Bourbaki vs Pragmatism
A methodological comparison through the multi-armed bandits problem
DRAFT

Sebastiano Ferraris*

April 6, 2020

In these pages we introduce the well known multi-armed bandits problem and we propose a solution with two different mathematical methodologies: a Bourbakist approach and a pragmatic one. The Bourbakist way is concerned with the mathematical foundation upon which a formal solution is derived in the shape of an axiomatic structure. In contrast, the pragmatist approach aims at finding the shortest path towards a solution, reducing the mathematical formalisms to the bare minimum. The article ends with a critical note, where pros and cons of each method are compared.

If you came across this article when searching for an introduction to the multi-armed bandit problem, and not a methodological comparison, you can ignore section 2. The code to create the figures and run a range of algorithms to solve the problem can be found at https://github.com/SebastianoF/MAB.

1 Multi-armed bandits problem

You have to repeatedly choose between K different possibilities, all choices cost the same, though each may or may not provide you with a cash reward. The reward is drawn from a probability distribution, *unknown* and different fro every possibility.

The problem of finding a strategy to maximise the reward in this setting is called multi-armed bandits (MAB) after the situation of playing repeatedly at a row of K slot machines (or single-armed bandit, as they have been called). If you have an initial amount of money of \$1000 and a costs of \$1 for each draw, than you have 1000 attempts to balance an exploration phase when estimating the unknown distributions of each arm,

*sebastiano.ferraris@gmail.com

with an exploitation phase, when the acquired knowledge is used for a gain the highest reward[1].

The MAB problem can be generalised to clinical or pre-clinical trials, control engineering, mechanical and software testing, stock market investments, behavioural modelling, dynamic pricing, and many more[2].

Goal of these pages is to introduce the problem and to provide a solution with two different methodologies:

- *Bourbakist.* This approach, named after the collective pseudonym of the group French mathematicians, who have developed it in 1934. The initial goal of Bourbaki was to re-write one of the most widely used analysis textbook of the time, the Goursat's *Cours d'analyse mathématique*, that was considered inadequate as some counterexamples had been overlooked[3]. To reach this goal and to aim at extending the possible generalisations of the problem, while keeping at the same time possible counterexamples under control also in other branches of mathematics, Bourbaki favours an axiomatic foundation based on set theory. In this approach exercises, examples and the history motivating the given problem are not provided to the reader.

- *Pragmatic.* In contrast with the Bourbakist approach, the pragmatic approach reduces the formalisation to its bare minimum, and it orient the effort of the reader into finding a solution of a problem, rather than the creation of a generalisable formal theory.

The MAB presented in this section, is formalised *a la Bourbaki* in the next one, and presented in the pragmatic way in section 3.

2 The Bourbakist perspective

Definition 2.1. Let $(\Omega, \mathcal{A})$ be a σ-algebra defined as a non-empty set Ω paired with a subset of its power set $\mathcal{A}$, containing the empty set and closed under numerable union and complement set. Let this be called *action space*. Let $\mathcal{I}_K = \{1, 2, \dots, K\} \subset \mathbb{N}$ be a set of indexes whose generic element k is called *arm*, by convention. Let $\mathcal{I}_T = \{1, 2, \dots, T\} \subset \mathbb{N}$ be another set whose elements are called, again conventionally, *time*. Let $\mathbb{R}_+$ the positive real axis including the zero.

The relationship between the above defined elements are given by a function A defined as:

$$
\begin{aligned}
A : \mathcal{I}_T \times \mathcal{A} &\longrightarrow \mathcal{I}_K \\
(t, \omega) &\longmapsto A(t, \omega) = A_t(\omega)
\end{aligned}
$$

[1]See Thompson [Tho33] for an early approach where the two arms are two medical treatments, Bellman [Bel56] where the problem is formulated in a Bayesian perspective for two arms, and the more recent Sutton [SB18], chapter 1, for a reinforcement learning perspective.

