

The Snowball Effect

Mukul Pal
mukul@alphablock.org

January 19, 2022

Abstract

If Intelligence was a dynamic, architectural, chaotic, non-linear mechanism, it would have moved perpetually between equilibrium and disequilibrium. This would have reduced information entropy, transformed disorder into order, would be normal and non-normal at the same time, lived in probabilistic informational states, which are always in a conflict but as a whole are aware of where they are headed.

The Snowball Effect

A draft of wind nudged a lucky snowflake to bond with another snowflake to make a chain. As it rolled down the mountain, the chain attracted more snowflakes, snowflake by snowflake, transforming into a snowball, a power to reckon, a momentous force, a phenomenon, that makes you wonder, was the snowflake lucky or did it persevere to be at the right time at the right place.

The Watchmakers

The history of science is replete with thought experiments. Herbert Simon, father of artificial intelligence, behavioral finance, and Nobel prize-winning economist (1976) talked about one such thought experiment in his 1962 paper, "The Architecture of Complexity" [1].

"There once were two watchmakers, named Hora and Tempus, who manufactured very fine watches. Both of them were highly regarded, and the phones in their workshops rang frequently new customers were constantly calling them. However, Hora prospered, while Tempus became poorer and poorer and finally lost his shop. What was the reason? ... The watches the men made consisted of about 1,000 parts each. Tempus had so constructed his that if he had one partly assembled and had to put it down to answer the phone say-it immediately fell to pieces and had to be reassembled from the elements. The better the customers liked his watches, the more they phoned him, the more difficult it became for him to find enough uninterrupted time to finish a watch ... The watches that Hora made were no less complex than those of Tempus. But he had designed them so that he could put together subassemblies of about ten elements each. Ten of these subassemblies, again, could be put together into a larger subassembly; and a system of ten of the latter subassemblies constituted the whole watch. Hence, when Hora had to put down a partly assembled watch in order to answer

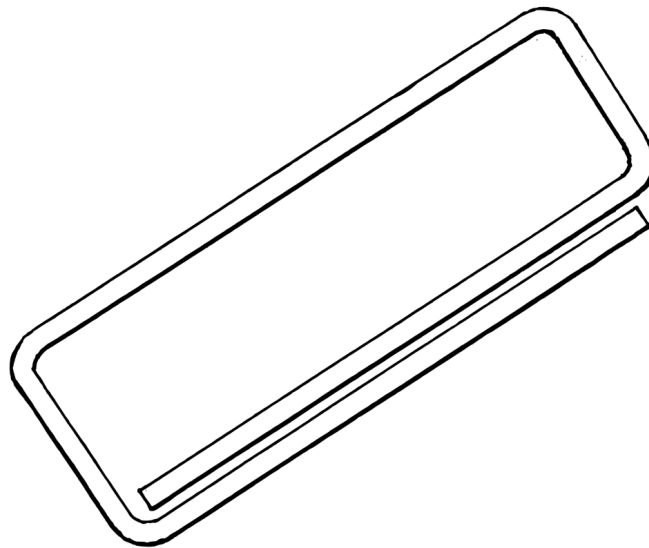
the phone, he lost only a small part of his work, and he assembled his watches in only a fraction of the man-hours it took Tempus.”

Simon emphasized how complexity was efficient because it was built from simple units, how chains came to be, and how the rich got richer (RGR). He was in search of a probabilistic mechanism to explain the rich get richer phenomenon.

Polya's Urns

Another thought experiment by George Pólya, a Hungarian Mathematician came up with the ‘Polya Urn’ [2] thought experiments, a type of statistical model used as an idealized mental exercise framework, unifying many treatments. In an urn model, objects of real interest (such as atoms, people, cars, etc.) are represented as colored balls in an urn or other container. In the basic Pólya urn model, the urn contains x white and y black balls. A ball is drawn randomly from the urn and its color is observed. It is then returned in the urn, an additional ball of the same color is added to the urn, and the selection process is repeated. The urn experiment is used to understand the evolution of the urn population and the sequence of colors of the balls drawn out.

Both Simon and Polya tried explaining the self-reinforcing property of RGR. How complexity endowed an object with longevity. The simplicity of these thought experiments is engaging and can be tried at home with a box of paper clips. Take a paper clip, join it with another paper clip put it back in the heap, and repeat the experiment. The first paper clip you picked up from the heap will eventually turn out to be the longest chain among many other competing paper clip chains.



