

# Regression Explanation and Statistical Autonomy

Joeri Witteveen

Section for History and Philosophy of Science  
Department of Science Education  
University of Copenhagen

*Accepted for publication in Biology & Philosophy*

## Abstract

The phenomenon of regression toward the mean is notoriously liable to be overlooked or misunderstood; regression fallacies are easy to commit. But even when regression phenomena are duly recognized, it remains perplexing how they can feature in explanations. This article develops a philosophical account of regression explanations as “statistically autonomous” explanations that cannot be deepened by adducing details about causal histories, even if the explananda as such are embedded in the causal structure of the world. That regression explanations have statistical autonomy was first suggested by Ian Hacking and has recently been defended and elaborated by André Ariew, Yasha Rohwer, and Collin Rice. However, I will argue that these analyses fail to capture what regression’s statistical autonomy consists in and how it sets regression explanations apart from other kinds of explanation. The alternative account I develop also shows what is amiss with a recent denial of regression’s statistical autonomy. Marc Lange has argued that facts that can be explained as regression phenomena can in principle also be explained by citing a conjunction of causal histories. The account of regression explanation developed here shows that his argument is based on a misunderstanding of the nature of statistical autonomy.

**Keywords:** regression toward the mean; regression explanation; statistical autonomy; statistical explanation; regression fallacy; reversion; heredity; Francis Galton

## 1 The puzzle of regression

In *Thinking, Fast and Slow*, Daniel Kahneman recounts how a flight instructor from the Israeli Air Force once tried to convince him that a well-established psychological principle about skill training is false (Kahneman, 2012). The instructor claimed that he had solid evidence for the unorthodox view that punishment works better than praise to improve skilled performance. Through his extensive experience with evaluating the performance of fighter pilots on executing acrobatic maneuvers, he had noticed a distinct pattern of changes in performance following his feedback. Pilots that he praised for performing exceptionally well tended to perform worse in the next round of exercises. In contrast, pilots that he gave an earful because of their subpar performance tended to do much better in the next round. He hypothesized that praise makes pilots lazy and erodes their skills, whereas punishment increases their determination and energizes them to improve their mastery of demanding flight maneuvers.

Kahneman was quick to realize that the flight instructor had fallen into a trap known as the *regression fallacy*. The pattern he had observed was not due to his feedback but presented a statistical effect: a manifestation of *regression toward the mean*. Since repeated performances of a highly-trained, skilled exercise will be imperfectly correlated—they will fluctuate stochastically around a pilot’s baseline level of skill—it is to be expected on statistical grounds that exceptionally good and bad performances will tend to be followed by more average ones. Hence, the instructor should have expected to discern an ebbing away from excellence and an upturn from underperformance in any case, regardless of his feedback. He was wrong to conclude that the observed pattern challenged the psychological orthodoxy.<sup>1</sup>

The manifestation of regression toward the mean is pervasive, as is the failure to recognize it. Examples abound of phenomena that once prompted a search for causal difference makers, but end up being explained statistically: from business success (Hotelling, 1933), to sports performance (Schall & Smith, 2000), to traffic accidents

---

<sup>1</sup> Note that this is not to say that the instructor *must* have been wrong about the effects of punishment and praise on fighter pilots. Perhaps his feedback did have the hypothesized effect on performance over and above the effect of regression toward the mean. The point is that his observations provide no evidence for this.

(Senn & Collie, 1988), to clinical studies (Morton & Torgerson, 2003), and many other phenomena in science and daily life (Nesselrode, Stigler, & Baltes, 1980; Smith, 2018). In addition to being easily overlooked, the phenomena of regression toward the mean are also prone to being misunderstood. Statisticians often emphasize the “subtlety and deceptive simplicity” of regression phenomena (Stigler, 1999, p. 186) and are aware of the paradoxically “profound and trivial” character of explanations that appeal to regression toward the mean (Senn, 1997, p. 101). Profound, because they can take the mystery out of otherwise puzzling patterns. Yet also trivial, since it seems that regression explanations do not reference empirical laws or causal facts.

Given its pervasiveness and curious character, it is surprising how little attention regression explanation has received in the philosophy of science literature on explanation. Traditionally, the literature on statistical explanation has focused on irreducibly indeterministic and individual-level events (e.g. Hempel 1965, Salmon 1971). Until recently, the only notable philosophical discussion of the population-level phenomena of regression was due to Ian Hacking. In *The Taming of Chance* (1990) and several other essays, Hacking argued that regression explanations have a *statistically autonomous* character. They appeal to statistical laws and concepts not merely to describe or predict population-level phenomena, but to account for their occurrence (Hacking, 1983; 1990; 1992).<sup>2</sup> Hacking stressed that statistical autonomy is a distinctly *explanatory* notion: it implies explanatory irreducibility, not theoretical or metaphysical irreducibility. In other words, a statistically autonomous explanation cannot be “deepened” by adducing details about the causal history of the explanandum phenomenon, even though that phenomenon is part of the causal structure of the world.

Hacking stopped short of elucidating what, exactly, gives a regression explanation its statistically autonomous character and how this yields a distinct explanatory payoff. Recent work in philosophy of science has attempted to fill this void, but with conflicting results. One approach, taken by André Ariew, Yasha Rohwer, and Collin Rice (henceforth: ARR) (Ariew, Rice, & Rohwer, 2015; Ariew, Rohwer, & Rice, 2017; Rice

---

<sup>2</sup> Peter Lipton (2004; 2009) also noted that an appeal to regression toward the mean can be explanatory and stated that regression explanations are statistical explanations. Yet he does not develop this point other than by citing the example from Kahneman.

et al., in press) has been to elaborate Hacking's account of regression as a statistically autonomous phenomenon. They have done this by providing a close philosophical analysis of a historical episode that (as Hacking already pointed out) features the first genuine regression explanation. This historical case is that of the Victorian polymath Francis Galton, from whose writings on heredity ARR aim to "extract some norms of scientific explanation" that shed light on the nature of regression explanation as such (Ariew et al., 2017, p. 64). Another recent study, by Marc Lange (2013; 2017), ends up denying that regression explanations are statistically autonomous. While Lange recognizes regression explanations as a variety of non-causal, "really statistical" explanations that identify a result "as an instance of some characteristically statistical phenomenon" (Lange, 2017, p. 196), he also claims that it is always possible (at least in principle) to provide a causal explanation for the same fact.

