What Follows from all that Data? Logic in the Methodology of Data-Intensive and AI-Driven Science

Hykel Hosni[*]
Logic Uncertainty Computation and Information (LUCI) Lab, Department of Philosophy, University of Milan
hykel.hosni@unimi.it

Jürgen Landes[†]
Munich Center for Mathematical Philosophy, LMU Munich
juergen_landes@yahoo.de

In honour of Dov Gabbay's 80th birthday

Abstract

There is no foreseeable future in which science is not about data and the inferences data license. For centuries, logic has been the tool to analyse inference. And yet, logic is vastly underappreciated in the current methodology of data-driven science, as we argue in this paper. We first outline two historical reasons behind this mismatch, then highlight the need to bridge it by examining a widely used form of scientific inference: Null Hypothesis Significance Testing. Finally, we argue that the question: what follows from data? is ripe to be tackled by logicians. We submit that this will help lay a sound methodological foundation for the practice of data-intensive and AI-driven science.

[*]Hykel Hosni acknowledges generous funding from the Italian Minister for University and Research (MUR) in the form of the FIS 2021 *Advanced Grant* "Reasoning with Data" (ReDa) G53C23000510001.

[†]Jürgen Landes reports financial support was provided by German Research Foundation project number 528031869.

1 Introduction

There is no foreseeable future in which science is not Data-Intensive and AI-Driven (DIAD) [34, 40, 55]. While the "transformative" promise of AI techniques is ubiquitous in science, it is particularly detectable in the biomedical sciences, spanning drug discovery [71] and diagnostics [11]. This is, of course, great news for healthcare [6]. Provided, that is, the key questions of transparency and trustworthiness in AI-driven science are addressed quickly and effectively enough [61, 68].

The logical community has responded impressively to the problem of transparency of AI systems, with the field of Formal/Abstract Argumentation Theory emerging as cornerstone of explainable AI [3, 5, 19, 47]. The logicians' involvement in formulating and addressing the challenge of trustworthy AI is relatively more recent [81], but it is delivering very promising results [16]. Those approaches typically combine logical and statistical methods, used to account for the stochastic aspects of data. There is, in short, no doubt that logicians are making an impact in this new AI spring [28].

Those recent developments add to the contributions made over the past four decades to the field of Knowledge Representation and Reasoning subfield of AI. Well-known success stories here include non-monotonic logics, paraconsistent logics, probabilistic and fuzzy logics, logics for multi-agent systems, logics for collaboration – the list is *very* long, see for example [18, 32, 57, 58, 75]. Much of the cutting-edge contribution of logic to AI today is inspired, if not grounded, on those logical systems. Those, in turn, largely developed starting from the 1970s in response to the pressing needs arising in computer science and symbolic AI. The kind of humus was pinned down in the inaugural editorial of the *Journal of Logic and Computation*

> We do not mean 'Logic' as it is now. We mean 'Logic' as it will be, as a result of the interaction with computing.[29]

The ensuing three decades witnessed the co-evolution of logic and computation, paving the way for a *turn towards the practical* [30] which challenged the academic boundaries of mathematics, computer science, philosophy, and law.

And yet, logic is vastly underappreciated in the methodology of data-driven *science*. As we noted in [44], and will argue in greater detail herein, this neglect is detrimental to both logicians and scientists. To be able to serve the methodology of data-intensive and AI-driven science, logic and logicians would benefit from undergoing yet another turn. This time towards the stochastic, the practical problem solving.

Timothy Gowers put forward a distinction between the mathematicians who work on theory and those that solve problems [63]. There is a very long and successful tradition of logicians solving problems in probabilistic logic, which starts with Boole and De Morgan, and includes [37, 38, 62, 64, 65, 80] as very influential frameworks. As we point out in [43] those are proposals which seek to extend classical logic with the ability to express and do inference with probabilities. This has led and keeps leading to deep mathematical results. It is however unlikely that those results will be directly applicable to methodological purposes of data-intensive and AI-driven science for the same sort of reasons classical logic won't do (see below). To do so, borrowing again from Gowers, we need more theory. As we argue below, we need to think rather carefully at the building blocks of data-driven inference, all the way down to what we think valid data-driven inference should (not) be. Another Fields medalist, David Mumford, has rather radical ideas on this score. He urges us to put random variables into the foundations of mathematics, thereby marking a *dawning age of stochasticity* [15]. Among the consequences of doing so, one needs to reject the continuum hypothesis. Theory (re)building need not go that far, of course, but Mumford's point brings to the foreground that our understanding of probability theory has profound consequences for foundational questions of logic. And probability, as Laplace pointed out very clearly [53], arises precisely where data and the lack thereof play an equally important, and intuition-defying, role in scientific inference.

Using Gowers' terms, we thus urge logicians to solve real-world (inspired) problems *and* for modern logical theory to be applicable in science, again.

