\*\* Pre-Print. Published article includes corrections.

**Learning from Non-Causal Models**

Francesco Nappo

\* Department of Mathematics, Politecnico di Milano

**Abstract:** This paper defends the thesis of *learning from non-causal models*: viz. that the study of some model can prompt justified changes in one’s confidence in empirical hypotheses about a real-world target in the absence of any known or predicted similarity between model and target with regards to their *causal* features. Recognizing that we can learn from non-causal models matters not only to our understanding of past scientific achievements, but also to contemporary debates in the philosophy of science. At one end of the philosophical spectrum, my thesis undermines the views of those who, like Cartwright (2009), follow Hesse (1963) in restricting the possibility of learning from models to only those situations where a model identifies some causal factors present in the target. At the other end of the spectrum, my thesis also helps undermine some extremely permissive positions, e.g., Grüne-Yanoff’s (2009, 2013) claim that learning from a model is possible even in the absence of any similarity *at all* between model and target. The thesis that we can learn from non-causal models offers a cautious middle ground between these two extremes.

**1. Introduction\***

One of the central questions in the epistemology of scientific modelling is the ‘problem of learning’ (Grüne-Yanoff 2013). This is the question of when and under what conditions the information gathered from a model may, if not license new inferences, at least justify changes in the degree of credence assigned to empirical hypotheses about real-world targets.[[1]](#footnote-1) Two positions on this problem emerge from the contemporary literature. According to a popular view, a version of which is defended by Cartwright (2009), one condition that must be satisfied in order for the study of a model to prompt learning about a real-world target is that there be similarities between model and target with regards to their *causal features*, i.e., with regards to the kinds of causal connections that obtain in each domain.[[2]](#footnote-2) For instance, tests on mice can justify increasing our confidence regarding the effectiveness of drug X in human populations, according to this view, only if we can expect humans to share with mice biological features (e.g., similar immune systems) that are causally connected to mice’s observed response to drug X.

In recent years, some theorists have begun to depart from this popular answer to the problem of learning. In particular, Grüne-Yanoff (2009, 2013) and Knuutila (2009) have claimed that some scientific models can prompt learning in the absence of any similarity with their targets *altogether*; because of this, Grüne-Yanoff (2009) refers to those models as ‘minimal’. Importantly, the claim that we can learn from minimal models would not only, if correct, invalidate accounts such as Cartwright’s that demand similarities in causal features as a condition on a model’s capacity to prompt learning, but would also overthrow the received wisdom about how epistemic agents reason from observations about a model to new hypotheses about a target. According to this received wisdom, which goes back to Hesse (1963; cf. also Giere 1988; Sudgen 2000; Fumagalli 2016), reasoning from a model consists in a kind of ‘analogical projection’ of the results obtained in a model onto the target, fueled by the consideration of relevant similarities and dissimilarities. By insisting that no similarities *whatsoever* are necessary for learning, defenders of learning from minimal models thus severe this commonly recognized link between learning from a model and reasoning from analogy.

The aim of this paper is to carve out an interesting middle ground between epistemologies (such as Cartwright’s) that restrict the possibility of learning from scientific models to only those cases where model and target resemble each other in causal features and epistemologies (such as Grüne-Yanoff’s) that are instead extremely permissive about the circumstances in which learning from models can occur. The thesis that I will defend is that we can learn from *non-causal* models: i.e., that the study of some models can justify changes in the credence assigned to an empirical hypothesis about some real-world target purely as the result of similarities in mathematical features and in the absence of any known or predicted similarity in causal features.[[3]](#footnote-3) The example that I will discuss is Galton’s (1877) use of his ‘quincunx’ machine to draw conclusions about trait distribution in population biology. In my view, the function that Galton assigns to his model is not the merely heuristic one of suggesting (in the psychological sense) novel hypotheses to be put to test. Instead, I will defend the claim that there is genuine learning involved in the study of Galton’s quincunx model, despite its bearing no resemblance in causal features with any current or past biological populations. Specifically, the hypotheses that Galton’s model justifies assigning additional credence to are claims about the processes of heredity obeying some distinctive and precisely describable statistical pattern.

The claim that we can learn from non-causal models is doubly interesting for the debate on the problem of learning. On the one hand, my thesis refutes those views that restrict the possibility of learning from models to causal models. Importantly, my criticism is broader in scope than the concerns raised by some authors, most notably Sudgen (2009), which target Cartwright’s (2009) account specifically. My objection extends to *any* account that appeals to similarities in causal features to explain a model’s capacity to prompt learning. On the other end, my thesis also helps undermine Grüne-Yanoff’s (2009) claim that we can learn from minimal models. This is because the examples used in support of the latter thesis can often be understood as other instances of learning from non-causal models. Moreover, a plausible account of how it is possible to learn from non-causal models can be provided. Conversely, as Fumagalli (2015, 2016) has persuasively argued, there are serious problems with the epistemological story that lies in the background of the thesis of learning from minimal models. Therefore, even though the thesis of learning from non-causal models and the thesis of learning from minimal models are compatible, I will argue that we should reject the latter as both conducive to epistemological mystery and as unnecessary to account for the case-studies that initially motivated it.

The discussion is organized as follows. Section two presents the contributions of Cartwright (2009), Hesse (1963) and Sudgen (2009) to clarify the basic tenets of the popular view according to which the use of a model can engender learning about real-world targets only if that use is underwritten by similarities in causal features. In section three, I will argue that Galton’s (1877) use of his quincunx model undermines that popular view. In section four, I will emphasize the novelty of my claim within the contemporary philosophical landscape and defend my reconstruction of Galton’s case-study from the objections. In section five, I will lay out my argument that, once the thesis of learning from non-causal models is accepted, the thesis of learning from minimal models loses its initial appeal. Using Schelling’s (1978) model for segregation as an example, I will argue that, if any learning at all is prompted by that model, it is in virtue of its being a non-causal model and not (as Grüne-Yanoff suggests) a minimal one. In section six, I will conclude with an overview of the ‘middle ground’ position that I favor.

**2. Learning from Causal Models**

A popular view about learning from models is summarized in Cartwright’s (2009) motto: “if no capacities then no credible worlds” (45). This is a way of stating the general idea that a model can be ‘credible’, and thus justify changes in the credence assigned to hypotheses about real-world targets, only if it is underwritten by similarities in causal factors – what Cartwright (1989) calls ‘capacities’. This condition on learning from models is interpreted in more and less restrictive ways by different authors. Cartwright (2009) allows for learning only when one *knows* (or can safely assume) that model and target share some causal factors. On her view, scientific models work as *isolators*: they prompt learning about a real-world target by offering a concrete demonstration of the kinds of effects that obtain in a situation where some causal factor known to be present in the target operates ‘without impediments’.[[4]](#footnote-4) Sudgen (2009), instead, has argued for a less restrictive account, whereby the sharing of some causal factor only needs to be a “credible candidate for truth” (18). This relaxed condition allows for learning from *constructions*: cases in which the sharing of causal factors is not antecedently known, but where from the recognition of significant similarities in some of the stable characteristics of model and target one reasons that the same causal factors operative in the model may also operate in the target.[[5]](#footnote-5)

Even though I believe that Sudgen is right that a less restrictive condition is needed, the argument that I intend to put forward in this paper has considerably broader scope than his. My view is that the motto ‘if no capacities then no credible worlds’ is problematic *regardless* of whether we interpret its meaning in the more restrictive way that Cartwright indicates or in the less restrictive way that Sudgen’s discussion urges. To help bring out the philosophical import of this argument in its full generality, it will be helpful to discuss how it affects a considerably permissive version of the motto ‘if no capacity then no credible worlds’. This is the influential account of learning from models originally proposed by Hesse (1963). Although Hesse’s contribution is somewhat dated, two features make it especially suited to illustrate the inadequacy of *any* epistemological position that appeals to similarities in causal features to explain a model’s capacity to prompt learning. First, her account countenances both cases of learning from ‘isolations’ and from ‘constructions’. Second, her account is more precisely articulated than any of its contemporary descendants in some respects that will be relevant to my argument. These gains in generality and precision will come in especially handy later on.