In this article, I develop a novel account of regression explanation by way of critiquing the recent accounts of ARR and Lange. I will show that while both accounts contain valuable insights, they also incorporate fundamental misunderstandings about the nature of regression phenomena. In case of ARR, I will argue that their philosophical analysis of Galton's explanatory breakthrough rests on an important misreading of the historical facts. In the lecture that ARR identify as the locus of Galton's first regression explanation, Galton in fact failed to recognize the phenomenon he was studying as a regression phenomenon. Ironically, ARR therefore attempt to extract an account of regression explanation from what happens to be a regression fallacy committed by Galton. Nevertheless, I will suggest that this flaw in ARR's account actually speaks in favor of adopting their historical approach to improving our philosophical understanding of regression explanation. For I will show that the recognition that Galton at first committed a regression fallacy enables us to neatly trace and analyze the conceptual shift he later made, and that led him to a breakthrough in recognizing and explaining the phenomena of regression toward to mean. After developing this historically-sourced philosophical account of regression and statistically autonomous explanation, I will put it to test by comparing it to Lange's conception of "really statistical" explanation. I will show that while Lange gives an essentially correct account of regression explanations in the abstract, he offers a flawed argument for denying their statistical autonomy and for claiming that causal explanations can explain the exact same facts—causally. The only

detailed example Lange provides of the equivalence between regression explanations and conjunctions of causal explanations is problematic. It conceals an important feature of the regression phenomenon that, once exposed, reveals that a conjunction of causes cannot account for it.

## 2 The riddle of reversion

Following Hacking's lead, ARR suggest that we can specify with remarkable precision when and where the first regression explanation was given. It was on the evening of Friday, February 9, 1877, at the Royal Institution in London, where Galton gave a public lecture on the phenomenon of "reversion to mediocrity": the tendency for offspring of parents with extreme characters to have character values that are closer to the population mean. Galton noted that the occurrence of parent-offspring reversion was hard to square with another curious phenomenon. Data from anthropological, biological, and geological studies all showed that successive generations of the same isolated population displayed "a perfect statistical resemblance" to each other under unchanged environmental conditions. Each such generation conformed to the "law of deviation from the average" and approximated (what we now call) a normal distribution with approximately the same mean and dispersion. The combination of these two phenomena, reversion and intergenerational stability of the character distribution, gave Galton the topic for his evening lecture. "Galton's aim in the 1877 essay was ... to explain why the distribution of characters is stable over the course of generations," given that "extremes appear to 'revert' to the population mean" (Ariew et al., 2017, p. 65).

ARR observe that Galton analyzed the problem of explaining the remarkable intergenerational stability of heredity in a particular way: as that of explaining the existence of a *balance* between component processes of heredity. "Galton writes that from the statistical point of view 'uniformity prevails' and that 'the processes of heredity are found to be so wonderfully balanced, and their equilibrium to be so stable, that they concur in maintaining a perfect statistical resemblance' (Galton 1877, p. 282)" (Ariew et al., 2017, p. 67). One of these component processes of heredity was the process of reversion. Galton called the other process "family variability": that which causes offspring of the same parent(s) to depart from their parental character value and to differ from each other. Explaining stability thus came down to explaining why reversion and

family variability kept each other exactly in check and “conspired to maintain an approximately normal distribution” (Ariew et al., 2017, p. 67).

ARR argue that the explanation Galton offered for the maintenance of this cross-generational balance was of an unexpected and unfamiliar sort:

We would expect the explanation for a biological phenomenon to reference biological mechanisms and processes. But, Galton’s explanation for this phenomenon is surprising because his explanation only makes reference to a mathematical result. The intergenerational stability of the distribution of variation is a deductive consequence of the distribution of variation of the previous generation ... [I]f the frequency of characters in the ensemble is normally distributed (under ordinary or equilibrium conditions) then it can be deduced that in the next generation there will be a normal distribution of the same mean and dispersion. Consequently, the exceptional characters will revert as a deductive consequence of the normal distribution. (Ariew et al., 2017, p. 64)

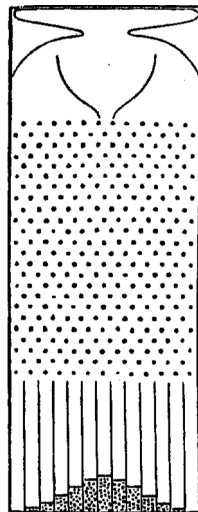
This outline of Galton’s explanation of stability and reversion mirrors the earlier account by Hacking, who similarly argued that Galton deduced the occurrence of stability and reversion from the presence of a normal distribution in the parental generation. He added, moreover, that Galton “did not strictly deduce it, but rather demonstrated it by the device of his shot-dropping machine, the quincunx, in which an analogy of this effect could be observed” (Hacking, 1990, p. 186). However, Hacking stopped short of discussing how this device worked and why it could be taken to simulate a mathematical result that could explain the target phenomenon of intergenerational stability. This is where ARR pick up the thread. They aim to give further expression to Galton’s novel kind of explanation, by providing a close analysis of what the quincunx showed and how it yielded explanatory insights.

## **2.1 Idealize, simulate, deduce**

The quincunx, pictured in Figure 1, derives its name from the geometric arrangement of pins, with each pin in the center being surrounded by four others, like five spots on the face of a die. When a shot of pellets is dropped into the quincunx at the top, each pellet is randomly diverted to the left or right at each row of pins, until it settles in one of the compartments at the bottom. ARR argue that Galton used this device to explain the

occurrence of stability and reversion in two steps.

His first step was to use the quincunx to demonstrate that the distribution of characters in real-world populations approximates a well-known mathematical phenomenon. If a large number of pellets was dropped into the quincunx they would form a distribution of an approximately normal shape, closely akin to the distribution of quantitative hereditary characters in a typical real-world population. ARR note that this suggested that, like the quincunx, real-world populations “involve numerous events that are statistically independent—and, therefore, will approximate the same patterns of large-scale behavior that are largely independent of the details of the individual-level events” (Rice et al., in press, p. 5). In other words, real-world populations appeared to meet certain “minimal material conditions” that made them amenable to investigation by way of studying a mathematical proxy: an idealized statistical model (Ariew et al. 2017, p. 67; Rice et al. in



**Figure 1.** An illustration of a quincunx similar to the one Galton brought to his lecture at the Royal Institution. (Figure 7 from Galton (1889).)

press, p. 6).