The rest of this paper is organised as follows. We next identify two potential reasons for the disconnect between scientists and logic (Section 2) and then zoom in briefly on the logical foundations of statistical inference (Section 3). We conclude with some reflections on the challenges that lie ahead (Section 4).

2 How working scientists lost contact with logic

Logic in vastly under-appreciated by working scientists and virtually absent from the methodology of data-driven science [44]. We briefly discuss two potential reasons for this state of affairs. The first is rooted in the enduring influence of Alfred Tarski's stance on the role of logic in the methodology of empirical science. The second originates in the very enthusiastic reception of hypothetico-deductive and falsificationist approaches within the sciences, resulting in a very narrow perspective on logic.

2.1 Tarski's restriction and its influence on logicians

Alfred Tarski, as reported by his student John Corcoran, purportedly referred to himself as "the greatest living sane logician." The qualification was, of course, instrumental to rank himself above Kurt Gödel. Be this as it may, Tarski is certainly one of the great modern logicians. He is also the author of the first popular-science book in logic, which shaped the subject and its image within the wider scientific landscape. Written in Polish in 1936, it appeared in German with the title *Einführung in die mathematische Logik und in die Methodologie der Mathematik* and it was turned into a textbook for the 1941 English translation, titled *Introduction to Logic and to the Methodology of Deductive Sciences*. As explained in the Preface, Tarski contrasts the latter with "the empirical sciences" which do not lend themselves purposefully to logical methods:

> I see little rational justification for combining the discussion on logic and the methodology of empirical sciences in the same college course [77, p. xiii].

This should not be read as Tarski dismissing outright the importance of empirical science. In a comment to a seminar quoted in Chapter 10 of [24] Tarski states:

> It would be more than desirable to have concrete examples of scientific theories (from the realm of the natural sciences) organized into deductive systems. Without such examples there is always the danger that the methodological investigation of these theories will, so to speak, hang in the air. Unfortunately, very few examples are known which would meet the standards of the present-day conception of deductive method and would be ripe for methodological investigations [...] The development of metamathematics, that is, the methodology of mathematics, would hardly have been possible if various branches of mathematics had not previously been organized into deductive systems.

Requiring axiomatisation as a necessary condition for the applicability of logic to scientific methodology, effectively means restricting the applicability of logic to metamathematics. Owing to Tarski's scientific and academic prominence, this view has been very influential in shaping the twentieth-century perception of the subject, in and outside logic. The recent [25] testifies the enduring importance of this perception.

However, as noted by Wilfried Hodges in his 2011 *Division of Logic, Methodology and Philosophy of Science* Presidential Address [41], not very many would go on doing the methodology of the empirical sciences as Tarski suggested. The birth and development of DLMPS owes much to Tarski, as recounted by [24], so his views shaped the way Logic, Methodology and Philosophy of science have been construed there, especially with respect to their lack of interaction. Indeed, inspection of the DLMPS tables of contents over the decades shows very little trace of logic being part of the methodology of "empirical" sciences. The emerging picture is rather one in which DLMPS provides a rich umbrella encompassing both "metamathematics" and the methodology of "empirical" sciences, but quite separately.

2.2 The Popper/Hempel influence on scientists

A second reason for the current underappreciation of logic in science may be rooted in the enthusiastic acceptance, by working scientists, of hypotetico-deductivist disconfirmation as the signature of the scientific method. In this picture, which has been incredibly influential in the construction of the self-image of generations of scientists [74], logic contributes only with *modus tollens* (Popper, 2005,p. 89). Interestingly, Ken Aizawa argues that Hempel may have exerted an even stronger influence than Popper in elevating *modus tollens* to the only logical snippet deemed relevant to experimental science [2]. This picture has percolated magnificently in the mindset of the methodologically conscious scientist: scientific hypotheses can only be *disconfirmed*. For confirmation relies, as Hempel notes, on the deductive fallacy of *affirming the consequent* [39]. Yet, as argued by Pólya, "affirming the consequent" is what scientists do when taking stock of (a positive) correlation [70]. To mark its importance in data-driven reasoning, Pólya rebrands the long-vituperated fallacy as the *fundamental pattern of induction*, on which we will come back soon. Some scientists seem to have taken notice, notably [8, 48], but they appear to be rare exceptions.

We have a logical blind spot: many working scientists are unaware of the rich variety of logical nuances that would provide a much more adequate methodological pillar in data-driven inquiry compared to classical validity. Aizawa suggests [1] that fixing this blind spot requires replacing the HD view of disconfirmation with a framework centred on *abductive reasoning*. This is good news, because practical logicians have contributed to this area consistently for decades, see [31, 56] for overviews. However, there are some usually neglected challenges, identified in Section 4 below, that must be addressed.