[2]See Bouneffouf [BR19] for a survey with a list of applications of the main algorithms solving the MAB problem.

[3]See Marmier [Mar14] for a brief history of Bourbaki and its impact in France and the rest of the world.

and by a function R, defined as:

$$\begin{aligned} \mathcal{R} : \mathcal{I}_T \times \mathcal{A} &\longrightarrow \mathbb{R}_+ \\ (t, \omega) &\longmapsto R(t, \omega) = \mathcal{R}_t(\omega) \end{aligned}$$

Let the former be called *action* and the latter be called *reward*. Let

$$\begin{aligned} R : \mathcal{I}_T \times \mathcal{I}_K &\longrightarrow \mathbb{R}_+ \\ (t, \omega) &\longmapsto R(t, k) = R_t(k) \end{aligned}$$

another function, defined as the only possible function making the diagram below commutative.

$$\begin{tikzcd} \mathcal{A} \arrow[r, "A_t"] \arrow[dr, "\mathcal{R}_t"] & \mathcal{I}_K \arrow[d, "R_t"] \\ & \mathbb{R}_+ \end{tikzcd}$$

for each $t \in \mathcal{I}$. Let R_t be called again *reward*, and the difference between $\mathcal{R}$ and R will be clear from the context.
We observe that $\mathcal{R}$ maps the events of the σ-algebra, while R maps the corresponding indexes. This is an analogous of the definition of probability respect to the one of random variable and probability density function, for when the real axis is restricted to $[0, 1] \subset \mathbb{R}$.

Now consider

$$\begin{aligned} \mathcal{Q} : \mathcal{I}_T \times \mathcal{A} &\longrightarrow \mathbb{R}_+ \\ (t, \omega) &\longmapsto \mathcal{Q}(t, \omega) = \mathcal{Q}_t(\omega) \end{aligned}$$

the *estimated reward of the action ω up to time t*, for $\omega = A_t^{-1}(k)$ for a fixed $k \in \mathcal{I}_K$, with the corresponding function $Q : \mathcal{I}_T \times \mathcal{I}_K \to \mathbb{R}$. It follows that $\mathcal{Q}$ is defined as an application of the mean value in a Lebesgue space over $(\Omega, \mathcal{A})$, that is now a Borel σ-algebra[4] as:

$$\mathcal{Q}_t(\omega) = \mathbb{E}\left[R_\tau(k) \mid k = A_\tau(\omega), \forall \tau \in \mathcal{I}_t\right] \qquad \omega \in \mathcal{A} \qquad t \in \mathcal{I}_T \tag{1}$$

and therefore

$$Q_t(k) = \mathbb{E}\left[\mathcal{R}_\tau(\omega) \mid \omega = A_\tau^{-1}(k), \forall \tau \in \mathcal{I}_t\right] \qquad t \in \mathcal{I}_T \qquad k \in \mathcal{I}_K \tag{2}$$

As $\mathcal{R}$ and R did, also $\mathcal{Q}$ and Q satisfies the commutativity of a diagram analogous to the one shown above. The notation can be simplified for brevity[5] to:

$$Q_t(k) = \mathbb{E}\left[R_t(k) \mid A_t = k\right] \tag{3}$$

[4] For a foundational perspective, see Bourbaki [Bou04a].

[5] The simplified notation is often the only notation appearing in engineering textbooks (e.g. Sutton [SB18]), although this would not allow the reader to understand the subtle formalisation of assigning to an event ω its index k. More tragically the simplified notation makes most of the concepts introduced so far pedantic and irrelevant.

where the mean value is for all the time indexes up to t and where the domain values of A_t is clear from the context.

