}(X) = \int_A X(A) dP(A) \quad (5)$$

For the second moment, we need an homologous condition of integrability on squares *discrepancies* $\because [X - E(X)] \in L^2(\Omega, A, dP)$, in similar terms.

$$Var_{dP}(X) = E_{dP}[X(A) - E(X)]^2 = \int_A [X(A) - E(X)]^2 dP(A) \quad (6)$$

We have now two candidate probabilizations for the product space: an arbitrary one here noted dQ , and the simple product $dP \times dP$ or dP^2 defined as $dP^2(A, B) = dP(A) \times dP(B)$. For the arbitrary case dQ , its *consistency* with dP may be questioned: they are *consistent* if the marginal probabilities on the product space are equal to the probabilities on the original space:

$$dP(A) = \int_{B \in A} dQ(A, B) = \int_{B \in A} dQ(B, A) \quad (7)$$

Back to the rewritten variance

As in section (1), the *extended variance* on $(\Omega \times \Omega, A \times A, dQ)$ will be:

$$ExtVar_{dQ}(\Delta X^2) = \frac{1}{2} \iint_{A \times A} \Delta X(A, B)^2 dQ(A, B) \quad (8)$$

In Brian [19] we have used these definitions and we arrived to the classical statistical toolbox of covariance, correlation, regression analysis, independence between random variables, and inequality named after Bienaymé-Tchebichev ("BTI"). This inequality insures a convenient convergence of discrepancies - the deviations between $X(A)$ and $E_{dP}(X)$ - towards 0. The French and the Russian mathematicians have established in parallel (in 1853 and 1867) in strict mathematical terms for countable number of cases the empirical property expressed by Jakob Bernoulli [20]: the relative frequency of the repetition of identical proofs converges towards its probability (for instance play a great number of times with non piped dices, the frequency to get one given face will be closer and closer to 1/6). This "law of large numbers" knew several different formulations all along three centuries. Classic BTI is as phrased as: for a variable with a finite variance, the probability to observe a deviation between the variable and its expectation reduces as the inverse of this deviation squared. Extended BTI may be phrased as such: for a variable with a finite extended variance, the probability to observe a deviation between the

^cIn French : «les «écarts au centre»».

two values of this random variable reduces as the inverse of this deviation squared. It is not the place here to develop these formal arguments, and their formulation thirty years ago, was may be too much austere and somehow unclear ^d.

The theorem proved in section (1) should be examined again. We consider:

$$X \in L^1(\Omega, A, dP)$$

$$[X - E(X)]^2 \in L^1(\Omega, A, dP)$$

$$\Delta X \in L^2(\Omega \times \Omega, A \times A, dQ)$$

For the extended variance, we have, according to its definitions (8):

$$ExtVar_{dQ} X = \frac{1}{2} E_{dQ}(\Delta X)^2 = \frac{1}{2} \iint_{A \times A} \Delta X(A, B)^2 dQ(A, B)$$

$$ExtVar_{dQ} X = \frac{1}{2} \iint_{A \times A} \left[X(A) - \overbrace{E_{dP}(X) + E_{dP}(X)}^= - X(B) \right]^2 dQ(A, B)$$

$$ExtVar_{dQ} X = \frac{1}{2} \iint_{A \times A} [X(A) - E_{dP}(X)]^2 dQ(A, B)$$

$$+ \frac{1}{2} \iint_{A \times A} [X(B) - E_{dP}(X)]^2 dQ(A, B)$$

$$- \iint_{A \times A} [X(A) - E_{dP}(X)][E_{dP}(X) - X(B)] dQ(A, B)$$

If dQ and dP are coherent as defined in (7), then the two first terms are twice:

$$\frac{1}{2} \int_{A \in A} [X(A) - E_{dP}(X)]^2 dP(A) = Var_{dP}(X)$$

Then:

$$ExtVar_{dQ} X = Var_{dP}(X) - \iint (X(A) - E_{dP}(X))(E_{dP}(X) - X(B)) dQ(A, B)$$

Probabilization as a Play Ground

For an integrable random value X (L^1 and L^2), the extended variance, the expectation of squared mutual deviations, is equal to the classical variance with a correction based on the probabilizations of the set of proofs, and the one of the product set of compared couples of proofs. This correction is a term I will call *warp*, coming from the distinction between the squared probabilizations on the original probabilistic space of single proofs on the one hand, and of the actual probabilisation on the product space of couples of proofs on the other hand. Hence we have definition (9) and then theorem (10), equivalent of (4), above.