Summing up, the restriction of logic to the methodology of deductive sciences on the one hand, and the extraordinary success of the Popper/Hempel view impacted negatively on the scope of logic and logical methods outside mathematics. As a consequence, the logical questions on data-driven inference were left to statisticians.

3 *Inference* in statistical inference

Around the same time Hempel was advocating for the primacy of classical logic, Richard Feynman was crystallizing the role of modus tollens in the methodology of science: "It doesn't make any difference how beautiful your guess is [..]. If it disagrees with experiment, it's wrong" [26]. We take no issue, of course, with the message delivered by the Feynman dictum. However, it may not be clear what it means for data to (dis)agree with a hypothesis, as pointed out long ago by Duhem and Quine. In fact, with the exception of very simple cases, it is *almost never* clear, as the one century-old debates on statistical significance clearly testify [7, 54, 59, 67, 82].

3.1 Case study: Null Hypothesis Significance Testing

There is a commonplace view to the effect that statistics *is* "the logic of science" [48, 49]. In an overview of R.A. Fisher's contributions to statistics, Efron says:

> Fisher believed that there must exist a logic of inductive inference that would yield a correct answer to any statistical problem, in the same way that ordinary logic solves deductive problems [21].

While this is a very logical way of putting it, likely borrowed from Boole's formulation of his *general method* [9], identifying criteria of validity for data-driven inference is not what Fisher (or any other statistician, for that matter) went on to do. Rather than being loosely inspired by logic, the logic of data-driven science should be logic; even better, a logical framework for the sciences.

To see why, take the procedure known as *Null Hypothesis Significance Testing* (NHST). Statisticians are often appalled by this acronym, because it juxtaposes the conflicting ideas of Fisher on one hand, and the Neyman–Pearson duo on the other. But given that NHST is what most working scientists do, it suits our present purposes very well.

The key idea behind NHST is quite entrenched in scientific thinking: *to assess a hypothesis, look at its consequences.* Specifically, we want to draw a conclusion about a statistical hypothesis of interest on the basis of the data it would generate, if true. So let H_0 be the statistical hypothesis of interest, referred to (after Fisher) as the *null hypothesis*, and denote the observed (low probability) data by D.

Specific instances of NHST are variations on the following pattern:

(NHST1) Data D are observed.

(NHST2) If H_0 were true, then data at least as improbable (according to some test statistic) as D would be very improbable.

(NHST3) Either we accept that very improbable events have occurred or we reject H_0.

(NHST4) Very improbable events do not occur.

(NHST5) Therefore, we reject H_0.

Since Fisher's [27] it has become customary to translate "very improbable", with $p < 0.05$. This threshold stuck, so 0.05 is indeed the most used *p-value*. Informally, this is the calculated (hypothetical) conditional probability of observing data as improbable as, or more improbable than D, given that H_0 is true. Hence, the purpose of NHST so construed is to let the data speak about H_0: a small-enough p-value being licence to *reject* H_0.

Given the observations in Section 2, it is unsurprising that concluding (NHST5) from the conjunction of the premisses (NHST1-NHST4) is often justified by invoking *modus tollens*. Here are a few notable examples: [17] speaks of "inductive modus tollens", [59] refers to "statistical modus tollens", whereas [72] tellingly notes that the analogy with *modus tollens* lends tests of significance their prominence in the "scientific method":

> [modus tollens] is at the heart of the philosophy of science, according to Popper. Its statistical manifestation is in [the] formulation of hypothesis testing that we will call 'rejection trials'. ([72], p.72)

Quite interestingly, this view is shared by prominent critics of NHST as well, see e.g. [76], more on this below.

The NHST pattern of inference can be mapped to something that looks like *modus tollens* by making the following assumptions:

(A1) Boolean implication is the adequate formalisation of the conditional statement featuring in the p-value.

(A2) The clause that the observed data given H_0 should be at least as improbable as D (as measured by some test statistics) can be omitted with no loss of generality;

(A3) Small (but non-zero) probability events are false.

With (A1-A3) in place we get the following instance of *modus tollens*:

(NHST'1) D

(NHST'2) $H_0 \to \neg D$

(NHST'3) Therefore, $\neg H_0$

Before going into some detail as to why (A1-A3) are rather unpalatable, note that (A1) and (A3) are compatible with the so-called *Fisher Disjunction*:

> The force with which such a conclusion is supported is logically that of the simple disjunction: *Either* an exceptionally rare event has occurred, *or* the theory of random distribution is not true [27].

The untenability of (A2) is easily ascertained by probabilistic means, and hence well-appreciated in the statistical literature. Its implausibility is in fact the reason why the awkward "as improbable or more improbable" bit cannot be omitted from a correct conceptualisation of the p-value [46]. So let's focus on the other two assumptions.