Definition 2.2. Let the *total reward* $Q_\infty : \mathcal{I}_K \to \mathbb{R}$ be the function

$$Q_\infty(k) = \mathbb{E}\left[\mathcal{Q}_t(\omega) \mid \omega = A_t^{-1}(k), \forall t \in \mathcal{I}_T\right] \qquad k \in \mathcal{I}_K \tag{4}$$

or with the simplified notation as:

$$Q_\infty(k) = \mathbb{E}\left[R_t(k) \mid A_t = k\right] \tag{5}$$

So far we have been considering the reward and the total reward for a fixed choice of k. We can vary $k \in \mathcal{I}_K$ in function of the time index. So let $\mathbf{k}$ an element of

$$\mathcal{I}_K^T = \underbrace{\mathcal{I}_K \times \mathcal{I}_K \times \cdots \times \mathcal{I}_K}_{T\text{-times}}$$

or equivalently a function from $\mathcal{I}_T$ to $\mathcal{I}_K$. Definitions 1 and 3 are so generalised to $Q_t : \mathcal{I}_K^T \to \mathbb{R}$ for $t \in \mathcal{I}_{\leq T}$, having defined $\mathcal{I}_{\leq T}$ any interval of positive integers between 1 and T, and

$$Q_t(\mathbf{k}) = \mathbb{E}\left[\mathcal{R}_\tau(\omega) \mid \omega = A_\tau^{-1}(\mathbf{k}_\tau), \forall \tau \in \mathcal{I}_t\right] \qquad \mathbf{k} \in \mathcal{I}_K^T \qquad t \in \mathcal{I}_T$$

and therefore

$$Q_\infty(\mathbf{k}) = \mathbb{E}\left[\mathcal{R}_t(\omega) \mid \omega = A_t^{-1}(\mathbf{k}_t), \forall t \in \mathcal{I}_T\right] \qquad \mathbf{k} \in \mathcal{I}_K^T$$

The mean value computed with a Lebesgue measure, over the Borel space generated as the sets of images [6] $\mathcal{R}_t(\omega)$ for all $\omega \in \mathcal{A}$ can be reformulated as:

$$Q_t(\mathbf{k}) = \frac{\sum_{\tau=1}^{t} R_\tau(\mathbf{k}_\tau)\mathbf{1}_{A_\tau=\mathbf{k}_\tau}}{\sum_{\tau=1}^{t} \mathbf{1}_{A_\tau=\mathbf{k}_\tau}}$$

where $\mathbf{1}_{A_\tau=\mathbf{k}_\tau}$ equals to 1 for when the event ω corresponding to $\mathbf{k}_\tau$ is mapped exactly to $\mathbf{k}_\tau$ through A_t, and 0 for any other event. Extending the time indexes up to infinity, and to justify the notation introduced above, where we used ∞ for a finite case, we have the following theorem:

Theorem 2.1. *Given a ring of infinite cardinality to which the time index t belongs, and an Hilbert module*[7] *to which the vector $\mathbf{k}$ belongs, it follows that*

$$Q_\infty(\mathbf{k}) = \lim_{T\to\infty} Q_T(\mathbf{k})$$

Proof. Direct consequence of the definition of Q_∞ generalised to the theorem hypothesis' extended structures. □

[6] We consider the definition under the accordance with the axiom of choice as in the ZFC axiomatic set theory, in order to avoid *virages dangereux*. See also Bourbaki [Bou04b] and [TZ82].

[7] An algebraic structure generalising Hilbert vector spaces over the now introduced ring of time indexes. See for example [Bou87].