Galton’s demonstration that heredity satisfied the minimal material conditions was consonant with his earlier theorizing about the material basis of heredity. Drawing on Darwin’s ideas about pangenesis, Galton had argued that heredity involved the sampling of large quantities of particulate hereditary particles (Galton, 1872). The quincunx helped

to show that one could understand mathematically why such sampling conditions would yield an approximately normal distribution of characters at the population-level, without knowing the mechanism of heredity at the level of the individual.

ARR argue that once Galton had shown that the distribution of characters could be modeled mathematically, he was ready to “turn a puzzle about biological heredity into a mathematical problem. The original question, why do processes of heredity concur to ‘maintain a stable statistical resemblance’? can be restated: how is the normal distribution maintained over the course of generations?” (Ariew et al., 2017, p. 67).

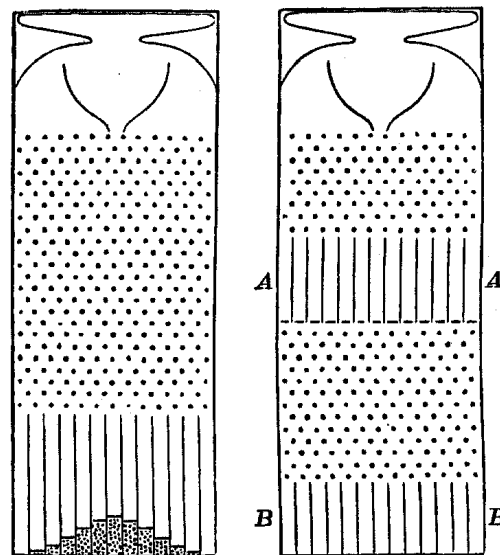
This brings us to what ARR identify as the second step of Galton’s explanation. They argue that Galton explained the intergenerational stability of heredity by showing that recurrence of the normal distribution with the same mean and dispersion was a mathematical result that could be derived from the idealized (mathematical) model.<sup>3</sup> ARR claim that Galton simulated this mathematical result using “a modified version of his quincunx” as depicted in Figure 2. This “two-stage” quincunx is essentially the original quincunx with an additional series of compartments squeezed in between the middle rows of pins. Each of these middle-level compartments contained a trapdoor underneath, which could be opened to release the pellets in the compartment. If all trapdoors were closed and a shot of pellets was released from the top, an approximately normal distribution would form at this mid-level. Opening a single trapdoor would cause a smaller normal distribution to form at the bottom, with its mean positioned exactly below the opened middle-level compartment. So far, nothing new: these effects were familiar to what could be shown using the original quincunx. The novelty of the two-stage quincunx consisted in its capacity to demonstrate the combined result of opening all middle-level compartments. Galton showed that the aggregate of small normal distributions that was thus produced at the bottom once again took the shape of a normal distribution.

---

<sup>3</sup> Although Ariew et al. (2017) present this as the final step, Rice et al. (in press) suggest that Galton’s explanatory schema included a third step: the interpretation of the modeled result as being applicable to the biological phenomenon by “justify[ing] the application of results obtained from highly idealized statistical models to real-world systems.” It is unclear what makes this final interpretation/justification step necessary, since the success of the first step already depended on the justification for the interpretation of the biological problem as a statistical one.



ARR claim that this simulation presented Galton with the solution to his biological problem about heredity (p. 68). If the middle-level compartments were taken to represent a parental distribution of characters and the bottom-level compartments a distribution of offspring characters, the two-stage quincunx showed how the stability of heredity could be derived as a purely statistical result, without making any assumptions about the mechanism of heredity. “The insight is that intergenerational stability could be explained by reference to the mathematical properties of the law of deviation, modeled by the quincunx which instantiates the minimal material conditions required for the law of deviation to hold” (Ariew et al., 2017, p. 68).



**Figure 2.** Galton’s “two stage” quincunx (on the right) next to his original quincunx (on the left). The two quincunxes contain the same number of rows of pins. (Figures 7 and 8 from Galton (1889).)

## 2.2 The problems with ARR’s account

A closer look at the two-stage quincunx suggests that there is something suspect about ARR’s account of Galton’s statistically autonomous explanation for the intergenerational balance of heredity. For a start, notice that the two-stage quincunx does not demonstrate any effect of parent-offspring reversion. Pellets that drop from the “parental” middle-level compartments to the bottom-level “offspring” compartments are not being pulled closer to the population mean. Instead, pellets will on average land in a compartment

exactly below the one they have been released from. Hence, “offspring” values are distributed normally around mean of their “parental” value, rather than being distributed around a value closer to the population mean. This suggests that instead of modeling the balance between reversion and family variability, the two-stage quincunx isolates the effect of family variability acting on its own. It models the creation of variation without representing how, in real-world populations, it is kept in check by the influence of reversion (as Galton thought).

But how could the two-stage quincunx demonstrate the intergenerational stability of heredity if it fails to model one of the processes that, according to Galton himself, is required to maintain stability? In other words, how could Galton have used this device to model the recurrence of a normal distribution *with the same mean and dispersion* if it fails to model the process that stops family variability from *increasing dispersion*? Another look at the two-stage quincunx shows that it indeed fails to produce the result that ARR allege it does. Contrary to what ARR claim, the distribution at the bottom level will not be statistically identical to the distribution at the middle level, but it will be rather more dispersed. It is easy to picture that when the pellets tumble down from the middle-level compartments, the impact of the pins below will further stretch the tails of the middle-level normal distribution and thus “flatten” the bell curve that appears at the bottom.

What has gone wrong here? Did ARR and Galton all fail to notice that the two-stage quincunx provides no demonstration—let alone an explanation—of the recurrence of a normal distribution with the same statistical properties? Or have ARR misrepresented Galton? The proceedings of Galton’s evening lecture settle the matter. They show that Galton neither demonstrated nor even referenced the two-stage quincunx in his lecture. He did sketch it in a letter to George Darwin (a son of Charles), with whom he had been discussing several drafts of his lecture, but then only to illustrate the isolated effect of family variability—exactly as one would expect.<sup>4</sup> However, the proceedings also show

---

<sup>4</sup> Letter from Francis Galton to George Darwin, 12 January 1877, Galton Papers, University College London (GALTON/3/3/7). The illustration of Galton’s two-stage quincunx that Ariew et al. (2017) include in their paper is taken from Stigler (1986), who reproduced it from Galton’s letter. ARR suggest that they are following Stigler’s analysis of how Galton explained intergenerational stability using this device. However, Stigler only asserts that Galton used it “to

that Galton did demonstrate a different device in his 1877 lecture. He did not name it, so let us call it his “balancing quincunx”. In contrast with the two-stage quincunx, the balancing quincunx *did* model the combined effects of family variability and reversion.

