Linköping Electronic Articles in
Computer and Information Science
Vol. ?(2002): nr ?

Leibniz and Boole on logic and probability

James Cussens

Department of Computer Science
University of York
York UK

Linköping University Electronic Press
Linköping, Sweden

http://www.ep.liu.se/ea/cis/2002/???/

Published on ??, 2002 by
Linköping University Electronic Press
581 83 Linköping, Sweden

Linköping Electronic Articles in
Computer and Information Science
ISSN 1401-9841
Series editor: Erik Sandewall

Typeset by the author using LaTeX
Formatted using étendu *style*

Recommended citation:
<Author>. <Title>. Linköping Electronic Articles in Computer and Information Science, Vol. ?(2002): nr ?. http://www.ep.liu.se/ea/cis/2002/???/. *??, 2002.*

This URL will also contain a link to the author's home page.

Abstract

Combined logical-probabilistic frameworks are very old; often the result of grand attempts to formulate a calculus of general reasoning. A careful account of the assumptions and achievements of these systems can help us address issues in contemporary logical-probabilistic frameworks. This paper aims to provide such an account drawing on the pioneering work of Leibniz and Boole. In Leibniz, we find the first account of epistemic probability in terms of possible worlds. In Boole, we see the extent to which a combined logical-probabilistic calculus is possible using a propositional representation.

1 Probability and Logic

In view of the alleged dichotomy between probability and logic, it is ironic that, as Hacking [9] notes, the word 'probability' was first introduced to denote something measurable in a logic book: the Port Royal *Logic* [1] of 1662, "the most successful logic book of the time", which "cast the mould for generations of future treatises" [9, p. 73].

The central argument of this paper is that probability and logic have been bedfellows from the birth of probability onwards; and that an acquaintance with the history of probability—a history which has led to our current conception of probability—helps greatly in addressing conceptual issues concerning their relationship.

Encouraging study of the history of probability to gain a better understanding of probability seems uncontroversial, but, in fact, it is not at all obvious that this historical knowledge is much help for current research in AI. Reading early work on probability one is struck by the authors' strong concern for issues (often theological ones) which just do not concern current research, either in AI or elsewhere. Here are a couple of examples which give a flavour of this.

Hacking [9] notes that the section of the Port Royal *Logic* dealing with probability contains an attack on the 'probabilism' of the Jesuits. This 'probabilism' asserts that when authorities conflict on some theological issue, one may follow some probable opinion. But here 'probable' is being used in the old sense of "being supported by testimony or writ of authority" [9]. So in practice, according to probabilism's critics, the probabilists would firstly decide on an expedient course of action and then "they would find some old text that could be interpreted as approval for that course of action" [9]. This approach was thus in stark contrast to the careful consideration of evidence advocated by the authors of the Port Royal *Logic*. Another example, from a later period, is Hume's famous argument of 1748 that if something was miraculous enough to be called a miracle, then whatever the evidence for it (for example, authoritative testimony) it must still remain improbable.

Such examples usefully demonstrate the intellectual environment into which early work on probability emerged, but few workers on the foundations of probability would now want to expend much effort attacking defunct ideas such as the Jesuits' probabilism, or thinking about miracles. However, there *are* elements of early work on probability which relate to current concerns more directly and the rest of this article will focus on such elements in the work of Leibniz and Boole.

Leibniz (1646–1712) and Boole (1815–64) belong to two quite separate eras in the history of probability. Leibniz was active just after the birth of probability which occurred around 1660. His role in the development of probability is summed up by Hacking

> He made no serious formal contribution to probability theory, but he had a lasting and profound interest in the subject.[1] Indeed he was the first philosopher of probability. He was the first to insist that probability theory can serve in a branch of logic comparable to the theory of deduction. [9, p.57]

[1] JC: It is surprising that a brilliant mathematician (co-founder of the calculus no less!) with such an interest in probability should make no real mathematical contribution. It would be interesting to find an explanation of this fact.