We now consider the value $\hat{k}$ that satisfies

$$\hat{k} = \arg\max_{k \in \mathcal{I}_K} Q_t(k) \qquad \forall t \in \mathcal{I}_T$$

for a constant value for each time index, as in the definition of Q_t given in 2. If we consider the possibility of varying the chosen arm k across time, and so if we are allowed to compare different images of the function A_t then $\hat{\mathbf{k}}$ is defined as

$$\hat{\mathbf{k}} = \arg\max_{\mathbf{k} \in \mathcal{I}_K^T} Q_t(\mathbf{k}) \qquad \forall t \in \mathcal{I}_T \tag{6}$$

Under the light of theorem 2.1, and with the given definitions, we can now call the defined vector $\hat{\mathbf{k}}$ the *solution of the generalised multi-armed bandits problem.*

3 The pragmatic perspective

Creating a Benchmark

To have a ground truth where to test a range of algorithms and parameters, we consider a benchmark dataset where the distribution of each arm is known. In this example we assume each arm normally distributed, with mean sampled from a uniform distributions in the interval $[-3, 3]$ and the standard deviation sampled form a normal distribution with median 2 and standard deviation 4. A representation of these distribution is plotted in figure 1. It is straightforward to adapt this benchmark to any other given distribution, and therefore to benchmark the algorithm under multiple assumptions.

Encoding the problem

We encode the rewards in a $T \times K$ matrix q, where $T = 1000$ is the time we will be pulling an arm and $K = 10$ is the number of arms. Its element $q_{t,k}$ represents the reward collected for having pulled the arm k at the time-point t. As we can pull only one arm at a time, there is only one known value for each column of the matrix. All the other values are initialised to nan.

ϵ-greedy algorithm

Bayesian algorithm

Non-stationary distributions

Epsilon greedy modified
Bayesian modified
Comparisons and benchmarking

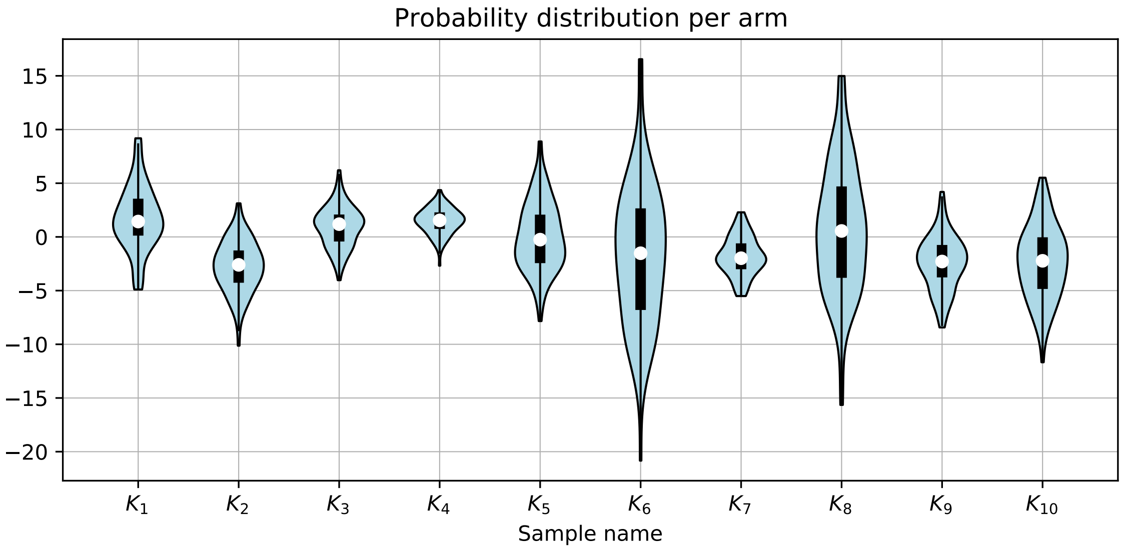

Figure 1: Example of a benchmark distribution of the reward for each one of the 10 arms.

4 Discussion

In this article, we considered the example of the multi-armed bandit problem to show the ineffectiveness of the Bourbachist approach in solving the problem.

Even if formally correct (excluding typos) and coherent (excluding Goedel[8]), the theory provided in section 2 has the effect of turning a relatively simple problem into a complicated maze of concepts.