^d One another mathematician I am not thanking here, member of the jury in 1986 said at the defense something like: «where is the use of all this work?». At that time, my motives were in the sociological points I'll comment at the end of the paper. I left the early mathematical arguments untouched. Recent discussions touching cognitive science have revived them. Here I am reducing the early memoir to what will be actually necessary for the final arguments, clarifying the mathematical proofs, and paraphrasing them in intuitive terms.

$$Warp_{dQdP}(X) = \iint [E_{dP}(X) - X(A)] \times [E_{dP}(X) - X(B)] dQ(A, B) \quad (9)$$

It is clear that if $dQ(A, B) = dP(A) \times dP(B)$, $Warp_{dQ} dP(X)$ is null, there is no warp at the concrete meaning of the term. Decomposing the extended variance or the classical variance according to formula (10) gives one term coming from the systematic comparison of all possible proofs, and one another coming from an arbitrary alteration of the set of possible couples to be taken into account, together with a measure of the alteration due to the distortion between the previous two.

$$ExtVar_{dQ} X = Var_{dP}(X) + Warp_{dQdP}(X) \quad (10)$$

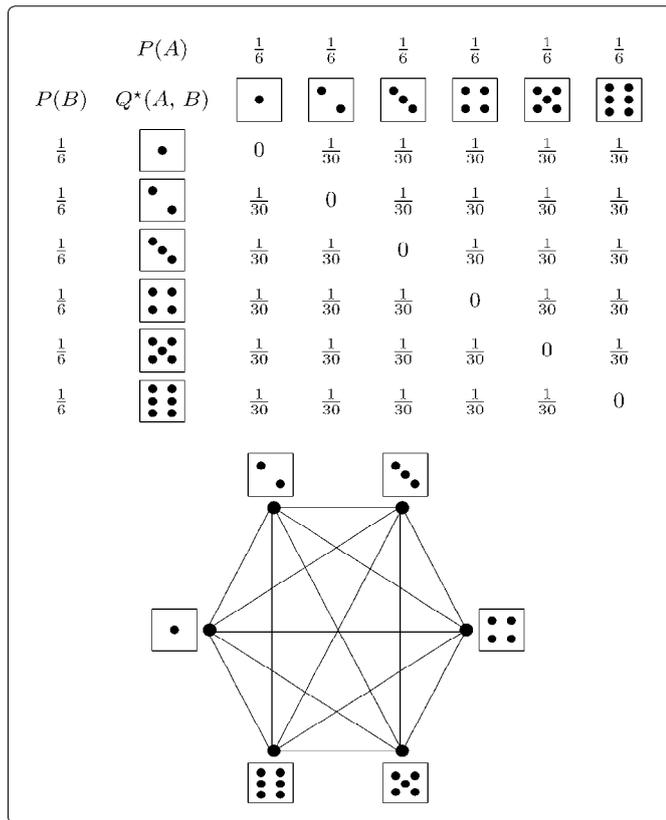
From an empirical standpoint the distinction between the two contributions to variability, the one due to systematic comparisons and the one due to some filtration on the couples to be compared is not easy. Usually statisticians use the systematic comparison even if the phenomenon itself encompasses some filter in its apprehension, this is the precise point where the discussion in sociology or cognitive science begins to be relevant. The previous abstract construction may be easily transposed in computer statistics (see the above mentioned example of statistical software). As a consequence, this mathematical frame work and its corresponding computing application give an area for experimenting extended variances and warps. In this paper the point is to introduce the concept of warp and to show its potential impacts. An example based on a fictional game may now be helpful.

Six Dices, Three Variances and Three Primes

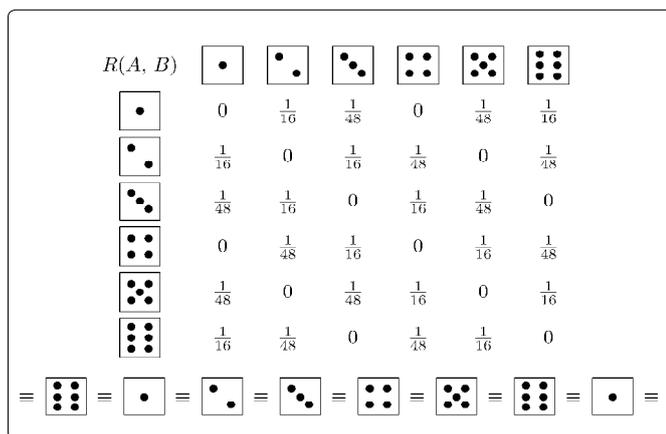
Three variances

For the sake of argument there will be here three different probabilizations of the product space for a set of six equiprobable cases and then three variances of the dQ type. They will all be compatible with the equiprobability of the six faces of a non piped dice: $dP = 1/6$. In the following definition matrixes all diagonals of dQ will be null: the proofs like (A, A) making $\Delta X = 0$ are not taken into account.

