For this situation the Charnes-Cooper-Schechter result may be stated as follows:

Theorem 0.53. *The linear fractional programming problem:*

$$\textit{optimize} \quad \frac{\mathbf{cx}}{\mathbf{ax}}$$

subject to

$$\mathbf{Ax} = \mathbf{b}$$
$$\mathbf{x} \geq \mathbf{0},\ \mathbf{ax} > 0$$

is equivalent to the linear programming problem:

$$\textit{optimize} \quad \mathbf{cy}$$

subject to

$$\mathbf{Ay} = t\mathbf{b}$$
$$\mathbf{ay} = 1$$
$$t, \mathbf{y} \geq \mathbf{0}.$$

Chapter 1

Early History

§1.1. Leibniz's vision

The following is an excerpt from Leibniz's *Nouveau essais* (*1962*, 466):[1]

> I have said more than once that there is need for a *new kind of logic* which would treat of degrees of probability. For what Aristotle has in his *Topics* is quite different. He is satisfied with arranging a few familiar rules according to common patterns; these could serve on the occasion when one is concerned with amplifying a discourse so as to give it some likelihood. No effort is made to provide a balance necessary to weigh the likelihoods in order to obtain a firm judgement. Anyone wishing to treat these matters would do well to examine games of chance; in general, I wish that some skillful mathematician would be willing to produce a detailed, systematic and extensive work on all the varieties of games, ...

The *Topics* of Aristotle, to which Leibniz is here referring, has been described as a handbook for the guidance of contestants in public debating contests. In contrast Leibniz, whose early training was in jurisprudence, is primarily interested in the problem of rationally adjudicating between opposing views or conflicting claims. In this connection, as our quotation indicates, Leibniz believed that the recently developed mathematical methods for calculating chances well worth looking at. He gives the example of

1. This is a modern edition. Although the *Nouveau essais* were first published in 1765, long after Leibniz's death, he had ideas for a *doctrina de gradibus probabilitatis* at least as early as 1670 (*Schneider 1981*, 204). These ideas are extensively described and discussed in *Couturat 1901*, 238–55, 552–55. *Schneider 1981* also covers this ground in less detail, but from the advantage of an additional 80 years of Leibniz scholarship.

two players, one winning if a 7, the other if a 9, is the outcome of a throw of a pair of dice. The likelihoods of winning are in proportion to the number of ('equally possible') outcomes favorable to the respective values. Aside from such simple examples Leibniz seems not to have pursued the matter any further.

The *Nouveaux essais* were written in Leibniz's mature period, the first decade of the eighteenth century. His much earlier baccalaureate thesis of 1665, *De conditionibus* (an improved version in *Specimen juris* of 1669), written before he became aware of the currently developing theory of probability, is a theoretical legal treatise on conditional rights.[2] It contains the suggestion of ordering conditional rights on the scale of 0 to 1 (*Couturat 1901*, 552. *Hacking 1975*, chapter 10, 'Probability and the law'). A right, for example to an estate or a throne, may not be absolute but dependent upon a condition. A statement of the right would be a conditional sentence whose antecedent embodies the condition. With regard to such statements Leibniz (*1971*, 420) presents the following schema:

Antecedent:	impossible	contingent	necessary
	0	$\frac{1}{2}$	1
Right:	non-existent	conditional	absolute

This indicates that a right is non-existent, conditional, or absolute depending on whether the antecedent is impossible, contingent or necessary. The fraction $\frac{1}{2}$ is used by Leibniz as standing for some unspecified value between 0 and 1. In other words Leibniz is allowing a right to have a range of values between 0 and 1, but he gives no indication of how such values are to be determined. There is perhaps some basis for Couturat's enthusiastic assertion (*1901*, 553):

> En effet, ces valeurs 0, θ [arbitrary fraction between 0 and 1], et 1 mesurent précisément la probabilité de ce droit dans les trois cas ... Ainsi Leibniz a entrevu ici, d'une part, le Calcul des probabilités, ...

Indeed, involved here is the idea of a numerical measure of the likelihood of the consequent of a (necessary) conditional on the basis of that of the antecedent—but only inchoately. So far as we know Leibniz never developed the idea. Jakob Bernoulli did.

2. Apart from the matter of interest to us, it contains a nice exposition of the conditional proposition and its principal logical features. Except for Couturat (*1901*, 553–54), this seems not to have been noticed by historians of logic.