Hesse’s articulation of the idea “if no capacities then no credible worlds” makes use of two distinct requirements: the *material* and the *causal* condition.[[6]](#footnote-6) The former demands that there be ‘observable’ similarities between model and target. What makes a similarity ‘observable’ at time *t* is that it is consists in two objects sharing a description that is countenanced by the scientific community’s vocabulary at *t*.[[7]](#footnote-7) Hesse’s causal condition, instead, specifies that the ‘observable’ similarities between model and target must be with respect to their causal features. More precisely, Hesse’s claim is that, in order for the study of a model with properties A, B, C and D to justify assigning additional credence to a target’s having property D, knowing that it has properties B and C but not A, (*i*) there must be some justification for believing that causal connections obtain between A, B, C and D in the model and *(ii)* it must be an open epistemic possibility, whose actual occurrence may be further supported by the known similarities, that causal connections of the same kind as the ones operative in the model also obtain in the target.

Two features of this account are worth emphasizing briefly. First, clause (*ii*) in Hesse’s causal condition allows for learning from a model when the sharing of a causal factor with the target is merely an epistemic *possibility*, which may be supported by the known similarities. As Hesse (1963) states this requirement: “the argument [from model’s having D to target’s having D] rests on the presumption that if AB is connected with D in the [model], then there is some *possibility* that B is connected with D, and that this connection will tend to make D occur with BC in the [target]” (85, my emphasis).[[8]](#footnote-8) Hesse’s formulation thus allows for learning to occur not only in situations where BC are shared causal factors responsible for D in the model, but also where BC are recognized as mere ‘effects’ of some causal factor D in the model and where (analogues of) BC have been observed in the target. The latter is precisely the type of situation in which a scientific model might prompt learning *qua* construction and not *qua* isolator of causal factors. Therefore, if any counterexamples to Hesse’s condition can be discovered, they would pose equally a problem for the condition on learning suggested by Sudgen (2009) and would be *a fortiori* counterexamples to more restrictive conditions such as Cartwright’s (2009).

The second feature of Hesse’s account worth noting here is its precision in distinguishing the material and the causal requirements on learning from models. This allows us to clearly discern two controversial theses regarding the problem of learning. Her material condition rules out learning from *formal* models: cases where the quantities of two otherwise incomparable domains can be considered “different interpretations of the same formal theory” (1963, 68), as when one notes that there is a one-to-one isomorphism between quantities related to heat transfer (source of heat, flow, temperature, etc.) and electrical ones (center of attraction, field, voltage, etc.).[[9]](#footnote-9) In these cases, there exist no ‘observable’ similarities between the corresponding quantities of model and target (e.g., heat flow is utterly unlike an electric field, temperature is utterly unlike voltage, etc.); what makes those quantities ‘analogous’ to one another is solely the fact that they play structurally similar roles in their respective theories. Hesse’s causal condition, instead, rules out learning from *non-causal* models: cases where the study of a model supports hypotheses about a target in the absence of any known or predicted similarity between model and target with regards to their causal features. Crucially, such non-causal models may display plenty of ‘observable’ similarities with their targets; what they lack are similarities in causal features. The difference between formal and non-causal models will turn out to be useful in section four.

The case of learning from a model that I will offer in the following section aims to show that neither Hesse’s causal condition nor (*a fortiori*) more restrictive conditions such as Cartwright’s (2009) adequately capture the conditions under which a scientific model can prompt learning. In order to approach this discussion with the necessary precision, however, one final issue must be addressed. This is the problem that the class of ‘similarities in causal features’ is bound to remain vague until an account of the causal relation is provided. Hesse’s way of getting around this issue is symptomatic of a difficulty that affects not only her account, but also all of its descendants, including Cartwright’s (2009). After taking remarkable care in distinguishing the material from the causal requirement on learning from models, Hesse practically evades the issue of defining causation by invoking “causal relations in some scientifically acceptable sense” (1963, 72). As a result, we are left to second-guess the intended meaning of ‘causal connection’. Appeals to paradigmatic cases are hardly sufficient in this case, given the wide variety of causal ascriptions in ordinary and scientific discourse: token-level ascriptions such as ‘the 8 ball was struck by the 5 ball’ differ in obvious respects from type-level ascriptions such as ‘smoking causes cancer’.

My proposal for regimenting the causal language is as follows. First, by ‘causal factor’ I will mean whatever property or event fills the *x*-variable in expressions of the form ‘*x* is (positively) causally connected to *y*’, as in ‘smoking is (positively) causally connected to cancer’. Second, the expression ‘(positively) causally connected’ will be understood as a tendency to co-occur that is *mediated by a causal law*. Even though defining ‘causal law’ is no straightforward matter, for the purpose of the following discussion it will suffice to note two features of those laws that make them special. Plausibly, causal laws are generalizations (e.g., all BC’s are D’s) that (a) underwrite ‘effective strategies’ (Cartwright 1989) and that (b) are discovered empirically. This initial characterization of what is distinctive about causal laws already puts us in a position to discern causal connections from two other kinds of tendencies to co-occur among properties or events: merely statistical correlations, such as rising water level in Venice and rising bread prices in Britain (cf. Sober 2001), and mathematical connections, such as having four sides and four angles. For mere statistical correlations arguably satisfy (b) but not (a): lowering the bread price in Britain is not an effective strategy to lower water levels in Venice. Also plausibly, most mathematical connections satisfy (a) but not (b): drawing a four-sided figure is an effective strategy for drawing a four-angled one, though the justification for this fact is not observational.

 The proposed characterization of the distinctive features of causal connections vis-à-vis other tendencies to co-occur is sufficient to state the main thesis of the next two sections: as a matter of fact, the connections among properties or events that some scientific models draw upon in order to prompt learning do not satisfy at least one of the abovementioned descriptions (a) and (b); for this reason, those connections are unlikely candidates for being counted among the causal connections of a given domain on *any* remotely plausible construal of the latter notion. It follows that (to adopt Cartwright’s 2009 terminology) there can be ‘credible worlds’ without ‘capacities’. Let’s consider an especially neat illustration of this fact from the history of science.