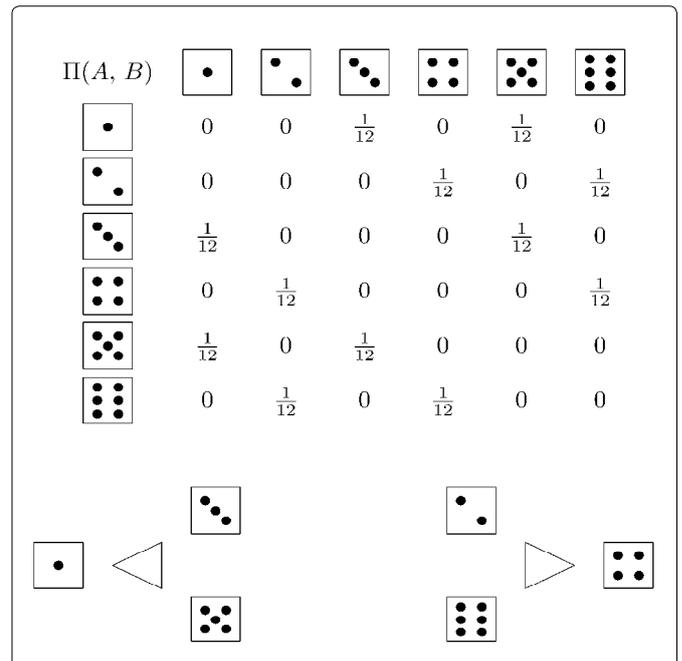
Systematic variance: It could have been $P^2 = 1/36$, but we have seen that $1/N(N-1)$ is better. So the most systematic and equiprobable account of all possible couples of proofs is given by $Q^* = 1/30$. After the matrix built as the next ones on the exposed rules that govern the weight of each couple in the calculus on the schematized network each segment represents a couple taken into account.



A ring type of variance: For the second considered extended variance $ExtVar_R$ the most probable couples taken into account are neighbours and, with a lower weight, next neighbours. Other couples are not computed. It is like computing variance following a ring.



Variance on a discriminated space: For the third variance $ExtVar_{II}$ couples are taken into account if the proofs of the dices display the same parity. Couples of unlike parities are excluded. The variance here is based on two distinct networks of proofs sketched according to the following matrix.



Three primes

Let us consider the following primes, X_1 , X_2 and X_3 defined as random variable on the set of proofs A and the probabilized space (Ω, A, dP) .

$P(A)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
Variables						
X_1	-3	1	2	-2	-1	3
X_2	2	3	1	-2	-3	-1
X_3	-3	1	-1	3	-2	2

The mathematical expectation of each random variable is null, and its classical variance for the systematic probabilisation of the product space is always 5.6. The extended variances and the warps for probabilization Π are given in the next table, and for the ring probabilization R in the second next:

Variables	$Var_{Q^*}(X)$	$ExtVar_{II}(X)$	$Warp_{\Pi dP}(X)$	$\frac{ExtVar_{II}}{Var_{Q^*}}$	$\frac{Warp_{\Pi dP}}{Var_{Q^*}}$
X_1	5.60	6.33	-0.73	113%	-13%
X_2	5.60	7.00	-1.40	125%	-25%
X_3	5.60	1.00	+4.60	18%	82%

Variables	$E_p(X)$	$ExtVar_R(X)$	$Warp_{R dP}(X)$	$\frac{Var_R}{Var_{Q^*}}$	$\frac{Warp_{R dP}}{Var_{Q^*}}$
X_1	0.00	6.96	-1.36	124%	-24%
X_2	0.00	3.50	2.10	63%	37%
X_3	0.00	6.62	-1.02	118%	-18%