The problem with (A1) is more subtle, and quite hard to spot unless one looks at it from a probability logic angle. Recall that the probability of a Boolean implication does not equal, in general, the "associated" conditional probability, i.e:

$$P(D \mid H) \neq P(H \to D),$$

where P is a probability function on the propositional language which includes H and D, and which is closed under the usual Boolean connectives. This obvious fact may easily go unnoticed to the working scientist as a consequence of the ambiguity of natural language. More precisely, it may be hard to see the difference between

- The probability of the data given the null hypothesis.
- The probability that the null hypothesis implies the data.

It is no mystery that implication is vastly misunderstood outside its mathematical usage.

(A3) amounts to a metaphysical assumption on the nature of probability. Sometimes it is referred to as the *Cournot Principle* after [14], and it lies at the heart of the frequentist view of probability underlying NHST [73]. Loosely put, the Cournot principle says that the (physical) meaning of probability is that very small-probability events do not happen. This relates to a more general question: *how small should* *a small probability be before it is rationally neglected*? Debated at least since Jacob Bernoulli's *Ars Conjectandi*, it isn't likely to be decided any time soon. So rather than untenable, the adequacy of (A3) seems not to be tackled meaningfully from the logical point of view. There is some irony in this, of course, in light of the rhetoric that views NHST as purely data-driven, and hence objective. An irony that of course has not escaped the attention of critics. See [36, Chapter 14] for a concise, witty overview by I.J. Good.

3.2 On the analogy between statistical inference and classical deduction

In light of the above discussion, it is curious that the statistical camp seems to be generally happy with justifying NHST through *modus tollens*. Indeed, some have strongly dissented, pointing to the failure of *modus tollens* as a strong line of criticism of p-value based significance [23, 50, 76]. Consider the following (now) standard example made popular by [69]:

(P1) Harold is a member of Congress;

(P2) If Harold is a US citizen, than he is most probably not a member of Congress;

(P3) Therefore, Harold is likely not a US citizen.

While premisses (P1-P2) may be easy to accept, the conclusion (P3) is not, since being a US citizen is a necessary condition to sit in the US Congress. Hence, the example shows a context in which (NHST'1-NHST'2) above may be satisfied, but not (NHST'3).

The 'Harold example' certainly suggests that there is something to be careful about in NHST inference, but it still concedes a lot to assumptions (A1) and (A3) above. For it shows the *classical* invalidity of NHST inference. However, similarly to what was noted above regarding (A1-A3), classical validity is hardly adequate to capture the "most probably" and the "likely" that appear in (P2) and (P3). As argued in detail in [4] turning to *non-monotonic consequence relations* pays nice dividends. In addition to expressing adequately the qualitative uncertainty conveyed by (P2) and (P3), suitably defined non-monotonic consequence relations are capable of vindicating some of the intuition linking conceptually NHST to *modus tollens*. Indeed, from the non-monotonic-logic point of view, the natural conclusion of (P1-P2) above is that *Harold is not a typical US citizen*.

The discussion above leads to the more abstract and general question as to whether data-driven patterns of reasoning can at all "borrow their validity", so to speak, from analogous classical patterns. Let's call this the Analogy Principle. We have seen that *modus tollens* is classically valid, but fails in contexts of interest to significance-based inference. But the Analogy Principle is unreliable also in the other direction. To see this, consider Pólya's *fundamental inductive pattern* mentioned briefly above:

$$\frac{\begin{array}{r} A \to B \\ B \end{array}}{A \text{ [becomes] more plausible.}} \tag{FIP}$$

As a scheme of inference, (FIP) has no direct counterpart in classical logic, but if one reasons by analogy and takes "A [becomes] more plausible" as "A [is] true", then (FIP) is precisely the deductive fallacy of "affirming the consequent" despised by Hempel.

So the Analogy Principle can mislead in both directions: data-driven inference may not get methodological support from analogous classical patterns of inference, and conversely, classically invalid patterns of inference may be crucial to their analogues for data-driven reasoning. No wonder then that one sees rather spectacular failures of the Analogy Principle in connection to the (mis)applications of NHST [42, 50, 66].

Taking stock from the failure of the Analogy Principle we may conclude that validity does not form a continuum spanning mathematical proof and data-driven inference. Indeed, it is well known that *modus tollens* need not be probabilistically valid [10], and in general fails to deliver a point-valued probability [79]. Hence, it is not advisable to trust a statistical pattern of inference because it can be articulated as the stochastic analogue of a classically valid one. Interestingly, this view resonates with Tarski's insistence that logic's proper role resides exclusively in the methodology of deductive sciences. However, one century on, both the scientific and the logical landscapes have changed dramatically. The coming of age of data-intensive and AI-driven science, with its ability to immediately impact society and ultimately our daily lives, compels us to devise a fail-safe "general method" for valid data-driven inference. The time is ripe for logicians to venture where Tarski saw no role for logic.