**3. Galton’s Quincunx**

The case-study that I will discuss in this section comes from Francis Galton’s work on hereditary traits, specifically from his article “Typical Laws of Heredity” (1877). What sparked Galton’s interest in the laws of heredity was the following puzzle. According to the data collected by Adolphe Quetelet, measurements of height in each generation of French, Belgian and American men yielded a normal curve of about the same mean and deviation as that of any previous generation. The same stability was observed for trait distributions of many animals and plants, as well as in the fossil record. Two aspects of these data were especially puzzling to Galton. The first is that, for any trait, its distribution in a population tends to approximate normality; in Galton’s words, the distribution shows conformity to the “law of deviation” (490): as the number of measurements increases, so the distribution tends to a smooth bell-shaped form that is symmetric around the mean.[[10]](#footnote-10) The other puzzling aspect is that, “so long as the external conditions remain unaltered” (1877, 492), the normal curves produced by different generations of a given population exhibit what Galton calls “statistical identity” (1877, 493), meaning that the two standard statistical parameters of mean and deviation are approximately the same.

Both aspects are perplexing. First, it is not obvious why the distribution of traits should be normal. This issue can be illustrated through Galton’s ‘quincunx’ (fig.1). In this machine, a large number of pellets are poured into a funnel from its top, they tumble upon various rows of spikes, and accumulate in vertical compartments at the bottom. A heap approximating a normal curve tends to be generated each time. This behavior is related to the quincunx’s setup: the rows are numerous enough and arranged in such a way that, for any spike encountered, there is about the same chance that a pellet will fall to the right as to the left. With this setup, it is a mathematical consequence of Galton’s law of deviation that the final distribution will tend to be normal, regardless of the pellets’ precise initial positions when dropped. As Galton notes, however, one would not expect the “host of petty disturbing influences” story that we appeal to when explaining the pellets’ final distribution to apply in population biology: “the processes of heredity that limit the number of children of one class.., diminish their resemblance to their fathers, and kill many of them, are not petty influences, but very important ones” (1877, 512).[[11]](#footnote-11)

Even more surprising is the fact that the distribution of a trait in *different generations* exhibits the same mean and deviation. As Galton notes: “If there be any who are inclined to say

 Fig. 1  Fig. 2 

*Fig. 1: Simple Quincunx. Pellets fall from a funnel on the top (represented by the U-shaped lines), they tumble upon rows of spikes (represented by the dots in the middle) and divide themselves in vertical compartments at the bottom (represented by the vertical lines), forming a curve at the bottom whose shape is represented in dark. Fig. 2: Modified Quincunx. The three curves depicted represent three stages of the machine. First stage: pellets are normally distributed in vertical compartments at the top. Second stage: pellets are made to fall through the inclined shoots (represented by the oblique lines). Third stage: pellets go through a series of rows of spikes as in the simple quincunx machine and accumulate at the bottom. The mean and deviation of the curves at stage one and three are the same.*

there is no wonder in the matter, because each individual tends to leave his like behind him, and therefore each generation must resemble the one preceding, I can assure that they utterly misunderstand the case” (1877, 493). He gives the example of giants, who, because of various factors, including that “giants marry much more rarely than medium men” and are less likely to “survive hardships, [since] their circulation is apt to be languid”, leave “fewer giants and more medium-sized men in the next generation” (1877, 493). Hence, the key to solve the puzzle of statistical identity must lie elsewhere. Galton found it partly with the help of some laborious tests in his garden. Measuring the size of seeds in different generations of sweet peas (which do not cross-fertilize, and where it can be safely assumed that size does not affect fertility and is not subject to significant selection), he was able to observe a curious ‘reversion’ effect: the offspring of heavier-than-average seed parents tended to be, on average, lighter than their parents; also, the offspring of lighter-than-average seed parents tended to be heavier than their parents.

With the sweet peas data in his hands, Galton returned to his quincunx for insight. Since that apparatus can replicate distribution curves with any degree of deviation by adding or subtracting rows of spikes, Galton devised a machine that, taking as input a large number of pellets arranged into a bell-shaped heap, yields another bell-shaped curve of an additional degree of dispersion. This is the bottom half of the apparatus of fig. 2, representing the effect of ‘family variability’, i.e., the departure of offspring of some family from the mean of the parent generation. On top of that machine, Galton placed another apparatus capable of replicating the reversion effect found in sweet peas. For each compartment at the top of the machine, a dedicated ‘inclined shoot’ is present. When trap-doors underneath those compartments are removed, the pellets falling through the shoots ‘revert towards mediocrity’: those initially positioned towards the two ends of the machine move closer to the center. The resulting curve still approximates normality, but is “more contracted in width…, and…more humped up in the middle” (1877, 513).

What Galton noted is that if one pre-arranges a large number of pellets into a normally distributed heap at the top of the modified quincunx, then, by letting the pellets first revert and then disperse (or vice versa, since the order is irrelevant), a heap at the bottom is generated whose shape “bears an exact resemblance to the heap from which we originally started” (513). In other words, the modified quincunx replicates statistical identity outcomes. Galton realized that this effect was, once again, a mathematical consequence of the new machine’s setup. The conclusion about population biology that he drew is quite general and is twofold:

1. First, when statistical identity is observed between two adjacent generations, the processes of heredity at play in the passage from the earlier to the later one (e.g., family variability, selection, etc.) are likely to have conformed to the law of deviation.[[12]](#footnote-12) This is surprising but not totally unexpected, since randomness is present in those processes, and since selective pressures are not always very strong under stable external conditions.
2. Second, a phenomenon of reversion towards mediocrity must mitigate the dispersive effect of variation during inheritance. Specifically, from reflecting on the relation between the angle of inclination of the shoots and the dispersion produced by the rows of spikes in the machine, Galton concludes that the magnitude of the reversion effect must be some linear proportion of the magnitude of dispersion. While the proportionality between regression and variation in actual biological populations may never be perfect in any single case, Galton argues that “equilibrium must at length ensue” (514) in the course of generations, thereby producing the statistical identity outcomes that we observe.

Galton’s conclusions are certainly bold given how limited his evidence about trait distributions was and how even more limited was his knowledge about the actual workings of heredity processes. Still, it is arguable that the study of his quincunx model manages to provide at least some degree of inductive support to 1) and 2). One crucial fact that Galton’s argument trades on is that an element of randomness is known to be involved in both shot-dropping and inheritance: where each pellet ends up falling and which traits a given individual will inherit from his or her parents are both subject to a significant amount of chance. With this piece of background knowledge, the systematicity with which (despite the randomness) statistical identity outcomes are observed both in Quetelet’s population data and in the modified quincunx makes the inference to Galton’s 1) and 2) rather plausible: roughly the same kind of setup that the modified quincunx machine instantiates somehow holds true of biological populations as well. Indeed, so much was Galton confident about the strength of his argument for 1) and 2) that he sentenced: “We now clearly see the way in which the resemblance of a population is maintained” (532). Hence, the thesis that the quincunx offers inductive support to Galton’s conclusions 1) and 2) possesses some degree of plausibility both in its own right and as an interpretation of what the historical Galton took himself to have accomplished through the use of his quincunx model.