Analysing the three primes

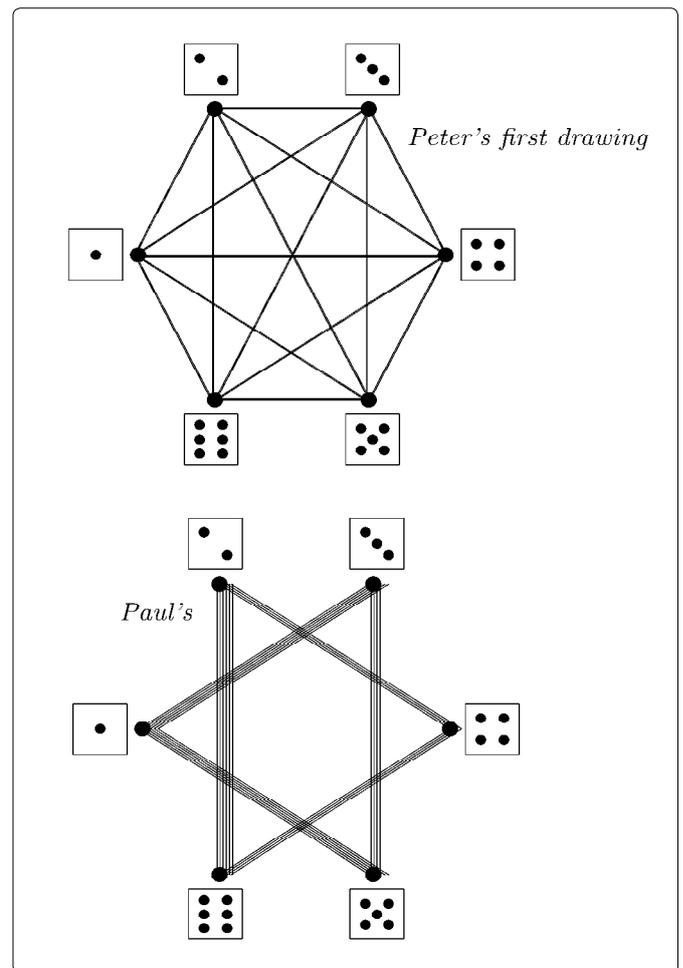
At the first glance, according to the two usual first moments, the three random variables are similar. Dices

are not piped. The three primes, the variables, could appear equivalent. But the implicate rules of systematic statistical computing, *VarP* and *ExtVarQ**, are challenged by the use of *ExtVarII* and *ExtVarR*. For instance the values of X_3 are always negative for odd proofs and positive for even proofs. Using *II*, 82% of the systematic variance is to be attributed to a warp due to parity. As an interpretation one may say that the extended variance based on *II* absorbs only 18% of the systematic variance of the primes, which appears as a combination of toss (even vs. odd) and dice, the first making around the four fifth of the systematic variability.

If a statistician would have compute the extended variance of X_2 with *R*, walking along the ring he would have obtained the lowest indicator of variability. He would have apprehended the probable space as stable, at least less variable as if he would have systematically compute the variance according to textbooks. Hence, probabilizations of the product space of squares proofs operate as a statistical frames of reference. With X_1 one sees that the standard variance is lower than the two others. Would the statistician work in financial business and follow Markowitz [21], he would expect to increase profits by way of variance minimization. Doing so, he would have been abused by the usual calculus and confronted to more volatile phenomena. For some decades there were similar mishaps in this domain for a few mathematical reasons, some other being discussed in Brian [22].

A fable based on variances

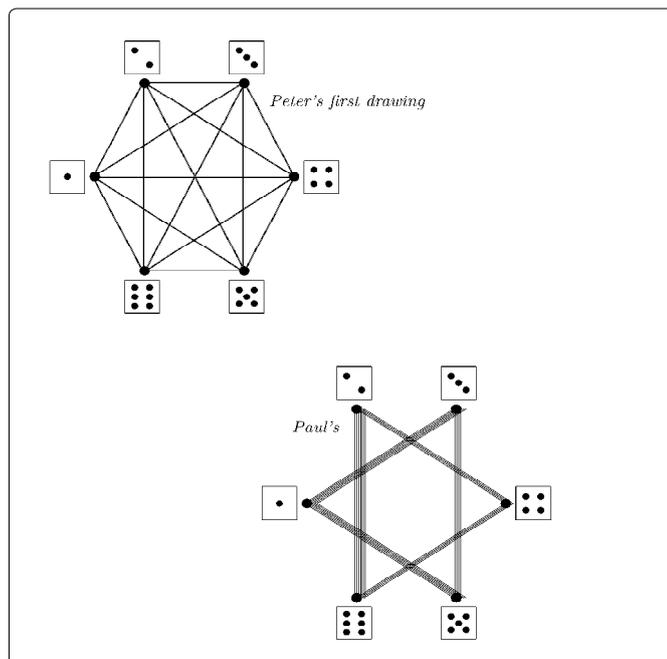
This fable may help the reader to be more familiar with probabilization switches. Peter and Paul have been protagonists of probabilist games for three centuries. Once they were in a mysterious place, both with the same amount of ducats in hands, walking separately from one room to the other. From time to time they met, but kept silence all along their thirty first moves, according to given instructions. The master of the game said that they would exit the place without any loss or gain. At each step each player arrived in front of doors, marked with one face of dice, where he could receive or pay a few ducats according to a permanent posted tariff. This transaction done he was authorized to go ahead towards a next door. The challenge was to draw the map of the labyrinth. All stations were alike. The door from which the player came had no handle and five other were open at the tariffed conditions. Peter decided to keep notes step by step. After being convinced that there were only six different doors, he experienced the thirty possible corridors between them. Properly speaking at the long run, he observed that the amount of money in his hands was more or less stable. Having been trained as a statistician, he computed the variance of one step