ARR’s mix-up of these devices is not without consequence. As we will see in the next section, the problem is not simply that ARR misidentified which device Galton used to forge a breakthrough in scientific explanation. The real problem is that they failed to notice that Galton did not use his balancing quincunx to provide *any* explanation for the occurrence of intergenerational stability and reversion—let alone a statistically autonomous one.

### 3 The riddle revisited

Before examining the balancing quincunx, let us briefly consider an experiment Galton carried out that gave him the ideas for its design.<sup>5</sup> In the mid-1870s, Galton enlisted several friends and acquaintances to help with an experiment on the heredity of weight in sweet peas. He had found that peas were suitable material for a study of heredity: they were prolific, didn’t cross-fertilize (making it easy to track parentage) and the impact of variable environmental conditions on their weight was small. Having obtained a large number of sweet pea seeds from a nursery garden, Galton marshalled them in order of their weight—from light to heavy—and obtained a roughly normal distribution. From this distribution he sourced nine sets of seeds, each composed of seventy seeds selected in the same way: for each degree of deviation from the mean from -3 to +3 degrees he picked ten seeds. He then mailed each set to one of his collaborators across the UK, together with detailed instructions on how to plant the seeds and when to harvest and return the produce. When the results came in several months later, Galton was most satisfied. They taught

---

provide an analogue proof that a normal mixture of normal distributions was itself a normal” (Stigler, 1986, p. 280–281).

<sup>5</sup> Although ARR briefly discuss this experiment, they fail to appreciate its import. They take the outcome of the experiment to be supported by the simulation of two-stage quincunx, rather than the balancing quincunx. Immediately following their discussion of the two-stage quincunx, they write: “This is the same result seen in the sweet pea breeding experiment,” and “The sweet pea experiment acted exactly in the way that the [two-stage] quincunx predicts” (Ariew et al., 2017).

him two important things about the operation of family variability and reversion.

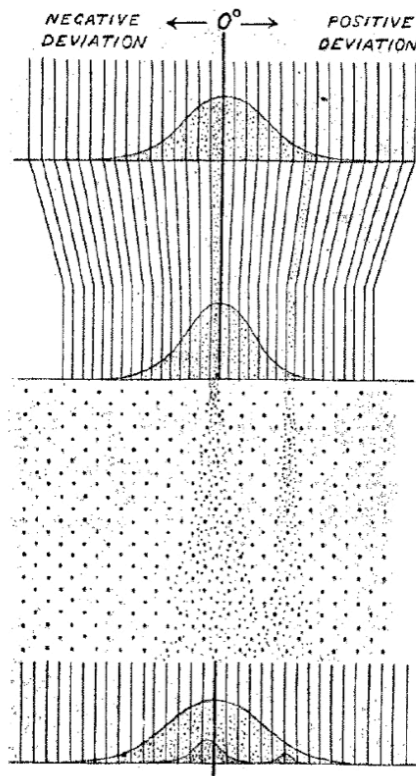
First, Galton reported having been “certainly astonished” to find that for each weight class of the seeds, the produce deviated from the parental value in a statistically identical manner. For each parental value, the offspring complied with the law of deviation and formed an approximately normal distribution with an identical dispersion. This provided compelling evidence that the effect of family variability could be modeled using a quincunxal pattern.<sup>6</sup> The other “great fact” that Galton reported, was that the process of reversion also conformed to the law of deviation and did so independently. The effect of reversion could be described statistically in its own right: “reversion followed the simplest possible law; the proportion being constant between the deviation of the mean weight of the produce generally and the deviation of the parent seed” (Galton, 1877, p. 291). In other words, reversion appeared to be pulling characters towards the mean with an intensity that was proportionate to their degree of deviation.

Based on these findings, Galton could sketch an idealized material model that showed how the combined effects of reversion and family variability ensured the stable recurrence of the normal distribution. This was the balancing quincunx, depicted in Figure 3. The inclined chutes in the top half of this device represent the pull of reversion on the normally distributed parental traits at the top. The quincunxal pins in the bottom half model show how the effect of family variability exactly counterbalances this reverting process. The combination of reversion and family variability thus generate an offspring distribution at the bottom with almost exactly the same mean and dispersion as the parental distribution at the top.<sup>7</sup>

---

<sup>6</sup> If the degree of deviation of offspring character values from the parental mean had varied with the parental character value, the uniform quincunxal pattern would not have been a good model of the action of family variability. Galton recognized this and reported that “if it had been otherwise, I cannot imagine, from theoretical considerations, how the typical problem could be solved” (Galton, 1877, p. 291).

<sup>7</sup> Galton noted that the order in which he modeled these two processes was arbitrary. It was only for modeling purposes that he needed to present the two processes as acting sequentially. Hence, the distribution in the middle is an artifact of the material simulation and has no real-world referent. It should not be mistaken for an intermediate ‘generation’.



**Figure 3.** Galton's balancing quincunx. (Figure 1 from Galton (1877), with slight modifications.)

### 3.1 Switching explananda

The balancing quincunx presented an innovative simulation of Galton's conception of heredity as a "wonderful balance". It showed how two processes whose actions could be *described* in statistical terms could account for another statistically *describable* phenomenon: the recurrence of a normal distribution with the same mean and dispersion. And yet, merely describing and simulating the actions of the component processes of heredity was not what Galton had said he was aiming for. As we saw in Section 2, Galton had begun his lecture by asking *why* the distribution of characters in an isolated population remained statistically identical in successive generations. "Why is it" he asked "when we compare two groups of persons selected at random from the same race, but belonging to different generations of it, we find them to be closely alike?" (Galton, 1877, p. 282). The balancing quincunx did not help answer *this* question. It showed *how* reversion and family variability could generate a recurring normal distribution on the assumption that they maintained an exact balance. But it did not answer *why* reversion and family variability were so precisely balanced. This remained a significant outstanding

question. It indeed seemed rather miraculous that the effects of two entirely independent processes would happen to be exactly counterbalanced.