However, for those interested in the connections between logic and probability, Leibniz is of more than just historical interest; as we shall see in Section 2 many of his innovations remain in use today.

Boole, discussed in Section 3 played an entirely different role. He was a mathematician, not a philosopher (although he had a profound influence on the philosopher Peirce). As he himself states:

> ...we are, for all really scientific ends, unconcerned with the truth or falsehood of any metaphysical speculation whatever [4, p.41]

Boole appeared on the scene after arguments on the nature of probability had taken very much second place to the development of the mathematical theory of probability. But, although he provides something of a counterpoint to Leibniz, in common with Leibniz he had useful things to say about the connections between logic and probability; the echoes of which we see in some contemporary work.

2 Leibniz

2.1 Leibniz's theory of probability[2]

> I have more than once said that we should have a new kind of logic which would treat of degrees of probability - Leibniz. Quoted in [11]

Leibniz is important because he was the first to construct a theory of epistemic probability with 'possible worlds' semantics; a common approach in contemporary AI research [10, 3]. Unlike many early workers on probability theory he did not begin by considering games of chance, or any other sort of repeatable randomised experiment—but started directly by considering probability as a measure of degree of belief. This is partly due to his legal background. The probability that, in some particular case, the defendant is guilty simply can not be understood in terms of some long-run frequency. This led Leibniz to consider probability in terms of degrees of proof rather than long-run frequencies. The other reason for Leibniz's focus on epistemic probabilities is simply that he was initially ignorant of other probabilistic work, only becoming acquainted with the work of 'Pascal, Huygens and others' during his long stay in Paris during 1672-6.

Leibniz's possible worlds semantics for probability was inextricably connected to his larger metaphysical concerns. For "Leibniz every possible world has some tendency to exist, and our world is the one with the greatest propensity" [9, p.137]. In this sense our world is the best of all possible worlds, a view Voltaire satirised in *Candide*, by putting it into the mouth of the Leibniz caricature Dr Pangloss.

The tricky business of constructing this set of possible worlds is, predictably, given over to God. However, the propensity for a world to exist is due to its own internal degree of 'possibility'—"The creatability of the things will correspond to the degree of possibility in the divine mind" [9, p. 138]. As Leibniz puts it "probability is degree of possibility" [9, p. 127]. The probability of a possible world determines the more humdrum probabilities such as those given by random experiments. As Hacking puts it:

[2]Throughout this section I am drawing on Hacking's account [9] of Leibniz's philosophy of probability

> Just as the possible worlds in the mind of God vie with one another for creation, so all the possibilities that we can distinguish will also have some propensity to be actual. [9, p. 139]

Leibniz used the word 'possible' to mean internally consistent, but also used it to mean 'having a propensity to occur'—and he did not distinguish the two. This conflation is embedded in his metaphysics, since Leibniz invites us:

> ... to contemplate consistent notions having more than mere internal consistency; they have a positive drive to come into being. The more they have of this, the more possible they are. 'The possible demands existence by its very nature, in proportion to its possibility, that is to say, its degree of essence.' [9, p. 138]

Hacking argues, rather convincingly, that Leibniz's definition of probability in terms of possibility (and that of later writers) is a "fairly essential feature of the developments of both concepts" [9, p.123]. This was because the objective/subjective duality of probability was already present in the older term 'possibility'.

There is now the issue of how *we* can discover the set of possible worlds and their attached probabilities. This can be achieved using Leibniz's grand scheme for a perfect logical language. Leibniz was obsessed with the idea of finding a basic set of ideas from which all other ideas could be mechanically generated by combination. In particular, the descriptions of all possible worlds could be so generated:

> The art of combinations would enable us to compute all descriptions of possible worlds that could be expressed with that stock of ideas. And the possible worlds so described would all have some propensity to exist. [9]

If we have this perfect logical language, then the correct prior is a uniform distribution over those possible worlds describable by it. Leibniz believes such perfection is only asymptotically achievable, but that our language can improve and more closely come to mirror the structure of the universe. At any point, we should still apply a uniform probability over the worlds describable by our imperfect language, but must realise that such a prior will be discarded as our language improves.