cost; it was very close to 5.6. Paul, taking notes too, by superstition or arbitrariness decided to accept the step transactions only if the new door was marked with a dice of the same parity as the previous one. To some point he got the idea that there were two analogous circuits each between three doors. After thirty moves, twelve between even doors, and eighteen between odd doors, the remaining sum in his hands was quite the one of the beginning. Paul stressed by the variability computed a variance of quite 7.0. Being both somehow exhausted after thirty runs from one door to the other with no gain, they decided to exchange their observations and results. The difference between the variances was striking. They decided to play one another round but both choosing at random the door to go. Recording and computing again they both arrived to results close to Peter's first one. They attributed the over volatility of Paul's first observations to his initial arbitrary method. After a brief discussion they compared their two first drawings where each line was a move between two doors (they are reproduced above). Even if they first disagreed, Paul arguing that he walked around a small hexagonal tower and Peter that he crossed a wide hexagonal perimeter, they concluded that their results were compatible from an architectural standpoint as if that they were both in corridors built

inside some defensive tower of the 17th-18th Centuries, may be some room in a tower at one extremity of Peter's and Paul's fortress in Saint-Petersburg.

In this fable the tariff is given by X_2 as previously defined. Peter walked in the labyrinth following the graph of $ExtVar_{Q^*}$. Paul arbitrarily followed $ExtVar_{IT}$. Before exchanging their results Peter walked alone in a relatively smooth labyrinth, and Paul facing an higher level of variability. As far as an experimenter collects comparative proofs according to the same frame of reference nothing can tell him he is actually in a standard space or in one fitting an unknown extended variance definition. In both cases, there is variability, higher or lower. To go beyond the frame, the warp must be improved. If now we imagine Peter and Paul in a business company tower playing with charts, and driven by Markowitz' principle of minimal variance, they would compare their results and prefer Peter's computing, if not $ExtVar_R$ giving 3.5 when here the yield stays at 0 whatever the computing rules for variability could be.



Conclusions

A brief historical remark

The reader may ask why the standard rules for computing variability are so strong. It comes from history of statistics. During the first three fourths of the 19th Century continental European statisticians were focussing on the average, the mathematical expectation. With the British school (Galton, Pearson, *etc.*) discrepancies became a major focus, with the average of their squares. Here came the variance and the standard deviation. In addition the two parameters of expectation and standard deviation were enough to feed the widely admitted law of dispersion drawn by Laplace and Gauss. So, in a few words, the predilection for usual variance comes from

the history of the commodities in calculus. Following the methodological debates between mathematicians, moral statisticians, and naturalists during the 19th Century, observing as an historian the uses of statistics in averaging observations, in defining types, and in evaluating variations, one can show that three different types of variability have been diversely confused (or taken one for the other) during this century: empirical variations, errors in measurements, variability due to randomness. Routinized statistical teaching generally provides a selected compilation of technical tricks. I alluded to the fact that 20th Cent mathematicians from various countries have improved the foundations of these tools. The definition of extended variance proposed here could not have been expressed without these improvements. One additional name should be added here. At the beginning of the 20th Century, the Italian statistician Gini (1884-1965) addressed the issue of variability, for instance in Gini [23]. In his dissertation, four years before, considering once again the puzzling stability of human sex ratio at birth, he arrived at some point to this conjecture. This index looks quite constant for the human species. But for other species the issue is open. This stability for humans could be due to the stability of their environments, as opposed to the environments of other species that could be taken for less stable. So, for Gini, variability could come from the environment or from the species themselves observed in a stable environment. Being a precursor of bio-statistics, Gini was the first author to have conceived a possible balance of statistical variability between a potentially variable environment and a variable phenomenon *per se*. Of course, Gini had Darwin in mind) but he did not argue on fitness and adaptation. A history and epistemology of 19th and 20th Centuries works at the border of Mathematics, moral Statistics, and Biology is given in Brian and Jaisson [24].