Galton's surreptitious substitution of a new explanandum ("How is the balance between the processes of heredity maintained?") for his original explanandum ("Why are the processes of heredity exactly balanced?") occurred at the end of the introduction of his lecture. Having just reiterated his opening remark that within each population there exists a "beautiful balance of the processes of heredity", Galton announces: "The outline of my problem for the evening is ... [to] find what the laws of heredity must then be to enable successive generations to maintain statistical identity" (p. 285). In other words, Galton was now asking what the balance of processes had to look like in order to ensure intergenerational stability. In a subtle rhetorical sleight of hand, he withdrew the more ambitious question why this stability existed.

The fact that Galton quietly abandoned the problem that he had announced he would solve already tells us Galton's 1877 lecture did not present the breakthrough in statistically autonomous explanation that ARR ascribe to this lecture. Galton neither explained stability "only by reference to a mathematical result" nor did he show that the phenomenon of reversion was "a consequence of this deduction" about stability. In reality, Galton did not offer *any* explanation for the occurrence of stability or reversion—let alone a statistically autonomous one. His only use of statistics was descriptive. Galton had shown that the presumed balance between hereditary processes could be described statistically, and that this statistical description could be illustrated with help of an analogical device. He did not use statistical concepts and theories to explain why this fortuitous balance existed.

### **3.2 A regression fallacy *avant la lettre***

In hindsight, it is not surprising that Galton failed to offer a statistically autonomous regression explanation in his 1877 lecture. After all, his very conception of heredity at the time betrays that he did not conceive of the intergenerational stability of heredity as the kind of phenomenon that would be amenable to a regression explanation. We have seen that Galton conceived of stability as being due to the combined actions of the (population-level) causal processes of reversion and family variability. In this regard, Galton's reasoning calls to mind the case of the flight instructor from the introduction. He similarly reasoned from the assumption that the manifestation of a reversion in skilled

performance required a distinct cause. It did not occur to the flight instructor that the phenomenon of reversion might not be underpinned by a reverting cause and instead be a purely statistical phenomenon. Galton made essentially the same error.

However, unlike the flight instructor, Galton realized that his analysis of reversion was unsatisfactory. He continued his research on the population-level phenomena of heredity in the years after his lecture at the Royal Institution, and by the mid-1880s he was ready to present a novel account. In 1885, in a Presidential Address at the British Association for the Advancement of Science meeting in Aberdeen, he presented an explanation for what he now called “regression” rather than “reversion”. As we will see next, this subtle change in terminology was accompanied by a watershed in his thinking. Aberdeen, not London, was the birthplace of regression explanation and statistical autonomy.

#### **4 Galton’s explanatory breakthrough**

Galton begun his address in Aberdeen by acknowledging that in his lecture at the Royal Institution eight years earlier he had failed to fully solve the puzzle about heredity that occupied him. He admitted to having been “blind to what I now perceive to be the simple explanation” (Galton, 1886, p. 1207). Before expounding his new explanation, Galton presented findings from a new and richer dataset on the heredity of human stature, while assuring his audience that his explanation was equally applicable to his earlier research on sweet peas, or to any other quantitative heritable trait for that matter.

Galton’s new data were based on measurements of 930 adult children and their respective parentages. After tabulating these data, Galton could tell with “unexpected coherence and precision” by how much offspring would tend to revert to the mean of their “mid-parentage” (p. 1207). The “mid-parent” was a new construct Galton introduced to account for the complications of sex in human heredity. It denoted “an ideal person of composite sex” that represented the weighted average of the contributions of the father and mother in heredity.<sup>8</sup> His charts with height data showed that “the height-deviate of the offspring is, on the average, two-thirds of the height deviate of its mid-parentage” (p. 1209). Having presented these findings, Galton was ready to offer his novel explanation

---

<sup>8</sup> Galton argued that the height of the mother needed to be multiplied by a factor 1,08 before taking the average. The details of this calculation and Galton’s defense of the mid-parent concept need not concern us here.

for the observed pattern in offspring values:

The explanation of it is as follows. The child inherits partly from his parents, partly from his ancestry. Speaking generally, the further his genealogy goes back, the more numerous and varied will his ancestry become, until they cease to differ from any equally numerous sample taken at haphazard from the race at large. Their mean stature will then be the same as that of the race; in other words, it will be mediocre. Or, to put the same fact into another form, the most probable value of the mid-ancestral deviates in any remote generation is zero. (Galton, 1886, p. 1209)

This explanation requires some unpacking in order to appreciate its significance. First, take the somewhat cryptic remark that each individual “inherits partly from his parents, partly from his ancestry.” This needs to be understood in the context of Galton’s theory of ancestral inheritance, which he developed more fully in other writings. Galton thought that although offspring receive about half their hereditary particles from each parent, they correspond to each parent in less than half of their “patent” or “personal” (phenotypically expressed) elements. The remaining proportion of an offspring’s patent elements would correspond to particles that had been “latent” (phenotypically unexpressed) in their parents, but patent in more distant ancestors. This was possible on the assumption that the status of hereditary particles as patent or latent could be reset in a reproductive event: patent particles could become latent and vice versa.

Galton realized that on this view of inheritance it was possible to conceive of an offspring’s phenotype as composed of a combination of random samples that had been drawn from previous generations in different proportions. The largest sample would be drawn from the two parents. For each more distant generation, the size of the sample would decrease, but the number of ancestors that was sampled from would increase. Given Galton’s assumption that the distribution of variation in past generations of the same population had been statistically identical to the present generation, it would be the case that the spectrum of variation in remote generations of ancestors would progressively approximate the distribution of variation in the whole population. And this, in turn, implied that random samples drawn from these remote generations would have an expected value that approached the mean. The expectation from the combination of all samples would therefore be for offspring traits to have a value closer to the mean than the



mid-parental value. Offspring trait values tended to be reverted relative to those of their parents.

#### 4.1 From reversion to regression

To fully appreciate the novelty of this explanation, consider how it compares to Galton's earlier thinking about reversion. We already saw that at the time of his London lecture, Galton reasoned that the phenomenon of reversion had to be underpinned by a causal process of reversion that acted uniformly and with equal strength on all individuals of the same size (as modeled by the inclined chutes in the balancing quincunx). The reason why reversion manifested itself as a mere statistical tendency, he thought, was that its effects were partially counteracted by the process of family variability. Hence, reversion-the-phenomenon could be described statistically, but it needed to be analyzed (and explained) causally: as due to the action of reversion-the-process.