Leibniz's concentration on epistemic probability, his metaphysical belief in an objectively (or rather God-determined) set of possible worlds with attached probabilities, and his focus on language all contribute to the logical nature of this theory of probability. It is also worth nothing that, like later proponents of a logical theory of probability, for Leibniz all probabilities are relative, i.e. conditional on the available evidence. This reflects Leibniz's legal background. However, on the crucial issue of how to construct a set of possible worlds and their attached probabilities, Leibniz differs from later workers on logical probability, as the next section shows.

2.2 Logical probability after Leibniz

What happened to Leibniz's "new kind of logic"? Howson states that it was never properly developed for a number of reasons:

> In retrospect, it is easy to see the factors that hindered it: firstly, rapid technical development, and exploration of the

> problem-solving power of the mathematical theory, were primary, relegating 'philosophical' investigation to a relatively low-priority task; secondly a satisfactory theory of *deductive* logic did not arrive until the beginning of the twentieth century; thirdly probability became a puzzlingly equivocal notion with two seemingly quite different aspects. [11, p. 138]

However, Howson fingers a fourth development as the central culprit guilty of throwing Leibniz's programme off course: the adoption of the, in fact, extra-logical Principle of Indifference into putatively logical theories of probability.

Using a Principle of Indifference to produce a 'logical' or 'objective' prior distribution over a set of alternatives [13, p. 167] is an attempt to squeeze more out of logic than it has to give. To borrow Russell's famous quote concerning postulation, I would say that using objective 'logical' probability distributions has all the advantages of theft over honest toil. We get a probability distribution for free, but there is nothing 'logical' about this distribution. The usual conception of logic is that it just does not tell you how to get started. It does not tell you anything you did not already know or had assumed (perhaps implicitly). As Howson insists it is about consistency not truth.

> Jaynes's objective priors do not exist. But it does not follow, as he, and Fisher, and people too numerous to mention (among them Bayesian personalists) think it follows, that without objective priors the Bayesian theory is constrained to be a record merely of whims of individual psychology. That quite fallacious inference has been possibly more damaging to rational methodology than any other. We have pressed the analogy with deductive logic, and we shall press it again. Deductive logic is the theory (though it might be more accurate to say 'theories') of deductively valid inferences from premisses whose truth-values are exogenously given. Inductive logic—which is how we regard subjective Bayesian theory—is the theory of inference from some exogenously given data and prior distribution of belief to a posterior distribution. . . . neither logic allows freedom to individual discretion: both are quite *im*personal and objective. [12, p.289-90]

Another related fallacy is the 'Popper-Miller argument' which asserts that

> Although evidence may raise the probability of a hypothesis above the value it achieves on background knowledge alone, every such increase in probability has to be attributed entirely to the *deductive connections* that exist between the hypothesis and the evidence. [15]

Here, again, there is an attempt to get more from deductive logic than it has to give—it is simply not true that 'deductive connections' are enough to explain probabilistic support [6].

2.3 The return to Leibniz?

The central distinction between Leibniz and these later attempts to formulate 'objective', 'logical' priors is summed up by Hacking. For Leibniz

> The prior distribution is applied not because our set is one among whose alternatives we are ignorant; it is a set such that by metaphysics we know each element has some propensity to exist. Relative to our finite knowledge we may be able to assign only a uniform distribution over possibilities, but we will slowly correct this, ... We are able to apply probability theory here not because of a principle of indifference applied to an *a priori* language, but rather by metaphysical ascription of propensities to a classification of possibilities based on learning, scholarship and experiment. [9, p.140-1]