On decomposing variance

We called twice a theorem, (4) and (10), a decomposition of variance between the systematic comparison of all possible couple of proofs and a possible warp due to an arbitrary filter on couple of proofs. The first and usual computing is encompassing a strong hypothesis: as D'Alembert wrote, this abstract hypothesis may not fit with the physical phenomena to be considered. From an interpretative standpoint the warp could be due to physical and phenomenological constraints, or from arbitrary computing rules.

On Durheim's coercion

In Durkheim's general sociology, the social fact is defined by a coercion of the external morphology of society on the internal physiology of representations among the social agents. This notion of coercion has been criticized

by authors troubled by the action of the outside on inside feelings or motives for instance. Going beyond methodological discussions held in the 1970's, it is necessary to focus on this notion of coercion and to be able to analyse it, Brian [25]. If we consider that the extended variance is the operator that expresses the agent's exposure to or apprehension of the external phenomenon, then must be traced its difference to what could be driven from systematic and usual variance. The variability from the agent's standpoint may largely neutralize the systematic variability of external morphology, as in the case of X_3 and ExtVar_{Π} . In front of risks, aversion or blindness are well known characteristic of some individual and collective behaviours. By extension one could consider aversion or coecity in front of volatility or of higher levels of variability. It suggests that following Durkheim's analysis, but extracting it from its raw moment paradigm, coercion of the external state of society on internal representations does not only operate by means of trends but by means of exposure to variability. If we add now that social division of labor - for instance its technical division in finance - affects together volatility (variability) and the circuits of information mobilized for its estimation we get that the morphology of the social world may be object of social transformation for social stakes related to variability.

On Bourdieu's field-habitus homology

Bourdieu's general sociology is derived from Durkheim's one, but structuralist, Fabiani [26]. The structure of the social field, is for Bourdieu the external morphology of the social world, as the habitus, on the agents' side, the structure of internalized representations. His theory of action is based on the analysis of the adequation or not of habitus to field. The discrepancies between objective chances offered by the field and subjective probabilities proper to the habitus are constant analytic keys. An hypothetic homology between habitus and fields reduces them to nothing. One of Bourdieu's principle is that with duration these discrepancies tend to disappear: «making bad luck good heart»: an habituation. The structure of the field in this theoretical framework is made of all possible distinctions, Bourdieu [27], this is why multivariate analysis with systematic variance was his privileged tool for displaying sketches of fields. With extended variances the principle of habituation does not touches only the distribution of chances of proofs, but of comparable couple of proofs. A matrix dQ expressing the probabilities of encounters or comparisons could be shaped by the morphology of the field (external constraints), or by the physiology of representations (the agents' predilec-

tions). Evolving geographical and social segregations provide examples for balances between external systems of mutual comparisons, and interiorized predilections in their accounts. So, «making bad luck good heart» does not only operates from objective chances onto subjective probabilities, but in the same manner from systematic variability onto subjective comparisons of proofs - onto subjective extended variance. This is habituation to a given differentiated external world^e. On the other, if social agents who share a common apprehension of possible encounters may project this common apprehension on the external world reworking its concrete morphology^f. With extended variance, we have a tool to formulate these sensibilities and to evaluate their impact on perception of variability, and to figure out the balance between exteriority and interiority.

On Berthoz' simplicity

For Berthoz [29], «simplicity is the set of solutions found by living organisms [human beings among others] that enable them to deal with complex information and situations, while taking into account past experiences and [trying] to anticipate future ones. Such solutions are new ways of addressing problems so that actions may be taken more quickly, more elegantly, and more efficiently». I would consider here that a high level of variability is a factor of complexity. Then adapting one's extended and selective variance in order to reduce apprehended variability is a simplex solution, or making it higher for competitive species could appear as a simplex option in a interspecies competition. It is not only first order probabilistic anticipations (2012, chap 7), (bets on proofs) but second order probabilistic anticipations (bets on couple of proofs). Here, there is a room for empirical research: how do elementary organisms like bacteria react to different levels of environmental variability? For more elaborated organisms, are there simplex strategies of reduction of apprehended variability? Or competitions based on simplicity (the paradigm of extended variances offering a rigorous experimental framework)? Beside the appearances of systematic external variability, the extended variances give access to the impact of computation rules and of peculiar apprehensions on variability, offering a room for balancing between the first and the two other, and for conceptual speculations touching general sociology, cognitive sciences and life sciences.