In offering his new account of ancestral inheritance, Galton abandoned this picture of heredity as a balance of distinct hereditary processes. The hypothesis of ancestral inheritance implied that heredity instead consisted of a single, structured sampling process that operated identically in any reproductive episode. But this also raised a pressing new question: if there was no reversion-the-process of which reversion-the-phenomenon was the result, then what sort of phenomenon was reversion? Galton realized that it had to be a purely statistical phenomenon—not the outcome of a causal process that could be described in statistical terms, but a statistical phenomenon as such: a statistical relation between two sets of scores. “Regression” was his name for the phenomenon of reversion after severing it from his earlier causal interpretation.

Note that this means that, ultimately, Galton's “simple explanation” for regression did not appeal to the process of ancestral inheritance itself, but to a statistical property of this process: the *imperfect correlation* between parental and offspring values. His account of ancestral inheritance showed that the expected value of offspring characters could be analyzed statistically as consisting of different components: a parental component whose expected value was the same for the offspring, and a series of variable ancestral components whose expected values were invariably closer to the population mean. This was Galton's statistically autonomous explanation of the phenomenon that had long puzzled him. He had explained one statistical phenomenon (regression toward the mean) by reference to another statistical phenomenon (correlation).

However, one might object that this explanation isn't truly autonomous of the causal facts. For it seems that there is a causal dimension in the background of Galton's statistical explanation of regression toward of the mean. Since his explanation applied to characters that "regressed" *from* their value in the parental generation, *to* their value in the offspring generation, one might argue that it relied on postulating some kind of "process of regression" after all. However, Galton's own further observations show why it would be a mistake to relapse into this causal-processual interpretation.

Looking again at his plots of parent and offspring heights, Galton noticed that apart from the fact that the character values of offspring with the same parental value tended to be closer to the mean, the character values of parents with the same offspring value also tended to be closer to the mean. The latter phenomenon made no sense on a causal interpretation, but it made perfect sense on a purely statistical interpretation of regression. As Galton explained in characteristically lucid terms, the explanation for regression of parents relative to offspring was that: "the number of individuals in a population who differ little from mediocrity is so preponderant that it is more frequently the case that an exceptional man is the somewhat exceptional son of rather mediocre parents, than the average son of very exceptional parents" (Galton, 1885a, p. 1210). In other words, the statistical properties of two distributions could not only help explain why there would be parent-offspring regression toward the mean, but also why there would be offspring-parent regression toward the mean. Both statistical phenomena were statistical effects of another statistical phenomenon: the imperfect correlation between two variables with identical, normally-distributed sets of scores.

## 4.2 The nature of statistical autonomy

At this point it is worth returning briefly to writings of ARR to call attention to one paragraph in which we do find of a glimmer of a recognition of the distinctly statistical nature of regression and its contrast with the biological notion of reversion. At the very end of their discussion of Galton's lecture from 1877, Ariew et al. (2017) make a brief reference to his later writings and cite the monograph *Natural Inheritance* (1889) in which Galton elaborated on his new ideas about regression. ARR note that in this work "Galton no longer calls the phenomenon 'reversion', which traditionally indicated an empirical phenomenon well known to Darwin's contemporaries. Rather, Galton calls it 'regression', Stigler [the statistician and historian of statistics Stephen Stigler] concludes:

‘Galton’s great insight from this new approach was that stability implied ... regression. [T]he entire puzzle was resolved by this one fundamental insight’ (Stigler, 2010, p. 477)” (Ariew et al., 2017, p. 68). While this is correct as such, it is surprising that ARR fail to recognize that this brief remark about regression is completely at odds with their preceding account of Galton on reversion. What is more, immediately following their quotation of Stigler, ARR slide back into this earlier analysis and rehearse their erroneous conclusion: “The insight is that intergenerational stability could be explained by reference to the mathematical properties of the law of deviation, modeled by the quincunx which instantiates the minimal material conditions required for the law of deviation to hold” (p. 68).

We can now see that this conclusion is wrong in at least three respects. First, we have seen that Galton did not give a statistically autonomous explanation of intergenerational stability, but instead assumed the existence of intergenerational stability to offer his statistically autonomous explanation of regression. Second, Galton did not use his quincunxes to help explain either stability or regression (by using them as simulations of mathematical models), but merely used them to describe and explore the properties of the normal curve. Third, ARR misidentify the random sampling process that instantiates the minimal material conditions as Galton’s explanans. In reality, Galton explained the manifestation of regression by showing that it was a consequence of sampling being asymmetric, thus establishing an imperfect (but non-zero) correlation between variables.

These are not separate, isolated errors. ARR’s failure to identify which phenomena Galton explained in a statistically autonomous way and what this explanation looked like are manifestations of a deeper misconception about what it means to give a statistically autonomous explanation. ARR emphasize repeatedly that Galton’s key conceptual move was “to treat the biological phenomenon to be explained as a mathematical problem”. Yet *treating* biological phenomena as a mathematical (statistical) ones is not the same as *conceiving* of them as statistical phenomena in and of themselves. We have seen that Galton’s statistically autonomous explanation was not an explanation of a causal phenomenon treated statistically, but an explanation of a purely statistical phenomenon—a relation between scores of a variable—in terms of another statistical phenomenon. Indeed, we have seen that Galton’s explanatory breakthrough consisted precisely in recognizing that a phenomenon he had previously assumed to be causal (and distinctly

biological) was in fact a purely statistical.

## **5 Statistical autonomy and “really statistical” explanation**

How does the understanding of regression explanation that we have arrived at by reexamining Galton’s researches compare to Lange’s account of regression explanation as “really statistical” explanation? At first gloss, it might seem that the difference between “really statistical” and “statistically autonomous” explanation is merely terminological. Lange correctly asserts that a regression explanation identifies a phenomenon as “fall out from the statistical character of the case”, with this statistical character being that of “a statistical relation rather than a perfect correlation between two variables” (Lange, 2017, p. 190). This is in line with our analysis of regression’s statistical autonomy: regression toward the mean is a statistical phenomenon that is explained by reference to another statistical phenomenon: imperfect correlation. However, as mentioned in the introduction, Lange ends up departing from the idea of statistical autonomy by denying that the explanandum of a regression explanation is irreducibly statistical. He claims that “the same fact as regression toward the mean explains” can be explained through “a conjunction of causal explanations” (p. 190). Moreover, he adds that these separate causal explanations “may themselves be statistical explanations,” albeit of a “fundamentally dissimilar” kind than really statistical explanations (p. 191). Lange does not give a name for this other variety of statistical explanations, but let us call them descriptive statistical (DS) explanations as opposed to really statistical (RS) explanations. DS explanations are statistical only in the sense of abstracting away from particular causal factors. Unlike RS explanations, DS explanations can therefore “be deepened by being supplemented with descriptions of the causal factors” (p. 192).