Although Leibniz proposes an objective prior for his 'new kind of logic' he makes it abundantly clear that *this prior is metaphysical*, he does not argue that such a prior can be merely formal. Interestingly, some recent work in AI on logical probability does not claim that there is one true prior. A good example is [2], where, for a given set of possible worlds (defined via some first-order language), three priors are considered:

> All three are applications of the principle of indifference, because they assign equal degree of belief to all basic "situations" consistent with the knowledge base. They differ because there are competing intuitions about what the basic situations are. [2]

These "competing intuitions" are essentially competing metaphysical views. Leibniz was also right to stress that the language we use has a decisive impact on the set of possible worlds. This is backed up by modern work using possible world semantics:

> the specific language chosen (for the knowledge base) is much more critical in inductive reasoning of the sort we consider than it is in traditional deductive reasoning. [2]

Although Leibniz contributed much to a logical conception of probability, I, for one, do not find his metaphysics compelling, so any return to Leibniz should only be partial. What might such a return look like?

Howson proposes to "go back to the beginning, and combine the intuitions of of Leibniz and J. Bernoulli with the conceptual apparatus of modern logic." [11, p. 140]. Howson does this by arguing that imposing mere *consistency* (as opposed to the *rationality* required by theories based on utility) of degrees of belief delivers the probability axioms in the same way that considerations of consistency lead to the rules of deductive logic:

> ...the probability axioms for epistemic probability are naturally, if not inescapably, interpreted as being of the same general species of consistency constraint as the rules of deductive logic, and to that extent *intrinsically* logical in nature. [11, p.151]

This is quite different from those approaches which use modal concepts to build logics of probability, such as those of Halpern and co-workers [10, 3] where the probability axioms are *taken for granted* and then added to first-order logic.

3 Boole

3.1 Boole's mathematical theory of logic and probability

Boole's 1854 book *An Investigation of the Laws of Thought, on which are founded the Mathematical Theories of Logic and Probabilities* [4] is of particular interest to researchers in AI seeking to combine logical and probabilistic reasoning, because it provides computational methods not just conceptual analysis.

Boole's distinction was to introduce a symbolic mathematical approach to both logic and probability. This innovation is presumably what led Bertrand Russell to make the exaggerated claim that "Pure mathematics was discovered by George Boole in his work published in 1854."

Boole is often credited as the founder of modern propositional logic, but his *Investigation* does not read like a contemporary introduction to that formalism. In contemporary propositional logic literals represent propositions but Boole begins by having them represent things:

> *1st. Literal symbols, as x, y, &c., representing things as subjects of our conceptions* [italics in original] [4, p. 27]

He then spells out just what 'things' are thus represented.

> *We are permitted, therefore, to employ the symbols x, y, z, &c., in the place of the substantives, adjective, and descriptive phrases subject to the rule of interpretation, that any expression in which several of these symbols are written together shall represent all the objects or individuals to which their several meanings are together applicable, ...* [italics in original] [4, p. 29-30]

So here Boole's literals represent *sets* (but Boole refers to them as *classes*); and Boole claims that such an interpretation is sufficiently powerful to represent natural language expressions. For example,

> if x alone stands for "white things", and y for "sheep," let xy stand for "white sheep" [4, p. 28].

He uses singleton sets to represent individuals:

> By a class is usually meant a collection of individuals, to each of which a particular name or description may be applied; but in this work the meaning of the term will be extended so as to include the case in which but a single individual exists, answering to the required name or description, ... [4, p. 28]

"Primary" propositions are "relations between things" and are expressed by set-theoretic equations. For example consider Boole's translation of the proposition "Cæsar conquered the Gauls". We would most likely represent this as a relation: conquered(Cæsar, the_Gauls), but this is not possible in Boole's system. According to Boole "Cæsar conquered the Gauls" asserts the connection between "Cæsar" and "One who conquered the Gauls". If we represent "Cæsar" by x and "One who conquered the Gauls" by y, the original sentence translates to $x = y$. This is an assertion of equality between two singleton sets; one whose only member is "Cæsar" and another whose only element is "One who conquered the Gauls".

