

Evidence amalgamation, plausibility, and cancer research

Marta Bertolaso¹  · Fabio Sterpetti¹ 

Received: 30 April 2017 / Accepted: 8 October 2017 / Published online: 20 October 2017
© Springer Science+Business Media B.V. 2017

Abstract Cancer research is experiencing ‘paradigm instability’, since there are two rival theories of carcinogenesis which confront themselves, namely the somatic mutation theory and the tissue organization field theory. Despite this theoretical uncertainty, a huge quantity of data is available thanks to the improvement of genome sequencing techniques. Some authors think that the development of new statistical tools will be able to overcome the lack of a shared theoretical perspective on cancer by amalgamating as many data as possible. We think instead that a deeper understanding of cancer can be achieved by means of more theoretical work, rather than by merely accumulating more data. To support our thesis, we introduce the analytic view of theory development, which rests on the concept of plausibility, and make clear in what sense plausibility and probability are distinct concepts. Then, the concept of plausibility is used to point out the ineliminable role played by the epistemic subject in the development of statistical tools and in the process of theory assessment. We then move to address a central issue in cancer research, namely the relevance of computational tools developed by bioinformaticists to detect driver mutations in the debate between the two main rival theories of carcinogenesis. Finally, we briefly extend our considerations on the role that plausibility plays in evidence amalgamation from cancer research to the more general issue of the divergences between frequentists and Bayesians in the

The article was jointly developed by both authors and the thesis commonly shared. Marta Bertolaso was mainly responsible for the writing of Sects. 3, 4, and 5. Fabio Sterpetti was mainly responsible for the writing of Sects. 1 and 2.

✉ Fabio Sterpetti
f.sterpetti@unicampus.it

Marta Bertolaso
m.bertolaso@unicampus.it

¹ Campus Bio-Medico University of Rome, Via Álvaro del Portillo, 21, 00128 Rome, Italy

philosophy of medicine and statistics. We argue that taking into account plausibility-based considerations can lead to clarify some epistemological shortcomings that afflict both these perspectives.

Keywords Cancer research · Evidence amalgamation · Plausibility · Probability · Somatic mutation theory · Tissue organization field theory

1 Introduction

Cancer research is experiencing what has been defined a period of ‘paradigm instability’ (Baker 2014), since there are two main rival theories of carcinogenesis which confront themselves, namely the somatic mutation theory (SMT) and the tissue organization field theory (TOFT) (see Sect. 3.1). Despite this theoretical uncertainty, the huge quantity of data that became available thanks to the improvement of genome sequencing techniques led to the development of new statistical tools. These tools, some authors think, will be able to overcome the lack of a shared theoretical perspective on cancer by amalgamating as many data as possible in order to give us the ‘right’ answers as outputs.

We think instead that a deeper scientific understanding of cancer may come by means of more theoretical work, rather than by merely accumulating more data to be statistically analyzed. Indeed, the main thesis of this article is that the role played by plausibility-based considerations in the development of statistical models across scientific disciplines has been underestimated or even neglected. This led to underestimate the ineliminable contribution of the epistemic subject to the development of statistical tools, and to the process of evidence amalgamation. Cancer research is a clear example of it. In this field, the relations between rival theoretical hypotheses on carcinogenesis and the recent development of sophisticated statistical tools form an intricate tangle. We think that our proposal can be fruitfully tested in such a context, and that it can contribute to identify some of the epistemological shortcomings that afflict the debate in this field.

This article is divided into three parts. In the first part (Sect. 2), we draw some consequences from the usually underappreciated platitude that statistics is developed in precisely the same way in which all other scientific disciplines are developed (Sect. 2.1). We consider some criticisms that have recently been moved to the so-called big data revolution to point out that the relevance of the knowing subject has not to be overlooked in accounting for statistics from an epistemological point of view (Sect. 2.2). Then, in order to better illustrate the role played by the epistemic subject in the context of statistical research, the analytic view of scientific theories development is presented (Sect. 2.3). The concept of plausibility is especially analyzed, in order to make clear the difference between that concept and the concept of probability (Sects. 2.4, 2.5, 2.6), and to disentangle the concept of subjectivity from that of arbitrariness (Sect. 2.7). A brief digression on the relation between the concept of probability and the concept of randomness concludes this part (Sect. 2.8). In the second part of the paper (Sect. 3), after having briefly illustrated the main rival conceptions of carcinogenesis (Sect. 3.1), the notion of personalized cancer medicine (Sect. 3.2), and the concept of driver mutations (Sect. 3.3), we address some issues in cancer research to test the

adequacy and fertility of the theoretical framework presented in the first part. More precisely, we focus on some of the computational tools that have been developed by bioinformaticists for searching driver mutations in cancer specimens (Sects. 3.4, 3.5), in order to highlight the role played by plausibility-based considerations in the development of statistical tools and in the assessment of theoretical hypotheses (Sect. 3.6). Finally, in the third part, we put the conclusions that can be drawn from our analysis in a broader context (Sect. 4). We think that our proposal may be of use to address a more general issue, which characterizes both the philosophy of medicine and the philosophy of statistics, and which is also crucial for the epistemological investigations of cancer research, namely the confrontation between frequentists and Bayesians on what is the more adequate way to conceive of evidence amalgamation (Sects. 4.1, 4.2, 4.3).

2 Statistics and uncertainty

Several definitions of statistics can be found in the philosophical literature (see for a survey [Bandyopadhyay and Forster 2011](#); [Romeijn 2017](#)). One of the most interesting way to conceive of statistics is the one advocated, among others, by [Lindley \(2000\)](#), according to which statistics is the study of uncertainty and statisticians are experts in handling uncertainty, who “developed tools, like standard errors and significance levels, that measure the uncertainties that we might reasonably feel” ([Lindley 2000](#), p. 294).

Although it may appear very general and quite uninformative, this way of defining statistics points out with immediacy the reason why statistics is nowadays so central in almost every science. Indeed, since scientific knowledge is usually regarded by scientists and philosophers as fallible, and so not certain, dealing with uncertainty in order to develop fallible knowledge is what scientists routinely do. Computational devices that may be of help in such an enterprise, as those developed by statisticians, are obviously deemed to be of great value and widely adopted by researchers and practitioners of many fields.

But, since uncertainty is related to fallibility, this way of defining statistics underlines also the fact that statistics has to face the same epistemological difficulties that are often thought to be of exclusively concern of other scientific disciplines. Indeed, there is a widespread perspective that takes “as given the statistical models we impose on data, and treats the estimated parameters of such models as direct mirrors of reality rather than as highly filtered and potentially distorted views” ([Goodman 2001](#), p. 295). Contrary to this perspective, statistics itself provides fallible knowledge, since it is the result of human efforts aimed at knowing and managing the world, as any other scientific discipline is. So, it cannot be regarded as a mere repository of reliable mathematical tools from which researchers of other disciplines can safely choose the most adequate tool in order to produce genuine and objective knowledge in their fields.

2.1 Statistics and the method of science

In order to stress the analogies between statistics and other scientific disciplines, it may be of use to consider also the definition of statistics given by [Romeijn](#), according to

which statistics “investigates and develops specific methods for evaluating hypotheses in the light of empirical facts” (Romeijn 2017). In this view, a method is called statistical “if it relates facts and hypotheses of a particular kind: the empirical facts must be codified and structured into data sets, and the hypotheses must be formulated in terms of probability distributions over possible data sets” (Ibidem). It is important to underline that in this line of reasoning statistics is not a merely mathematical discipline, since it concerns the relationship between *facts* and *hypotheses* which (usually at least) are not mathematical in character. On the other hand, it is undeniable that statistics is a throughout mathematized discipline, since it relies on a specific branch of mathematics, namely probability theory, in order to estimate whether a given hypothesis is confirmed by the facts. This makes the commonness between statistics and other scientific disciplines even more transparent: every scientific discipline develops models, which rely in some way or another on mathematics, in order to better understand its object of inquiry. Statisticians, relying on probability theory, develop models to better understand how hypotheses relate to facts in uncertain contexts, models that in their turn can be used to construct models in other disciplines.

To sum up: there are two crucial steps that have to be performed in order to develop statistical models that may help us in dealing with uncertainty: (1) empirical facts must be codified into *data* sets, and (2) hypotheses must be formulated in terms of *probability* distributions. These two steps are what makes mathematics applicable to worldly facts in the context of statistical research. In what follows, we will argue that these steps involve the human knowing subject in an ineliminable way, in the sense that these steps cannot be made human-independent in any relevant sense.

2.2 The big data approach

To better see this point, i.e. that model-building cannot be made (or regarded as) human-independent, even in the case of statistics, or in the case of those disciplines whose models strongly rely on statistics, consider the so-called big data revolution. Supporters of this revolution usually maintain the view that constructing theories is an unnecessary effort, since it may be replaced by big data analysis. For example, according to Anderson, we are at “the end of theory,” because “the data deluge makes the scientific method obsolete” (Anderson 2008). The “new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world” (Ibidem). In this perspective, we “can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find pattern where science cannot” (Ibidem). So, Anderson concludes, the old “approach to science—hypothesize, model, test—is becoming obsolete” (Ibidem).

There are at least two main problems with this approach.¹ The first problem is *how* to discriminate among the huge number of correlations that data analysis may pick out in the vast sea of available data. In other words, the question is: How can we evaluate the significance of those correlations? Even granting that statistical algorithms may

¹ For a survey of the challenges that the big data approach has to face, see Fan et al. (2014).

reliably find patterns or correlations, this does not guarantee that those patterns or correlations are significant. If we do not possess some theory to identify a criterion to discriminate among those correlations, we will be unable to determine whether or not the finding of any new correlation represents a genuine instance of scientific progress, i.e. an amplification of our knowledge. This is the problem of *spurious* correlations.² And it is a big problem for big data supporters. Indeed, they claim that the more data we have, the more theory is unnecessary to produce new knowledge. But it has been demonstrated in a paper recently published by Calude and Longo (2016b), that the more data we have, the more spurious correlations we may find among our data. And so, the more we need a theory to discriminate the significant correlations among all the correlations individuated by our algorithms. In a nutshell, they base their argument, among other things, on Ramsey theory, i.e. the branch of combinatorics which investigates the conditions under which order must appear. If we restrict our attention to mathematical series, more precisely to arithmetic progressions, Ramsey theory investigates the conditions under which an arithmetic progression must appear in a string of numbers.

Calude's and Longo's analysis hinges on Van der Waerden's theorem, according to which for any "positive integers k and c there is a positive integer γ such that every string, made out of c digits or colours, of length more than γ contains an arithmetic progression with k occurrences of the same digit or colour, i.e. a monochromatic arithmetic progression of length k " (Calude and Longo 2016b, p. 11).

For example, if we take a binary string of x digits, digits can be either '0' or '1'. Take '0' and '1' to be the possible colours of those x digits, i.e. $c = 2$. From Ramsey theory, we know that there will be a number γ such that, if x is bigger than γ , that string will contain an arithmetic progression of length k such that all k digits of that progression are of the same colour, i.e. either all the k digits are '0' or all the k digits are '1'.³

Consider now a database D , where some kind of acquired information about some phenomenon P is stored. We want to investigate the correlations among the data stored in D in order to increase our knowledge of P :

² On what spurious correlations are, cf., e.g., Dellsén (2016, p. 78): "Suppose we have two variables V1 and V2 that are known on independent grounds to be unrelated, causally and nomologically. Let us further suppose that we learn, i.e. come to know, that there is some specific statistical correlation between V1 and V2—e.g. such that a greater value for V1 is correlated with a greater value for V2." The latter correlation represents an instance of spurious correlation, i.e. a correlation between two variables which is not due to any real relation between them. Such a correlation does not convey any information on the correlated variables, nor on some other relevant aspect of the world, so it is useless, irrelevant, or worse, it may lead us astray, if we do not correctly identify it as spurious.

³ In this case (i.e. $c = 2$), if we have $k = 3$, then $\gamma = 8$. To see this, consider the following sequence of binary digits of length 8: 01100110. This string contains no arithmetic progression of length 3, because the positions 1, 4, 5, 8 (which are all '0') and 2, 3, 6, 7 (which are all '1') do not contain an arithmetic progression of length 3. However, if we add just one bit more to that string (i.e. if we add either '1' or '0'), we obtain the following two strings: 011001100 and 011001101. Both these strings contain a monochromatic arithmetic progression of length 3. Consider 011001100: positions 1, 5, 9 are all '0'. Consider 011001101: positions 3, 6, 9 are all '1'. More generally, it can be proved that if a string contains more than 8 digits, it will contain a monochromatic arithmetic progression of length 3.

In full generality, we may consider that *a correlation of variables in D is a set B of size b whose sets of n elements form the correlation [...]*. In other words, when a correlation function [...] selects a set of n -sets, whose elements form a set of cardinality b , then they become correlated. Thus, the process of selection may be viewed as a colouring of the chosen set of b elements with the same colour—out of c possible ones. [...]. Then Ramsey theorem shows that, given any correlation function and any b , n and c , there always exists a large enough number γ such that any set A of size greater than γ contains a set B of size b whose subsets of n elements are all correlated. (Calude and Longo 2016b, p. 12).⁴

Calude and Longo prove that the larger is D , the more *spurious* correlations will be found in it. In other words, when our stock of available data increases, most of the correlations that we can identify in it are spurious. Since large databases *have to* contain arbitrary correlations, owing to the size of data, *not* to the nature of data, the larger the databases are, the more the correlations in such databases are spurious. Thus, the more data we have, the more difficult is to extract meaningful knowledge from them.⁵

The second problem that afflicts the ‘big data revolution’ view is that it overlooks the fact that data are not abstracted from the world in neutral and objective ways. There is always “a theory or hypothesis which guides observation and experiment, and generally data-finding” (Cellucci in press, Sect. 1). The big data revolution view overlooks also the fact that the very algorithms used for data analysis are based on some theory or another. Theories and previous knowledge are in some sense incorporated in the design of algorithms when they are developed. Thus, “it is illusory to think that statistical strategies may automatically discover insights without presupposing any theory” (Ibidem).