The paper clip patented by Johan Vaaler in 1899 and 1901.

The chronology of RGR began in 1845 with the 'Branching Process'. Irénée-Jules Bienaymé was the first to explain mathematically the observed phenomenon that family names, both among the aristocracy and among the bourgeoisie, tend to become extinct [3]. The branching process was later explained by Antoine Augustin Cournot in 1843 [4], by Francis Galton and Henry William Watson in 1875 [5], Nikolay Semyonov in 1954 [6] explained the branching process in chemical reactions and Per Bak talked about the branching process as Self organized criticality in 1999 [7].

The branching process was looked at as a gambler's problem by Cournot.

"The gambler at the beginning of the game buys a ticket and in the second round uses all the money (if any) that he won in the first round to buy new tickets. The game continues so that all the money, won in the preceding round, is used to buy tickets in the next round. What is the probability that the gambler will eventually go bankrupt?"

The simplicity of the RGR experiments was their ability to transform one thought experiment into another. While Bak looked at real economics as a pile of sand as every discretion was discrete. The sand grains, which move during the original toppling, were equivalent to sons. The sandpile was a branching process where you keep adding a grain of sand to the hill till it breaks down. These moved grains could cause further toppling, resulting in the motion of more grains, which were equivalent to grandsons, and so on. The total number of displaced grains is the size of the avalanche and was equivalent to total offspring in the branching process.

In 1896 Vilfredo Pareto, father of microeconomics in "Cours d'économie politique" [8] was the first to look at the income distribution patterns in Italy and how 80% of the wealth was owned by 20% of the rich Italians. The father of microeconomics has a well-deserved right to have the RGR phenomenon sometimes called the Pareto curve. D. G. Champernowne in 1953 gave mathematical proof to Pareto's income distribution in 'A Model of Income Distribution' [9].

"The forces determining the distribution of incomes in any community are so varied and complex, and interact and fluctuate so continuously, that any theoretical model must either be unrealistically simplified or hopelessly complicated. We shall choose the former alternative but then give indications that the introduction of some of the more obvious complications of the real world does not seem to disturb the general trend of our conclusions."

In 1907 A.A. Markov moved the RGR thought beyond branching and income distribution. He extended the weak law of large numbers and the central limit theorem to certain sequences of dependent random variables forming special classes of what are

now known as Markov chains in “Extension of the law of large numbers to dependent variables” [10].

Markov showed that probability does not always converge towards a certain average, but sometimes diverges. He applied his chains to the distribution of vowels and consonants in A. S. Pushkin’s poem “Eugeny Onegin” offering another illustration of the rich get richer phenomenon. How for the most popular words, the probabilistic chains increase in length proportionally. Eggenberger and Polya in 1923 wrote about Markov’s model and reformulated the idea as Polya’s urn. [11]

Though Markov was the first to talk about probabilities and words, it was not until 1916 that the subject of ‘Computational Linguistics’ started taking shape. Estoup, J.B., was the first to study a structure in man-created texts. In ‘Gammes sténographique’ [12] he explained how the structure of the language remains stable across different languages, and to some extent can be traced out as well in certain animal communication systems. Intelligence worked regardless of the species and obeyed certain rules of information theory. The subject of computation linguistics was further developed by Condon in 1928 in “Statistics of Vocabulary” [13], followed by Kingsley Zipf “Principle of least effort” published in 1949 [14], Herbert Simon in 1955 [15], P.A.P. Moran in 1958 [16], Samuel C. Bradford in 1934 [17] and S. Price in 1965 [18] re-expressed the idea by replacing words by genes and words by the number of articles in scientific journals. The RGR in language was extensively published and cited.

“Consider a book that is being written, and that has reached a length of N words. With probability α the $(N+1)$ st word is a new word – a word that has not occurred in the first N words. With probability $1-\alpha$, the $(N+1)$ st word is one of the old words. The probability that it will be a particular old word is proportional to the number of its previous occurrences.”

Herbert Simon

Moran replaced words with genes and showed that if at each step one selected at random gene dies and it is replaced by a random gene from the pool and add a gene of the same type, what was the probability that the added gene can mutate?

1922 was about botany. J C Willis, a botanist suggested that the area occupied by a given species (of plants) at any given time in any given country in which there occur no well-marked barriers depends upon the age of that species in that country [19]. In other words, the older the species is, the wider is its range. Another principle derived by the extension of that of age and area is ‘size and space’. A group of large genera would occupy more space than a group of small genera. It followed that size went with age and the ‘laws’ of age, size, and area explained distribution.