ARR have responded to Lange by denying that RS and DS explanations are genuinely different kinds of explanation. They argue that on closer examination there is no substantive difference between the explanantia of DS and RS explanations and conclude that these are different formulations of essentially “the same kind of explanation” (Ariew et al., 2017, p. 71; also see Rice et al., in press, p. 20–21). The reason why Lange mistakenly concludes that they are distinct, is that he “ignores the roles of what we’ve been calling ‘approximation’ and ‘minimal material conditions’ in the background assumptions of both his causal and really statistical explanation” (Ariew et al., 2017, p.

71). This appeal by ARR to approximation and mathematical idealization as a unifying features of DS and RS explanations already suggests that something is amiss with their response. ARR appear to be claiming that RS explanations can be assimilated to DS explanations since both *treat* causal phenomena statistically. But we have seen that this is a mistake. RS explanations do not explain causal phenomena by interpreting them statistically, they explain by showing that what may initially appear to be a causal phenomenon is in fact a purely statistical effect.

In what follows, I provide a different analysis of what is wrong with Lange's account of the dual nature of statistical explanation. I submit that the problem with his analysis is not (as ARR submit) that the explanantia of DS and RS explanations are the same, but rather that their explananda are different. In other words, I argue that Lange is wrong to think that facts that can be explained by appeal to regression toward the mean can also be explained causally.

### **5.1 The explananda of DS and RS explanations**

Consider the example Lange gives to support his claim that DS and RS explanations can explain the same phenomena in fundamentally different ways. Suppose we toss a fair coin 100,000 times and assess the relation between consecutive runs of 20 tosses. A run is defined as sharing its final 10 tosses with the next run. Because of this overlap between runs, the expected number of heads in a run depends on the run that precedes it (excluding the first run). For example, a run that is preceded by a run with 18 heads will again tend to contain more than 10 heads, but probably fewer than 18. Lange argues that we can explain this result with either an RS (regression) explanation or a DS explanation: "This result is explained by regression toward the mean. It is also explained by the coin's 50% chance of landing heads on any given toss (independently of the outcomes of other tosses): if we compute the chance that a run with an exceptionally high number of heads will be followed by a run with fewer heads, we will find the chance to be high (For instance, the chance that a run of 20 heads will be followed by a run with fewer heads is just the chance of at least one tail's appearing in the next 10 tosses, which is  $1 - (0.5)^{10} = 1,023/1,024$ )." Lange contends that the latter, DS explanation, is causal, because it appeals to the coin's propensity to produce a "heads" outcome. Such an explanation is causal "in the broad sense ... that it works by describing the relevant features of the result's causal history: the chance of any toss's landing heads and the way that those

chances combine to yield the chance that an extreme run will be followed by a less extreme run” (Lange, 2017, p. 191).

Before considering whether Lange is right that these two explanations explain the same phenomena, let us articulate precisely what this explanatory target is. The question is not just why a run with an above-average number of heads will tend to be followed by a run with fewer heads, but why a run with an above-average number of heads will tend to be followed by a run with fewer, *but more than the average* number of heads. The latter explanandum is the phenomenon that is particular to overlapping runs. (Non-overlapping runs also show a tendency for above-average runs to be followed by runs with fewer heads, but not by runs with more than the average number.)

This clarification of the explanandum tells us that Lange’s rendering of the DS explanation is at least incomplete. The DS explanation for why a run will tend to have a number of heads *closer* to the average cannot just appeal to “the chance of any toss’s landing heads and the way that those chances combine to yield the chance that an extreme run will be followed by a less extreme run,” but also needs to refer to the overlap between runs. This overlap is part of the relevant history that explains the outcome and could perhaps be construed causally as the “inheritance” of coin tosses from the previous run. Lange implicitly adopts this causal interpretation in his example. By presenting a calculation based on “the next 10 tosses” to explain the expected outcome of the next sequence of 20 tosses, he quietly appeals to (but omits mention of) the role of the “inherited” 10 tosses for the causal explanation.

However, I submit that even with these clarifications of how a DS explanation explains, it cannot be maintained that DS explanations causally explain the *same* phenomena that RS explanations explain statistically. To see why, it will be instructive to consider in further detail the case Lange singles out, of explaining the tendency for a run with 20 heads to be followed by a run with a number of heads closer to the average. I will argue that this particular choice of example conceals an important complication that, once exposed, reveals how the explananda of DS and RS explanations come apart.

The complication is related to the orderings (or permutations) of tosses for an initial run with a certain number of heads. An initial run with a combination of 20 heads and 0 tails is peculiar in having only one ordering: there is only way to throw 20 heads in 20 trials, and that is by throwing heads on each and every trial. Any run that follows will

therefore contain at least 10 heads (inherited from the previous run) and have an expected number of 15 heads. Compare this to an initial run with a different number of heads, say, 14. Since there are many possible orderings for a run with 14 heads, we cannot say precisely by how many heads it will with overlaps with the next run. This matters when it comes to predicting and explaining the tendency of the next run. For example, compare an initial run that produced 14 heads in this order THHHTHHHTHHHTHHTTHHH with an initial run that produced 14 heads in this order THHTTHTTHTHHHHHHHHHH. A run that follows the run with the first ordering will tend to have a lower number of heads. (Since the initial run has 7 heads in the last 10 tosses, the expected number of heads in the next run is 12.) In contrast, a run that succeeds the second run will tend to have a higher number of heads. (Since the second run contains 10 heads in the last 10 tosses, the expected number of heads in the next run is 15.) Thus, whether any *particular* run with 14 heads will tend to be followed by a run with more or fewer heads depends on the order in which the heads appeared.