Using sets as his basic units allows Boole to represent universally quantified statements easily. Instead of translating "All swans are white" into

$\forall x : swan(x) \rightarrow white(x)$ Boole can have $s \subseteq w$ where s is the set of all swans and w the set of all white things; where both s and w are subsets of some "universe of discourse" [4, p. 42]. Instead of representing "Bob is a swan" as $swan(Bob)$, Boole can write $b \subset s$ where b is the singleton set containing Bob. (Note though that Boole does not actually use subset notation, having a preference for equations. $y \subseteq x$ is written as $y = vx$, where v is "a class indefinite in every respect" [4, p. 61] except that $vx \neq \emptyset$.) Boole shows how Aristotelian syllogisms can be effected in his logic, and also refutes the scholastic notion that syllogism is the only form of deduction. Despite superseding scholastic logic he is generous enough to state that it "cannot be altogether unworthy of attention" [4, p. 242].

Although Boole introduces his calculus by having literals refer to things (i.e. sets), he notes at the outset that

> Logic is conversant with two kinds of relations,—relations among things, and relations among facts [4, p.7]

Propositions concerning relations between things (e.g. that Cæsar is the same thing as "one who conquered the Gauls") are called by Boole 'Primary Propositions'. 'Secondary Propositions' "concern[], or relate[] to, other propositions regarded as true of false" [4, p.159]. By Chapter 11 of his *Investigation*, he is ready to investigate these Secondary Propositions. Boole claims that the symbols for Secondary Propositions are "subject to the same laws of combination as the corresponding symbols employed in the expression of Primary Propositions" [4, p.165]. This claim is unconvincingly supported by drawing a connection between a proposition X and the extent of time x (i.e. the set of instants) for which X is true. So, for example, to express the Secondary Proposition "The proposition X is true", we denote the time for which proposition X is true by x and hence express our original Secondary Proposition by the Primary Proposition $x = 1$. This Primary Proposition states that the set of instants for which X is true is equal to the set of all instants.

Having advanced logic by his mathematical approach, Boole turns his attention to probability. He begins by outlining the relationship between the two (and blowing his own trumpet!).

> Speaking technically, we must be able to express the event whose probability is sought, as a function of the events whose probabilities are given. Now this explicit determination belongs in all instances to the department of Logic. Probability, however, in its mathematical acceptation, admits of numerical measurement. Hence the subject of Probabilities belongs equally to the science of Number and to that of Logic. In recognising the co-ordinate existence of both these elements, the present treatise differs from all previous ones [4, p. 13]

Boole tended to be too sanguine concerning the extent to which deductive relationships determined probabilistic ones. He notes that:

> ...there exists a definite relation between the laws by which probabilities of events are expressed as algebraic functions of the probabilities of other events upon which they depend, and the laws by which the logical connexion of the events is itself expressed [4, p. 16]

But this "definite relation" is less close than he thinks; and, in fact, Boole goes even further than Popper and Miller in arguing that deductive relations determine probabilities.

The problem is that Boole, despite certain hedges, has a basically subjective interpretation of probability. This leads him to assert that we can infer independence whenever we can not infer dependence:

> The events whose probabilities are given are to be regarded as independent of any connexion but such as is either expressed, or necessarily implied, in the data; and the mode in which our knowledge of that connexion is to be employed is independent of the nature of the source from which such knowledge has been derived. [4, p. 256-7]

Boole seems to be aware that such an inference is at least controversial. He uses his "can't find dependence thus there's independence" argument in an example using an urn model, and defends it thus:

> To meet a possible objection, I here remark, that the above reasoning does not require that the drawings of a white and a marble ball should be independent, in virtue of the physical constitution of the balls. The assumption of their independence is indeed involved in the solution, but it does not rest upon any prior assumption as to the nature of the balls, and their relations, or freedom from relations, of form, colour, structure, &c. It is founded upon our total ignorance of all these things. Probability always has reference to the state of our *actual* knowledge, and its numerical value varies with varying information. [4, p. 262-3]

Boole is not entirely consistent in his interpretation of probability. As the previous quote indicates, probability is generally defined by Boole in terms of partial knowledge, but as he acknowledges this does not explain stable long-run frequencies:

> So also the theory of probabilities contemplates the numerical measure of the circumstances upon which expectation is founded; and this object embraces the whole range of its legitimate applications. The rules which we employ in life-assurance, and in the other statistical applications of the theory of probabilities, are altogether independent of the *mental* phenomena of expectation. They are founded upon the assumption that the future will bear a resemblance to the past; that under the same circumstances the same event will tend to recur with a definite numerical frequency; not upon any attempt to submit to calculation the strength of human hopes and fears. [4, p. 244-5]

So, here it is the actual circumstances of an event, not just our partial knowledge of these circumstances that determine a probability. This ambiguity led Keynes [13, p. 167] to argue that Boole had two inconsistent notions of independence: the standard one and one where independence equals "no known dependence" notion. Keynes describes this contradiction as Boole's "central error" and provides an interesting condensed account of reactions to it. Here is a condensed account of that condensed account. Apparently Wilbraham pointed out Boole's contradiction in a review (also written in 1854) of Boole's book. Boole did not understand Wilbraham's criticism and "replied hotly, challenging him to impugn any individual results [13, p. 167]. Most interestingly, Boole returned to the issue in another paper where he appeals to the Principle of Indifference to support his approach.

3.2 Boole and 'propositional' representations

The bulk of the probabilistic material in Boole's *Investigation* applies the mathematical methods he has developed for what is essentially propositional logic to probabilistic inference. A desired probability value is derived from given probability values (and some dubious assumptions of independence). In some cases, not all the required probability values are given in which case the desired probability value cannot be returned, only the interval in which it lies. Deductive relations provide constraints between probabilities.

But, for the purposes of this paper, this technical work is of less interest than the choice of representation Boole adopts. Although Boole concentrates on a set-theoretic ('Primary') interpretation of his literals, it is clear that his logic can also be interpreted as a propositional logic (of 'Secondary' Propositions). Hopefully, this is uncontroversial, even if Boole's own argument for the connection is unconvincing.

Before examining the extent to which Boole's representation succeeds in representing probabilistic inference, it is worth making a small detour to consider the application of propositional representations in AI. Considerable care must be taken when describing a probabilistic model (or indeed a machine learning technique) as 'propositional'. Propositional logic is really a very limited formalism; and falls far short of what is required to formalise probabilistic inference or any machine learning algorithm. To even formalise basic arithmetic on natural numbers we need first-order logic augmented with arithmetic postulates. For an account of this (including an axiomatisation of first-order Peano arithmetic) see, for example, [14]. However, once we have added the axioms of first-order Peano arithmetic to first-order logic it is very hard work to find a "fact of genuine mathematical interest" [14] that does not follow. It is hard work, but not impossible: Paris and Harrington succeeded in 1977. However, I am not aware of any truths of probability theory which are known not to follow.

Nonetheless finite probability distributions, such as those defined using Bayesian nets (BNs) with finite discrete variables, are often described as 'propositional'. The reasoning is that each sample point represents a proposition stating how the world is—in the case of BNs a sample point is a full joint instantiation of the random variables in the BN. Since all we have done is to add a probability distribution over these propositions, then the resulting representation is still propositional or at least 'essentially propositional'.

Such a description does indeed capture the *atomic* nature of sample points in finite probability distributions. There is no necessary relation between them, and the probability function does not 'look inside' them.