Reduced to its simplest expression Willis's belief was that evolution was not haphazard but proceeded according to certain definite RGR law. He held the view that natural selection could not be regarded as an adequate working explanation of evolution, but the law of 'Dichotomous Divergent Mutation' (an RGR expression) according to which the larger groups of organisms came about as the result of sudden and considerable changes, rather than by the gradual accumulation of small variations. He theorized that the larger groups or categories of organisms originated first, and only subsequently differentiated into what was now their constituent genera and species. Evolution for Willis worked 'downwards' and not 'upwards', as was inherent in the theory of natural selection.

George Udny Yule [20] replaced genera with particle and species with energy to showcase newer relationships between age and energy. With the probability distribution of age, you could find the probability distribution of the energy.

The Psychophysical law [21] extended the analogy to brightness and power. The exponent is about $1/3$ i.e. if we increased the luminous power of the light source 8 times it would appear only two times brighter. $(8)^{1/3}$

In 1941 P.J. Flory illustrated RGR in molecular distribution. In "Molecular size distribution in three-dimensional polymers" [22] he explained how polymeric substances are distinguished at the molecular level from other materials by the concatenation of atoms or groups to form chains, often of great length. In the attributes of long molecular chains, the skeletal bonds of the molecular chain were thus likened to the steps in a random walk in three dimensions, the steps being uncorrelated one to another. He showed how the molar volume should be proportional to the mean-square separation of the ends of the chain.

A monkey stole the magic

The magic-like properties of RGR have captured the scientists like a child, jumping with joy, clapping, and saying, "let's do it again". The fun lasted for 150 years and pulled in a generation of thinkers, in awe of the experiment, starting from computation linguists, to statisticians, to mathematicians, to physicists, to chemists, to psychologists. There were few left untouched by the paper clip experiment and all it needed was a monkey to steal the magic. George Miller, father of cognitive psychology and cognitive science used the analogy of a monkey [23] with a typewriter and re-explained the RGR mathematics to prove that the magic was not in the human language but in the information structure. There was no magic in the frequency and proportion of Zipf's law which stated that the most frequent word would occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word, and so on.

We may continue to be enamored by the RGR laws, the law of Pareto, the 80-20 rule, or whatever name we may give it, a monkey with a typewriter has put the generational

experiment in question, not because of it being a monkey but because of the law's universality that the magic had nothing to do with humans, monkeys, crystals or polymers, the expression just happens.

Champernowne's forgotten findings

When a subject is as old as 1845 and a few generations of larger-than-life scientists are connected in a chain of confirming thought processes, an RGR thought experiment, you need something beyond courage to show that the structure had cracks. You need to start with an acceptance that Science is not just about standing on the shoulder of giants but also about being able to question. Fortunately for me, there was always a multiplicity of competing thoughts, oscillating in gravity, as the weight shifted from one end of the scale to the other.

Champernowne published work on what is now called the Champernowne constant in 1933, whilst still an undergraduate at Cambridge. In 1948, working with his old college friend Alan Turing, he helped develop one of the first chess-playing computer programs, Turochamp [24]. Both Champernowne and Turing were good chess players and Turochamp certainly could not have given them a good game, so Champernowne taught his wife the basic moves so that Turochamp might be tested against a total beginner.

With such a brilliant start to his undergraduate career, one might have expected Champernowne to carve out for himself a career as a top mathematician. However, one of his lecturers at Cambridge was John Maynard Keynes who quickly spotted Champernowne's potential and interest in economics. Advised by Keynes, Champernowne completed his mathematics degree in two years and then embarked on a second degree in economics. From October 1934 his work in economics was supervised by Keynes. He ended up with two first class degrees, one in mathematics and the other in economics.

Champernowne's 1953 mathematical proof for Pareto's income distribution highlighted some divergent findings.

First: The model removed the assumption that there was the lowest income range and set up conditions that lead to a two-tailed distribution, one for the poor and one for the rich. In actual income distributions, Pareto's law was not even approximately obeyed for low incomes. If the logarithm of income was measured, the frequency distributions found in practice were not J-shaped, but single-humped and moderately symmetrical. Two sets of distributions for two extremities in the income data set.

Second: There was some degree of regularity regarding a tendency for the lowest incomes to shift upwards by rather more ranges on the average than the high incomes.