The relevance of this difference between runs with 20 heads and runs with (say) 14 heads becomes clear when we consider Lange's claim that a *conjunction* of DS explanations can *causally* explain the same fact as an RS explanation explains statistically. If Lange is right about this, then the conjunction of causal explanations for why a large sample of runs with different orderings of 14 heads will tend to be followed by a certain number heads must constitute a further causal explanation for why runs with 14 heads as such tend to be followed by runs with a more average number of heads. But why, we should now ask, is this conjunction of causal (DS) explanations supposed to constitute a distinct *causal* explanation in its own right? It would imply that the individual DS explanations for why certain runs with 14 heads will tend to be followed by runs with more heads is *causally* at odds with the conjunctive DS explanation for why runs with 14 heads as such tend to be followed by runs with fewer than 14 heads. Yet, it obviously false to say that the tendency of runs with 14 heads as such "counteract" the tendencies or runs with certain particular orderings in a causal manner. Therefore, the conjunction of causal tendencies that feature in the individual DS explanations cannot itself be a (higher-level) causal tendency; it rather presents a weighted average of the tendencies of runs with particular orderings. The conjunction only becomes explanatory once we recognize that the result it identifies is to be expected on different grounds: on *statistical*

grounds. The only reason why the conjunction of causal histories of a great many runs with different orderings of 14 heads will show that, on average, they are followed by runs with a more average number of heads, is the population-level statistical fact that the “parent” and “offspring” runs are imperfectly correlated. Thus, it is the setup of overlapping runs—with each run being identical to the preceding in one component (the first 10 tosses) and different in another component (the last 10 tosses)—that creates a statistical relation that explains the outcome.

This is bad news for Lange’s claim about the explanatory potential of DS explanation. It suggests that although we can use a conjunction of DS explanations to identify the explanandum phenomenon of an RS explanation, we cannot explain it using that conjunction—let alone explain it causally. The only way to explain the tendency for non-average values of an imperfectly correlated variable to be followed by more average values is to provide an RS (regression) explanation.

## 6 Conclusion

I opened this article with an example of a regression fallacy. In the discussion that followed, I showed that regression fallacies can be stepping stones for understanding how regression explanations work and for appreciating their statistically autonomous nature. Galton’s regression fallacy *avant la lettre* was my prime example. In 1877, Galton showed that reversion could be described statistically but mistakenly assumed that it would need to be analyzed and explained causally. By 1885, Galton realized that the assumption that explanations need to be causal had been the reason for his failure to explain reversion. It finally struck him that reversion was a relation between scores of a variable rather than the effect of a reverting process. What had appeared to be the population-level manifestation of a biological process was in fact a population-level statistical phenomenon that called for a population-level statistical explanation.

This reexamination of Galton’s breakthrough has helped to show where other recent philosophical studies of regression explanation have taken a wrong turn. We have seen that ARR’s attempt at articulating the autonomous statistical nature of regression explanation failed because they conflated reversion and regression. In case of Lange, we have seen that although he provided an accurate characterization of regression explanation as such, he failed to recognize that what makes it really statistical is that it



explains a phenomenon that cannot also be explained causally. *Really* statistical explanation is statistically *autonomous* explanation.

## Acknowledgements

I thank the audience of the Videnskabsteori Seminar at the Niels Bohr Institute and my colleagues in the Section for History and Philosophy of Science for helpful comments and suggestions. This work was supported by a Veni research grant from the Netherlands Organisation for Scientific Research (NWO), grant number 275-20-060.

## References

- Ariew, A., Rice, C., & Rohwer, Y. (2015). Autonomous-statistical explanations and natural selection. *The British Journal for the Philosophy of Science*, 66(3), 635–658.
- Ariew, A., Rohwer, Y., & Rice, C. (2017). Galton, reversion and the quincunx: The rise of statistical explanation. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 66, 63–72.
- Galton, F. (1872). On blood-relationship. *Proceedings of the Royal Society*, 20, 294–402.
- Galton, F. (1877). Typical laws of heredity. *Proceedings of the Royal Institution*, 8, 282–301.
- Galton, F. (1886). Presidential Address, Section H, Anthropology. *Report of the British Association for the Advancement of Science*, 55, 1206–1214.
- Galton, F. (1889). *Natural Inheritance*. London: Macmillan and Co.
- Hacking, I. (1983). The autonomy of statistical law. In N. Rescher (Ed.), *Scientific Explanation and Understanding* (pp. 3–19). Lanham, MD: University Press of America.
- Hacking, I. (1990). *The Taming of Chance*. Cambridge, UK: Cambridge University Press.
- Hacking, I. (1992). Statistical language, statistical truth and statistical reason: the self-authentication of a style of scientific reason. In E. McMullin (Ed.), *The Social Dimension of Science* (pp. 130–157). Notre Dame, IN: University of Notre Dame Press.
- Hempel, C. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.
- Hotelling, H. (1933). Review of *The Triumph of Mediocrity in Business*, by Horace

- Secrist, *Journal of the American Statistical Association*, 28(184), 463–465.
- Kahneman, D. (2012). *Thinking, Fast and Slow*. London: Penguin Books.
- Lange, M. (2013). Really statistical explanations and genetic drift. *Philosophy of Science*, 80(2), 169–188.
- Lange, M. (2017). *Because Without Cause*. Oxford: Oxford University Press.
- Lipton, P. (2004). *Inference to the Best Explanation*. London: Routledge.
- Lipton, P. (2009). Causation and explanation. In: H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *The Oxford Handbook of Causation*. Oxford: Oxford University Press.
- Morton, V., & Torgerson, D. J. (2003). Effect of regression to the mean on decision making in health care. *BMJ*, 326(7398), 1083–1084.
- Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin*, 88(3), 622–637.
- Rice, C., Rohwer, Y., & Ariew, A. (in press). Explanatory schema and the process of model building. *Synthese*.
- Salmon, W. (1971). Statistical Explanation. In: W. Salmon, (Ed.), *Statistical Explanation and Statistical Relevance*, (pp. 29–87). Pittsburgh: University of Pittsburgh Press.
- Schall, T., & Smith, G. (2000). Do baseball players regress toward the mean? *The American Statistician*, 54(4), 231.
- Senn, S. (1997). Editorial -- Regression to the mean. *Statistical Methods in Medical Research*, 6(2), 99–104.
- Senn, S. J., & Collie, G. S. (1988). Accident blackspots and the bivariate negative binomial. *Traffic Engineering and Control*, 29(3), 168–169.
- Smith, G. (2018). *What the Luck?* London: Bloomsbury Publishing Plc.
- Stigler, S. M. (1986). *The History of Statistics: The Measurements of Uncertainty Before 1900*. Cambridge, MA: Harvard University Press.
- Stigler, S. M. (1999). *Statistics on the Table*. Cambridge, MA: Harvard University Press.
- Stigler, S. M. (2010). Darwin, Galton and the Statistical Enlightenment. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(3), 469–482.