Similarly, a machine learning algorithm (such as a decision tree inducer) which learns rules of the form $type = swan \rightarrow colour = white$ is generally called propositional. This is because even though machine learning generalises from instances (e.g. "Bob is a white swan") to general rules (e.g. "All swans are white") it is possible to use the *Single Representation Trick* to represent both instances and rules in propositional logic. But the Trick does not always come off. As Flach points out:

> ...even if concept learning from examples proceeds from instances to sets of instances, it can be formalised by employing a simple propositional language that fails to recognise the difference. However, it should be made clear that something is lost along the way. For instance, consider a situation in which some of the examples are incomplete: the values of some attributes

> are left unspecified. It may happen that one of the missing attributes is crucial for explaining the classification of those instances—in fact, it is possible that two instances receive the same incomplete description, while one is classified positively and the other negatively! Clearly, in the above setting this would immediately result in logical inconsistency, because the two instances are described by syntactically equal expressions. Such inconsistencies can be avoided by employing a predicate-logical language, since such a language allows us to mention the instance explicitly. [7, p. 39]

So how did Boole get away with a propositional representation for logical and probabilistic reasoning? Firstly, as the title of his book makes explicit, Boole's theory is a *mathematical* theory. He takes sets, numbers and algebraic operations for granted, he is not seeking to construct some formal (perhaps logical) framework within which to justify his use of these basic mathematical objects. It is instructive to compare Boole's contribution to Frege's programme. Frege's view was that arithmetic (although not geometry) can be derived from logic—a doctrine known as *logicism*—and his 1879 *Begriffschift* was an attempt to establish this. Boole, on the other hand took mathematics as given and attempted to derive logic by deducing " *the laws of the symbols of Logic from a consideration of those operations of the mind which are implied in the strict use of language as an instrument of reasoning*".

Secondly, Boole is able to talk about both individuals and properties of individuals by the simple expedients of (i) conflating an individual a and the singleton set containing that individual $\{a\}$ and (ii) equating a property with the set of all individuals that have that property. Of these two moves, the latter is embedded in the standard semantics for first-order logic [14, p.51], but the latter is now regarded as fallacious. Thirdly, Boole simply allows himself enough distinct literals so that even relational statements can be given a propositional gloss. His kludgy translation of "Cæsar conquered the Gauls" shows the sort of problems this will lead to. If we were to translate "Cæsar conquered the Britons" in the same way we have to consider a new singleton set whose only element is "One who conquered the Britons". The fact that conquering is going on in both cases is completely obscured. Also it would be difficult to connect these statements to other about Gauls or Britons.

So summing up, we see that many of the pros and cons of propositional representations are apparent in a book written in 1854. On the negative side, as Flach puts it, "something is lost along the way", in particular a clear distinction between instance and predicate. On the positive side, Boole's *Investigation* was a breakthrough and the very simplicity of his representation—his reductionism where everything is reduced to a set—must surely be partly responsible for his progress.

4 Conclusions

I would like to argue that both Leibniz and Boole have useful things to say to those working on logic and probability in contemporary AI. Taking Boole first, his disrespectful attitude towards logic, where mathematics is used to analyse logic rather than vice-versa is appealing (to me).

The logicist-inspired view that a logical foundation is essential, or at least desirable, for new formalisms is still with us—for what can provide a firmer foundation than the laws of logic? And if an approach does not fit within a standard logic, then it is necessary to move the goalposts and invent a new logic with different laws, a common approach in AI and computer science in general:

> ... new theories of reasoning are increasingly marketed under the general title 'logic' as the information technology revolution gets under way (whereas even twenty years ago most logic texts were produced in university philosophy departments, now probably most are produced in computer science departments). [11, p. 149]

Of course, formalising reasoning systems generally clarifies the properties of a system and, by and large, these new formalisms *do* have many of the characteristics of a logic (separation between syntax and semantics, rules of inference, etc) and hence merit the title logic. The thing is not to use the title logic to confer spurious authority. Hacking hits the right note:

> It will be useful to dispel the idea, popular among non-logicians, that logic is a kind of fixed immutable science. In fact, like every other human endeavour, it is just a patchwork quilt whose patches do not meet very well, and which are continually being torn up and restitched. [8, p. 34]

Another useful point to take from Boole is that probabilistic inference is just deduction with numbers, there is no funny business going on. Or, at least, there need be funny business going on as long as we stick to deducing probabilities from existing ones, rather than inducing them from Principles of Indifference or the like.