2.3 Statistics and the logic of discovery

Let us now briefly address the issue of the identity of the method used in statistics and in other sciences. At this regard, we think that it is important to untangle some issues that are usually conflated. Indeed, usually statistics is used to confirm hypotheses, i.e. statistics is primarily used in the context of justification, not in the context of discovery. The distinction between those two contexts may confound our reflections on statistics, since it may lead us not to take into due consideration, in some sense to ‘hide’, the process of discovery that led to the development of statistical theories. And this may affect our ideas on what statistics is.

⁴ It is important to stress that the nature of the *correlation function* is irrelevant: it can be completely arbitrary, i.e. in no way related to the nature of the data stored in the database.

⁵ Cf. Calude and Longo (2016b, p. 6): “it is exactly the size of the data that allows our result: the more data, the more arbitrary, meaningless and useless [...] correlations will be found in them.” It may be interesting to note that, in order to derive their result, Calude and Longo define ‘spurious’ in a more restrictive way than usually is done. According to them, “a correlation is spurious if it appears in a ‘randomly’ generated database” (p. 13). Details can be found in Calude and Longo (2016b). In any case, this does not impinge on the considerations that follow.

So, even if we use some statistical tool (s) to confirm some hypothesis (h) in a given scientific domain (D), we have to keep in mind that s has been developed through a process of discovery in its turn. In other words, even if in a *justificatory* context with respect to the hypothesis h in the domain D we use s , we have not to forget that s in its turn has been produced and assessed in a different scientific domain (S), namely the statistical field, where it played the role that h plays in D .

Taking into account the process of discovery is relevant in order to understand the ineliminable role played by the knowing subject in the development of science. Thus, taking into account the process of discovery of statistical theories may help us to recognize the ineliminable human epistemic ‘coefficient’ that the statistical tools we use introduce in our research.

Unfortunately, according to many authors, while there may be a logic of confirmation, since confirmation can be formalized, there cannot be a logic of discovery, since discovery processes cannot instead be formalized (for a survey, see [Schickore 2014](#)). For example, Popper states that “there is no such thing as a logical method of having new ideas, or a logical reconstruction of this process” ([Popper 2005](#), p. 8). The problem with this view is that it equates the intelligibility of a given reasoning process with the possibility of formalizing that process, i.e. the possibility of making that process algorithmically reproducible, and thus mechanizable. This approach leaves out from the perimeter of rational analysis and understanding both (1) the inferential paths of discovery that are not algorithmically describable (e.g. the process of hypotheses production through non-deductive inferences); and (2) the non-algorithmic constituents of those processes that are thought to be algorithmically describable (think of the indispensable role that emotional circuits and subconscious inferences play in making us able to experience the ‘sense of certainty’ that we associate with valid deductive reasonings, see [Rigo-Lemini and Martínez-Navarro 2017](#)). But the fact that those elements cannot be formalized does not mean that they are irrational, nor that they cannot be analyzed at all.

Moreover, the asymmetry between discovery and confirmation is unjustified. As Putnam states, if we follow Popper and claim that there is no logic of discovery, because observations do not lead to theories “in a mechanical or algorithmic sense,” then, “in that sense, there is no logic of testing, either” ([Putnam 1975](#), p. 268). The idea that there can be a logic of confirmation because, since confirmation can be described in purely deductive terms, there can be an algorithmic method for confirmation, is unjustified.⁶ Indeed, algorithms do not exhaustively account for all that is relevant to the process of hypotheses confirmation. Just as “there is no algorithmic method of discovery, there is no algorithmic method of testing. Indeed, by the undecidability theorem, there is not even an algorithmic method for testing whether a formula is logically valid or not” ([Cellucci 2017a](#), p. 144). So, either one admits that there cannot even be a logic of confirmation, or one should accept the idea that there can also be a logic of discovery.

In the last decades, the idea of developing a logic of discovery has been mainly conceived as the attempt to develop a logic of inductive inferences in terms of proba-

⁶ Cf. [Cellucci \(2017a, p. 142\)](#): “Methods can be divided into algorithmic and heuristic. An algorithmic method is a method that guarantees to always produce a correct solution to a problem. Conversely, a heuristic method is a method that does not guarantee to always produce a correct solution to a problem.”

bility calculus (Howson and Urbach 2006). The main problem of this approach is that it is mainly aimed at showing the validity and consistency of probabilistic inductive inferences in the face of classical deductive logic. But in so doing, the probabilistic view of the logic of discovery becomes analogous to the deductivist view: it cannot take into account (and say something relevant about) some characteristic features of the process of discovery, namely how we *produce* and *appraise* new hypotheses. Those features cannot be straightforwardly formalized, nor can they adequately be described in probabilistic terms (more on this below). In this view, hypotheses production is just taken as a *datum*, something prior and external to a logic of discovery, precisely in the same way the process of hypotheses production is regarded as external to a logic of confirmation by those who deny that there can be a logic of discovery.

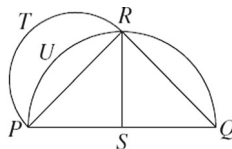
Underlying these points does not mean to deny the theoretical relevance and practical usefulness of formal approaches to confirmation or induction. It is only meant to stress that there may be some relevant theoretical insights in considering the role of non-formalizable components of reasoning when addressing the issue of whether knowledge ampliation can be regarded as (or made) human-independent.

At this regard, an interesting proposal aimed at modeling the process of scientific development is the analytic view of theory development (see Cellucci 2013, 2016, 2017a), according to which knowledge is increased through the analytic method.⁷ In this view, “to solve a problem one looks for some hypothesis that is a sufficient condition for solving it. The hypothesis is obtained from the problem, and possibly other data already available, by some non-deductive rule, and must be plausible [...]. But the hypothesis is in its turn a problem that must be solved, and is solved in the same way” (Cellucci 2013, p. 55).⁸

⁷ For the differences that exist among the analytic method, the analytic-synthetic method, and the axiomatic method, see Cellucci (2013).

⁸ The origin of the analytic method may be traced back to the works of the mathematician Hippocrates of Chios and the physician Hippocrates of Cos, and was firstly explicitly formulated by Plato in *Meno*, *Phaedo* and the *Republic*. As an example of the analytic method, consider the solution to the problem of the quadrature of certain lunules provided by Hippocrates of Chios:

Show that, if PQR is a right isosceles triangle and PRQ , PTR are semicircles on PQ , PR , respectively, then the lunule $PTRU$ is equal to the right isosceles triangle PRS .



To solve this problem, Hippocrates of Chios states the following hypothesis:

(B) Circles are as the squares on their diameters.

Hypothesis (B) is a sufficient condition for solving the problem. For, by the Pythagorean theorem, the square on PQ is twice the square on PR . Then, by (B), the semicircle on PQ , that is, PRQ , is twice the semicircle on PR , that is, PTR , and hence the quarter of circle PRS is equal to the semicircle PTR . Subtracting the same circular segment, PUR , from both the quarter of circle PRS and the semicircle PTR , we obtain the lunule $PTRU$ and the triangle PRS , respectively. Therefore, “the lunule” $PTRU$ “is equal to the triangle.” [Simplicius, *In Aristotelis Physicorum libros Commentaria*, A 2, 61]. This solves the problem. But hypothesis (B) is in its turn a problem that must be solved (Cellucci 2013, p. 61).

Assessing the *plausibility* of any given hypothesis is crucial in this perspective. But how plausibility has to be conceived? The interesting suggestion made by the analytic view is that, in the ultimate analysis, the plausibility of a hypothesis is assessed by a careful examination of the arguments (or reasons) for and against it.

Let's try to better illustrate this point. According to this view, in order to judge over the plausibility of a hypothesis, the following 'plausibility test procedure' has to be performed: (1) "deduce conclusions from the hypothesis"; (2) "compare the conclusions with each other, in order to see that the hypothesis does not lead to contradictions"; (3) "compare the conclusions with other hypotheses already known to be plausible, and with results of observations or experiments, in order to see that the arguments for the hypothesis are stronger than those against it on the basis of experience" (Ibidem, p. 56). If a hypothesis passes the plausibility test procedure, it can be temporarily accepted. If, on the contrary, a hypothesis does not pass the plausibility test, it is put on a 'waiting list', since new data may always emerge, and a discarded hypothesis may successively be re-evaluated.⁹ Thus, according to the analytic view of method, what in the ultimate analysis we really do in the process of scientific knowledge ampliation, is producing hypotheses, assessing the arguments/reasons for and the arguments/reasons against each hypothesis, and provisionally accept or refute such hypotheses.

It is important to stress here the difference between the concept of probability and the concept of plausibility. Indeed, as Kant points out, "plausibility is concerned with whether, in the cognition, there are more grounds for the thing than against it" (Kant 1992, p. 331), while probability measures the relation between the winning cases and possible cases. This means that plausibility involves a comparison between the arguments for and the arguments against, so it is not a mathematical concept. Conversely, probability is a mathematical concept (see Cellucci 2013, section 4.4).

This distinction is relevant here because it allows us to better illustrate the epistemological import of statistics. As we have already noted, when a statistician develops a statistical model s of some worldly domain D , she formulates the relevant empirical D -facts in terms of a data set e , and the hypotheses relative to those facts in terms of probabilities distributions h_i over e . This account may give us the impression that a statistician deals only with probabilities, and probabilities may well be interpreted in a robust objective way, independent of the knowing subject who develops s .¹⁰ In other words, since probability distributions may be claimed to be fully determined by the way the world is, and statistics deals mainly with probabilities and empirical facts, it may seem that statistics does nothing more than "translating" in terms of probabilities what the world dictates to us. And so that nothing which is relevantly dependent on the human knowing subject is added by the statistical tools to what we model through them.

⁹ For a more detailed confrontation of the concept of plausibility with some related (but distinct) concepts, such as truth, probability, and warranted assertibility, see Cellucci (2017a, Chapter 9).

¹⁰ On the possible interpretations of probabilities, see Gillies (2000). Basically, probabilities may be regarded as 'objective' or 'subjective'. Cf. e.g. Djulbegovic et al. (2011, p. 309): "'objective probability' is believed to reflect the characteristics of the real world, i.e., the probability somehow relates to the physical property of the world or a mechanism generating sequences of events [...]. On the other hand, 'subjective probability' is believed to represent a state of mind and not a state of objects [...]."

But the fact is that, independently of what interpretation of probability one prefers, this is not the all story. Indeed, as we have already stressed above, the theories that a statistician uses in order to build s (e.g. the theories she relies on when she translates empirical facts into data sets, or she derives a probability distribution for rival hypotheses from our knowledge of D , or she decides what inferences can be legitimately drawn from the amalgamated evidences for any h), are instances of human scientific knowledge, and so they have been produced in their turn in the way described by the analytic model of theory development. That is, every statistical theory or technique can be regarded as a hypothesis that has been produced in order to solve a problem in the statistical field of inquiry. This hypothesis may have been retained and accepted by statisticians since it passed the plausibility test procedure. This means that conclusions have been deduced from the hypothesis and have been compared with each other, in order to see that they do not lead to contradictions, and then that conclusions have been compared with other statistical hypotheses already known to be plausible, in order to see that the arguments for the hypothesis are stronger than the arguments against it. In this kind of evaluative process, hypotheses are assessed referring to their *plausibility*, which is not just a matter of probabilities, and this makes the role played by the human knowing subject epistemologically ineliminable.

2.4 Probability and plausibility

It is important to better explain why the process of knowledge ampliation cannot simply be accounted for in terms of probabilities, and so the reason why relying on the analytic view of scientific progress may be of use in this context.

When we produce a hypothesis to solve a problem, we do it by some non-deductive inference rules (e.g. induction, analogy, etc.). Non-deductive inference rules are indeed fallible, i.e. they are not truth-preserving, since they may lead us to incorrect conclusions, even if the premises are regarded as true. But they are also ampliative, i.e. they add something to what is already known that may be necessary in order to solve the problem we want to solve (Goodman 1999).

Why not to use only deductive rules, which are truth-preserving, in the process of knowledge ampliation? Because deductive rules are non-ampliative, i.e. they lead us to conclusions that are correct, but which are in some sense already contained in the premises (Ibidem). So, in many cases, and certainly in the most interesting and difficult ones, non-deductive rules would not allow us to solve the problem we need to solve. Thus, in the problem-solving process, we tentatively produce new hypotheses applying some non-deductive rules to our data and background knowledge. Now the question is: Why we do not deal with such hypotheses assigning them probabilities in some objective way, instead of referring to their *plausibility* in order to accept or refute them?

The problem is the same that spans across many debates in the philosophy of science, e.g. in the debate over scientific realism. In order to assign a probability p with some degree of objectivity to the hypothesis h that we produced to solve some problem A , we should know the space of all the other possible hypotheses that may be formulated

in order to solve A.¹¹ But knowing such space is normally impossible. If we could be able to know with certainty the space of all the possible solutions to A, we could systematically examine all of them and pick out the best one. There could be no more doubt about the hypothesis we selected, and this would make our knowledge certain, i.e. forever unrevisable. Indeed, if we can know the space of *all* the possible solutions to a given problem, we can also know whether we have exhausted the space of all the possible alternatives to a given solution. And this implies that no other alternative could ever appear, not even in the future. So, our hypothesis could be safely said to be *unrevisable*. Unfortunately, there is no way to construct the space of all the possible alternatives to any given hypothesis. Since probability is “a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible” (Laplace 1951, p. 7), in order to effectively calculate the probability of a hypothesis, we have to know the denominator, i.e. the number of all the cases possible. But in many cases, we do not know (and perhaps we cannot know) the number of all the cases possible. Thus, if plausibility were to be understood in terms of probability, we could not be able to evaluate the plausibility of all those hypotheses for which we are unable to determine the set of all the possible rival alternatives. But we routinely evaluate the plausibility of that kind of hypotheses, so it cannot be the case that probability is equivalent to plausibility.

Moreover, that plausibility has to be distinguished from probability clearly appears by considering the fact that there are hypotheses that are plausible, but which, according to probability theory, have zero probability, while there are hypotheses that are implausible, but which have non-zero probability.¹² Thus, contrary to Pólya (1941), we should conclude that the calculus of plausibilities does not obey the same rules as the calculus of probabilities, and that plausibility has to be distinguished from probability.