The extremes behaved differently. The bottom was more reverting than the top, which expressed relatively weaker momentum.

Third: The paper observed a continuous downward curvature in the Pareto curve.

“For example, a rich man's income must be allowed some risk through death or misadventure of being degraded to a lower range in the following year; but the incomes in the lowest range cannot by definition be allowed this possibility. There is a noticeable tendency recently for the Pareto curves of the United Kingdom and other countries to curve very slightly downwards at the tail.”

Fourth: Champernowne warned that although the models discussed threw some light on the reasons why approximate obedience of Pareto's law is so often found in actual income distributions, they did not throw much light on the mechanism determining the actual values observed for Pareto's exponent. Now, it was Champernowne along with Simon, who started wondering about the mechanism for RGR.

The RGR Challengers

In her book, *Complexity: A Guided Tour*, Melanie Mitchell, 2009 mentions how too many phenomena are being described as power-law or scale-free. Data used by Laszlo Barabasi [25] and colleagues for analyzing metabolic networks came from a web-based database to which biologists from all over the world contributed information. Such biological databases, while invaluable to research were invariably incomplete and error-ridden. Several networks previously identified to be “scale-free” using curve fitting techniques were later been shown to in fact have non-scale free distributions. Considerable controversies over which real-world networks are scale-free.

Evelyn Fox Keller mentioned that the current assessments of the commonality of power laws (RGR laws) are probably overestimated [26]. Preferential attachment (RGR) was not necessarily the one that naturally occurs in nature. There turn out to be many ways of constructing power laws, and every single of them was right. It was not obvious how to decide which ones are the mechanisms that are causing the power-law mechanism in the real world. Even for scale-free networks, there were many possible causes for power-law degree distributions in networks.

Back to 1738

Now that we have some food for thought, we can go back and look at the giants from the other side of the generational spectrum. The poor get richer (PGR) side, the version referred to as normality. The search for a mechanism could not have happened back in 1738. Mathematics had to evolve, probability had to be defined, printing press technology had to establish, and statistics had to become mainstream. Without all these steps we could not have had the luxury of debating what works, the RGR or the PGR.

"Its appearance is so frequent, and the phenomena in which it appears so diverse, that one is led to the conjecture that if these phenomena have any property in common it can only be a similarity in the structure of the underlying probability mechanisms."

Herbert Simon

It was in 1738 that Abraham de Moivre [27] did the earliest work on normal distribution by approximating the rule for the binomial coefficients. Since there was no probability Moivre could not have been expected to understand the concept of the probability density function.

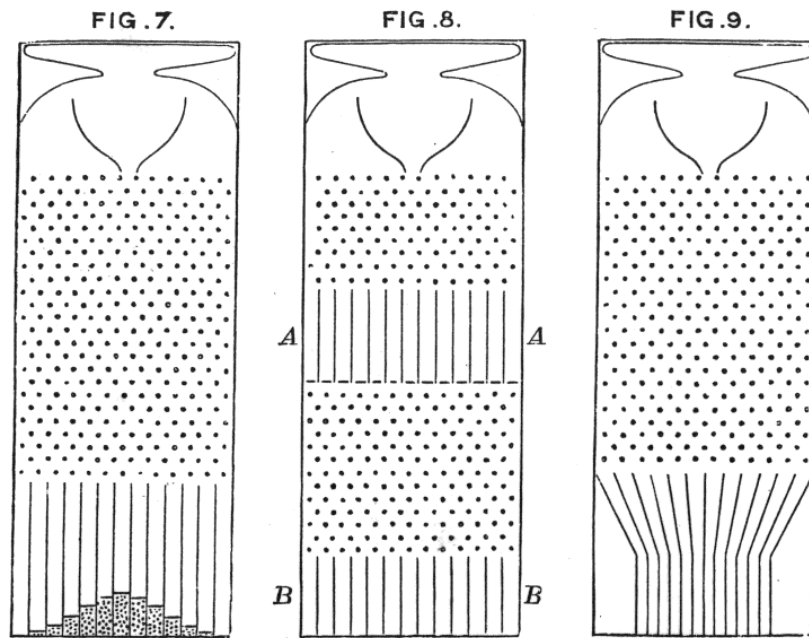
In 1774 Pierre-Simon Laplace [28] proved the fundamental central limit theorem, which emphasized the theoretical importance of the normal distribution.

In 1809 Robert Adrain [29], an Irish mathematician published two derivations of the normal probability law, simultaneously and independently from Gauss.