The use of Principles of Indifference, on the other hand, is an attempt to get something (a probability distribution) for nothing. But, unsurprisingly, the temptation to claim an alleged free lunch is not restricted to philosophers. In machine learning, the claim that 'simpler' hypothesis are *a priori* more probable is sometimes made—often due to a misinterpretation of valid mathematical arguments concerning 'Occam algorithms'. This view has been correctly criticised by Buntine [5], amongst others. A probability distribution, whether uniform or with a bias towards simplicity cannot be axiomatic.

Leibniz, although he did not give a fully satisfactory account of the propensities of his possible worlds to exist, did, at least, realise that a distribution over possible worlds cannot be axiomatic. We do not need to share Leibniz's metaphysics to agree with him on this point.

References

[1] Antoine Arnauld and Pierre Nicole. *La logique, ou l'art de penser.* Bobbs-Merrill, 1662. The English title is the *The Art of Thinking* or the *Port-Royal Logic*. First published in 1662, the edition published by Bobbs-Merrill is a 1964 English translation of the 5th edition of 1683. The Bobbs Merrill edition only credits Arnauld with authorship; I have followed Hacking in also acknowledging Pierre Nicole.

[2] Fahiem Bacchus, Adam Grove, Joseph Y. Halpern, and Daphne Koller. From statistics to belief. In *Tenth National Conference on Artificial Intelligence (AAAI-92)*, pages 602–608, 1992.

[3] Fahiem Bacchus, Adam J. Grove, Joseph Y. Halpern, and Daphne Koller. From statistical knowledge bases to degrees of belief. *Artificial Intelligence*, 87:167–215, February 1997.

[4] George Boole. *An Investigation of the Laws of Thought, on which are founded the Mathematical Theories of Logic and Probabilities*. Macmillan, 1854. Reprint by Dover, 1958.

[5] Wray Buntine. Myths and legends in learning classification rules. In *Proceedings of the Eighth National Conference on Artificial Intelligence (AAAI-90)*, pages 736–742. AAAI Press, 1990.

[6] James Cussens. Deduction, induction and probabilistic support. *Synthese*, 108(1):1–10, July 1996.

[7] Peter Flach. *Conjectures: an inquiry concerning the logic of induction*. PhD thesis, Katholieke Universiteit Brabant Tilburg, 1995.

[8] Ian Hacking. *Logic of Statistical Inference*. Cambridge University Press, Cambridge, 1965.

[9] Ian Hacking. *The Emergence of Probability*. Cambridge University Press, 1975.

[10] Joseph Y. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311–350, 1990.

[11] Colin Howson. The logic of Bayesian probability. In David Corfield and Jon Williamson, editors, *Foundations of Bayesianism*, volume 24 of *Applied Logic Series*, pages 137–159. Kluwer, Dordrecht, 2001.

[12] Colin Howson and Peter Urbach. *Scientific Reasoning: The Bayesian Approach*. Open Court, La Salle, Illinois, 1989.

[13] J. M. Keynes. *A Treatise on Probability*. Macmillan, London, 1921.

[14] Moshé Machover and John Bell. *A Course in Mathematical Logic*. North-Holland, Amsterdam, 1977.

[15] Karl R. Popper and David Miller. Why probabilistic support is not inductive. *Philosophical Transactions of the Royal Society of London*, 321:569–591, 1987.