2.5 Plausibility and the role of the knowing subject

We think it is important to better point out why what we said so far leads us to conclude that the role played by the human knowing subject is ineliminable even in a context such as statistics. The point is that if it were possible to produce new hypotheses through some *deductive* method, and assign to these hypotheses an objective *probability* value, then our method could be regarded as an algorithmic method and could be mechanized.

¹¹ This is the problem of the unconceived alternatives, see below Sect. 2.6.

¹² See Cellucci (2013, Chapter 20). An example of plausible hypotheses that have zero probability are all the plausible hypotheses derived by an Induction from a Single Case (ISC). On the classical concept of probability as ratio between favorable and possible cases, a conclusion obtained by (ISC) has zero probability when the number of possible cases is infinite. An example of implausible hypotheses that have non-zero probability are implausible hypotheses that have been obtained by Induction from Multiple Cases (IMC). Consider the hypothesis that all swans are white. Until the end of the seventeenth century, “all swans observed were white. From this, by (IMC), it was inferred that all swans are white. But in 1697 black swans were discovered in Western Australia.” Since then, the hypothesis that all swans are white is highly implausible. But, this contrasts with the fact that, “on the classical concept of probability, a conclusion obtained by (IMC) has non-zero probability when the number of possible cases is not infinite”, and such “is the case of the hypothesis that all swans are white” (Cellucci 2013, p. 335).

In this way, the ampliation of knowledge could be made human-independent in a relevant sense.

Indeed, if it were possible to produce new knowledge through some deductive method applied to already established knowledge, knowledge ampliation would be a trivial and routine task.¹³ In fact, there is an algorithm for enumerating all deductions from given premises. The algorithm “can be said ‘to proceed like Swift’s scholar, whom Gulliver visits in Balnibarbi, namely, to develop in systematic order, say according to the required number of inferential steps, all consequences and discard the *uninteresting* ones’ (Weyl 1949, p. 24). Given enough time and space, the algorithm will enumerate all deductions, from given premises” (Cellucci 2017a, p. 138). Thus, knowledge ampliation would be a routine task. But there is a wide consensus that this is not the way in which knowledge is really amplified in scientific practice, and that knowledge ampliation is not a routine, nor a trivial task. Indeed, while there is an algorithm to enumerate all deducible consequences from given premises, there is no algorithm for discovering new hypotheses through non-deductive inferences. Moreover, if it were possible to extend our knowledge by applying a deductive method to some given premises already established, given that deduction is non-ampliative, this would amount to say that those given premises will never be modified, and that all our current and future knowledge will rest on the very same set of prime premises. But this view is not really able to account for all the cases of knowledge ampliation in which our already established knowledge is insufficient to solve a problem, and so new hypotheses (i.e. premises) need to be introduced. For example, “when Cantor demonstrated that to every transfinite cardinal there exist still greater cardinals, he did not deduce this result from truths already known [...], because it could not be demonstrated within the bounds of traditional mathematics. Demonstrating it required formulating new concepts and new hypotheses about them” (Ibidem, p. 310). So, even in the case of mathematical knowledge, which is usually regarded as the paradigm of certain knowledge, new knowledge is not acquired by merely deductive methods from already established results. A fortiori, these considerations can be applied to the case of natural sciences.

Consider now the possibility of knowing the space of all possible alternatives to a given hypothesis. If it were possible to know the space of all possible alternatives to any given hypothesis h , we could assign objective probabilities to each alternative hypothesis h_i . This would amount to say that the procedure of hypotheses evaluation can always be performed through an algorithmic method and could be mechanized. Indeed, we can develop an algorithm that enumerates all the possible alternative hypotheses to a given hypothesis h , assigns to each of them the relative objective probability, calculates the likelihood of each hypothesis, and then picks out the one which displays the higher likelihood. This would render the process of hypotheses evaluation a trivial task, in the sense that this process could be made human-independent in a relevant sense. Indeed, if probabilities are regarded as objective, i.e. they reflect the way the world is, and knowledge ampliation can be pursued through an algorithmic method, i.e. a

¹³ For an opposite view, see Musgrave (2011). For a criticism of Musgrave’s view, see Cellucci (2017a).

mechanizable one, the role played by the human knowing subject in the development of scientific knowledge may well be said to be deniable.

2.6 Probability and the problem of the unconceived alternatives

It is important to clarify the reason why it is usually impossible to know the space of all possible alternatives to a given hypothesis in the process of knowledge ampliation. Indeed, knowing the space of all possible alternatives is necessary in order to assign *objective* probabilities to each hypothesis, i.e. to consider the value of those probabilities as determined by the way the world really is. Consider a standard six-faces dice. We know that there are precisely six possible outcomes for one throw of that dice. The space of possibilities is completely determined in advance by the symmetries of the system, and this allows us to assign probabilities to the possible outcomes. But usually in science when we try to solve a problem and produce new knowledge we are not in such a position. We do not know in advance the space of relevant possibilities for the given phenomenon we want to explain. Nor we know the exact configuration of the space of all the possible alternative hypotheses that can be formulated in order to explain that phenomenon. If the space of possible theoretical alternatives to a given hypothesis h is not determinable in advance, we cannot safely claim to have exhaustively searched that space, found that h is the hypothesis that best explains the phenomenon under investigation, and so that we should trust h , because it is confirmed by the eliminative inferential procedure we performed. This is the problem of the unconceived alternatives.

This problem has been in recent years fiercely stressed by [Stanford \(2006\)](#), in his defense of the instrumentalist attitude towards science, according to which we should refrain to commit ourselves to the existence of theoretical entities, because historical record of science shows us that we humans routinely failed to conceive all the possible alternatives to a given theoretical hypothesis h at any given time t .¹⁴

¹⁴ Many replies have been elaborated in the last decade to [Stanford \(2006\)](#) (see, e.g., [Magnus 2006](#); [Saatsi et al. 2009](#); [Ruhmkorff 2011](#)); see [Saatsi et al. \(2009\)](#) for Stanford's rejoinder to some criticisms; see [Rowbottom \(2016\)](#) for an interesting extension of Stanford's line of reasoning. Here we will focus on [Mizrahi's \(2016\)](#) attack to Stanford's view, because, as a reviewer suggested, it puts into question the very coherence of Stanford's position, so it risks questioning the validity of all those positions which similarly rely on the problem of the unconceived alternatives, as the one we advocate for in this paper. [Mizrahi \(2016\)](#) develops an argument against Stanford's view according to which, if (1) one accepts Stanford's argument against scientific realism, and (2) it is possible to adopt Stanford's own line of reasoning in the field of philosophy, then Stanford's position is self-debunking. Indeed, according to [Mizrahi \(2014\)](#), it is possible to construct an argument, which is analogous to Stanford's argument against scientific realism, and so it is not easily refutable by those who accept Stanford's argument, according to which we should not believe our current philosophical theories, because history of philosophy shows that philosophers routinely failed to conceive of serious objections to their theories. Call Mizrahi's Stanford-like argument for philosophy MA. Now, according to Mizrahi's argument against Stanford, if Stanford's position is a philosophical position, then we should not trust it, precisely because it is a philosophical position, given that according to MA we should not trust philosophical theories. Many criticisms can be raised to Mizrahi's approach, but we cannot analyze them all here for reason of space. What can be briefly pointed out is that Mizrahi's argument against Stanford crucially relies on MA. Now, it is MA which is a self-defeating argument. Indeed, if one maintains MA, one is clearly advocating for a philosophical position, i.e. one is committing oneself to a given philosophical theory. But, according to MA itself, we should not trust philosophical theories. So, MA

Before Stanford's proposal, analogous concerns were made by van Fraassen,¹⁵ in his criticism of the inference to the best explanation (Fraassen 1989), and by Sklar (1981), who considered both the case of the inference to the best explanation and the case of confirmation theories.

Here, for the sake of brevity, we will consider just the case of confirmation theories, which are usually developed in terms of probabilities. A clear formulation of the problem of the unconceived alternatives in this context can be found in Sklar (1981):

Consider Bayesian strategies for confirmation theory. Here we must distribute a priori probabilities over all the alternative hypotheses to be considered. If there is only a finite set of hypotheses we have in mind, this is easy to do [...]. But if we must keep in mind the infinite and *indeterminate* class of all possible hypotheses, known and unknown, how can we even begin to assign a priori probabilities to those few hypotheses [...] we do have in mind [...]? (Sklar 1981, p. 19).

We will follow (and simplify a bit) Rowbottom (2016) to better illustrate this point. In Bayesian theories of confirmation, the confirmation of a given hypothesis h is equal to its conditional probability given some evidence e :

$$P(h, e) = P(h)P(e, h)/P(e)$$

where $P(h, e)$ is the conditional probability, $P(h)$ and $P(e)$ are the prior probabilities respectively of h and e , and $P(e, h)$ is the likelihood. In this approach, the prior probability of e must be determined considering all the alternatives to h . Indeed, $P(e)$ decomposes as follows:

$$P(e) = P(h)P(e, h) + P(\sim h)P(e, \sim h)$$

and $P(\sim h)P(e, \sim h)$ in its turn decomposes in:

$$P(\sim h_1)P(e, \sim h_1) + \dots + P(\sim h_n)P(e, \sim h_n)$$

where the set of all the possible alternatives to h is $\{\sim h_1, \dots, \sim h_n\}$.

Theories are considered to be highly confirmed provided that $P(\sim h)P(e, \sim h)$ is low, i.e. when the probability assigned to the negation of the proposed hypothesis is

Footnote 14 continued

is self-defeating. If we consider now Mizrahi's argument against Stanford's position, it is easy to see that also this argument falls victim of MA's self-defeatingness. Indeed, Mizrahi's argument against Stanford conveys in its turn a philosophical position, which implies a commitment to a given philosophical theory. But, according to MA, we should not trust philosophical theories. Thus, since Mizrahi's argument against Stanford crucially rests on MA, if (1) one takes MA to be a cogent argument, then one should not trust Mizrahi's argument against Stanford, because Mizrahi's argument against Stanford rests on a philosophical theory, and according to MA we should not trust any philosophical theory; if (2) one takes MA to be a self-defeating argument, i.e. a non-cogent argument, then Mizrahi's argument against Stanford cannot even take off the ground, since it rests on a self-defeating argument.

¹⁵ Cf. Schupbach (2011, p. 119, fn. 2): "Such scenarios correspond to van Fraassen's best of a bad lot objection as well as what Stanford (2006) calls 'the problem of unconceived alternatives'."

low. Confirmation theorists call $\sim h$ the ‘catchall hypothesis’, i.e. the hypothesis that incorporates all the alternatives to h .

To sum up, to confirm a hypothesis h , we have to assign $P(e)$; to assign $P(e)$, we have to be able to estimate $P(e, \sim h)$; and to estimate $P(e, \sim h)$, we have to be able to construct the set of all the alternatives to h and assign a prior probability to each of those alternatives.

The impossibility to actually construct the set of all the possible alternatives to a given hypothesis has been clearly stated by Salmon:

At any given stage of scientific investigation, the catchall is the disjunction of all of the hypotheses we have not yet conceived. What is the likelihood of any given piece of evidence with respect to the catchall? This question strikes me as utterly intractable; to answer it we would have to predict the future course of the history of science. (Salmon 1990, p. 329)

Salmon’s solution to the problem of unconceived alternatives for confirmation theory is to consider, when evaluating the confirmation of a given hypothesis h , only the actually conceived alternatives to h .

This is clearly an example of a plausible (and pragmatic) theoretical choice, since it allows us to produce an estimation of the confirmation of a hypothesis, although a provisional and revisable estimation. But this choice certainly cannot be justified by making reference to its *probability*. This kind of theoretical choices can be proposed, evaluated and accepted by pondering the arguments for and against it, i.e. by assessing their *plausibility*. This is just an example of the fact that in the process of theory construction and knowledge ampliation we do not deal merely with probability-based considerations. Rather, we have to resort to plausibility-based considerations. The process by which we evaluate this kind of considerations cannot be made algorithmic, and so the process of knowledge ampliation cannot be mechanized.

2.7 Subjectivity and arbitrariness

We are aware of the possibility that many would be unwilling to concede a role to the concept of plausibility in knowledge ampliation, since this concept is subjective in character. Indeed, there has been the tendency in the last decades to equate ‘subjectivity’ and ‘arbitrariness’, and many scholars tried to avoid the latter by denying any role to the former (Gelman and Hennig (2017)). Obviously, there is a sense in which the attempt to avoid subjectivity in pursuing knowledge has a positive meaning, namely when it means avoiding personal biases. For example, as Bird (2017) clearly points out, the adoption of systematic methodologies allowed clinical medicine to become a science, exactly because systematic methodologies allowed to eliminate (or at least minimize) personal biases from medical practice. But not any subjective element in the process of knowledge production can be regarded as a bias-producer. The risk is that in some circumstances the quest for objectivity ends in merely hiding some of the subjective components of the process of knowledge ampliation.

We think that the analytic view of theories, by making reference to the notion of plausibility as defined above, may be of help in the attempt to counter this tendency

and untangle the notion of subjectivity from that of arbitrariness. Indeed, some authors seem to think that if knowledge were not objective (i.e. if it does not leave out any subjective element), then knowledge would be arbitrary, and so there would be no real knowledge at all. Contrary to this perspective, in the analytic view the presence of some subjective components cannot be avoided, since the process of plausibility evaluation of the hypotheses cannot be made algorithmic, nor can it be ruled out. In this perspective, there would be no knowledge only if the hypotheses we deal with in the process of knowledge ampliation were arbitrary. But they need not be arbitrary, they must be plausible, i.e. the arguments for them have to be stronger than the arguments against them. If the plausibility evaluation of hypotheses is carefully conducted, even if this process cannot be formalized, it nevertheless cannot be regarded as arbitrary, since it is constrained in several rational ways (e.g. by the need of checking whether contradictions can be derived from a given hypothesis, and whether conclusions that can be derived from a given hypothesis are consistent with other hypotheses already judged to be plausible, etc.). Thus, knowledge may well be possible, even if some subjective elements enter the process of knowledge production.