 However, Galton’s argument for 1) and 2) is not underwritten by similarities in causal features. After all, the argument does not work by suggesting that the same kinds of causal connections that obtain in the quincunx (e.g., the specific stories of how pellets, owing to the force of gravity, fall from the top of the machine and end up in a bell-shaped heap) are also likely to obtain in biological populations. Rather, his argument works by suggesting that the same *mathematical connections* that obtain between the machine’s setup and its tendency to generate statistical identity outcomes are also likely to obtain in biological populations. Those connections are not ‘causal’ because they are not mediated by causal laws. As Galton recognizes, “the law of deviation is purely numerical” (495). Similarly for his reversion law: reversion towards mediocrity is a purely statistical effect, resulting merely from the imperfect correlation between the variables being subject to measurement.[[13]](#footnote-13) Like the law of deviation, the reversion law holds no matter what kinds of events the imperfectly correlated variables stand for, regardless of how specifically those events are causally connected, and indeed whether they are causally connected at all.[[14]](#footnote-14) Hence, the connections on which the quincunx model draws are ostensibly non-causal.

In the next section, I will provide a fuller defense of my diagnosis of Galton’s case-study as a case of learning from non-causal models. This will also be an occasion to distinguish my thesis from other superficially related but ultimately distinct theses in the neighborhood. In this way, I hope to more clearly display both the plausibility and the novelty of my conclusion.

**4. Learning from Non-Causal Models**

The aim of this section is to defend what is by far the most controversial and novel claim of this paper, namely that the study of Galton’s quincunx model can engender *learning*. The reason for this focus is easy to state. Putative examples of what here I have been referring to as ‘non-causal models’ in science already exist in the contemporary literature. Plausible candidates have been identified by, e.g., Hausman (1992) in economics, by Batterman (2002) in physics and by Pincock (2012) in biology. However, these non-causal models have typically been assigned either a broadly ‘explanatory’ or a broadly ‘heuristic’ function, depending on the state of knowledge that is assumed. Specifically, a broadly explanatory function has been attributed to non-causal models whose assumptions are already known (or assumed) to hold for the target (e.g., Batterman 2002, p. 26). In conditions of incomplete information, instead, the function of non-causal models has been reduced to that of useful heuristics: merely the capacity to suggest (in the psychological sense) testable hypotheses (Hesse 1963, p. 88; Hausman 1992, p. 221; Pincock 2012, p. 489).[[15]](#footnote-15) My claim, instead, is that Galton’s model fulfils an ‘inductive’ and not a merely heuristic role: viz., that the study of Galton’s model prompts genuine learning*,* despite Galton’s lacking full information about hereditary processes and despite his model’s being non-causal. In order to facilitate the defense of this crucial claim, it will be helpful to first clarify how it differs from two other theses that one can find defended in the contemporary literature.

First, my thesis is distinct from the claim, which I find doubtful but that Fraser (forthcoming) has defended, that one can learn from ‘formal’ models, i.e., that the study of a model can prompt learning in the absence of any ‘observable’ similarity with its target, purely as the result of structural correspondences among otherwise distinct properties: e.g., the correspondence between ‘temperature’ in fluids and ‘voltage’ in currents. While it is true that, on my reconstruction, Galton’s argument works by projecting resemblances in mathematical features from the quincunx onto biological populations (namely, the quincunx’s ‘setup’ whereby statistical identities outcomes can be expected to obtain with high frequency), I deny that the similarities on which the quincunx model draws are purely formal. In particular, it seems to me that the distribution curves that one finds by analyzing Quetelet’s data resemble those that the modified quincunx generates in various ‘observable’ respects: e.g., in being bell-shaped and in having the same mean and deviation as some earlier or later distribution. In my view, part of what gives plausibility to Galton’s conclusions lies precisely with the fact that those properties are considered genuine respects of similarity even *before* the argument for 1) and 2) is put forward; as such, the resemblances qualify as ‘material’ in Hesse’s sense.

Second, my thesis of learning from non-causal models is related to, but distinct from, Bartha’s (2009) examples of learning from ‘explanatory’ analogies. These are cases where the connections between the known and the merely predicted similarities between model and target are not strictly speaking connections of cause and effect, but of deductive entailment. Bartha’s main example is Priestley’s (1767) argument for the inverse-square electrostatic law. In brief, Priestley had noticed a remarkable similarity between the result of a recent experiment by Benjamin Franklin, demonstrating the absence of an electric field inside a closed conducting spherical shell, with the consequence of a theorem proved by Newton in the *Principia*, which is that any point inside a hollow sphere of uniform density must be subject to no gravitational acceleration. From this striking resemblance, Priestley reasoned (correctly) that electrostatic forces may well obey a similar inverse square law as gravity: “May we not infer from [Franklin’s] experiment, that the attraction of electricity is subject to the same laws with that of gravitation, and is therefore according to the square of the distances […]?” (1767, 732).

As Bartha notes, Priestley’s argument relies on an ‘explanatory’ analogy because the relation between Newton’s law being inverse square and the gravitational ‘null effect’ is not causation but deductive entailment (given the initial conditions). At least in one sense, then, Priestley’s argument resembles Galton’s: just as Galton realized that the tendency of his modified quincunx to generate statistical identity outcomes must be a mathematical consequence of its setup, from which he inferred that statistical identity in biological populations may well result from an analogous setup, so Priestley noticed that a null effect in the gravitational case is entailed by an inverse square law of gravity, from which he inferred that the electrostatic null effect may well result from an analogous inverse square law. However, there is a crucial difference between the two cases. It is that Newton’s law of gravity, while relating to the observations by entailment relations, mediates causal connections between the distribution of masses and the attractive forces at play in a given system. We appeal to this connection when explaining why a point of mass *m* is subject to gravitational attraction towards the center of the Earth. In this sense, Newton’s law is a causal law. Conversely, the laws of deviation and reversion invoked in Galton’s example are not causal.[[16]](#footnote-16) This is why I regard Galton’s model as genuinely ‘non-causal’, but I am not inclined to say the same about Priestley’s gravitational model.

Of course, there are other ways of understanding the terms ‘formal’, ‘explanatory’ and ‘non-causal’ which support different categorizations of the models just mentioned. For instance, it may be argued that on a narrower understanding of ‘material similarities’, the resemblances between the modified quincunx’s statistical identity outcomes and those observed in Quetelet’s population data do not count as ‘observable’, making Galton’s quincunx model a purely formal one. Alternatively, it may be argued that on a narrower understanding of ‘similarities in causal features’, Priestley’s reasoning to the inverse square electrostatic law already qualifies as learning from a ‘non-causal’ model. These alternative categorizations are fine as far as the present argument goes, so long as we recognize that they are obtained by *narrowing* the meaning of ‘material similarities’ and ‘similarities in causal features’. This is exactly the opposite of the strategy that I have taken in this paper, which is to identify a plausible instance of the thesis of learning from non-causal models when the notions of ‘material similarities’ and ‘causal connections’ are understood in the *broadest* possible way: when ‘material’ similarities include those between the modified quincunx outcomes and the population data, and when ‘causal connections’ go so far as to include entailments mediated by causal laws.