4 Challenges

The coming of age of data-intensive and AI-driven science raises questions revolving around a key one: *what follows from stochastic data*? Addressing it involves taking a step back, and asking *logical* questions about statistical inference. The previous Section demonstrates that we must be cautious before we leap. We conclude this paper by offering some reflections on two rather distinct challenges inherent in making this leap.

4.1 Choosing a good trade-off between context-dependent and formal nature of validity

Suppose we agreed to play dice, no money involved. The first who rolls three "ones" wins. If you win on the first try, we'd be witnessing a 1/216 probability event. Would this be evidence that the dice are loaded? NHST would say so, with rather strong "significance". But an event of this improbable kind is certainly not impossible. Second round, and you score again. The probability of you winning twice in a row now gets quite small at 1/46656. This is roughly four times smaller than the chance of a (natural) triplet birth. Again not impossible, but uncommon enough to justify jumping to the conclusion that maybe the dice *are* loaded after all. Doing so is reasoning along the lines of NHST. And so do particle physicists when claiming discoveries [20]. Famously, the Higgs Boson was announced with a "5 sigma" significance, which roughly corresponds to a p-value of around 1 in 3.5 million. Those are just two of very many circumstances in which significance inference delivers unquestionable results. But apparently minor variations on the NHST pattern of inference soon lead to very questionable outcomes [13, 45]. So, what appears to be the same pattern of inference seems to be sometimes valid and sometimes not. This is a clear call for logical scrutiny.

Defining the meaning of *formality* may not be straightforward. However, one rather uncontroversial view holds formality to be pinned down by closure under uniform substitution, sometimes also referred to as *structurality*. While it is a characteristic property of classical logic, closure under substitution fails for many logics for practical reasoning, notably non-monotonic logics [57, P. 14] and paraconsistent logics [32, P. 76]. However, since classical logic is manifestly unfit for purpose, a trade-off between formality and concreteness must be identified.

It is well-known from the psychology of reasoning that, while we may struggle with basic inferential patterns in decontextualized tests, we do very well in dialogical, argumentative contexts. The key difference being that in the latter we are genuinely trying to evaluate or convey practically directed information [60]. So, if we are interested in providing logical criteria for the validity of data-driven inference, we must strike a balance between:

- offering criteria that are applicable beyond the single, specific, case;

- ensuring that some context-specific aspects of the data-driven inference we are interested in must be represented.

It turns out that the terms of this trade-off have a lot to do with opposing statistical methodologies. Discounting some inevitable and very human attachment to one's own ideology, it is quite easy to see scientific communities naturally clustering around certain statistical methodologies. The kind of data that is typically available in a given field plays a big role in matching probabilistic ideology and research contexts. This may account for the fact that, say geneticists rarely identify as "Bayesians", whereas the strongest opposition to the NHST hard-core frequentism comes from the social sciences [82]. So it is the nature of the data-generating mechanisms that provides the context for data-driven reasoning. And this is what probably marks the difference between apparently similar, but logically distinct, patterns of data-driven inference.

Creating families of logical systems capable of adapting to the context-depending demands of practical reasoning is the core methodology of Labelled Deductive Systems (LDS). While we are not aware of any LDS framework motivated by the problem of accounting for data-driven inference, a recent retrospective [33] demonstrates the applicability of LDS to the analysis of legal evidence, which is a particularly regimented case of data-driven inference. Similarly, the logical framework for data-driven reasoning recently put forward in [4] seems to be promising by striking a reasonable balance between formality and context-dependence. The formality of the concept of validity is provided by a *blueprint consequence relation*, whose intended semantics is built on the idea that data can reject, to some degree, and possibly by mistake, any well-formed (statistical) hypothesis. Then, different data-generating contexts give rise to distinct consequence relations, fine-tuning the "blueprint".

4.2 Communicating across academic disciplines and societal sectors

Let us close with a challenge which is hardly ever on the logician's agenda, but which is crucial nonetheless: finding an effective way to communicate the criteria for valid data-driven reasoning across disciplines and, ideally, across sectors.

First, as noted in Section 2, working scientists have very limited familiarity with the subject. This owes to the fact that most scientists acquire only indirect training in logic, chiefly through introductory chapters in mathematics textbooks. Those typically cover rudimentary classical logic, often expressed in the form of naive set theory. This results in scientists typically identifying logic with an informal version of classical logic, very much in line with the Popper/Hempel view recalled above. Of course, we can't expect, say, a translational biologist to get acquainted with the formal detail of logics for data-driven reasoning. So, the challenge arises to make the results which will (hopefully) spring by addressing the logical issues raised above *usable* for the translational biologist. It is no small challenge, of course, but we can look at statistical software for inspiration. The computational turn in statistics made data-intensive, and then AI-driven inference, unprecedentedly accessible [22]. Modelling the construction of data-driven scientific knowledge with abstract argumentation theory stands out as a promising route to achieving that [52].