We propose that not any aspect of reasoning is reducible to probability calculus, and that this does not imply that those aspects which are not captured by the rules of probability are irrational. In this perspective, some aspects of our reasoning remain argumentative and inferential in character.¹⁶

2.8 Probability and randomness

A brief digression on the relation between the concept of probability and the concept of randomness may be useful to conclude this section by recapitulating some of the issues we addressed. Moreover, the analysis of such relation will allow us to provide an argument to support our thesis. Indeed, the claim that the process of knowledge ampliation cannot be adequately accounted for in terms of probabilities can be clarified by reflecting on the relation between the concept of probability and the concept of randomness.

As we said, statistics deals with uncertainty. Randomness can informally be conceived of as unpredictability, i.e. a lack of those correlations which are able to guide

¹⁶ On this issue see Pollock (1983). Although his approach is distant from the view advocated here, there is nevertheless some similarity between the two. For instance, Pollock points out the impossibility of equating what he calls ‘epistemic probability’ and ‘statistical probability’. His conception of ‘epistemic probability’ is more akin to what we call ‘plausibility’ than to what is usually meant with ‘probability’. According to him, ‘statistical probability’ is that kind of probability “about which we can learn by discovering relative frequencies, counting cases” (Pollock 1983, p. 236). On the contrary, the “epistemic probability of a proposition is the degree to which it is warranted” (Ibidem). In this view, a proposition is deemed warranted by a careful examination of the reasons for and against it: “a person is justified in believing P just in case he has adequate reason to believe P [...], and he does not have any defeaters for it at his immediate disposal” (p. 233). Moreover, Pollock clearly denies that arguments evaluation can be represented in probabilistic terms. Indeed, Pollock also explicitly denies the possibility of equating ‘epistemic probability’ and ‘subjective probability’. In his view, even if we follow the Bayesians and consider probability as expressing a person’s ‘degree of belief’ on the truth of a given proposition, we cannot equate probability assignment and arguments evaluation, mainly because we cannot reasonably impose on arguments evaluation the rules of probability calculus.

our predictions, so randomness can be regarded as a source of uncertainty. Probability is the tool we use to manage uncertainty. Thus, randomness and probability are deeply related. Our point is the following: if randomness is theory-dependent, and probability can be regarded as a measurement of randomness, the process by which we select a given theory in the first place cannot adequately be accounted for in terms of probability-based considerations. Let's unpack this claim a bit.

According to Calude and Longo (2016a), randomness is “unpredictability with respect to the intended theory and measurement” (p. 266). In this view, probability is a measurement of randomness,¹⁷ and randomness is unpredictability deriving from theoretical assumptions. So, the probability values that we assign to the set of possible outcomes of an event in a given domain are dependent on our theoretical commitments. Thus, in order to assign probability values, we *have* to previously make a theoretical choice. Since the choice of the theoretical framework we decide to deal with is indispensable in order to assign probability values, this choice cannot in its turn be made by relying on probability-based considerations. Otherwise a regression is lurking. Indeed, if we commit to a given theoretical framework, say T_a , in order to assign probability values in the A-domain, and, if in order to pick out T_a from the set T of similar but not equivalent theoretical frameworks, i.e. T_a, T_a^*, T_a^{**} , etc., we rely on probability-based considerations, this means that we can assign a probability value to every member of T, i.e. T_a, T_a^*, T_a^{**} , etc. This also means that we can do that because we have already chosen another theoretical framework, say F_t , which allows us to assign probability values in the T domain. Now we have to account for how we chose F_t among similar but not equivalent theories in set F. And so on.

Thus, we have to choose a theoretical framework, which will allow us to assign probability values in the domain of interest, in some different way, i.e. without relying on probability-based considerations. This choice, we claim, is made by relying on plausibility-based considerations. Indeed, reasons and arguments that support different theoretical frameworks can be assessed even if we are unable to coherently assign probability values to rival theoretical frameworks. And evaluations made by relying on plausibility-based considerations are fallible and revisable. It is important to underline this point, because it stresses that our proposal is able to account for certain cases in a more satisfying way than the rival hypothesis, i.e. that the choice of the relevant theoretical framework is made by relying on probability-based considerations. Consider theory change, or information update. Since when we deal with plausibility, we evaluate whether the arguments for are stronger than the arguments against a given hypothesis, we may form judgments which can be revised if new relevant information is provided, since new arguments for or against a given hypothesis may be elaborated in the light of this new information, and thus our plausibility-based judgement about that given hypothesis may change. On the contrary, if the theoretical choice were made on probability-based considerations, and probability is deemed objective, how could we account for the phenomenon of theory change, which may lead to changes in the probability values assigned to a given domain?

¹⁷ Cf. Calude and Longo (2016a, p. 273): “Randomness plays an essential role in probability theory, the mathematical calculus of random events. Kolmogorov axiomatic probability theory assigns probabilities to sets of outcomes and shows how to calculate with such probabilities.”

It is also important to stress that, once the theoretical choice is made, probability values can be assigned and computed in rigorous way. In this sense, probability may well be regarded as ‘objective’, i.e. non-arbitrary. Two distinct epistemic subjects, if they take the same theoretical framework, will obtain the same probability values for any given domain. This accounts for the reliability and ‘objectivity’ of probabilistic and statistical reasoning in scientific inquiry.

Finally, the adoption of our perspective may be of help in accounting for the case in which two epistemic subjects, relying on plausibility-based considerations, make different theoretical choices. To explain this, it is not necessary to search for some arbitrary factors which prevent their reasoning from being rational.¹⁸ It may often suffice to consider the different problems that they have in mind when assessing the arguments for and against a given theory. In a given context, the arguments for the choice of a given theory may be stronger than the arguments for the choice of that theory in another context of inquiry. If, on the contrary, theoretical frameworks were chosen by means of probability-based considerations, and probability is objective, i.e. it is determined by the way the world is, how could we account for the possibility that different theories may be chosen in different context to solve different problems?

Crucial to this defense of our thesis is the assumption that randomness is not absolute, that “it depends on (and is relative to) the particular theory one is working on” (Calude and Longo 2016a, p. 263). But, why should we think that randomness is theory-dependent? This point is crucial, because if randomness is absolute, and probability is regarded as an objective measurement of randomness, it may be claimed that plausibility-based considerations are irrelevant for the process of knowledge amplification. Rather, we should prefer probability-based considerations.

Thus, one may be tempted to search for an abstract and theory-independent definition of randomness. Mathematics is a good candidate as the place where to search for, since dealing with mathematics allows one to avoid the issue of measurement, which, in principle, introduces a degree of epistemic uncertainty that may be regarded as a cause of the theory-dependence of randomness (Calude and Longo 2016a). The search for randomness in mathematics is usually conducted by analyzing binary sequences, the simplest infinite mathematical objects. So, in this line of reasoning, in order to prove that there is theory-independent randomness, one has to investigate whether there are true random infinite sequences. The problem is that such a kind of ‘pure’ randomness cannot be proved to exist in mathematics. Calude and Longo (2016a) clearly illustrates this point. If we confine “to just one intuitive meaning of randomness—the lack of correlations—the question becomes: *Are there binary infinite sequences with no correlations?*” The answer is in the negative, so the search for ‘theory-independent’ randomness is doomed to fail: “*there is no true randomness*” in this abstract sense (p. 272). It is interesting to note that to prove this statement, Calude and Longo rely on the very same result of combinatorics that we have illustrated above when dealing with spurious correlations. In this case, the point is that, by Van der Waerden’s Theorem, *every infinite binary sequence contains arbitrarily long monochromatic arithmetic*

¹⁸ It is important to note that we are not denying that external factors may relevantly affect theoretical choices, we are just stressing that our perspective is able to account for relevant non-arbitrary epistemic factors that may lead to theoretical divergence or theory change in a rational way.

progressions. So, we know that there cannot exist binary infinite sequences *with no correlations*.¹⁹ Thus, there is no ‘true’ or absolute randomness, and randomness can well be regarded as unpredictability in the intended theory. We share Calude’s and Longo’s view on randomness (2016a), according to which “randomness is not in the world nor it is just in the eyes of the beholder, but it pops out at the interface between us and the world by theory *and* measurement” (p. 265). This view fits our idea that the epistemic role played by humans in the process of knowledge ampliation is ineliminable, and that, despite its being subjective in nature, it needs not be arbitrary.

3 Uncertainty and cancer research

Many of the issues discussed so far can be found combined together in cancer research. Moreover, in this field the issue of how conceiving of uncertainty is not only related to the debate on the statistical tools used in medical research, it is also related to the even more basic assumptions that one has about what role uncertainty plays in biology, and so about what are the very basic principles of biology (Longo et al. 2015; Longo 2017; Zbilut and Giuliani 2008). We cannot fully address this topic here, but it may be interesting to look at cancer research in order to show that: (1) the way we use our statistical tools in medicine cannot be neutral with respect to our more basic theoretical commitments; (2) the more adequate way to account for how one chooses some basic theoretical commitments rather than some rival ones is by describing that choice in terms of plausibility. Given the vastness of this theme, here, for illustrative purpose, we will focus on the debate about what is the current theory that best explains carcinogenesis, and on few related issues which affect the development of the so-called personalized cancer medicine.

3.1 The somatic mutation theory and the tissue organization field theory

First of all, let us briefly set the stage. Currently, there are two main competing views on carcinogenesis, namely the somatic mutation theory (SMT) and the tissue organization field theory (TOFT) (Bertolaso 2016; Sonnenschein and Soto 2016; Baker 2015a; Bedessem and Ruphy 2015; Rosenfeld 2013; Soto and Sonnenschein 2011, Longo 2017).

SMT represents the mainstream view, and its main tenets can be summarized as follows: (1) cancer is derived from a single somatic cell that has accumulated DNA mutations; (2) cancer is a disease of cell proliferation; (3) quiescence should be actually considered the default cellular state (Baker 2015a). Corollaries of SMT are: (1) mutations are needed for carcinogenesis; and (2) the analysis of genetic instability may be the key to individuate any cancer’s cause. On the contrary, according to TOFT, which is a minoritarian perspective, cancer is a tissue-based disease, not a cell disease. The main tenets of TOFT can be summarized as follows: (1) carcinogenesis represents a problem of tissue organization; (2) proliferation and motility are the default state of

¹⁹ For a demonstration that no infinite sequence passes all tests of randomness, so that ‘true randomness’ does not exist, see Calude (2002).

all cells; (3) cancer arises from the disruption of interactions among cells and adjacent tissue. Corollaries of TOFT are: (1) mutations are not needed for carcinogenesis; and so, (2) genetic instability is mainly a byproduct of carcinogenesis (Baker 2015a).²⁰

It is important to stress that in this paper we are mainly interested in using this debate as a case study to show the relevance of theoretical disagreement and plausibility-based considerations in the development of statistical tools and in the interpretation of their results, i.e. in processes of evidence amalgamation and science advancement. So, we are not concerned here with taking sides in the SMT-TOFT debate, since this would require a wider and quite different analysis (see Bertolaso 2016).²¹ We will use the debate between SMT supporters and TOFT supporters to point out that often scientific disputes are driven by plausibility-based considerations, and so that they cannot be accounted for merely in terms of probability and empirical confirmation.

3.2 Personalized cancer medicine²²

The majority of cancer patients is usually treated on the basis of large random clinical trials in the general population of a specific tumor type. As a result, “a considerable number of patients are exposed to often highly toxic treatment, with only a small subset of these patients having benefit” (Cirkel et al. 2014, p. 417). In the past years, new DNA sequencing techniques have revolutionized the identification of somatic muta-

²⁰ It may be objected that the existence of hereditary cancer is a strong evidence for SMT. Supporters of TOFT usually specify that “a distinction should be made about the types of cancers that appear in the clinic; there are ‘sporadic’ cancers and hereditary ones. ‘Sporadic’ cancers represent over 95% of the cancers in humans. On the other hand, inherited cancers (less than 5% of total cancers) are a discrete subclass, mediated by germline mutations that have a distinct natural history, mostly appearing in early childhood and/or young adults [...]. While the DNA mutations in this latter type of cancers are present in all cells of the organism, tumors mostly appear in one or a few organs” (Soto and Sonnenschein 2011, p. 333). Since, according to TOFT, cancer is ‘development gone awry’, in this view hereditary cancers are regarded as ‘inherited inborn errors of development’ (Soto and Sonnenschein 2011).

²¹ It may be objected that SMT and TOFT, despite their diversity, are not genuine rival theories, because they are not *really* incompatible. This issue is in fact strongly debated (see Bedessem and Rupy 2017, 2015; Bizzarri and Cucina 2016; Bertolaso 2016; Rosenfeld 2013). However, this issue does not impinge on our argumentation. We observed the disputes that are going on in the field and tried to represent the situation. And, in fact, many of the participants in the SMT vs. TOFT debate *do* see those theories as really incompatible. For example, Bizzarri and Cucina (2016) state that SMT and TOFT are *irreconcilable* theories. According to them, “*irreconcilability* depends on radical divergence existing among basic premises [...]. Copernican theory was *irreconcilably different* from the Ptolemaic one, given that the central place in the solar system was occupied by the Earth in the latter and the Sun in the former. It is obviously impossible to support at the same time these two opposing hypotheses by constraining them into a ‘unified’ cosmological model. By analogy, SMT and TOFT cannot be merged because the premises on which those frameworks rely are incompatible: the default state of the cell can be considered either quiescence (SMT), or proliferation (according to TOFT). The two default states cannot be operational at the same time” (Bizzarri and Cucina 2016, p. 232). Now, whether SMT and TOFT are *really* incompatible is not relevant here. Indeed, we do not aim at solving the dispute on which is the best theory between SMT and TOFT, nor we aim at solving the dispute on whether SMT and TOFT are really incompatible. We just try to highlight how this confrontations rest on theoretical commitments, whose adoption is better accounted for in terms of plausibility rather than probability. We thank an anonymous reviewer for having raised this issue.