The discussion over the possible categorizations of Galton’s case-study serves as a useful entry-point for the defense of my most important claim, namely that Galton’s model is capable of prompting learning. In the previous section, I have argued for this claim by appealing to its independent plausibility as well as to Galton’s own testimony. The considerations advanced in this section strengthen my case in two ways. First, noting that Galton’s argument draws upon ‘material’ similarities helps distinguish my thesis from the controversial idea that we can learn from formal models. The latter thesis is often thought to be subject to a triviality worry: since arbitrary isomorphisms between any two domains can be cooked up at will, confirmation of empirical hypotheses would be too easy to achieve if we allowed purely formal models to provide inductive support (Hesse 1963, 49).[[17]](#footnote-17) The fact that Galton’s model is not formal makes it invulnerable to this sort of concern.[[18]](#footnote-18) Second, the plausibility of my reconstruction is further enhanced by the comparison with Priestley’s argument. After all, the two cases differ only with respect to the nature of the connections that they draw upon: in Priestley’s case, they are entailment relations mediated by causal laws; in Galton’s case, they are entailment relations mediated by mathematical theorems. Therefore, if one accepts that there is some learning in Priestley’s case, there is at least some rational pressure to think the same about Galton’s.

Two final considerations about Galton’s case-study bring the odds squarely in favor of my reconstruction. The first is that, by showing that recurrent statistical identity outcomes can be obtained in a domain as simple as that of pellets, Galton’s model shows that appeals to a divine interest for species stability or systematic faults in the data collection process are neither sensible nor necessary to explain the surprising statistical identity outcomes observed in population biology. At the very least, then, Galton’s model is not inductively inert in that it disconfirms some of the rival accounts of the phenomenon of statistical identity in population biology. Second, Galton’s model also fulfills an evidential role by serving as a vehicle of generalization. As previously mentioned, Galton had already observed an effect of reversion towards mediocrity in his experiments on sweet peas. Yet inferring from the sweet pea data that trait distributions in populations of plants and animals obey the ‘laws of heredity’ that Galton proposes is quite an inductive leap. Plausibly, Galton’s model offers independent support for the generalization, by suggesting that the observed statistical identity outcomes may have less to do with proprietary features of the biological populations in which they obtain, and rather more to do with the broadly statistical features of those cases, features that may be replicated by a mechanical apparatus where nothing even remotely like a process of heredity is occurring.

In order to conclude this discussion, one final objection must be dealt with. In light of the previous arguments, one might concede that there is some learning involved in Galton’s case-study. However, one might insist that, unlike with what happens with causal models, the only kinds of hypotheses that the study of Galton’s model may confirm are claims about what might *possibly* occur, and not what *actually* is the case: viz., that a reversion effect might possibly take place, but not that it actually took place in any specific case of inheritance. My response to this objection is two-fold. First, even if the point were to be conceded (which I am not), it seems to me that an interesting notion of learning would be instantiated in Galton’s case-study – so long as by ‘possibility’ we do not merely mean logical possibility. Second, the burden on those who pose the objection is to give some principled reason for thinking that the distinction between confirmable and un-confirmable hypotheses must coincide with the distinction between hypotheses about what is actually and what is possibly the case. I struggle to find any intuitive motivation for this view. For instance, the hypothesis that the statistical identity outcomes observed in many plants and animals are *frequently* due to the combined effect of variation and reversion is a hypothesis about what actually occurs, despite not being about any specific instance of inheritance. I see no reason why the study of Galton’s model could not confirm this hypothesis (to some extent) just as it confirms hypotheses about what might possibly occur.

To sum up, the arguments in this section help me reassert that Galton’s reasoning to the laws of deviation and reversion in population biology from the study of the modified quincunx is indeed a positive instance of the thesis of learning from non-causal models. As a way of illustrating the novelty of this claim, I have distinguished it from three other superficially similar but ultimately distinct theses in the neighborhood: first, the thesis that non-causal models can be useful ‘heuristically’; second, the thesis of learning from ‘formal’ models; third, the thesis of learning from ‘explanatory’ models. In the next section, I will distinguish my thesis from an alleged fourth class of scientific models, namely ‘minimal’ models. Even though some authors have claimed that such minimal models are capable of prompting learning, I will argue that, once the thesis of learning from non-causal models is accepted, the thesis that we can learn from minimal models loses much of its initial appeal. Let’s consider this argument in detail.

**5. Minimal Models?**

According to Grüne-Yanoff (2009, 2013) and Knuutila (2009), learning from a scientific model can sometimes occur even in the absence of any known or predicted similarity *whatsoever* between model and target. A putative example of such ‘learning from minimal models’ comes from Schelling’s (1978) work in economics.[[19]](#footnote-19) While segregating behavior in cities was often attributed to blatant racism, Schelling (1978) proposed a new model to show that segregation could be the result of a less questionable factor: a mild preference to not be a minority group in one’s own neighborhood. Using randomly distributed dimes and pennies over a checkerboard, Schelling demonstrated that an analogue of segregation among dimes and pennies is quite invariably generated by repeated applications of the following rule: whenever a token (dime or penny) has less than one-third of its neighbors of its own type, move it to the nearest empty space on the board where at least one-third of its neighbors are of the same type. Banking on the dissimilarities between the checkerboard model and actual cities, Grüne-Yanoff (whose claims will be my target here) concludes that no resemblances whatsoever are necessary for learning.[[20]](#footnote-20)

Grüne-Yanoff’s startling claims have already been the subject of critical attention elsewhere, with especially significant contribution by Fumagalli (2015, 2016). My aim in what follows is to build on Fumagalli’s criticisms with the aim of outlining an alternative proposal about how to understand the models that Grüne-Yanoff discusses. In brief, my proposal is that the most plausible of Grüne-Yanoff’s case-studies should not be understood as *minimal* models, lacking any similarity or other world-linking relation *at all* to any real-world target; rather, they should be understood as other instances of *non-causal* models, lacking any similarity with real-world targets with regards to their causal features.[[21]](#footnote-21) Of course, to defend this proposal in full detail would require examining all the putative cases of minimal models mentioned by Grüne-Yanoff.[[22]](#footnote-22) These include Ainslie’s (2001) feedback model of self-control in psychology and Güth’s (1995) indirect evolutionary model for the reproduction of reciprocity preferences in sociology. For reasons of space, below I will focus on defending my claim in the case of Schelling’s checkerboard model, which Grüne-Yanoff takes to be an especially representative one for his thesis. Even though the discussion below will remain largely programmatic, I hope that it will be sufficient to illustrate the weaknesses of Grüne-Yanoff’s epistemological position.

I will start by outlining two worries for the thesis of learning from minimal models that have been raised by Fumagalli (2016). The first targets Grüne-Yanoff’s claim that Schelling’s model “lacks any similarity” (2009, 83) with real-world targets. This claim is questionable because there would seem to be several respects in which the checkerboard model resembles actual cities: for instance, in their division into neighborhoods, in the possibility of relocating elsewhere, etc.[[23]](#footnote-23) Furthermore, Fumagalli raises the concern that, by denying that similarities are necessary to prompt learning, Grüne-Yanoff’s thesis leads to epistemological mystery: for it becomes unclear in virtue of what features of those models (other than their resemblance to their targets) it is possible to learn from them. Grüne-Yanoff’s idea that minimal models prompt learning partly because they “present a relevant possibility” (2009, 97) is insufficient: he needs to provide an explication of ‘relevant possibility’ that makes no appeal to similarities with real-world targets. It is not clear that there is any plausible explication of the requisite sort. As Fumagalli writes: “it is hard to see on what grounds these models’ possibility results may foster justified changes in confidence in hypotheses about real-world targets unless one provides convincing reasons or evidence (e.g. analogies […]) to think that what is possible (or necessary) in the worlds posited by such models is also possible (or necessary) in the targeted real-world situations” (2016, 439).