In 1823 Carl Friedrich Gauss published his monograph "Theoria combinationis observationum erroribus minimis obnoxiae" [30] where among other things he introduced several important statistical concepts, such as the method of least squares.

In 1860 James Clerk Maxwell [31] demonstrated that the normal distribution is not just a convenient mathematical tool but occurred in natural phenomena.

In 1873 Charles S. Peirce [32] defined the term 'normal' as not the average of what actually occurs, but of what would, in the long run, occur under certain circumstances. In 1894 Francis Galton [33] created the bean board which was called the first generator of normal random variables and demonstrated the normal distribution.



The quincunx, as drawn by Sir Francis Galton.

RGR was vulnerable and Normality could not be dismissed

David N Esch [34] in the Journal of Investment Management addressed the non-normality facts and fallacies. The author reinitiated the century-old debate by suggesting that normal efficient models could not be simply rejected i.e. market rationality could not be just junked.

Power-law (RGR) distributions were not alone. The power law was ubiquitous, but they were not the only form of broad distribution. Gaussian distributions (PGR and rich get poorer i.e. RGP) tend to prevail when events are completely independent of each other. As soon as you introduce the assumption of interdependence across events, Paretian distributions tend to surface because positive feedback loops tend to amplify small initial events. For example, the fact that a website has a lot of links increases the likelihood that others will also link to this website.

Barabasi admitted that it's tough to explain why sometimes preferential attachment fails and a second-mover like Google overtook the first-mover like Yahoo. He called it the luck factor. In other words, the 'Rich' is supplanted by the 'Not Rich' or 'Poor' because of luck.

It's hard to challenge lack of luck as a factor, but such thinking does not take us to a mechanism thinking. For science to progress, we needed to think of intelligence as a mechanism.

The Height Mechanism

1886 was the year, Francis Galton wrote the “Regression to mediocrity” [35] paper in which he explained how society did not just become taller. Nature preserved a balance by making children of tall parents a bit shorter, trending towards group average, and the children of short parents, a bit taller trending upwards towards the group average (Upwards and Downwards). This was the first paper after 150 years of work done on normal distribution which indicated a movement, a dynamism, in a group, towards the mean. The paper gave birth to the modern idea of mean reversion i.e. the idea of PGR and RGP.

PROCESS THROUGH WHICH THE DISTRIBUTION OF STATURES, IN SUCCESSIVE GENERATIONS OF THE SAME PEOPLE, REMAINS UNCHANGED.

Height in inches.	Deviation in inches.	Statistical Distribution of Statures in the several Systems of							
		GENERATION I.		MID-PARENTS.		GENERANTS.		GENERATION II.	
73	—	2	Concentrates.		Concentrates.		Disperses.	2	Upper Quartile.
72	+ 4	6		2		1		6	
71	+ 3	10		6		7		10	
70	+ 2	15		16		13		15	
69	+ 1	17		26		29		17	MEDIAN.
68	0	17		26		29		17	
67	— 1	15	Concentrates.	16	Concentrates.	13	Disperses.	15	Lower Quartile.
66	— 2	10		6		7		10	
65	— 3	6		2		1		6	
64	— 4	2						2	
Total	100		100		100		100	
Probable derivation	...	1·7		1·2		0·8		1·7	

NOTE.—The cases are symmetrically disposed above and below the common mean value of 68½ inches. The Upper and Lower Quartiles are the values that in each case divide the number of cases above the Median or mean value, and those below it, respectively into equal parts. Thus in each column there are (1) 25 cases per cent. above the Upper Quartile, (2) 25 cases between the Upper Quartile and the Median, (3) 25 cases between the Median and the Lower Quartile, (4) 25 cases below the Lower Quartile. The difference between either Quartile and the Median is technically called the “Probable” deviation.

Language Learning Mechanism

Zipf in his book, “the path of least resistance” talked about a language learning mechanism with two opposing forces. The speaker’s force tended to reduce the size of the vocabulary to a single word by unifying all meanings behind a single word. Opposed to this unification is the second force, the listener’s force that would tend to increase the size of the vocabulary to a point where there will be a distinctly different word for each different meaning, a force of diversification. Language learning was constantly subject to the opposing forces of unification and diversification.