²² Some authors prefer to speak of ‘precision’ medicine rather than ‘personalized’ medicine. For a brief reconstruction of the opinions on this terminological issue, see Katsnelson (2013). Since this distinction does not impinge on our thesis, here we will use those terms as interchangeable.

tions in genomes, and their decreasing costs made these techniques widely available. These advances “hold promise for precision medicine, or precision oncology, where a cancer treatment could be tailored to a patient’s mutational profile” (Raphael et al. 2014, p. 1). So, the progresses made in molecular biology and the ineffectiveness of traditional pharmacological treatments developed on the basis of random trials encouraged investments in personalized cancer medicine, although it was widely recognized that this approach may face serious challenges (Tannock and Hickman 2016; Ow and Kuznetsov 2016).

Personalized cancer medicine can be regarded as laying at the intersection of (1) the big data approach and (2) traditional SMT-inspired therapeutic strategy.

As regard to (2), as already noted, according to SMT, “tumors originate from a single cell. Cancer is initiated and subsequently evolves by inactivating tumor suppressor genes and acquiring multiple mutations that activate oncogenic pathways” (Cirkel et al. 2014, p. 418). In this view, the best way to pharmacologically attack cancer is by inhibiting the molecular paths that are responsible for cancer growth, which are specific for every cancer. The idea is that it is possible to discriminate the mutations that are responsible for the insurgence of a specific kind of cancer by individuating the presence of some specific biomarkers in cancer specimens, and then calibrate the most adequate therapy for that specific kind of cancer. The most adequate therapy will be those drugs that fare better in selectively inhibiting the essential metabolic and signaling paths associated to the crucial mutations of such cancer. In this way, the argument goes, we will be able to disrupt the molecular paths specific of cancer cells without affecting and disrupting those paths that are essential for normal cells. Personalized cancer medicine moves along this traditional line of reasoning, but aims at improving it by tailoring therapy to patient’s genetic specificity.

As regard to (1), the crucial elements in recent development of personalized cancer medicine are usually considered to be (a) the availability of large amount of omics data, and (b) the availability of big data analytics to manage and interpret those data (Ow and Kuznetsov 2016; Raphael et al. 2014).²³ For some authors, that personalized medicine relies on big data and big data analytics almost amounts to a paradigm shift in medicine (Chen and Snyder 2013). For example, Talukder states that genetic “data analysis is mostly *hypothesis driven*; whereas, genomic data analysis is always exploratory and *hypothesis creating*” (Talukder 2015, p. 203). In accordance with the big data approach, some scholars regard the possibility of statistically analyzing a huge amount of available data coming from a given domain as able to reduce the need for developing theoretical hypotheses in order to advance the research in that domain (Stevens 2013).²⁴ In this view, by merely searching for correlations in databases through an exploratory algorithmic procedure, *hypotheses* can be ‘created’, and *knowledge* may be established (Gagneur et al. 2017). In this perspective, the use

²³ Cf. Ow and Kuznetsov (2016, p. 1): “Big data analytics is the process of examining large data sets containing heterogeneous patient sub-populations and a wide variety of data types [...]. Big data analytics aims to uncover hidden patterns, unknown correlations, complex trends, [...], as well as other useful features.”

²⁴ Cf. Stevens (2013, pp. 65–66): “Bioinformatics can be understood [...] as a kind of neo-Baconian science in which hypothesis-driven research is giving way to hypothesis-free experiments and data collection.”

of data which derive from DNA sequencing techniques “introduced a component of data-driven [...] science into evidence-based medicine” (Talukder 2015, p. 203).

In what follows, we do not aim at criticizing personalized cancer medicine, nor bioinformatics. We just try to show that it may be epistemologically misleading to focus on data analytics when one deals with theoretical disagreement. In our view, the reliance on the big data approach can lead one to neglect the role of plausibility-based considerations in the development of scientific research, and so to overlook the fallibility and revisability of one’s assumptions. Indeed, the big data approach relies on data analytics techniques, which essentially work by searching databases for correlations through some given algorithms (Ow and Kuznetsov 2016). The problem is that “big data approaches [...] fail to provide conceptual accounts for the processes to which they are applied. No matter their ‘depth’ and the sophistication of data-driven methods [...], in the end they merely fit curves to existing data” (Coveney et al. 2016, p. 1). This way of conceiving of research as data-driven may also lead one to think that scientific advancement can be mechanized and made algorithmic.²⁵ Moreover, conceiving of research as data-driven may lead one to think that one’s inquiry is independent from any specific theoretical hypothesis, and so that the data one produces and collects are model-independent.²⁶ And regarding some data as model-independent may lead one to think that those data can safely be used to independently confirm some theoretical hypothesis over some rival hypothesis.

Contrary to this view, we argue (and we will try to illustrate this point in the next sections) that hypotheses cannot be created, nor evaluated algorithmically. Nor can data be regarded as completely model-independent (Mazzocchi 2015; Allen 2001). If one neglects the role of plausibility-based considerations, one risks being not even aware of the possibility that one is not exploring alternative possible and (possibly) plausible research pathways (Baker 2017). Neglecting alternative hypotheses may lead one to mistake the ‘absence’ of alternatives for the confirmation that the path that one is actually exploring is directly dictated to one by the way the world really is, while this path is instead strongly dependent on one’s theoretical assumptions, which *may* be wrong.

3.3 Driver mutations and passenger mutations

As noted above, SMT and TOFT support different hypotheses on carcinogenesis. Those hypotheses imply different consequences, which are relevant for the development of clinical approaches. For instance, according to SMT, cancer progression is

²⁵ Cf. Stevens (2013, p. 69): “the computer becomes the crucial tool: efficiency is a product of bioinformatic statistical and data management techniques. It is the computer that must reduce instrument output to comprehensible and meaningful forms. The epistemological shift associated with data-driven biology is linked to a technological shift associated with the widespread use of computers.”

²⁶ On the idea of ‘hypothesis-free’ or ‘data-driven’ science, see Stevens (2013, Chapter 2); see also Chen and Snyder (2013). On explanatory data analysis, cf. Bassett et al. (1999, p. 54): “Knowledge discovery by exploratory data analysis is a ‘bottom up’ approach in which the data are allowed to ‘speak for themselves’ after a statistical [...] procedure is performed.” Cf. also Brown and Botstein (1999, p. 33): “this process is not driven by hypothesis and should be as model-independent as possible.”

a unidirectional and mostly irreversible process, i.e. once a cell has become a cancer cell it cannot reverse to a normal condition, while according to TOFT carcinogenesis is not a unidirectional process, rather it may be reversible (Rosenfeld 2013). Those discrepancies are due to the different role assigned to mutations in carcinogenesis. Since according to SMT mutations are responsible for cancer insurgence, and the very insurgence of cancer leads to increasing mutations rate because it disrupts the cell control mechanisms, so those mutations accumulate rapidly, once the ‘genetic program’ of tumor cells has been deteriorated in such a way, there is no way to remedy, reverse the process and ‘reprogramme’ the genome of tumor cells. In this view, the gene-level context is predominant in determining the fate of tumor cells.

On the contrary, according to TOFT mutations in somatic cells are not the cause of cancer insurgency, they are consequences of the disruption of communicating and regulatory paths at the tissue level, e.g. among somatic cells, stroma cells, and extra-cellular matrix. In this view, mutations are regarded, on the one hand, as byproducts of carcinogenesis, and, on the other hand, as neutralizable in most cases by a well-functioning tissue. In other words, if cancer cells are put in the context of a normal tissue, in which communicating and regulatory paths are not disrupted, despite the accumulated mutations, those cells may stop being malignant.²⁷ In this view, the tissue-level complex context is predominant in determining the fate of tumor cells. Obviously SMT and TOFT embrace two distinct perspectives on the role that genes play in biology, and so on the relevance of mutations to carcinogenesis (Longo et al. 2015; Longo 2017).

Since SMT and TOFT start from so divergent assumptions, from which so divergent empirical consequences can be drawn, one may be tempted to adjudicate between these two rival hypotheses on carcinogenesis on the basis of which theory is the most confirmed by evidences. Indeed, if they are genuine scientific hypotheses, their claims should be empirically verifiable (Soto and Sonnenschein 2011).

But things are not so easy. The point is that the search for the empirical confirmation of a given theory is not always equivalent to the search for the confirmation of that theory over some rival theory. Indeed, the pursuit of empirical confirmation of a given theory is often not really independent from the theory itself. This means that data cannot safely be said to be model-independent, so one cannot easily use data to independently confirm some theoretical hypothesis over some rival hypothesis. For instance, deciding whether or not some set of ‘evidences’ genuinely confirms a given hypothesis can be dependent on whether one already accepted that very hypothesis among one’s theoretical commitments in the first place. Consider again personalized cancer medicine. According to SMT, personalized cancer medicine represents the future of cancer research. According to TOFT, this way of searching for cancer remedies will be ineffective as it is currently proposed. It may seem reasonable to someone to claim that, since SMT and TOFT are rival theories, and they support two radical different stances on the very same issue, namely personalized cancer medicine, we could empirically verify which stance on personalized cancer medicine is the correct one by examining empirically verifiable consequences of each stance’s assumptions.

²⁷ This hypothesis has been empirically confirmed, see Soto and Sonnenschein (2011), Baker (2015a, b) and Bizzarri and Cucina (2016).

In this view, evaluating whether (some) central tenets of personalized cancer medicine are sound and empirically confirmed may give support to the claim that SMT is the right way of conceiving of carcinogenesis.

Now, in the case of personalized cancer medicine, there is at least a central claim of this approach which may appear *prima facie* easily empirically verifiable: the existence of *driver* mutations and the possibility of identifying them in tumor specimens. Indeed, high-throughput “DNA sequencing is revolutionizing the study of cancer and enabling the measurement of the somatic mutations that drive cancer development” (Raphael et al. 2014, p. 1). We will concentrate on this issue in the following.

In a cancer genome, “there often exist hundreds or thousands of various types of mutations” (Zhang et al. 2014, p. 244). But, along the line of reasoning supported by SMT, only a small subset of these mutations can be regarded as responsible for carcinogenesis. Indeed, cancer is thought to undergo a process of Darwinian selection, in which mutations are usually neutral, and so do not confer any ‘advantage’. Only rarely some mutations confer some kind of ‘advantage’, and so are selected for.

In cancer research, these selected mutations are called *driver* mutations (Stratton et al. 2009). A mutation is called a *driver* mutation if it is “directly implicated in carcinogenesis by its ability to confer a growth advantage to tumor cells”, while a mutation is called a *passenger* mutation if “it does not confer a growth advantage to tumor cells and, therefore, will not contribute to the development of cancer” (Zhang et al. 2014, pp. 244–245). Thus, in this perspective identifying driver mutations from the “background of passenger mutations is critical for understanding the molecular mechanisms of carcinogenesis and for identifying prognostic and diagnostic markers as well as therapeutic targets” (Ibidem, p. 245).

3.4 Computational approaches for the identification of driver mutations

Unfortunately, distinguishing driver mutations from passenger mutations proved very challenging (Tokheim et al. 2016). Raphael and colleagues, for example, state that “distinguishing driver from passenger mutations solely from the resulting DNA-sequence change is extremely complicated, as the effect of most DNA-sequence changes is poorly understood, even in the simplest case of single nucleotide substitutions in coding regions of well-studied proteins” (Raphael et al. 2014, p. 7). Nevertheless, in recent years, thanks to the increasing availability and affordability of DNA sequencing techniques, different computational approaches to identify somatic mutations in cancer genome sequences and to distinguish driver mutations from random passenger mutations have been developed by bioinformaticists (Dimitrakopoulos and Beerwinkler 2017; Tokheim et al. 2016; Merid et al. 2014; Raphael et al. 2014; Zhang et al. 2014).

Is it possible to reach some shared consensus on whether SMT is objectively confirmed by the statistical tools developed to identify driver mutations? As we have seen above, usually statistical tools are thought to allow us to calculate the degree of confirmation that some evidences confer to a given hypothesis. The main problem in the case of driver mutations is that if we closely inspect the computational tools developed to identify these mutations, things seem to go the other way around, i.e. it is the assumption of a given hypothesis on carcinogenesis that is necessary in order to

make sense of a huge amount of messy data and select what data can be regarded as evidences for that hypothesis.

Let's try to clarify this point. There are mainly three different kinds of approaches to individuate driver mutations in DNA sequences: (1) identifying recurrent mutations; (2) predicting the functional impact of individual mutations; (3) assessing combinations of mutations using pathways, interaction networks, or statistical correlations (Dimitrakopoulos and Beerenwinkel 2017; Merid et al. 2014; Raphael et al. 2014; Zhang et al. 2014). Since all these approaches have to face the same theoretical difficulty that we aim to point out, here it will suffice to focus on just the first approach aimed at detecting driver mutations, namely identifying recurrent mutations. The rationale behind this approach is that, even if each cancer sample has undergone an independent evolutionary process, the mutations that drive the progression of the same tumor type should appear more frequently than expected by chance across patient samples.

In this perspective, recurrence may be revealed at different levels of resolution, from individual nucleotide, or codon, to protein level, or to the whole gene, or even to a pathway (Raphael et al. 2014). For the sake of simplicity, and brevity, here it will suffice to focus on those approaches which deal with just one level of resolution, namely statistical tests for genes with recurrent single-nucleotide mutations.