^eIn French it is the sociologist Maurice Halbwachs' concept of «*remaniement des traditions* [of established representations]», see Brian [28].

^fNow, from the same author the «*remaniement des lieux* [of places]», *ibid*.

A tribute to Jean Le Rond D’Alembert

The issue of the probabilization of the product space was touched no less than two hundred sixty three years ago in D’Alembert [30]. One of the best mathematician of the time, he ridiculed himself for the eyes of the contemporaneous geometers, Brian [6,7]. But the argument was serious, at least enough to make so that the two best probabilists of the next parisian generation, Condorcet and Laplace, have later shaped the integral calculus of probabilities as a response to their master’s doubts, opening a new era called “The probabilistic Revolution”, Krüger, *et al.* [31]. D’Alembert point was the following: if when one tosses twice a coin, the rule is that the player wins by getting “head”. The usual calculus told to count the number of possible cases, 4, the number of beneficial cases, 3, and the usual bet is 3 against 1. But the Geometer added that, if at the first toss ”head” would appear, the game would stop immediately. Therefore only 3 actual possible cases should be taken into account: 1 for the first toss giving “head” and 2 the second flip. Then the number of beneficial cases should be 2, one for the first winning toss, and one for the second. Hence, for D’Alembert, the bet should be 2 against 1. He was as close as possible to the physics of the phenomenon, two consecutive tosses, respecting each concrete step of the game. Doing so, he proposed to exclude one couple of proofs, something that troubled the intuition of a symmetric game. The metaphysics of calculus was provocative. We can express his view in our terms.

$P \times P_{usual}(A, B)$ <table style="margin-left: 20px; border-collapse: collapse;"> <tr> <td style="border: 1px solid black; padding: 2px; margin-right: 10px;">C</td> <td style="padding: 2px 10px;">1/4</td> <td style="padding: 2px 10px;">1/4</td> </tr> <tr> <td style="border: 1px solid black; padding: 2px; margin-right: 10px;">P</td> <td style="padding: 2px 10px;">1/4</td> <td style="padding: 2px 10px;">1/4</td> </tr> <tr> <td style="padding: 2px 10px;">$P(A)$</td> <td style="padding: 2px 10px;">1/2</td> <td style="padding: 2px 10px;">1/2</td> </tr> </table>	C	1/4	1/4	P	1/4	1/4	$P(A)$	1/2	1/2	$Q_{D'Alembert}(A, B)$ <table style="margin-left: 20px; border-collapse: collapse;"> <tr> <td style="border: 1px solid black; padding: 2px; margin-right: 10px;">C</td> <td style="padding: 2px 10px;">1/3</td> <td style="padding: 2px 10px;">1/3</td> </tr> <tr> <td style="border: 1px solid black; padding: 2px; margin-right: 10px;">P</td> <td style="padding: 2px 10px;">0</td> <td style="padding: 2px 10px;">1/3</td> </tr> <tr> <td style="padding: 2px 10px;">$P(A)$</td> <td style="padding: 2px 10px;">1/3</td> <td style="padding: 2px 10px;">2/3</td> </tr> </table>	C	1/3	1/3	P	0	1/3	$P(A)$	1/3	2/3
C	1/4	1/4																	
P	1/4	1/4																	
$P(A)$	1/2	1/2																	
C	1/3	1/3																	
P	0	1/3																	
$P(A)$	1/3	2/3																	

Acknowledgements

A first sketch of this argument was in Brian [19], p. 25-32. Its tuning was facilitated by exchanges with M Métivier, L Lebart and B Ycart in 1980-1985. P Bourdieu was a permanent interlocutor from 1980 until 2001, and later M Barbut or C Walter, around 2005. More recently, A Berthoz drew the author’s attention to cognitive sciences. The author thanks them all but claims full responsibility for his imprudence and imperfections. The author thanks to the Journal reviewer for his rigorous and suggestive comments. In 2017, Prof. Brian is directeur d’études at EHESS (PSL*Research University Paris).

References

1. Prigogine I, Nicolis G (1971) Biological order, structure and instabilities. *Q Rev Biophys* 4: 107-148.

2. Nikolay Perunov, Robert A Marsland, Jeremy L England (2016) “Statistical physics of adaptation”. *Phys Rev X* 6.

3. CW Gardiner (1984) *Handbook of stochastic methods for physics, chemistry and the natural sciences*. (2nd edn), Springer, Newyork.