Gaussian (PGR) morphing into Paretian (RGR)

“Gaussian distributions can morph into Paretian distributions under two conditions – when tension increases and when the cost of connections decreases. In our globalizing economy, tension rises as competitive intensity increases and as business landscapes evolve faster than the capacity of most organizations to adapt. At the same time, costs of connections are rapidly decreasing as public policy shifts towards the freer movement of goods, money and ideas and rapid improvements in the price-performance of IT infrastructures dramatically reduce the cost of information transmission. Bottom line: Paretian distributions become even more prevalent.”

Andriani and McKelvey [36]

In the papers, ‘Power laws, Pareto distributions and Zipf’s law, M. E. J. Newman (2006) [37] and N. Jan, L. Moseley, T. Ray, and D. Stauffer is the fossil record indicative of a critical system? [38] the authors explain how Pareto and Galton could be reconciled.

“The two distributions can be reconciled. For example, if we consider one of the most famous systems in theoretical physics, the Ising model of a magnet. In its paramagnetic phase, the Ising model has a magnetization that fluctuates around zero. Suppose we measure the magnetization ‘m’ at uniform intervals and calculate the fractional change ‘ $\delta = (\Delta m)/m$ ’ between each successive pair of measurements. The change ‘ Δm ’ is roughly normally distributed and has a typical size set by the width of that normal distribution. The $1/m$, on the other hand, produces a power-law tail when small values of m coincide with large values of Δm .”

Natural systems are replete with phase changes like the Ising model [39]. And since aspects of the same natural systems can exhibit the two respective distributions, this suggests that both distributions could not only co-exist in natural systems but also could be linked with its dynamic nature.

So where did the pin come from?

How can we assume that nature operates like one heap of paper clips? What if there were a finite set of pins and for every pin that attracts newer pins, there were heaps that were repelling pins from its chains, a kind of detachment (preferential detachment). Nature could make us think it was a box of pins making chains, but it is more complex and built from states. There is a state which attracts pins and a state which repels pins. A state that grows at the expense of the state that decays. Galton’s mechanism of nature may have observed the boundaries of a mechanism, but he did not explore what was happening inside this group that powered the forces of unification in height.

The two forces of reversion and diversion operate together. As easy as this may sound, nature can’t be considered an isolated system. New York does not become the busiest

city because it attracts people and visitors, it becomes a bigger and busier city because some cities see a declining population and popularity. This dual-state simplicity that something has to give for something to have has been ignored by the generation of great minds of the RGR thought experiments. Maybe it was their time to observe and not question the observation.

A reconciliation between the two ideas, the RGR and the PGR, or the talk of a mechanism that generates the two, is essential scientific research, as it changes how we look at statistics, physics, mathematics, chemistry, psychology, history, computational linguistics, cancer, climate, stock markets, anything which deals with unpredictability, uncertainty along with snowballing trends, criticality and complexity.