Several methods have been designed to find single-nucleotide recurrent mutations. But they all share the same core principle. Indeed, the fundamental calculation “in all these approaches is to determine whether the observed number of mutations in the gene is significantly greater than the number expected according to a background mutation rate (BMR)” (Raphael et al. 2014, p. 7). It is not difficult to recognize here a standard way to statistically detect a significant deviation from expected results. But in this context, this is a key point. Indeed, the BMR “is the probability of observing a passenger mutation in a specific location of the genome” (Ibidem). From the BMR and the number of sequenced nucleotides within a gene, “a binomial model can be used to derive the probability of the observed number of mutations in a gene across a cohort of patients” (Ibidem).²⁸

The problem is: If we are searching for a way to identify driver mutations, i.e. to distinguish driver from passenger mutations, how can the BMR, i.e. the probability that a *passenger* mutation can be found in a specific location of the genome, be already estimable? In fact, it is estimated on the basis of previously acquired biological knowledge. Indeed, those who develop statistical models for detecting single-nucleotide recurrent mutations incorporate in their models some features of passenger mutations. For instance, they assume, among other things, that “BMR is not constant across the genome, but depends on the genomic context of a nucleotide [...] and the type of mutation”, that “the BMR of a gene is correlated with both its rate of transcription

²⁸ Cf. Raphael (2014, p. 7): “Using the background mutation rate (BMR) and the number n of sequenced nucleotides within a gene (g), the probability (Pg) that a passenger mutation is observed in g is given by $Pg = 1 - (1 - \text{BMR})^n$. Since somatic mutations arise independently in each sample, the occurrences of passenger mutations in g are modeled by flipping a biased coin with probability pg of heads (mutation). Thus, if somatic mutations have been measured in m samples, the number of patients in which gene g is mutated is described by a binomial random variable $B(m, Pg)$ with parameters m and Pg . From $B(m, Pg)$, it is possible to compute the probability that the observed number or more samples contain passenger mutations; this is the P value of the statistical test”.

[...] and replication timing”, and that the “BMR is also not constant across patients” (Ibidem). The estimated BMR greatly affects the identification of recurrent mutations, and so the identifications of driver mutations. This means that different methods for identifying recurrently mutated genes, since they may diverge in their estimation of the BMR, can (and in fact *do*) diverge in the identification of driver mutations (Ibidem; Tokheim et al. 2016). But this is not the problem we would like to focus on.

The big epistemological problem, as hinted above, is that these methods for identifying driver mutations assume data relative to the expected frequency of passenger mutations, data which can be produced only by assuming that driver and passenger mutations actually exist and can be distinguished, an assumption that is based in its turn on the very hypothesis that should be confirmed, namely SMT. Indeed, driver mutations can be regarded as such only by assuming SMT.²⁹ If one tries to detect driver mutations in order to confirm SMT, it would be circular to incorporate in one’s method for detecting driver mutations a frequency distribution which is developed relying on the assumed existence of driver mutations.

On the contrary, computational methods for detecting driver mutations, in order to confirm SMT over TOFT, should be able to show us that the set of all detectable mutations in a cancer genome can be unambiguously assigned to two distinct subsets, namely the sets of driver and passenger mutations, in a principled way independent of the way driver mutations are defined by SMT. If instead driver mutations are identified because in the set of all the detected mutations in a cancer genome they are the most recurrent, and, according to SMT, the most recurrent mutations cannot but be driver mutations, we have no independent reason to claim that SMT is confirmed by the computational methods developed to identify driver mutations.

Similar problems afflict the other main strategies that have been developed by bioinformaticists for detecting driver mutations, namely predicting the functional impact of individual mutations, and assessing combinations of mutations. Indeed, all these approaches “assume that *a priori* information [...] will help to distinguish passenger from driver mutations” (Ibidem, p. 9). But “one important bias in the methods that predict cancer genes is” precisely “the direct or indirect incorporation of prior knowledge” (Dimitrakopoulos and Beerenwinkel 2017, p. 12).³⁰ Such already avail-

²⁹ The fact that some data can be regarded as relevant to the estimation of the frequency distribution of passenger and driver mutations only if one assumes SMT can be clearly seen by considering that if one assumes TOFT, then the very same data (i.e. the detected mutations in a given cancer genome) cannot be regarded as instances of ‘driver’ mutations, simply because according to TOFT somatic mutations are not the cause of cancer. If one does not assume SMT, by searching for mutations one can at most display correlations of recurrent mutations in cancer cells, but one cannot prove that those mutations are the cause of carcinogenesis, and so one cannot disconfirm TOFT. Indeed, TOFT does not deny the existence of mutations in the genome of cancer cells. It denies that these mutations are the cause of cancer insurgence.

³⁰ It may be objected that the so called *de novo* approaches, which aim at the identification of driver mutations by statistically analyzing combinations of mutations in networks and pathways (this method belongs to the third kind of approaches to individuate driver mutations listed above) are less prone to this criticism, because they do not incorporate previous knowledge about genes associated with well-studied cancer pathways. But this objection is inadequate. Indeed, even *de novo* approaches are not independent from crucial assumptions that are not neutral with respect to what hypothesis on carcinogenesis is adopted. In order to identify novel combinations of mutations or mutated genes, “it would be ideal to test all possible combinations for recurrent mutations across a cohort of cancer patients, but such a *de novo* approach is impractical. For example, there are more than 10^{29} possible sets of eight genes in the human genome,

able information is often constituted and interpreted assuming SMT, and so assuming the distinction between driver and passenger mutations. At this regard, Baker maintains that current bioinformatic methods identify driver mutations “in terms of their likelihood of being driver mutations, but do not prove the existence of driver mutations [...]. Gold standards for evaluating bioinformatics predictions of driver mutations [...] are based on postulated driver mutations and not on unambiguously established driver mutations” (Baker 2015b, p. 1).

The case of the search for the identification of driver mutations seems to conform to the above proposed distinction between plausibility and probability. Relying on statistical tools we can well estimate the probability that a given mutation is a driver mutation. And we may also claim that such probability is objective, because it is calculated by relying on empirical reliable findings. But, for instance, the theoretical decision of considering some mutations as instances of driver or passenger mutations in order to estimate the BMR, i.e. to interpret these findings in accordance to SMT, cannot be accounted for in terms of probability, it may instead be accounted for in terms of *plausibility*, i.e. in terms of the assessment of the arguments for and against this hypothesis.

3.5 The search for driver mutations and big data

The confidence on the possibility of developing effective treatments based on personalized cancer medicine is driven in part by an optimistic attitude towards the increasing availability of large amount of data. Many thought that, despite the divergences between our current theories of carcinogenesis, conflating in some statistical algorithm the ‘deluge of data’ coming from ‘omics’ researches would have allowed us to derive the right diagnoses and prognoses. On the contrary, as we have seen above (Sect. 2.2), an enormous quantity of data may put pressure on the theoretical assumptions that currently dominate a research field. In other words, unless we possess a powerful and adequate theoretical perspective on the phenomenon we are analyzing, the deluge of data will probably not improve our understanding of that phenomenon. The risk is to be confused.

As we have already noted above, when the stock of data increases, even the number of spurious correlations increases. This is what happens in the case of the search for

Footnote 30 continued

which is both too many to evaluate computationally and too many hypotheses to test while retaining statistical power” (Raphael et al. 2014, p. 12). In order to overcome this difficulty, *de novo* approaches try to identify driver mutations by searching for genetic aberrations which are both (1) highly recurrent, and (2) mutually exclusive, i.e. they do not compare in different pathways. So, only if one accepts the “hypotheses that each tumor has relatively few driver mutations [...] and these driver mutations perturb multiple cellular functions in different pathways [...], one can conclude that a tumor rarely possesses more than one driver mutation per pathway”, and so that “when examining data across cancer samples, driver pathways [...] correspond to mutually exclusive sets of genes” (Ibidem). But these assumptions are not independent from prior knowledge. Indeed, they relies on SMT, since they presuppose the existence of driver mutations, and even a precise hypothesis about their frequency, presuppositions which cannot be based on nothing but some kind of prior knowledge. Thus, *de novo* approaches do not really differ from other kinds of approaches developed to identify driver mutations, and cannot be said to be independent from previous knowledge. So, this objection is inadequate.

the identification of driver mutations. Lawrence and colleagues describe the situation as follows: many “international projects are aimed at creating a comprehensive catalogue of all the genes responsible for the initiation and progression of cancer”; these studies “involve the sequencing of matched tumour-normal samples followed by mathematical analysis to identify those genes in which mutations occur more frequently than expected by random chance”; but the fundamental problem with cancer genome studies is that “as the sample size increases, the list of putatively significant genes produced by current analytical methods burgeons into the hundreds”, and the list “includes many implausible genes [...], suggesting extensive false positive findings that overshadow true driver events” (Lawrence et al. 2013, p. 214).

In our view, this means that in order to achieve a deeper understanding of a given phenomenon, what is really needed is more theoretical work and the production of more plausible hypotheses to be tested, rather than the mere production of more data. The process of evidence amalgamation necessarily requires some theoretical hypotheses to work properly, and those hypotheses are produced and accepted on the basis of plausibility-based considerations.

At this regard, Weinberg, one of the most influential authors of SMT, seems to have an ambiguous position. Indeed, if, on the one hand, he claims that the “data that we now generate overwhelm our abilities of interpretation”, and we “lack the conceptual paradigms and computational strategies for dealing with” these data, on the other hand, he seems to think, as supporters of big data usually think, that the key for the advancement of our understanding is the development of more powerful statistical models, i.e. the development of new tools to amalgamate the biggest quantity possible of evidences, models which will give us the right answers. He states that “we don’t know how to integrate individual data sets, such as those deriving from cancer genome analyses, with other, equally important data sets, such as proteomics”, and that this is frustrating, because it is “becoming increasingly apparent that a precise and truly useful understanding of the behavior of individual cancer cells and the tumors that they form will only come once we are able to integrate and then distill these data” (Weinberg 2014, p. 271).

We think instead that the development of new *theoretical* hypotheses to be assessed both by means of plausibility-based considerations and empirical confirmation, is still the key issue in the advancement of science. We also think, as we hope to have made clear, that also the so often invoked development of new statistical tools proceeds in this way, and so that it is not independent from human theoretical efforts and plausibility-based considerations.

3.6 Plausibility and the debate between SMT and TOFT

We would like to conclude this section with some more general considerations. It may be objected that even if there is some difficulty in adjudicating between SMT and TOFT by merely relying on computational tools for identifying driver mutations, if we consider all the evidences available, we should be able to reach a shared conclusion. But it seems not to be the case. Indeed, if it were possible to objectively collect and amalgamate all relevant evidences available from cancer

research, put data in some statistical model for theory confirmation, and objectively assign some probability distribution to the rival hypotheses, we would be able to clearly assess what theory is empirically more confirmed between SMT and TOFT.

But, again, things are more complicated than that. Recently, there have been some important confrontations on what is the best hypothesis on carcinogenesis between SMT and TOFT (see e.g. [Bedessem and Ruphy 2015, 2017](#); [Bizzarri and Cucina 2016](#); [Baker 2015a, b](#); [Kaye 2015](#); [Soto and Sonnenschein 2011](#); [Sonnenschein and Soto 2011](#); [Vaux 2011a, b](#)). Although they often consider the very same set of evidences, different scholars draw almost opposite conclusions on which hypothesis is the most confirmed by those evidences.

For example, [Bizzarri and Cucina \(2016\)](#) think that SMT is *not* indirectly confirmed by the results of the therapeutic strategy aimed at targeting specific relevant mutations in the treatment of chronic myelogenous leukemia:

Evidence arguing for the irrelevance of mutations as a target for therapeutic management comes from studies performed on chronic myelogenous *leukemia*. It has been claimed that the abnormal fusion tyrosine kinase BCR-ABL acts as an ‘oncogene’ and is deemed the key-initiating factor in myelogenous neoplastic transformation. Inhibition of the corresponding oncoproteins by means of tyrosine kinase inhibitor (TKI) has indeed lead to significant short-term beneficial responses, yet without achieving any benefit in terms of long-term survival. This latter failure has been ascribed to the fact that a reservoir of cancer stem cells still proliferates because they lack the alleged targeted-mutated gene and they are therefore insensitive to the TKI [...]. Thus, accordingly to this rationale, myeloid cells would become transformed by an oncogene that curiously is absent among the cancer stem cell population from which cancer is thought to arise. ([Bizzarri and Cucina 2016](#), p. 223).

On the opposite side, [Vaux \(2011a\)](#) think that precisely the same clinical example provides the best empirical support currently available to SMT:

the most dramatic support [to SMT] comes from the clinic. CML [i.e. chronic myelogenous *leukemia*] is caused by a chromosomal translocation that generates the Philadelphia chromosome and activates the BCR-ABL fusion oncoprotein. If CML were instead due to changes in tissue organization, it ought not respond to imatinib, an inhibitor of the ABL kinase, yet CML responds extraordinarily well to imatinib [...]. Furthermore, in cases that develop resistance to the drug, additional mutations are found to the *bcr-abl* gene in sub-clones of the leukemia cells [...], indicating that not only development of CML, but also drug resistance, is due to sequential DNA mutations arising in somatic cells, in accordance with SMT. ([Vaux 2011a](#), p. 343).

As already stated, we do not aim at solving the dispute between SMT supporters and TOFT supporters here. What we aim at pointing out by focusing on such a theoretical disagreement is that this kind of theory/hypothesis assessment may be better accounted for in terms of plausibility-based considerations rather than in terms of probability-based considerations. Given that often the authors that diverge on such

theoretical issues consider almost the same empirical data, if they should have adopted a theoretical stance relying on probability-based considerations, and probability were objective and simply dictated to us by the way the world is, these authors should have arrived at the same conclusions.

But authors do not arrive at the same conclusions. Nor they justify the theoretical stance they adopt by making reference to probability-based considerations. Rather, these authors provide *arguments* and *reasons* to support their favored hypothesis, they do not merely provide a probability-based estimation of the degree of empirical confirmation of that hypothesis. It may be objected that conclusions diverge because theory assessment follows a Bayesian path, and given that different authors start from different prior probabilities assigned to rival hypotheses, they arrive at different conclusions. But even accepting such a Bayesian framework, the point now is: Why different authors assign different prior probabilities to rival hypotheses? Again, if it were possible to objectively assign these probability values, conclusions should not be divergent. But conclusions *do* diverge. This means that priors probabilities are not objectively assigned, i.e. they are not assigned by means of a (potentially mechanizable) procedure which can univocally determine each prior in an uncontroversial way. Thus, it seems fair to suppose that scholars assign prior probabilities to rival hypotheses by relying on plausibility-based considerations.

Probability and plausibility do not stand opposed, they are not rival concepts, yet they are distinct concepts, and we think that understanding how they are related may allow us to better understand how evidences are produced and amalgamated in scientific research.