On the contrary, the pragmatic approach of section 3 show that there is no need of taking a functional perspective and to link it to several mechánemas developed in entirely different contexts. These are not just irrelevant in reaching a solution, they are also suffocating any possible creativity when facing a slightly different problem, as modulating the unknown distributions over time. If in doubt on this point the reader to continue the formalisation in this direction, to see how many pages and new definitions and diagrams are needed. We also challenge the reader to implement the code to solve the problem having only the functional and algebraic definitions at hand rather than relying on the matrix point of view. None of the given definitions in the Bourbaki approach had provided any hints on how to really solve the problem, in a numerical sense, as, in conformity with the Bourbaki style, no examples had been provided.

[8] To this regard, in the article *the ignorance of Bourbaki* [Mat92], Mathias noticed that the attempt of grounding the whole corpus of mathematics in an axiomatic sense have happened after, and in a conscious effort of ignoring the Goedel incompleteness theorems.

Other examples

Despite many, and all of them more experienced than the author, have express their negative view about the Bourbakist's mathematics[9], the Bourbaki approach is still widespread if not predominant across the mathematical community.

There are in fact numerous examples of practical problems, whose pragmatic approach had been ruined by an overformalisation[10] similar to the short one here proposed. These are all useful examples, in particular to anyone who may believe, that the author had been overzealous in writing section 2, to give on purpose a negative light upon this the methodology.

The most notable example is the optimal transport (OT) theory. From being a pragmatic methodology of solving a class of optimisation problems, it had become a 500 pages book underpinned by a great amount of measure theory and Lebesgue spaces, perfectly irrelevant to solve any instance of an optimal transport problem, when the reader will have to face one. Comparing one of the original optimal transport theory presentation by Hitchcock [Hit41] and the formalised one by Villani [Vil03] is possible to see the extent of the issue. The first one is easy to read, understand, implement and possibly extend in several directions by anyone having an highschool mathematical education. The second one is a seemingly uncreated maze of unassailable interlinked concepts, requiring few years of academic studies only to grasp the first few pages, with no advantages in finding a numerical solution to the problem.

A second notable example of the Bourbaki effect, is in the domain of medical image registration. Here the aim is to solve the problem of finding the non-rigid deformation or metamorphosis between anatomies. The problem originates from the anatomical studies of shapes growth by D'Arcy Thompson [d$^+$42] and pragmatically extended amongst others in Modersitzki [Mod04]. The Bourbaki over-formalized branch can be found in works like [You10], where we have to wait for 12 chapters before seeing what had motivated the formal mathematical theory developed until there. In this case too, there is no explanation of what are the advantages of the axiomatic approach respect to the pragmatic one, appeared half a century earlier.

The reader may say, for this specific case, that the pragmatic approach [Mod04] does not use neither diffeomorphisms nor reproducing Kernel Hilbert spaces. And this is true, although it is not proved that these two mathematical devices can provide more accurate results than their pragmatic counterparts when implemented in practice. Instead it is true that they are computationally slower.

[9] Amongst the many: Arnold [Arn98], De Finetti [DFN08], Lockhart [Loc09], the already cited Mathias and its follow up [Mat98], Velupillai [Vel12]. And even the article by Marmier [Mar14], that sees the Bourbaki movement with enthusiasm and appreciation for their work and impact in the society, begs some questions when it says "Living mathematics is being done elsewhere in the world, in connection with new problems arising from other disciplines or other human activities".

[10] Or *assiomattisation*, as Bruno de Finetti [DFN08] would have said playing on the Italian word *matti*, meaning crazy.

The Bourbaki pattern

Other than the branches of mathematics mentioned above, there are others that have been touched by Bourbachism, such as algebraic topology, fuzzy logic and topological data analysis. Their detachment from the practical set of problems that had them originated, had turned them into convoluted axiomatic buildings, where it is not possible to find any breakthrough result in the perspective of solving the initially given problem.