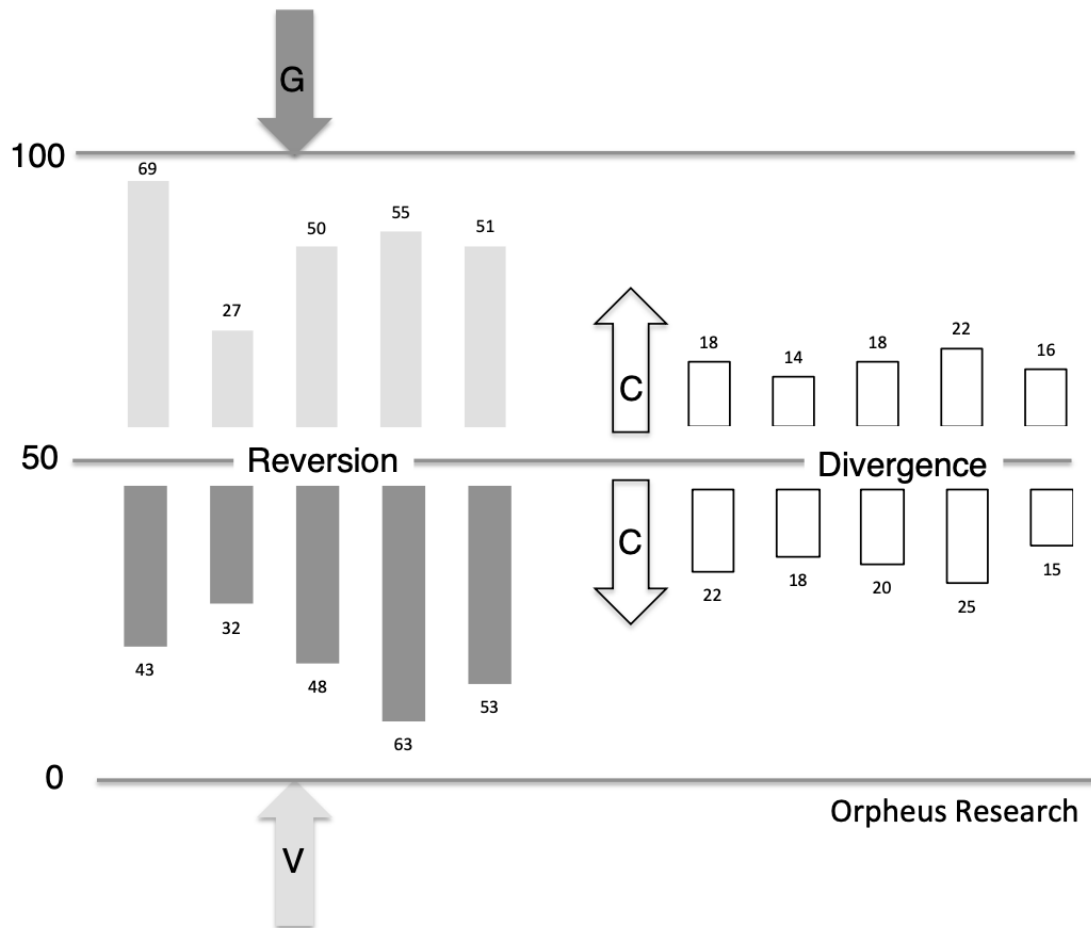
The search for a mechanism is a tall aspiration, which is married to transforming economics and finance back to its roots in fluid dynamics. It was another chain that started from Joseph Valentine Business's fluid dynamics [40] and Markov that inspired Louis Bachelier [42], who was lost in history and then found by Paul Samuelson [43] which created modern finance.

Reversion – Diversion Mechanism

In 2015, I went about recreating the Galton's height experiment in the "Mean Reversion Framework" [44] assuming the S&P 500 was the population and the return series of each stock were generations of smart agents (stocks). I created a set of quintiles based on relative price performance. The bottom quintile (V), the middle three quintiles (C), and the top quintile (G).

According to Galton, the extremes tend to reverse. This means in my experiment both the V and G should have expressed reversion. And since divergence is a needed requirement or an expected balancing force, the rankings around the C middle at 50 rankings should have expressed divergence. I took five global market benchmarks and tested them for 15 years of data for 20, 40, 60, 125, 250, 500, 750, 1000, 1250, 1500, 2500, 3750-day periodicities.

The stock prices were ranked on quarterly performance and ranked on a scale of 0 to 100 and then tested for absolute percentage of components moving from V to G bin and percentage of G component bins moving back to V. This was a test for extreme reversion. The C was tested for divergence from average 50 rankings back to V and G. The results confirmed reversion and divergence.



The Snowball Effect

Intelligence is a mechanism that is dynamic, architectural, chaotic, non-linear, flits perpetually between equilibrium and disequilibrium, reduces information entropy, transforms disorder into order, is normal and non-normal [45] at the same time, lives in probabilistic informational states like RGR-RGP-PGR-PGP [46], which are always in a conflict but as a whole are aware of where they are headed.

The snowball effect was never just about a snowflake becoming a snowball, it was always about the snowball breaking down back into snowflakes, forcing the snowflake to start its journey back up to the mountain top, where it could start the process all over again and we could all wonder, was the snowflake lucky or a persevering genius.

Bibliography

[1] Simon H., "The Architecture of Complexity", Proceedings of the American Philosophical Society, Vol. 106, No. 6. (Dec. 12, 1962), pp. 467-482