4 Plausibility and evidence amalgamation in medicine

Let us conclude this article by considering our proposal from a broader perspective. Philosophical investigations on cancer research are nested in the philosophy of medicine. So, we would like to address some more general epistemological issues that are nevertheless central for the philosophical analysis of cancer research. In what follows, we briefly put the thesis we argued for in this article in this broader context.

There are at least two main conceptions of statistics, namely classical statistics and Bayesian statistics.³¹ These two ways of conceiving of statistics are also often associated to two distinct ways of conceiving of probabilities: supporters of classical statistics usually adopt a *frequentist* perspective on probability, while supporters of Bayesian statistics usually adopt a conception of probability as *degree of belief*. The conception of probability one adopts may affect the evaluation of one's inquiry. We argue that, despite their divergences, neither frequentists nor Bayesians give a complete representation of how evidences are amalgamated in medicine, and that considering the role that plausibility-based considerations play in the process of evidence amalgamation can give some insights on this issue.

³¹ For a detailed illustration of the main views of statistics, see Romeijn (2017).

4.1 Random clinical trials

To better see this point, consider one of the most important issues in which statistics and medicine intersect each other, namely the validation of the efficacy of drugs and treatments. On this issue, scholars are divided into two main positions: on the one hand, there is the dominant position, inspired by the ideas of the Evidence Based Medicine (EBM),³² according to which the gold standard of confirmation in medical research are Random Clinical Trials (RCTs) (see e.g. [Papineau 1994](#)); on the other hand, there is the position advocated by those who criticize RCTs from a Bayesian perspective (see e.g. [Worrall 2007b](#)).

These two ways of conceiving of clinical trials rest on two different conceptions of probability. Indeed, supporters of RCTs usually adopt a frequentist interpretation of probability, while Bayesians usually adopt an interpretation of probability as degree of belief. The differences between these two conceptions of probability are reflected in the way these approaches to clinical trials consider evidences be amalgamable. Some authors speak of ‘evidence elitism’ with regard to the supporters of RCTs, and of ‘methodological pluralism’ with regard to the Bayesians. According to EBM, randomized trials are the only truly reliable source of evidence in clinical testing. Other sources of evidences may well be taken into consideration in order to decide how to act in the absence (or in the impossibility of performing) RCTs. But this does not mean that evidences coming from other sources can be amalgamated with those deriving from RCTs in order to draw a conclusion on the drug (or treatment) we are evaluating through a RCT. In this case, evidences coming from the RCT have to be preferred ([Worrall 2007a](#)).

4.2 Frequentist approaches to clinical trials

But how exactly the way probability is conceived affects the way clinical trials are conceived? Consider RCTs. As already noted, supporters of RCTs adopt frequentism, and in frequentism probabilities are relative frequencies of empirical events.

In this view, probability of an event e is defined as the limit of the proportion, as n increases, of

$$f = k/n$$

“where f is the frequency of occurrence of the relevant event, k is the number of times the event occurs in n repetitions of the experiment” ([Djulgovic et al. 2011](#), p. 309). Ideally, when n goes to infinite, f gives us the objective probability of e . In other words, in this perspective probability is equivalent to the frequency of an event, and this frequency is determined by the way the world is. This is clearly an objective view of probability, which may seem perfectly suitable for all those who aim at an objective knowledge of the world.

³² For a survey on EBM, see [Bluhm and Borgerson \(2011\)](#).

But the devil is in the details, and things are not so easy for frequentists as they may *prima facie* appear. The main problem with this conception of probability is precisely that it equates probabilities and frequencies. This idea leads to several difficulties, which all derive from the same theoretical problem: if probabilities are frequencies, then in order to calculate the *objective* probability of a given event e , we should try to replicate the very same experiment as many times as possible, and see how many times e occurs. Indeed, if probabilities are frequencies, in order to estimate the probability of e , we should estimate the frequency of e . Obviously, even granting that probability is objective, if we deal with very limited sets of trials, our estimation of the frequency of e may strongly diverge from the real value of such frequency. Thus, according to the law of large numbers, in order to secure our confidence in the objectivity of the probability assigned to e , we should be able to perform a huge number of replications of the same trial to better estimate the frequency of e .

Consider a fair coin. If you toss it ten times, there is a high probability that you will not obtain a score of 5 heads and 5 tails, i.e. the probability values that probability theory predicts in this case. Frequencies approximate theoretical probability values only if replications tend to infinity. So, if we try to test the equiprobability of heads and tails empirically, we will approximate the theoretical value only in the (very) long run. In the coin example, we will probably do better if we toss the coin 10,000 times. And even better if we toss the coin 100,000,000 times.

Consider now RCTs. The rationale behind this experimental design is that random assignment of patients may neutralize biases and confounders (Teira 2011). This would allow us to derive the objective probability of the hypothesis we are testing.

There are two main (and related) epistemological problems with this perspective. The first is how to determine whether the sample of population we select in our trial is sufficiently similar to the target population. As we have seen, in a small sub-set of occurrences, our estimation of the frequency of an event may strongly diverge from the frequency of that event relative to the whole set. In clinical trials, this divergence may be due to relevant differences in the distributions of the relevant factors in the study population and the target population.³³ An analogous issue is establishing whether the populations assigned to the different arms of the trial are equivalent.³⁴ Indeed, a trial is biased “if (whether or not we know it) there is some difference between the experimental and control groups” (Worrall 2007a, p. 993).

Randomization and replication should be the keys here (Howson and Urbach 2006). If there are say n factors that may be relevant and lead to biases or confounders in evaluating the efficacy of a drug, a random assignment of patients is thought to be able to distribute those n factors so that the frequencies of those n factors in the study population approximate the frequencies of those n factors in the target population. Indeed, it is known that non-random assignment of patients to the distinct arms of a trial may induce bias and confounding.

³³ This is the so-called problem of the external validity, see Worrall (2010).

³⁴ Cf. Howson and Urbach (2006, p. 183): “Clinical trials typically involve two groups of subjects, all of whom are currently suffering from a particular medical condition; one of the groups, *the test group*, is administered the experimental therapy, while the other, *the control group*, is not [...]”

The problem is that even random assignment may well produce a study population which is significantly divergent from the target population (Worrall 2010). To balance all the possible biases and confounders we have to add replication to randomization, i.e. we need to increase the number of study sub-populations selected through random assignment. When the number of random sub-populations grows, the probability that the mean value of the n factors in those populations will approximate the value of the n factors in the target population increases. But n may be constituted by known factors as well as unknown ones. This means that we could safely claim that the target population is well represented in our study population and relevant factors are well balanced, so that biases and confounders are prevented and neutralized, only in two cases: (1) if we already know with certainty all the relevant factors that should be considered for evaluating a given drug or treatment; (2) if the number of members of the target population and the number of sub-populations go to infinity. Both conditions usually do not obtain. The same reasoning applies to the issue of establishing whether the trial is biased in the sense that there is some difference between the experimental and control groups. How can we compare those groups and safely claim that they are equivalent if there may be some unknown factors that may be not equivalently distributed among them?

The point is that, as well as there is no way to claim that it is not possible that some unconceived alternative to a given hypothesis will appear, there is no way to rule out the possibility that there may be some not yet known relevant factors for the case under investigation.³⁵ Since our knowledge of what are the relevant factors is limited and fallible, even if those factors are actually of a finite number, in order to claim with certainty that *all possible* biases and confounders have been neutralized, we should be able to managing infinite populations and replications. But managing actual infinity in empirical domains is prevented to us humans. So, we cannot know with certainty whether the frequency of event e that we estimated in our trial is *really* objective, i.e. whether or not it approximates the frequency value of e in the whole population.

The second main epistemological problem with RTCs is more straightforwardly connected with the issue of replicability. As we have seen, in frequentism it is the number of replications that gives us reasons to think that the observed frequency approximates the objective probability of a given event. Unfortunately, RCTs usually cannot be too large for economical and ethical reasons (Worrall 2007a, b). But there is also a theoretical difficulty. A randomized clinical trial can never be *really* replicated. This is due to the fact that we cannot take our sample of patients, and after a first round of treatment, re-randomize it and start the trial again (Worrall 2007b).³⁶

This analysis of the epistemological difficulties that afflict RCTs is intended to point out how the *objectivity* of the results obtained through RCTs, which is usually invoked

³⁵ Cf. e.g. Worrall (2007b, p. 472): “there can be no estimate of how closely balanced a particular real trial is with respect to any unknown factor—this is so by definition, since the unknown factor is unknown!”

³⁶ Cf. Worrall (2007b, p. 472): “there is also an epistemological issue about whether any repeated random trial would be comparable to the initial one. If a particular patient in the study receives, say, the ‘active drug’ on the first round, then, since this is expected to have some effect on his or her condition, the second randomization would not be rigorously a true repetition of the first. The second trial population, though consisting of the same individuals, would, in a possibly epistemically significant sense, not be the same population as took part in the initial trial.”

as the main reason to adopt RCTs, can be maintained only if some epistemic decisions are taken on how and (to what extent) less idealized conditions can be accepted for a trial in a certain context. We maintain that these decisions are taken by performing a plausibility-based analysis of the context under investigation, and are informed by previous knowledge of relevant facts.³⁷

4.3 Bayesian approaches to clinical trials

As we have already noted, Bayesians make several criticisms of RCTs. The most relevant are: (1) the claim that RCTs deal with objective probabilities hides many epistemic decisions that are instead taken in the actual development of RCTs; (2) RCTs do not accept relevant kinds of evidences that should instead be taken into consideration in the process of evaluation of a drug or treatment.

Those criticisms have their roots in the different way of conceiving of probability that Bayesians adopt. Indeed, they usually regard probability as a measure of degree of belief, rather than a frequency. In this view, probabilities are in the ultimate analysis related to states of mind and not (directly, at least) to states of objects. According to many Bayesians, adopting a Bayesian strategy in the evaluation of drugs and treatments would allow us to validate the efficacy of a drug or treatment in minor time and at a minor cost.

Let's briefly consider the epistemological difficulties that the Bayesians have to face. Roughly, they can be reduced to the main one: the issue of *prior probabilities* assignment.³⁸ This is a crucial issue, because the assignment of different priors leads to different results in the calculation of conditional probabilities.

There is a huge amount of literature on this issue (see [Howson and Urbach 2006](#); [Williamson 2010](#)). What is undeniable is that there is not a principled way to assign prior probabilities which is widely accepted and may be *really* deemed to be objective. Indeed, since we are dealing here with a 'degree of belief' conception of probability, we cannot rule out *the possibility* that different priors may be assigned by different subjects to the very same hypothesis. According to critics of the Bayesian approach, such a subjective view of probability would introduce an unacceptable degree of subjectivity in our evaluative process, and this would make this process arbitrary.

Moreover, in this view priors reflect *current* knowledge, so they do not reflect the way the world really is, but the degree of our knowledge of it. Thus, even if we put aside the role of subjectivity in prior probabilities *assignment*, and adopt *objective*

³⁷ Cf. Worrall (2007a, pp. 1000–1001): “randomisation does not free us from having to think about alternative explanations for particular trial outcomes and from assessing the plausibility of these in the light of ‘background knowledge’.”

³⁸ On the different Bayesian perspectives on prior probabilities, cf. Williamson (2010, p. 2): “All Bayesian epistemologists hold that rational degrees of belief are probabilities, consistent with total available evidence, and updated in the light of new evidence by Bayesian conditionalization. Strict subjectivists (e.g. Bruno de Finetti) hold that initial or *prior* degrees of belief are largely a question of personal choice. Empirically based subjectivists (e.g. Howson and Urbach [...]) hold that prior degrees of belief should not only be consistent with total evidence, but should also be calibrated with physical probabilities to the extent that they are known. Objectivists (e.g. Edwin Jaynes) hold that prior degrees of belief are fully determined by the evidence.”

Bayesianism (Williamson 2010), according to which prior degrees of belief are fully determined by the evidence, the problem of the unconceived alternatives is still there. This means that the problem of subjectively assigning some prior degree of belief is just moved one step back. Indeed, prior probabilities can be assigned in an objective way only if we could affirm to know all the possible outcomes. But since our current knowledge is contingent and fallible, we cannot exclude that there may be other possible outcomes that we do not know yet. Thus, we cannot calculate the probability of each possible outcome in a *truly* objective way. As we have seen above, we can proceed by making plausibility-based considerations on the relevant data and our previous knowledge, and assigning prior probabilities accordingly. This means that the objective Bayesian may well assign a prior probability value p to a given hypothesis h on the basis of the set e of empirical evidences for p that are currently available. She may also claim that p is fully determined by e . But what is the prior degree of belief p_1 that we should assign to the hypothesis h_1 that evidences collected in e are reliable, and so that we can safely rely on them in order to determine p ? If we think that p_1 may be fully determined in its turn by another set of evidences e_1 , we risk ending in a regress. So, at some point at least, some prior probabilities assignment cannot be maintained to be objectively and fully determined by evidences alone.

In other words, since prior probabilities assignment cannot avoid resorting to plausibility-based considerations, even if Bayesian strategies of testing can account for all the relevant evidences that enter the evaluative process of practitioners, they are nevertheless unable to dictate an uncontroversial way of assigning values to the prior probabilities relative to such evidences.

To sum up, plausibility-based considerations play a relevant role both in frequentist and Bayesian approaches to clinical trials (although this role is neglected by both these approaches), because these approaches deal with a context afflicted by the problem of the unconceived alternatives. We think that taking into account the role of plausibility-based considerations can contribute to clarify some epistemological shortcomings that afflict both frequentist and Bayesian perspectives.

5 Conclusion

In this article, we firstly introduced the analytic view of theory development and illustrated the concept of plausibility to some extent in order to make clear in what sense plausibility and probability are distinct concepts. We used the concept of plausibility to point out the ineliminable role played by the epistemic subject in the process of evidence amalgamation and in the process of theory assessment. Then, we moved to address a central issue in current cancer research, namely the relevance of computational tools developed by bioinformaticists to detect driver mutations in the debate between the two main rival theories of carcinogenesis, namely SMT and TOFT. Finally, we briefly extended our considerations on the role that plausibility plays in evidence amalgamation from cancer research to the more general issue of the divergences between frequentists and Bayesians in the philosophy of medicine and statistics. We argued that considering the role played by plausibility-based considerations may lead to clarify some of the epistemological shortcomings that afflict both these perspectives.