It is at this point evident the existence of a pattern in the Bourbaki method: a problem that had been solved by 50 years or more, often in a very elegant way, is taken under the observation of axiomatisers and transformed into a list of axioms and theorems where no echo of the original problem that motivates the theory has to be found. Examples and counterexamples are removed, every reference to the history of the theory's development is eradicated too.

Remarkably, also Einstein had noted that *Since the mathematicians have invaded the theory of relativity, I do not understand it myself any more* [Sch51]. And it is difficult to find what are the new results that the Bourbaki formalisation have added to the theory of relativity.

In fact, within the Bourbaki pattern comes the fact that the formalised version of the proposed theory adds no value. No new results having any practical relevance can be found in the work by Bourbaki and followers.

There is only one notable exception to this rule that we could find. The work Mochizuki [Moc12] is claimed to have solved the *abc conjecture*. Worth of a pantomime, according to the mathematicians who had tried to follow the passages, "the proof is too impenetrable to be understood"[11].

Why is Bourbaki still around?

What the examples proposed in these pages intend to show, from a qualitative and empirical evaluation, is that the over-axiomatic version of a theory has several negative aspects respect to its pragmatic counterpart.

Nonetheless, these pages are not written to accuse Bourbaki. In fact, if all so negative, then the Bourbaki approach would not be still around any more. In reality, this approach is not only predominant in the universities. It is also greatly valued, prized, and appreciated by academics[12].

It seems that the over-axiomatization of a theory developed outside academia by mathematicians and scientist chasing practical problems is worth every recognitions, including the Fields Medal. Highly valued and cited uber-bourbachists works are for example Grotedniek [Gro11] and the four volumes of Mochizuki [Moc12][13].

[11] As stated in the paper appeared in Nature, Castelvecchi [Cas15].

[12] To this regard, the active reader is welcome to copy-paste section 2 and to extend it, in the Bourbakist style, into an article whose title could be something like *The multi-armed bandits for the working mathematician*, to see the consequences of it upon the mathematical community.

[13] It may be interesting to notice the similarity between this work and the outcome of the randomized generator of maths paper at thatsmathematics.com/mathgen/.

So what would justify the fact that the Bourbaki approach is still considered and widely thought? Below are itemized all the reasons that the author could find.

- *The outcome of the Bourbaki approach is inaccessible to neophytes*[14], and it allows the existence of professors of mathematics who can not code or do not have any ability to solve any problem in practice. This seems to be the main reason that justifies the continuation of the Bourbachism, as neophytes usually are the people who have to decide where the public money goes and professors the ones who receives it.

- *The Bourbaki approach is good for the ego.* The pleasure of owning an elegant notebook filled with mathematical formulae written with a well thought after handwriting is a most sublime one. Unfortunately, after empirical evidence we can assure that this very same pleasure is a great obstacle in acquiring knowledge. The notebook owner, when facing a problem, will inevitably tend to abandon any scientific method and will bend the problem to the behold solution. To this regard it would be interesting to measure the dopamine level of a mathematician trying to solve a difficult problem respect to the mathematician tweaking axioms to make the theory more elegant.

- *Bourbachism detaches completely mathematics from reality*, so it makes impossible to have any objective measurement of the quality or value of the work produced, besides the readers' opinion. This is a very unscientfic feature, and again a very useful one for whoever must defeat more skilled competitors within the academic environment.

  The advocates of the superiority of pure mathematics[15] have even arrived at accusing the mathematics that has anything to do with reality of being more prone of making mistakes[16]. We can be reassured by the fact that under every circumstance, the reality is adamant to persist in being what it is, and it is at this point an interesting exercise to imagine someone advocating for having medicine and engineering detached from reality.

[14] An attempt of turning an over-formalised Bourbakist theory into something accessible for the layman, can be found in Villani [Vil12]. Interestingly enough, it seems that the only way the author could popularize his theory was by omitting the definitions and proofs while leaving only the main mathematical formulas alongside some autobiographical events. The result is even less accessible than the original work, proving that mathematicians can also be successful surrealist artists.