Building on Fumagalli’s criticisms, I propose an alternative interpretation of Schelling’s model that preserves its distinctiveness. On my view, if there is any learning at all from the checkerboard model, it is in virtue of its being a *non-causal* model. The way in which segregating patterns among pennies and dimes are produced are, after all, very unlike (causally speaking) any real-world processes of segregation: dimes and pennies do not have ‘preferences’, and they do not ‘choose’ to relocate, etc.[[24]](#footnote-24) What matters for the purpose of Schelling’s argument is that the checkerboard instantiates the kind of setup from which it follows (by a mathematical theorem) that analogues of segregating patterns will occur with high frequency.[[25]](#footnote-25) The logic of the argument, on my reconstruction, is therefore similar to Galton’s: from the similarities between the segregation outcomes observed in the checkerboard and in cities, an inference is made that segregating patterns in cities may be the result of an analogous setup, which can be realized by people’s mild preferences not to be a minority in one’s neighborhood. Of course, an argument that the checkerboard model engenders learning would be required at this point to establish my alternative diagnosis. I will not offer this argument here. My aim here is to make the purely dialectical point that, if one is tempted *at all* by the arguments that Grüne-Yanoff (2009, 2013) puts forward in support of his thesis of learning from minimal models, one should consider embracing my diagnosis of the checkerboard model as a non-causal model instead.

Let me mention two reasons for thinking that my understanding of the checkerboard model is at least more plausible than Grüne-Yanoff’s. The first is that it fits the case-study better. Grüne-Yanoff is committed to denying there are any similarity relations *at all* between the checkerboard model and any actual cities, which, as Fumagalli persuasively argues, is implausible. A weaker and more plausible claim is to say that, while there may be various similarities between the segregating patterns observed in the checkerboard and in actual cities, they are not similarities with respect to the kinds of causal connections that are known or predicted to obtain in each domain. Among other things, the non-causal nature of the model helps explain why Schelling takes it to be so widely applicable: as he puts it, the model can teach us something about the segregating patterns observed among “whites and blacks, boys and girls, officers and enlisted men, students and faculty” (138). Plausibly, this wide applicability has to do with its describing behaviors that are relatively stable across systems, despite those systems differing significantly in their respective causal features. This makes for an important difference with many cases of inferences from ‘causal’ models, such as extrapolations from mice to humans. In those cases, the presence of similarities with respect to causal features (e.g., resemblances in the processes that underlie the assimilation of chemicals) is crucial to the plausibility of those inferences.

The second reason for favoring my interpretation of the checkerboard model over Grüne-Yanoff’s is that it does not lead to epistemological mystery. By denying that any similarity *whatsoever* is necessary to learn from a model, defenders of the thesis of learning from minimal models severe a commonly recognized link between learning from a model and reasoning from analogy. This is the link that most contributions to contemporary epistemology of scientific modeling (beginning with Hesse 1963) rely upon in order to explain the capacity of scientific models to prompt learning. Defenders of the thesis of learning from minimal models thus find themselves in the uncomfortable position of having to provide an alternative epistemological story about how we could learn from minimal models which makes no appeal to analogy. Conversely, my claim that Schelling’s case instantiates learning from non-causal models is compatible with the traditional picture that takes learning from a model to consist in a kind of ‘analogical projection’, fueled by relevant similarities and dissimilarities. What must be modified in that picture is only the specific account of what it takes for some similarities or dissimilarities to be *relevant* to drawing conclusions about a target. Whereas for Hesse and many of her followers (e.g., Cartwright 2009) a similarity with respect to B is relevant only if B bears appropriate *causal* connections to the property D that is predicted to hold for the target, my account adds *mathematical* connections (i.e., tendencies to co-occur mediated by mathematical theorems) among the legitimate relevance relations. This seems to me a natural extension of the traditional framework, well-grounded in the history of scientific modelling.

Together, the two arguments just outlined strongly support the main conclusion of this section: we should reject the thesis of learning from minimal models as both conducive to epistemological mystery and as unnecessary to account for the case-studies that initially motivated it. Ultimately, my thesis that we can learn from non-causal model offers a more secure epistemological basis for those who wish to recognize the contribution that some highly idealized models (such as Schelling’s checkerboard model for segregation) make towards the confirmation of novel hypotheses in the various branches of empirical research.

**6. Conclusion**

My aim in this paper has been to carve up a middle position between two factions in the contemporary debate over the problem of learning from scientific models: on the one hand, the position of those who, like Cartwright 2009, take learning from models to be possible only when model and target resemble each other in their causal features; on the other hand, the position of those who, like Grüne-Yanoff (2009), take learning from a scientific model to be possible even in the absence of any similarity whatsoever between model and target. My thesis that we can learn from non-causal models shows that neither view is ultimately tenable: the former because (as Galton’s case-study shows) it imposes too restrictive a condition on learning from scientific models; the latter because, by rejecting the intuitive link between learning from a model and reasoning from analogy, it leads to epistemological mystery and is in any case unnecessary to account for the case-studies that initially motivated it. All things considered, the middle ground position that I have indicated stands out as the most reasonable option.

That there is no mystery about the capacity of non-causal models to prompt learning is best illustrated by comparison with the way ordinary causal models work. After all, in both cases the features that one projects in reasoning from the known properties of a model to the predicted properties of a target are either the causal or the mathematical connections that are operative in the model. These are modally robust connections (more robust than mere statistical correlations) which hold between non-arbitrary, ‘observable’ features of the model. There remain, of course, some delicate questions regarding how to formulate in a more rigorous manner my condition on learning from models, which I cannot address here. In particular, I cannot possibly do justice to Norton’s (forthcoming) worry that any proposed account of learning from models must necessarily be vague and hence useless. Moreover, a separate discussion would be necessary to respond to Bartha’s (2009) claim that merely statistical correlations can underwrite defeasible arguments from models (see also my fn. 8). These and similar issues are tasks for future work. In the meantime, let’s be content with the realization that we can learn from non-causal models.

**References**

Ainslie, George, 2001, *Breakdown of the Will*, Cambridge: Cambridge University Press.

Ariew, Andre’, Rohwer, Yasha, and Rice, Collin, 2017, “Galton, Reversion and the Quincunx”, *Studies in History and Philosophy of Science Part C*, 66: 63-72.

Banerjee, Abhijit, 1992, “A Simple Model of Herd Behavior”, *The Quarterly Journal of Economics*, 107 (3): 797-817.

Bartha, Paul, 2009, *By Parallel Reasoning*, New York: Oxford University Press.

Batterman, Robert, 2002, “Asymptotics and the Role of Minimal Models”, *The British Journal for the Philosophy of Science*, 53: 21-38.