References

- Allen, J. F. (2001). In silico veritas. Data-mining and automated discovery: The truth is in there. *EMBO Reports*, 2(7), 542–544.
- Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*, 16(7), 16.
- Baker, S. G. (2014). Recognizing paradigm instability in theories of carcinogenesis. *British Journal of Medicine and Medical Research*, 4(5), 1149–1163.
- Baker, S. G. (2015a). A cancer theory kerfuffle can lead to new lines of research. *Journal of the National Cancer Institute*, 107(2), 405. doi:10.1093/jnci/dju405.
- Baker, S. G. (2015b). Response [to: Kaye, F.J. (2015)]. *Journal of the National Cancer Institute*, 107(5), djv061. doi:10.1093/jnci/djv061.
- Baker, S. G. (2017). The questionable premises underlying the search for cancer driver mutations and cancer susceptibility genes. *Organisms*, 1(1), 3–4.
- Bandyopadhyay, P. S., & Forster, M. R. (2011). Philosophy of statistics: An introduction. In P. S. Bandyopadhyay & M. R. Forster (Eds.), *Handbook of the philosophy of science. Volume 7. Philosophy of statistics* (pp. 1–50). Amsterdam: Elsevier.
- Bassett, D. E., Jr., Eisen, M. B., & Boguski, M. S. (1999). Gene expression informatics—It's all in your mine. *Nature Genetics*, 21(1 Suppl.), 51–55.
- Bedessem, B., & Ruphy, S. (2015). SMT or TOFT? How the two main theories of carcinogenesis are made (artificially) incompatible. *Acta Biotheoretica*, 63(3), 257–267.
- Bedessem, B., & Ruphy, S. (2017). SMT and TOFT integrable after all: A reply to Bizzarri and Cucina. *Acta Biotheoretica*, 65(1), 81–85.
- Bertolaso, M. (2016). *Philosophy of cancer. A dynamic and relational view*. Dordrecht: Springer.
- Bird, A. (2017). Systematicity, knowledge, and bias. How systematicity made clinical medicine a science. *Synthese*. doi:10.1007/s11229-017-1342-y.
- Bizzarri, M., & Cucina, A. (2016). SMT and TOFT: Why and how they are opposite and incompatible paradigms. *Acta Biotheoretica*, 64(3), 221–239.
- Bluhm, R., & Borgerson, K. (2011). Evidence-based medicine. In F. Gifford (Ed.), *Handbook of the philosophy of science. Volume 16. Philosophy of medicine* (pp. 204–238). Amsterdam: Elsevier.
- Brown, P. O., & Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nature Genetics*, 21(1 Suppl.), 33–37.
- Calude, C. (2002). *Information and randomness. An algorithmic perspective*. Berlin: Springer.
- Calude, C., & Longo, G. (2016a). Classical, quantum and biological randomness as relative unpredictability. *Natural Computing*, 15(2), 263–278.
- Calude, C. S., & Longo, G. (2016b). The Deluge of spurious correlations in big data. *Foundations of Science*. doi:10.1007/s10699-016-9489-4.
- Cellucci, C. (2013). *Rethinking logic. Logic in relation to mathematics, evolution, and method*. Dordrecht: Springer.
- Cellucci, C. (2016). Models of science and models in science. In E. Ippoliti, F. Sterpetti, & T. Nickles (Eds.), *Models and inferences in science* (pp. 95–122). Cham: Springer.
- Cellucci, C. (2017a). *Rethinking knowledge. The heuristic view*. Dordrecht: Springer.
- Cellucci, C. (in press). Theory building as problem solving. In E. Ippoliti & D. Danks (Eds.), *Building theories*. Cham: Springer.
- Chen, R., & Snyder, M. (2013). Promise of personalized omics to precision medicine. *WIREs Systems Biology and Medicine*, 5(1), 73–82.
- Cirkel, G. A., Gadellaa-van Hooijdonk, C. G., Koudijs, M. J., Willems, S. M., & Voest, E. E. (2014). Tumor heterogeneity and personalized cancer medicine: Are we being outnumbered? *Future Oncology*, 10(3), 417–428.
- Coveney, P. V., Dougherty, E. R., & Highfield, R. R. (2016). Big data need big theory too. *Philosophical Transactions of the Royal Society A*, 374(2080), 1–11.
- Dellsén, F. (2016). Scientific progress: Knowledge versus understanding. *Studies in History and Philosophy of Science*, 56, 72–83.
- Dimitrakopoulos, C. M., & Beerenwinkel, N. (2017). Computational approaches for the identification of cancer genes and pathways. *Wiley Interdisciplinary Reviews. System Biology and Medicine*. doi:10.1002/wsbm.1364.

- Djulbegovic, B., Hozo, I., & Greenland, S. (2011). Uncertainty in clinical medicine. In F. Gifford (Ed.), *Handbook of the philosophy of science. Volume 16. Philosophy of medicine* (pp. 298–356). Amsterdam: Elsevier.
- Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review*, 1(2), 293–314.
- Gagneur, J., Friedel, C., Heun, V., Zimmer, R., & Rost, B. (2017). Bioinformatics advances biology and medicine by turning big data troves into knowledge. *Informatik Spektrum*, 40(2), 153–160.
- Gelman, A., Hennig, C. (2017). Beyond subjective and objective in statistics. *Journal of the Royal Statistical Society: Series A*, 180(4), 1–31.
- Gillies, D. (2000). *Philosophical theories of probability*. London: Routledge.
- Goodman, S. N. (1999). Toward evidence-based medical statistics. 1: The P value fallacy. *Annals of Internal Medicine*, 130(12), 995–1004.
- Goodman, S. N. (2001). Of P-values and Bayes: A modest proposal. *Epidemiology*, 12(3), 295–297.
- Howson, C., & Urbach, P. (2006). *Scientific reasoning. The Bayesian approach* (3rd ed.). Chicago, La Salle: Open Court.
- Katsnelson, A. (2013). Momentum grows to make ‘personalized’ medicine more ‘precise’. *Nature Medicine*, 19(3), 249.
- Kant, I. (1992). *Lectures on logic*. Cambridge: Cambridge University Press.
- Kaye, F. J. (2015). RE: A cancer theory kerfuffle can lead to new lines of research. *Journal of the National Cancer Institute*, 107(5), djv060. doi:10.1093/jnci/djv060.
- Laplace, P. S. (1951). *A philosophical essay on probabilities*. New York: Dover Publications [1st French ed.: 1814].
- Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, 499(7457), 214–218.
- Lindley, D. V. (2000). The philosophy of statistics. *Journal of the Royal Statistical Society: Series D*, 49(3), 293–319.
- Longo, G. (2017). The biological consequences of the computational world: Mathematical reflections on cancer biology. [arXiv:1701.08085v2](https://arxiv.org/abs/1701.08085v2).
- Longo, G., Montévil, M., Sonnenschein, C., & Soto, A. M. (2015). In search of principles for a theory of organisms. *Journal of Biosciences*, 40(5), 955–968.
- Magnus, P. D. (2006). What’s new about the new induction? *Synthese*, 148(2), 295–301.
- Mazzocchi, F. (2015). Could big data be the end of theory in science? A few remarks on the epistemology of data-driven science. *EMBO Reports*, 16(10), 1250–1255.
- Merid, S. M., Goranskaya, D., & Alexeyenko, A. (2014). Distinguishing between driver and passenger mutations in individual cancer genomes by network enrichment analysis. *BMC Bioinformatics*, 15, 308. doi:10.1186/1471-2105-15-308.
- Mizrahi, M. (2014). The problem of unconceived objections. *Argumentation*, 28(4), 425–436.
- Mizrahi, M. (2016). Historical inductions, unconceived alternatives, and unconceived objections. *Journal for General Philosophy of Science*, 47(1), 59–68.
- Musgrave, A. (2011). Popper and hypothetico-deductivism. In D. Gabbay, H. Stephan, & J. Woods (Eds.), *Handbook of the history of logic. Volume 10. Inductive logic* (pp. 205–234). Amsterdam: North-Holland.
- Ow, G. S., & Kuznetsov, V. A. (2016). Big genomics and clinical data analytics strategies for precision cancer prognosis. *Scientific Reports*, 6(36493), 1–13.
- Papineau, D. (1994). The virtues of randomization. *The British Journal for the Philosophy of Science*, 45(2), 437–450.
- Pollock, J. L. (1983). Epistemology and probability. *Synthese*, 55(2), 231–252.
- Pólya, G. (1941). Heuristic reasoning and the theory of probability. *The American Mathematical Monthly*, 48(7), 450–465.
- Popper, K. R. (2005). *The logic of scientific discovery*. London: Routledge.
- Putnam, H. (1975). *Mathematics, matter and method. Philosophical papers* (Vol. 1). Cambridge: Cambridge University Press.
- Raphael, B. J., Dobson, J. R., Oesper, L., & Vandin, F. (2014). Identifying driver mutations in sequenced cancer genomes: Computational approaches to enable precision medicine. *Genome Medicine*, 6(5), 1–17.

- Rigo-Lemini, M., & Martínez-Navarro, B. (2017). Epistemic states of conviction. A conceptualization from the practice of mathematicians and neurobiology. In U. E. Yolcottzin (Ed.), *Understanding emotions in mathematical thinking and learning* (pp. 97–131). London: Academic Press.
- Romeijn, J.-W. (2017). Philosophy of statistics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/archives/spr2017/entries/statistics/>.
- Rosenfeld, S. (2013). Are the somatic mutation and tissue organization field theories of carcinogenesis incompatible? *Cancer Informatics*, 12, 221–229.
- Rowbottom, D. P. (2016). Extending the argument from unconceived alternatives: Observations, models, predictions, explanations, methods, instruments, experiments, and values. *Synthese*. doi:10.1007/s11229-016-1132-y.
- Ruhmkorff, S. (2011). Some difficulties for the problem of unconceived alternatives. *Philosophy of Science*, 78(5), 875–886.
- Saatsi, J., Psillos, S., Winther, R. G., & Stanford, K. (2009). Grasping at realist straws. *Metascience*, 18(3), 370–379.
- Salmon, W. C. (1990). The appraisal of theories: Kuhn meets Bayes. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2, 325–332.
- Schickore, J. (2014). Scientific discovery. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. <http://plato.stanford.edu/archives/spr2014/entries/scientific-discovery/>.
- Schupbach, J. N. (2011). *Studies in the logic of explanatory power*. Ph.D. thesis, University of Pittsburgh, School of Arts and Sciences, Department of History and Philosophy of Science, Pittsburgh.
- Sklar, L. (1981). Do unborn hypotheses have rights? *Pacific Philosophical Quarterly*, 62(1), 17–29.
- Sonnenschein, C., & Soto, A. M. (2011). Response to in defense of the somatic mutation theory of cancer. *BioEssays*, 33(9), 657–659.
- Sonnenschein, C., & Soto, A. M. (2016). Carcinogenesis explained within the context of a theory of organisms. *Progress in Biophysics and Molecular Biology*, 122(1), 70–76.
- Soto, A. M., & Sonnenschein, C. (2011). The tissue organization field theory of cancer: A testable replacement for the somatic mutation theory. *BioEssays*, 33(5), 332–340.
- Stanford, P. K. (2006). *Exceeding our grasp: Science, history, and the problem of unconceived alternatives*. New York: Oxford University Press.
- Stevens, E. (2013). *Life out of sequence. A data-driven history of bioinformatics*. Chicago: The University of Chicago Press.
- Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*, 458(7239), 719–724.
- Talukder, A. K. (2015). Genomics 3.0: Big-data in precision medicine. In N. Kumar & V. Bhatnagar (Eds.), *Big data analytics* (pp. 201–215). Cham: Springer.
- Tannock, I. F., & Hickman, J. A. (2016). Limits to personalized cancer medicine. *The New England Journal of Medicine*, 375(13), 1289–1294.
- Teira, D. (2011). Frequentist vs. Bayesian clinical trials. In F. Gifford (Ed.), *Handbook of the philosophy of science. Volume 16. Philosophy of medicine* (pp. 255–297). Amsterdam: Elsevier.
- Tokheim, C. J., Papadopoulos, N., Kinzler, K. W., Vogelstein, B., & Karchina, R. (2016). Evaluating the evaluation of cancer driver genes. *Proceedings of the National Academy of Sciences*, 113(50), 14330–14335.
- van Fraassen, B. C. (1989). *Laws and symmetry*. Oxford: Oxford University Press.
- Vaux, D. L. (2011a). In defense of the somatic mutation theory of cancer. *BioEssays*, 33(5), 341–343.
- Vaux, D. L. (2011b). Response to the tissue organization field theory of cancer: A testable replacement for the somatic mutation theory. *BioEssays*, 33(9), 660–661.
- Weinberg, R. (2014). Coming full circle—from endless complexity to simplicity and back again. *Cell*, 157(1), 267–271.
- Weyl, H. (1949). *Philosophy of mathematics and natural science*. Princeton: Princeton University Press.
- Williamson, J. (2010). *In defence of objective Bayesianism*. Oxford: Oxford University Press.
- Worrall, J. (2007a). Evidence in medicine and evidence-based medicine. *Philosophy Compass*, 2(6), 981–1022.
- Worrall, J. (2007b). Why there's no cause to randomize. *The British Journal for the Philosophy of Science*, 58(3), 451–488.
- Worrall, J. (2010). Evidence: Philosophy of science meets medicine. *Journal of Evaluation in Clinical Practice*, 16(2), 356–362.

- Zbilut, J. P., & Giuliani, A. (2008). Biological uncertainty. *Theory in Biosciences*, *127*(3), 223–227.
- Zhang, J., Liu, J., Sun, J., Chen, C., Foltz, G., & Lin, B. (2014). Identifying driver mutations from sequencing data of heterogeneous tumors in the era of personalized genome sequencing. *Briefing in Bioinformatics*, *15*(2), 244–255.