- [2] Pólya, G. "Ueber eine Eigenschaft des Gauss'schen Fehlergesetzes". In: Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928. vol. 6. pp. 63–64.
- [3] Bienaymé, I. J., "De la loi de la multiplication et de la durée des familles", 1845
- [4] Cournot, A. A., "Exposition de la théorie des chances", 1843
- [5] Watson, H. W., Galton, F., "On the probability of the extinction of families", Journal of the Anthropological Institute 4, 138, 1875
- [6] Semenov, N., "Some Problems of Chemical Kinetics and Reactivity", 1954
- [7] Bak, P. "How Nature Works: the Science of Self-Organized Criticality", 1999
- [8] Pareto, V., "Cours d'économie politique", 1896
- [9] Champenowne, D. G., "A Model of Income Distribution", The Economic Journal, Vol. 63, No. 250, pp. 318-351, 1953
- [10] Markov, A.A., "Extension of the Law of Large Numbers to Dependent Variables", Izv. Fiz.-Mat. Obshch. Kazan. Univ., ser. 2, vol. 15, pp. 135-156, 1906
- [11] Eggenberger, F. and Pólya, G. "über die Statistik verketteter", Vorgänge. Z. Angew. Math. Mech., 3, 279-289., 1923
- [12] Estoup, J.B., "Gammes sténographique", 1916
- [13] Condon, E., "Statistics of Vocabulary", 1928
- [14] Zipf, K., "The Principle of Least Effort", 1949
- [15] Simon, H.A., "On a Class of Skew Distribution Functions", Biometrika, 1955
- [16] Moran, P.A.P., "Random processes in genetics" Proc. Cambridge Philos. Soc. 54, 60-72., 1958
- [17] Bradford, S. C., "Sources of information on specific subjects. Engineering, 26, p. 85–86. 1934
- [18] Price, D. S., "A general theory of bibliometric and other cumulative advantage process", Journal of American Society for Information Science 27, 292, 1976

- [19] Willis, J.C., Yule, G.U., "Some statistics of evolution and geographical distribution in plants and animals, and their significance", *Nature*, 109, 177, 1922
- [20] Yule, G.U. "A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S.", *Philosophical Transactions of the Royal Society of London B*, 213, 21., 1925
- [21] Stevens, S.S., "Neural events and the psychophysical law", *Science* 170, 1043-1055, 1970
- [22] Flory, P.J., "Molecular size distribution in three-dimensional polymers", *J.Phys.Chem*, 1941
- [23] Miller, G. A., "Some Effects of Intermittent Silence", *The American Journal of Psychology*, Vol. 70, No. 2., pp. 311-314, 1957
- [24] O'Connor, J. J., Robertson, E. F., "Biography of David Gawen Champenowne", *School of Mathematics and Statistics, University of St Andrews, Scotland*, 2014
- [25] Barabasi, A.L., Albert, R., "Emergence of scaling in random networks", *Science*, 286 509. 1999
- [26] Keller, E. F., "Making Sense of Life: Explaining Biological Development with Models, Metaphors and Machines", 2002
- [27] de Moivre, A., "The doctrine of chances: or, a method for calculating the probabilities of events in play; W. Pearson, 1718
- [28] Laplace, P. S., "Mémoire sur la probabilité des causes par les événements", 1774
- [29] Adrain. R., "Research concerning the probabilities of the errors which happen in making observations", *The Analyst, or Mathematical Museum*. Vol. I, Philadelphia: William P. Farrand and Co., 1808
- [30] Gauss, C. F., "Theoria combinationis observationum erroribus minimis obnoxiae", 1823
- [31] Maxwell, J. C., "V. Illustrations of the dynamical theory of gases. — Part I: On the motions and collisions of perfectly elastic spheres". *Philosophical Magazine*, 1860
- [32] L. Welby, "Writings of Charles S. Peirce", 1873
- [33] Galton, F. "Natural Inheritance", 1894

- [34] Esch, D. N. "Non-Normality facts and fallacies," JOIM, 2010
- [35] Galton, F, "Regression to mediocrity", 1886
- [36] Andriani, P., McKelvey, B., "Redirecting organization science toward extreme events and power laws", 2007
- [37] Newman, M. E. J., "Power laws, Pareto distributions and Zipf's law", 2006
- [38] Jan, N., Moseley, L., Ray, T., Stauffer. D., "Is the fossil record indicative of a critical system?", Advances in Complex Systems, 1999
- [39] Ising, E., "Beitrag zur Theorie des Ferromagnetismus", Z. Phys., 1925
- [40] Businessq, J. V., "Théorie de l'écoulement tourbillonnant et tumultueux des liquids", 1897
- [42] Bachelier, L., "University of Paris", 1900
- [43] Samuelson, P., "Foundations of Economic Analysis", 1946
- [44] Pal, M., "Mean Reversion Framework", SSRN, 2015
- [45] Simkin, M.V., Roychowdhury, V.P., "Re-inventing Willis", Physics Reports, 2006
- [46] Pal, M., "How Physics solved the wealth problem", SSRN, 2017