Batterman, Robert, and Rice, Collin, 2014, “Minimal Model Explanation”, *Philosophy of Science*, 81(3): 349-376.

Cartwright, Nancy, 1989, *Nature’s Capacities and Their Measurement*, Oxford: Clarendon Press.

Cartwright, Nancy, 2009, “If No Capacities Then No Credible Worlds. But Can Models Reveal Capacities?”, *Erkenntnis*, 70: 45-58.

Franklin, Benjamin, 1941, *Benjamin Franklin’s Experiments*, ed. I. B. Cohen, Cambridge, MA: Harvard University Press.

Fraser, Doreen, forthcoming, “The Non-Miraculous Success of Formal Analogies in Quantum Theories”, in French, S. and Saatsi, J. (eds.), *Scientific Realism and the Quantum*, New York: Oxford University Press.

Fumagalli, Roberto, 2015, “No Learning from Minimal Models”, *Philosophy of Science*, Proceedings of the 24th Biennial Meeting of the Philosophy of Science Association.

Fumagalli, Roberto, 2016, “Why We Cannot Learn from Minimal Models”, *Erkenntnis*, 81(3): 433-455.

Galton, Francis, 1877, “Typical Laws of Heredity”, *Nature*, 15 (5): 492-533.

Giere, Ronald, 1988, *Explaining Science: A Cognitive Approach*, Chicago: Chicago University Press.

Goodman, Nelson, 1955, *Fact, Fiction and Forecast*, Cambridge, MA: Harvard University Press.

Grüne-Yanoff, Till, 2009, “Learning from Minimal Economic Models”, *Erkenntnis*, 70(1):81-99.

Grüne-Yanoff, Till, 2013, “Appraising Models Non-Representationally”, *Philosophy of Science*, 80(5): 850-861.

Güth, Werner, 1995, “An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives”, *International Journal of Game Theory*, 24: 323-344.

Hausman, Daniel, 1992, *The Inexact and Separate Science of Economics*, Cambridge, UK: Cambridge University Press.

Hesse, Mary, 1963, *Models and Analogies in Science*, Notre Dame: University of Notre Dame Press.

Knuutila, Tarja, 2009, “Isolating Representations vs. Credible Constructions? Economic Modelling in Theory and Practice”, *Erkenntnis*, 70: 59-80.

Lange, Marc, 2014, “On ‘Minimal Model Explanations’: A Reply to Batterman and Rice”, *Philosophy of Science*, 82 (2): 292-305.

Lange, Marc, 2016, *Because Without Cause*, New York: Oxford University Press.

Mäki, Uskali, 1992, “On the Method of Isolation in Economics”, *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 26: 316-351.

Mäki, Uskali, 2009, “MISSing the World. Models as Isolations and Credible Surrogate Systems”, *Erkenntnis*, 70(1): 29-43.

Nguyen, James, 2019, “It’s Not a Game: Accurate Representation with Toy Models”, *The British Journal for the Philosophy of Science*, available online.

Norton, John, forthcoming, *The Material Theory of Induction*, chapter on “Analogy” available online.

Pietsch, Wolfgang, 2019, “A Causal Approach to Analogy”, *Journal for General Philosophy of Science*, 50 (4): 489-520.

Pincock, Christopher, 2012, “Mathematical Models of Biological Patterns. Lessons from Hamilton’s Selfish Herd”, *Biology and Philosophy*, 27: 481-496.

Priestley, Joseph, 1767, *History and Present State of Electricity*, London: Dodsell, Johnson & Cadell.

Reutlinger, Alexander, Hangleiter, Dominik, and Hartmann, Stephen, 2017, “Understanding (with) Toy Models”, *The British Journal for the Philosophy of Science*, 69(4): 1069-1099.

Schelling, Thomas, 1978, *Micromotives and Macrobehavior*, New York: Norton.

Sober, Elliott, 2001, “Venetial Sea Levels, British Bread Prices, and the Principle of Common Cause”, *The British Journal for the Philosophy of Science*, 52: 331-346.

Steel, Daniel, 2007, *Across the Boundaries: Extrapolation in Biology and Social Science*, New York: Oxford University Press.

Stigler, Stephen, 2016, *The Seven Pillars of Statistical Wisdom*, Cambridge, MA: Harvard University Press.

Sudgen, Robert, 2000, “Credible Worlds: The Status of Theoretical Models in Economics”, *Journal of Economic Methodology*, 7: 1-31.

Sudgen, Robert, 2009, “Credible Worlds, Capacities and Mechanisms”, *Erkenntnis*, 70: 3-27.