Second, the curious scientist, policy-maker or ideal (in our view!) citizen who occasionally enjoys the popular science book, will only find logic-as-metamathematics titles on the shelves. Communicating the golden age of mathematical logic with the extraordinary contributions of Cantor, Russell, Hilbert, Gödel, Tarski, and Turing, certainly makes for good storytelling which is moreover naturally in tune with what most readers expect. Again, we take no issue with the narrative recounting how logicians ascended the highest peaks of human intellect. Yet, we badly need logical methods to enter inter-sectoral dissemination. Scientists, policy-makers, and curious citizens must have a chance to appreciate the role of logic in shaping the methodology of data-intensive and AI-driven science. Hence, the popular science bookshelves must hold logic books that extend well beyond the foundations of mathematics. Eugenia Cheng's [12] and Adam Kucharski's [51] are recent and successful books that go in this direction. Many more are needed.

The history of the subject is filled with logical types thriving on practical problems. I.J. Good's [35] reports on previously classified material of the work done by Alan Turing at Bletchley Park. Among other things, Good describes how his line manager worked out from scratch the (odds form of the) Bayes rule while trying to crack Enigma. A few years later, on 20 February 1947, Turing delivered a lecture at the London Mathematical Society, reporting on the construction of the Automated Computing Engine (ACE), where he argues extensively for giving priority to large enough and accessible enough data storage:

> I have spent a considerable time in this lecture on this question of memory, because I believe that the provision of proper storage is the key to the problem of the digital computer, and certainly if they are to be persuaded to show any sort of genuine intelligence much larger capacities than are yet available must be provided. [78]

Three quarters of a century later, the historical events have been turned into an award-winning movie (The Imitation Game) and the digital computer with proper storage (measured in terabytes) has become ubiquitous. We contend that after solving the challenges of building a digital computer and storing large amounts of data the next challenge is to *determine what follows from all that data.*

5 Dedication

It is an honour and a pleasure to offer this note in celebration of Dov Gabbay's 80th birthday. Dov's remarkable *approach* to logic has been a great inspiration to us since our days as graduate students. It is an inspiration that continues to this day.

References

[1] K. Aizawa. *Compositional Abduction and Scientific Interpretation : A Granular Approach.* Cambridge University Press, Cambridge, 2025.

[2] K. Aizawa. Disconfirmation is not Modus Tollens. *The Reasoner*, 19(2):26–32, 2025.

[3] K. Atkinson, P. Baroni, M. Giacomin, A. Hunter, H. Prakken, C. Reed, G. Simari, M. Thimm, and S. Villata. Toward Artificial Argumentation. *AI Magazine*, 38(3):25–36, 2017.

[4] P. Baldi, E. A. Corsi, and H. Hosni. A logical framework for data-driven reasoning. *Logic Journal of the IGLP*, 33, 2025.

[5] P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors. *Handbook of Formal Argumentation.* College Publications, Milton Keynes, 2018.

[6] K. Batko and A. Ślęzak. The use of Big Data Analytics in healthcare. *Journal of Big Data*, 9(1), 2022.

[7] Y. Benjamini, R. D. De Veaux, B. Efron, S. Evans, M. Glickman, B. I. Graubard, X. L. Meng, N. Reid, S. M. Stigler, S. B. Vardeman, C. K. Wikle, T. Wright, L. J. Young, and K. Kafadar. The ASA president's task force statement on statistical significance and replicability. *Annals of Applied Statistics*, 15(3):1084–1085, 2021.

[8] F. L. Bookstein. *Measuring and Reasoning.* Cambridge University Press, Cambridge, 2014.

[9] G. Boole. *An investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities.* Walton & Maberly, London, 1854.

[10] R. Bosley. Modus Tollens. *Notre Dame Journal of Formal Logic*, (1):103–111, 1979.

[11] T. Chanda et al. Dermatologist-like explainable AI enhances melanoma diagnosis accuracy: eye-tracking study. *Nature Communications*, 16(1), 2025.

[12] E. Cheng. The Art of Logic How to Make Sense in a World that Doesn't, 2018.

[13] A. Clayton. *Bernoulli's Fallacy. Statistical Illogic and the Crisis of Modern Science.* Columbia University Press, New York, 2021.

[14] A.-A. Cournot. *Exposition de la théorie des chances et des probabilités.* Hachette, Paris, 1843.

[15] D. Mumford. The Dawning of the Age of Stochasticity. In V. Arnold, M. Atiyha, P. Lax, and P. Mazur, editors, *Mathematics: Frontiers and Perspectives*, pages 197–218. International Mathematical Union, USA, 2000.

[16] F. A. D'Asaro, F. A. Genco, and G. Primiero. Checking trustworthiness of probabilistic computations in a typed natural deduction system. *Journal of Logic and Computation*, 2025.

[17] M. Dickson and D. Baird. Significance Testing. In *Handbook of the Philosophy of Science, vol 7: Philosophy of Statistics*, volume 7, pages 199–229. Elsevier, 2011.