4. Euler Leonhard (1778) *Observationes in praecedentem dissertationem illustris Bernoulli. “Diiudicatio maxime probabilis plurium observationem discrepantium atque verisimillima induction inde formanda”*. Euler to the Petersburg Academy. *Acta Academiae Scientarum Imperialis Petropolitinae*, 24-33.

5. Dodge Yadolah (2007) *Statistique. Dictionnaire encyclopédique*. Springer-Verlag, Paris.

6. Eric Brian (1994) *La Mesure de l’État. Administrateurs et géomètres au XVIII^e siècle*. Paris.

7. Eric Brian (1996) *L’Objet du doute. Les articles de D’Alembert sur l’analyse des hasards dans les quatre premiers tomes de l’Encyclopédie. Recherches sur Diderot et sur l’Encyclopédie* 21: 163-178.

8. Maurice Fréchet, Maurice Halbwachs (1924) *Le Calcul des probabilités à la portée de tous*. Dunod, Paris.

9. Banet TA, Lebart L (1984) *Local and Partial Principal Component Analysis (PCA) and Correspondence Analysis (CA)*. *Compstat* 113-118.

10. Lebart Ludovic, Morineau Alain (1985) *Spad-1985*, Saint-Mandé. CISIA.

11. Kunihiro Baba, Ritei Shibata, Masaaki Sibuya (2004) *Partial correlation and conditional correlation as measures of conditional independence*. *Australian and New Zealand Journal of Statistics* 46: 657-664.

12. Ola Larson, Nahum Sonenberg, Robert Nadon (2010) *Identification of differential translation in genome wide studies*. *PNAS* 107: 21487-21492.

13. Anders Hald (1998) *A History of mathematical statistics from 1750 to 1930*. Wiley, New York.

14. Stephen M Stigler (1999) *Statistics on the table. The history of statistical concepts and methods*. Harvard University Press, Cambridge.

15. Theodore M Porter (2004) *Karl Pearson: The scientific life in a statistical age*. Princeton University Press.

16. Jacques Neveu (1970) *Bases mathématiques du calcul des probabilités*. (2nd edn), Paris, Masson.

17. Fréchet Maurice (1955) *Les Mathématiques et le concret*. Presses universitaires de France, Paris.

18. Condorcet (1784) *Mémoire sur le calcul des probabilités: I. Réflexions sur la règle générale qui prescrit de prendre pour valeur d’un évènement incertain, la probabilité de cet évènement multipliée par la valeur de l’évènement lui-même*. *Mémoires de l’Académie royale des sciences pour l’année*. Imprimerie royale, Paris, 707-720.

19. Éric Brian (1986) *Techniques d’estimations et méthodes factorielles: exposé formel et application aux traitements de données lexicométriques*. PhD., University Paris XI.

20. Jakob Bernoulli (1713) *Ars conjectandi, Opus posthumum*. Basileae, Thurnisii.

21. Harry Markowitz (1952) *Portfolio selection. The Journal of Finance* 7: 77-91.

22. Éric Brian (2009) *Comment tremble la main invisible*. Springer-Verlag, Paris.
23. Gini Corrado (1912) *Variabilità e Mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche*. Bologna, Tipografia di Paolo Cuppin.
24. Éric Brian, Marie Jaisson (2007) *The Descent of Human Sex Ratio at Birth: a Dialogue between Mathematics, Biology and Sociology*. Springer, Netherlands.
25. Éric Brian (2012) Où en est la sociologie générale? *Revue de Synthèse* 133: 401-444.
26. Jean-Louis Fabiani (2016) *Pierre Bourdieu. Un structuralisme héroïque*. Seuil.
27. Pierre Bourdieu (1979) *La Distinction. Critique sociale du jugement*. Minuit 668.
28. Éric Brian (2008) Portée du lexique halbwachsiens de la mémoire. In: Maurice Halbwachs, *La Topographie légendaire des évangiles en Terre sainte. Étude de mémoire collective*. Presses universitaires de France, 113-146.
29. Alain Berthoz (2012) *Simplexity: simplifying principles for a complex world*. Yale University Press.
30. D'Alembert (1754) *Croix ou pile (Analyse des hasards)*. *Encyclopédie ou Dictionnaire raisonné* 513-514.
31. Lorenz Krüger, Gerd Gigerenzer, Mary S Morgan (1987) *The Probabilistic Revolution, Volume 2*. MIT Press, Cambridge.