1. Cf. Grüne-Yanoff’s (2013): “Learning from a model M… is constituted by a change in confidence in certain hypotheses, justified by reference to M” (2). In what follows, I will use ‘confidence’ and ‘degree of credence’ interchangeably. When a model justifies assigning some *additional* degree of credence to an empirical hypothesis, I will sometimes use the equivalent expression ‘provides inductive support’. [↑](#footnote-ref-1)
2. Other advocates of this view include Hesse (1963), Sudgen (2000, 2009), Steel (2007) and Pietsch (2019). [↑](#footnote-ref-2)
3. The expression ‘non-causal’ is borrowed from discussions of non-causal explanations in science (e.g., Lange 2016). While this paper is concerned with confirmation and not explanation, there are suggestive parallels between the two topics, some of which are discussed in footnote 21 (cf. also Reutlinger et al. 2017). [↑](#footnote-ref-3)
4. Mäki (1992) is another proponent of the view of scientific models as isolators. In response to Sudgen’s (2000) criticism, however, Mäki (2009) takes a more conciliatory position on the problem of learning. [↑](#footnote-ref-4)
5. Cf. Sudgen (2009) on Banerjee’s (1992) herd model: “The effect of herding in the model is similar to that of herding in the real world. From the similarity in effects, we are invited to infer […] similar causes” (10). [↑](#footnote-ref-5)
6. I am skipping over some of the subtleties regarding Hesse’s notion of ‘inductive support’ which is afforded by the use of models. On her view, the generic notion of inductive support that is at stake (before, as it were, a probabilistic theory of confirmation is invoked) is a comparative notion of one hypothesis (based on a model) being *more reasonable* than another (not based on a model). The probabilistic rendering in terms of ‘justifying additional credence’ to a hypothesis is discussed on pp. 112-3 in her chapter ‘The Logic of Analogy’ (1963). [↑](#footnote-ref-6)
7. Among other things, this condition purports to rule out that similarities in gerrymandered respects, such as those of the grue/bleen variety made famous by Goodman (1955), count as evidence for other, merely predicted similarities between a model and a target. For Hesse, those respects of similarity are illegitimate in inductive arguments since they do not belong to the vocabulary of any current or past science. [↑](#footnote-ref-7)
8. Bartha (2009, 43) misinterprets Hesse’s (1963) causal condition when he raises as a problem for her account Franklin’s argument that lightning is attracted by pointed metal rods, based on the fact that the ‘electrical fluid’ is so attracted and that “electrical fluid agrees with lightning in these particulars: Giving light. Color of the light. Crooked direction. Swift motion. Being conducted by metals…” (1941, 334). Granted, when proposing this analogy, Franklin had little knowledge of the causal features of either electricity or lightning. Yet Hesse’s causal condition does not require knowledge: “the use of analogical argument does not presuppose that the actual causal relation is known” (1963, 84). Hesse’s causal condition only requires that there be some antecedent reason for expecting causal relations to underwrite the correlations observed in the source. When interpreted in this way, it is far from obvious that Franklin’s case poses a problem for her causal condition. [↑](#footnote-ref-8)
9. One of Hesse’s (1963) favorite examples of this class of models is “the formal analogy between elliptic membranes and the acrobat’s equilibrium, both of which can be described by Mathieu’s Equation” (69). [↑](#footnote-ref-9)
10. Today this law is known as the ‘central limit theorem’. I will continue using Galton’s terminology to avoid confusion when quoting him. The same applies to ‘reversion towards mediocrity’ that will be mentioned momentarily: in contemporary statistics, this phenomenon is known as ‘regression towards the mean’ [↑](#footnote-ref-10)
11. Here I disagree with Ariew et al. (2017), according to which the quincunx of fig.1 “provides justification for his statistical assumption [viz. “that hereditary characters approximate the normal distribution”]” (70). As will be discussed below, it is the resemblance between Quetelet’s data with the more exceptional outcome of the machine of fig.2 (absent from their reconstruction) that yields confirmation to the statistical assumption. [↑](#footnote-ref-11)
12. If more than one process of heredity is at play, but they all conform to the law of deviation, then their combined result must necessarily conform to the law of deviation as well. Galton calls this statistical theorem (that the sum of independent normal variates is itself normal), the “law of the sum of two fallible measures” (1877, 533). [↑](#footnote-ref-12)
13. Its approximately linear character, whereby reversion is proportional to the magnitude of variation, results instead from the fact that the distribution of the parent generation is bell-shaped. This is in turn to be explained in the same way, that is, by the fact that the heredity processes leading to that generation’s distribution obeyed the law of deviation. Galton’s account is therefore ‘recursive’: assuming one generation’s distribution is normal, a story is given about what enables “*successive* generations to maintain statistical identity” (1877, 493). [↑](#footnote-ref-13)
14. It may be objected that in 1877, when “Typical Laws of Heredity” was published, Galton took himself to have singled out a causal law of reversion towards mediocrity. According to Stigler (2016), Galton realized that the reversion law does not describe a “pull” towards the ancestral type, but is rather a purely statistical effect, only when he was able to compare (a few years after 1877) the data concerning the heights of children in relation to their parents with the data about the height of brothers in the same family. However, first, it doesn’t follow from the law of reversion being understood as a causal law that Galton’s model satisfies the causal condition, since the law of deviation, which is also needed for the derivation of statistical identity, was by Galton’s lights in 1877 “purely numerical” (495). At most, the case-study is a hybrid. Second, it seems to me disingenuous for a defender of the causal condition on learning to rely so much on these historical contingencies. Regardless of Galton’s oversights in 1877, his example makes it plausible that some non-causal models can prompt learning. [↑](#footnote-ref-14)
15. I am including in the ‘heuristic’ category the use of models for “conceptual exploration” discussed in Hausman (1992). His idea is that some models are studied in order to learn facts about the models themselves – whether or not this information is in turn projectable onto real-world targets. I am also including in the broadly ‘heuristic’ category the use of non-causal models to support “weak conclusions” (i.e., plausibility claims) about the target and that of motivating “adherence to a research program” discussed in Pincock (2012). [↑](#footnote-ref-15)
16. Note that the distinction between causal and non-causal laws that I am drawing is orthogonal to the distinction between deterministic and probabilistic laws. Thus, the fact that ‘randomness’ is involved in Galton’s laws but none is involved in Newton’s law of gravity plays no role in my argument. Priestley’s case-study would still qualify as ‘causal’, on my account, even if Newton’s law made use of probabilities. [↑](#footnote-ref-16)
17. This argument targets ‘formal models’ in Hesse’s specific sense: i.e., purely structural isomorphisms without any ‘observable’ similarities. Note, however, that the triviality objection is still plausible under Bartha’s (2009, p. 195) slightly more liberal construal of the notion of a formal model. [↑](#footnote-ref-17)
18. At the very least, defenders of the thesis of learning from non-causal models do not face the burden of having to show what (if anything) is wrong with the triviality objection just mentioned. [↑](#footnote-ref-18)
19. Other authors (e.g. Batterman 2002; Batterman and Rice 2014) use ‘minimal models’ in a less specific way to refer to ‘highly idealized’ models. Here I will use ‘minimal models’ in Grüne-Yanoff’s specific sense. [↑](#footnote-ref-19)
20. Another seemingly related thesis (which I will not discuss here) is Nguyen’s (2019) view that some models can “adequately represent” a real-world target without being “similar” (2) to them. Unfortunately, the author does not provide an explication of ‘similarity’ that would clarify the content and distinctiveness of his thesis. [↑](#footnote-ref-20)
21. While in this paper I have deliberately avoided the topic of scientific explanation, I should note here that there is an interesting parallel between Grüne-Yanoff’s claim that minimal models can engender learning and Batterman and Rice’s (2014) claim that there exist ‘minimal model explanations’ in science. On their view, some scientific models can be used to explain properties of real-world systems without having *any* features in common with those systems. For a critical discussion, whose negative upshot nicely matches the ones I will advance below with regards to Grüne-Yanoff’s thesis, cf. Lange’s (2014) reply to Batterman and Rice. [↑](#footnote-ref-21)
22. I am not committed to claiming that all of Grüne-Yanoff’s case-studies are non-causal models. Indeed, some of them may even be ‘ordinary’ causal models– so long, of course, as they are not *minimal* models. [↑](#footnote-ref-22)
23. Note that, given any real-world target, there exists an arbitrary one-to-one isomorphism with the properties of the model that makes the latter (at least) a ‘formal’ model, i.e., a model bearing merely formal similarities with the target. Here, however, I am concerned with the allegation that there are no ‘observable’ similarities between Schelling’s checkerboard model and actual cities. Grüne-Yanoff’s (2009) argument that modellers often do not specify their intended targets is insufficient to establish the absence of such similarities since, as Fumagalli (2016, p. 440) notes, the intended targets may be left implicit by the modelers. [↑](#footnote-ref-23)
24. Of course, we can speak as if dimes and pennies have ‘preferences’ not to be a minority in their neighborhood, and as if those preferences ‘cause’ the segregated patterns in the checkerboard (Sudgen 2000 sometimes speaks in this way). However, that would not be enough to show that the checkerboard resembles actual cities in their causal features. This is because we would not be inclined to accept that dimes and pennies have ‘preferences’ *before* the analogy with actual cities was introduced. Hence whatever similarities result from adopting talk of ‘preferences’ for dimes and pennies would be purely ‘formal’ and not ‘material’. [↑](#footnote-ref-24)
25. Cf. Sudgen (2000): “suppose we read Schelling as claiming that *if* people lived in checkerboard cities, and *if* people came in just two colours, […] and *if*…, and *if*… (going on to list all the properties of the model), *then* cities would be racially segregated. That is not an empirical claim at all: it is a theorem.” (17). [↑](#footnote-ref-25)