[18] D. Dubois and H. Prade. The three semantics of fuzzy sets. *Fuzzy Sets and Systems*, 90(2):141–150, 1997.

[19] P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.

[20] D. A. Van Dyk. The role of statistics in the discovery of a higgs boson. *Annual Review of Statistics and Its Application*, 1(October 2013):41–59, 2014.

[21] B. Efron. R. A. Fisher in the 21st century: Invited paper presented at the 1996 R. A. Fisher lecture. *Statistical Science*, 13(2):95–122, 1998.

[22] B. Efron and T. J. Hastie. *Computer Age Statistical Inference.* Cambridge University Press, Cambridge, 2016.

[23] R. Falk and C. W. Greenbaum. Significance Tests Die Hard: The Amazing Persistence of a Probabilistic Misconception. *Theory & Psychology*, 5(1):75–98, 1995.

[24] A. Burdman Feferman and S. Feferman. *Alfred Tarski: Life and Logic.* Cambridge University Press, Cambridge, 2004.

[25] F. Ferrari and M. Carrara. *Logic and Science.* Cambridge University Press, Cambridge, 2025.

[26] R. Feynman. *The Character of Physical Law.* MIT Press, Cambridge, 1965.

[27] R. A. Fisher. *Statistical Methods and Statistical Inference.* Oliver & Boyd, Edinburgh, 1956.

[28] T. Flaminio and H. Hosni. Logics for the new AI spring. *International Journal of Approximate Reasoning*, 170, 2024.

[29] D. M. Gabbay. Editorial. *Journal of Logic and Computation*, 1(1):1–4, 1990.

[30] D. M. Gabbay and J. Woods. The practical turn in logic. In D. M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, 2nd Edition, Vol 13*, volume 13, pages 15–122. Springer, 2005.

[31] D. M. Gabbay and J. Woods. *The reach of Abduction: Insight and Trial.* Elsevier, Amsterdam, 2005.

[32] D. M. Gabbay and J. Woods, editors. *Handbook of the History of Logic Vol 8. The Many Valued and Non Monotonic Turn in Logic.* North-Holland, Amsterdam, 2007.

[33] D. M. Gabbay and J. Woods. The Law of Evidence and Labelled Deduction: Ten Years Later. In G. Casini, L. Robaldo, L. van der Torre, and S. Villata, editors, *Handbook of Legal AI*, pages 364–434. College Publications, London, 2022.

[34] A. Gandomi and M. Haider. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2):137–144, 2015.

[35] I. J. Good. A. M. Turing's Statistical Work in World War II. *Biometrika*, 66(2):393–396, 1979.

[36] I. J. Good. *Good thinking: The foundations of probability and its applications.* The University of Minnesota Press, 1983.

[37] R. Haenni, J.-W. Romeijn, J. Williamson, and G. Wheeler. *Probabilistic Logics and Probabilistic Networks.* Springer, Dordrecht, 2011.

[38] T. Hailperin. *Sentential Probability Logic. Origins, Development, Current Status, and Technical Applications.* Lehigh University Press, Bethlehem, 1996.

[39] C. Hempel. *Philosophy of Natural Science.* Prentice Hall, Upper Saddle River, 1966.

[40] T. Hey, K. Butler, S. Jackson, and J. Thiyagalingam. Machine learning and big scientific data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378(2166), 2020.

[41] W. Hodges. DLMPS – Tarski's vision and ours. In B. Loewe and D. Sarikaya, editors, *60 Jahre DVMLG*, pages 103–119. 2022.

[42] S. G. Hofmann. Fisher's fallacy and NHST's flawed logic. *The American psychologist*, 57(1):69–70, 2002.

[43] H. Hosni and J. Landes. Logical perspectives on the foundations of probability. *Open Mathematics*, 21(1), 2023.

[44] H. Hosni and J. Landes, editors. *Perspectives on Logics for Data-Driven Reasoning.* Springer, Cham, 2025.

[45] C. Howson. *Hume's Problem: Induction and the Justification of Belief.* Oxford University Press, Oxford, 2003.

[46] R. Hubbard and M. J. Bayarri. Confusion Over Measures of Evidence (p's) Versus Errors (α's) in Classical Statistical Testing. *American Statistician*, 57(3):171–178, 2003.

[47] A. Hunter, S. Polberg, N. Potyka, T. Rienstra, and M. Thimm. Probabilistic Argumentation: A Survey. In D. Gabbay, M. Giacomin, G. R. Simari, and M. Thimm, editors, *Handbook of Formal Argumentation*, volume 2, chapter 7, pages 397–444. College Publications, London, 2021.

[48] E. T. Jaynes. *Probability Theory: The Logic of Science.* Cambridge University Press, Cambridge, 2003.

[49] H. Jeffreys. *Scientific inference.* Cambridge University Press, Cambridge, 1931.