[15] It is worth noticing that the academic dichotomy pure/applied maths is something that can not be found in mathematics before Bourbaki. The reader is invited to classify any of the work of Archimedes, Gauss, Euler or any other great names in the history of pre-modern mathematics as pure or applied mathematics with no ambiguities.

[16] Gros [GST19] had found how even professional mathematicians can be mislead by reality. And they had leveraged on this most surprising fact, for advocating to increase the detachment. After concluding that "[...] we can't reason in a totally abstract manner", instead of suggesting to take into account the reality in the mathematical practice they suggested a move towards the opposite direction: "We have to detach ourselves from our non-mathematical intuition" [Sum19].

- *The Bourbaki inspired work are more generalizable.* While this appear to be true at first glance, we claim that on the contrary a Bourbaki theory is less generalisable. Simply adding a new or different assumption to the problem settings would require to re-write almost from scratch all the axiomatic building to include the new input in a generalisable manner. The pragmatic way provides handle on the problem that are easy to be tuned or re-adapted to a slightly different one.

- *Bourbachism binds the concepts in formal structures avoiding paradoxes and counterexamples.* This is a very valid point in favour of the Bourbakist method, as the concerns of mathematics with counterexample have led to various discoveries, from Galois theory, to Fractals, and the search for counterexample is a valuable tool to have a better understanding when exploring the limits of mathematical knowledge[17]. Although, the problem of the Bourbaki approach is the over-concern with counterexamples originating from theoretical considerations not related in any way to the sought solution. In section 2 we attained an algorithm that solves the problem, having never faced pathological counterexamples despite not relying on σ-algebras, Borel spaces, Lebesgue measures or even without explicit use of functions. What we came across were only some practical malices of the craft that are never learned by whoever is limiting themselves to the Bourbaki presentation of the problem.

About this last point in particular, there is an (in)famous case where over-concerned researchers had been mislead by a misquoted counterexamples: the medical imaging paper by Lorenzi [LP13] claims that there is no bijective correspondence between the space of the tangent vector fields and the one of their integral curves. This claim is made after citing a very peculiar counterexample found in one of the Milnor research papers, involving the tangent space of the circle.

Interestingly enough the conditions for the counterexample to happen are never met in the applications presented in the paper. The considered case is of $\mathbb{R}^3$, whose tangent space is $\mathbb{R}^3$ itself and the bijective correspondence is guaranteed by the theorem of existence and uniqueness for ODEs.

5 Conclusion

The comparison proposed in this paper aims at showing how important is the mathematical approach in achieving a possible solution of a problem, and how choosing the wrong approach, would end up in never reaching a practical solution. The paper shows the over-formalised approach first, and the pragmatic approach afterwards.

With the pragmatic way applied to the multi-armed bandits, an algorithmic solution can be quickly explained, implemented and tested. All this without calling into play Borel or Lebesgue, as done in the counterpart.

[17] For example Procesi [Pro77], Mandelbrot [Man83] and Gelbaum [GO03].

In the discussion we challenged the usefulness of the Bourbaki approach outside its circle of academics. The Bourbaki generalisations are not practical and the counterexamples that are excluded by the overformalisation do not arise in practice. the author also implies that the mindset of the Bourbaki educated student can even obstaculate the pathway leading to a solution of a practical problem.

Despite the author is convinced that students of the generations to come will be obliged to learn the Bourbaki sophistications to get a degree, and to unlearn them as quickly as possible to be of any use to society, this paper hopes to show that the Bourbaki method is around for academic reasons, and it is not what mathematical practice is about or should be, which is the application of the scientific method to understand the reality and to find the best algorithm that solves a given problem.

References

[Arn98] Vladimir I Arnol'd. On teaching mathematics. *Russian Mathematical Surveys*, 53(1):229, 1998.