[50] J. Krueger. Null hypothesis significance testing: On the survival of a flawed method. *American Psychologist*, 56(1):16–26, 2001.

[51] A. Kucharski. *Proof: The Uncertain Science of Certainty.* Profile Books, London, 2025.

[52] J. Landes, E. A. Corsi, and P. Baldi. Knowledge Representation, Scientific Argumentation and Non-monotonic Logic. In H. Hosni and J. Landes, editors, *Perspectives on Logics for Data-Driven Reasoning*, pages 155–179. Springer, 2025.

[53] P.-S. Laplace. *A Philosophical Essay on Probabilities.* Dover Publications, 1825.

[54] B. Lecoutre and J. Poitevineau. *The Significance Test Controversy.* Springer, 2nd edition, 2022.

[55] S. Leonelli and N. Tempini, editors. *Data Journeys in the Sciences.* Springer, Cham, 2020.

[56] L. Magnani, editor. *Handbook of Abductive Cognition.* Springer, Cham, 2023.

[57] D. Makinson. *Bridges from classical to nonmonotonic logic.* College Publications, London, 2005.

[58] P. Marquis, O. Papini, and H. Prade. *A Guided Tour of Artificial Intelligence Research 1: Knowledge Representation, Reasoning and Learning.* Springer, Cham, 2020.

[59] D. G. Mayo. *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars.* Cambridge University Press, 2018.

[60] H. Mercier and D. Sperber. Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(02):57–74, 2011.

[61] L. Messeri and M. J. Crockett. Artificial intelligence and illusions of understanding in scientific research. *Nature*, 627(8002):49–58, 2024.

[62] Z. Ognjanovic, M. Raskovic, and Z. Markovic. *Probability Logics - Probability based formalization of uncertainty reasoning.* Springer, 2016.

[63] P. T. Gowers. The Two Cultures of Mathematics. In V. Arnold, M. Atiyha, P. Lax, and P. Mazur, editors, *Mathematics: Frontiers and Perspectives*, pages 65–78. International Mathematical Union, USA, 2000.

[64] J. B. Paris. *The uncertain reasoner's companion: A mathematical perspective.* Cambridge University Press, Cambrdige, 1994.

[65] J. B. Paris and A. Vencovská. *Pure Inductive Logic.* Cambridge University Press, Cambridge, 2015.

[66] A. G. Patriota. Is NHST logically flawed? Commentary on: "NHST is still logically flawed". *Scientometrics*, 116(3):2189–2191, 2018.

[67] Y. Pawitan. Defending the P-value. *arXiv, 2009.02099*, 2022.

[68] J. Pearl. Radical empiricism and machine learning research. *Journal of Causal Inference*, 9(1):78–81, 2021.

[69] P. Pollard and J. T. Richardson. On the Probability of Making Type I Errors. *Psychological Bulletin*, 102(1):159–163, 1987.

[70] G. Pólya. *Mathematics and Plausible Reasoning: Patterns of plausible inference.* Princeton University Press, Princeton, 1954.

[71] S. Rodriguez, C. Hug, P. Todorov, N. Moret, S. A. Boswell, K. Evans, G. Zhou, N. T. Johnson, B. T. Hyman, P. K. Sorger, M. W. Albers, and A. Sokolov. Machine learning identifies candidates for drug repurposing in Alzheimer's disease. *Nature Communications*, 12(1), 2021.

[72] R. Royall. *Statistical Evidence: A likelihood paradigm.* Chapman and Hall, London, 1997.

[73] G. Shafer. Testing by betting: A strategy for statistical and scientific communication. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 184(2):407–431, 2021.

[74] M. Singham. *The Great Paradox of Science*. Oxford University Press, Oxford, 2020.

[75] L. Sonenberg. Logics and collaboration. *Logic Journal of the IGPL*, 31:1024–1046, 2023.

[76] D. Szucs and J. P. A. Ioannidis. When null hypothesis significance testing is unsuitable for research: A reassessment. *Frontiers in Human Neuroscience*, 11, 2017.

[77] A. Tarski. *Introduction to Logic and the Methodology of Deductive Sciences (1941)*. Oxford University Press, Oxford, fourth edition, 1994.

[78] A. M. Turing. Lecture to the London Mathematical Society on 20 February 1947. In *Mechanical Intelligence, vol 1*, volume 1. North Holland, 1992.

[79] C. G. Wagner. Modus Tollens Probabilized. *British Journal for Philosophy of Science*, 55:747–753, 2004.

[80] J. Williamson. *In Defence of Objective Bayesianism*. Oxford University Press, Oxford, 2010.

[81] J. M. Wing. Trustworthy AI. *Communications of the ACM*, 64(10):64–71, 2021.

[82] S. Ziliak and D. McCloskey. *The Cult of Statistical Significance*. University of Michigan Press, 2008.

Received August 2025