[Bel56] Richard Bellman. A problem in the sequential design of experiments. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 16(3/4):221–229, 1956.

[Bou87] Nicolas Bourbaki. Topological vector spaces, elements of mathematics. 1987.

[Bou04a] Nicolas Bourbaki. Integration. Springer Berlin, 2004.

[Bou04b] Nicolas Bourbaki. Theory of sets. Springer Berlin, 2004.

[BR19] Djallel Bouneffouf and Irina Rish. A survey on practical applications of multi-armed and contextual bandits. *arXiv preprint arXiv:1904.10040*, 2019.

[Cas15] Davide Castelvecchi. The biggest mystery in mathematics: Shinichi mochizuki and the impenetrable proof. *Nature News*, 526(7572):178, 2015.

[d$^+$42] W Thompson d'Arcy et al. On growth and form. *On growth and form*, 1942.

[DFN08] Fulvia De Finetti and Luca Nicotra. *Bruno de Finetti: un matematico scomodo*. Belforte, 2008.

[GO03] Bernard R Gelbaum and John MH Olmsted. *Counterexamples in analysis*. Courier Corporation, 2003.

[Gro11] Alexandre Grothendieck. Some aspects of homological algebra. 2011.

[GST19] Hippolyte Gros, Emmanuel Sander, and Jean-Pierre Thibaut. When masters of abstraction run into a concrete wall: Experts failing arithmetic word problems. *Psychonomic bulletin & review*, 26(5):1738–1746, 2019.

[Hit41] Frank L Hitchcock. The distribution of a product from several sources to numerous localities. *Journal of mathematics and physics*, 20(1-4):224–230, 1941.

[Loc09] Paul Lockhart. *A mathematician's lament: How school cheats us out of our most fascinating and imaginative art form.* Bellevue literary press, 2009.

[LP13] Marco Lorenzi and Xavier Pennec. Geodesics, parallel transport & one-parameter subgroups for diffeomorphic image registration. *International journal of computer vision*, 105(2):111–127, 2013.

[Man83] Benoit B Mandelbrot. *The fractal geometry of nature*, volume 173. WH freeman New York, 1983.

[Mar14] Anne-Marie Marmier. On the idea of 'democratisation','modern mathematics' and mathematics teaching in france. *Lettera Matematica*, 2(3):139–148, 2014.

[Mat92] Adrian RD Mathias. The ignorance of bourbaki. *Mathematical Intelligencer*, 14(3):4–13, 1992.

[Mat98] ARD Mathias. Further remarks on bourbaki (a reply to criticism by professor sanford l. segal of the essay "the ignorance of bourbaki".). *Preprint*, 7, 1998.

[Moc12] Shinichi Mochizuki. Inter-universal teichmüller theory i: Construction of hodge theaters, 2012.

[Mod04] Jan Modersitzki. *Numerical methods for image registration.* Oxford University Press on Demand, 2004.

[Pro77] Claudio Procesi. *Elementi di teoria di Galois.* Decibel, 1977.

[SB18] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction.* MIT press, 2018.

[Sch51] P Schilpp. Albert einstein, philosopher scientist, the library of living philosophers, vol. vii. 1951.

[Sum19] Article Summary. Expert mathematicians stumped by simple subtractions. https://www.sciencedaily.com/releases/2019/07/190710103231.htm, 2019.

[Tho33] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

[TZ82] Gaisi Takeuti and Wilson M Zaring. Introduction to axiomatic set theory. Springer, 1982.

[Vel12] K Vela Velupillai. Bourbaki's destructive influence on the mathematisation of economics. *Economic and Political Weekly*, pages 63–68, 2012.

[Vil03] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

[Vil12] C. Villani. Théorème vivant. 2012.

[You10] Laurent Younes. *Shapes and diffeomorphisms*, volume 171. Springer, 2010.