§1.2. Jakob Bernoulli—probability logic via the fates of gamblers

Bernoulli's *Ars conjectandi 1713*, left unfinished at his death and published posthumously, contains four parts. Part I is an annotated presentation of Huygens' *De ratiociniis in ludo aleae*, part II is a development of mathematical properties of permutations and combinations, and part III states and solves a variety of problems concerned with players expectations in games of chance. Part IV, entitled 'Use and application of the preceding theory in civil, moral and economic affairs', is famous for containing the law of large numbers. However there are no realistic applications of the sort mention in its title—presumably it had been Bernoulli's intention to include some. The material which is the subject of this and the next section is contained in chapter 3 of part IV. Yet, despite its innovative character, it has been largely neglected. For example, Todhunter (*1865*, 70–71) devotes three short paragraphs to it; van der Waerden, in the introduction to Volume 3 of Bernoulli's collected works, doesn't mention it in his description of *Ars conjectandi*; Maistrov in *1974* quotes one sentence from it. Not until we come to *Hacking 1974* and *Shafer 1978* do we have substantial discussion; however the conclusions we come to will be distinctively different from theirs.

In this chapter 3, 'On various kinds of arguments, and how their weights are estimated in order to compute the probability of things', Bernoulli investigates the 'force of proof' of an argument and the degree of certainty of an opinion or conjecture on the basis of arguments for it. As with Leibniz's conditional rights, these investigations involve conditional sentences and the likelihood of their consequents. With Bernoulli, however, we find a substantial advance beyond Leibniz's strongly expressed hope. Conceivably Leibniz may have had some influence on Bernoulli's ideas, for they did correspond with each other. But as far as our topic is concerned, there seems to be little evidence for it. For example, Bernoulli's repeated requests to Leibniz for a copy of his *De conditionibus* finally ended when Leibniz simply acknowledged that he had none. (See *Schneider 1981b*, 212.)

Bernoulli's treatment distinguishes three kinds of arguments which produce opinion or conjecture: those which exist necessarily and imply [the conjecture] contingently (*necessariò existunt & contingenter indicant*), those which exist contingently and imply necessarily, and those which both exist and imply contingently. These notions he illustrates by the following examples (*1713*, 217–18).[3]

3. We avail ourselves here, and elsewhere in this section, of the excellent translations in *Shafer 1978*. However, Shafer uses 'to prove' as a translation for both Bernoulli's *indicare* and *probare*. We believe 'to imply' more suitable for most instances of *indicare* and accordingly have made the substitution in our transcriptions. (This is not trivial

My brother has not written me for a long time; I am not sure whether his indolence or his business is to blame; also I fear he might in fact have died. Here there are three arguments concerning the interrupted writing: indolence, death, and business. The first of these exists necessarily (by a hypothetical necessity, since I know and assume my brother to be lazy), but implies contingently, for it might have happened that this indolence did not keep him from writing. The second exists contingently (for my brother may still be among the living), but implies necessarily, since a dead man cannot write. The third both exists contingently and implies contingently, for he might or might not have business, and if he has any it may not be so great as to keep him from writing. Another example: I consider a gambler who, by the rules of a game, would win a prize if he threw a seven with two dice, and I wish to conjecture what hope he has of so winning. Here the argument for his winning is a throw of the seven, which implies it necessarily (by a necessity from the agreement entered into by the players) but exists only contingently, since other numbers of points can occur besides seven.

The term 'argument' has a number of meanings, two of which are closely related: the word is used (i) as a substantive representing a statement (a premise) which, if true, serves to establish or justify another statement (the conclusion), and also (ii) a notion involving two statements when there is a deduction (argumentation) from one to the other. The items which in the preceding quotation Bernoulli refers as arguments ('indolence, death, and business', and 'throw of a seven') show that he is there using the term in the first sense since he refers to arguments as implying [the conclusion] necessarily, or contingently; and when he says an argument 'exists' (note, separate from 'and implies') we take it to mean that the argument-as-premise is true. But in his description of the three types of argument both components, argument-as-premise and argumentation, are involved. (However Bernoulli doesn't concern himself with deductions but with the implication relation between premise and conclusion.) Using A for the premise and C for the conclusion we may characterize Bernoulli's three

as it makes for a major difference between Shafer's and our interpretation of Bernoulli's ideas.) For want of a better word we have kept Shafer's 'to exist' for Bernoulli's *existere* although the Latin has an active sense (to arise, come forth, appear). Likewise we use Shafer's 'thing' for *res* but with the meaning not of 'an object' but in the sense of 'the matter, affair or circumstance'. The page numbers of Bernoulli's *1713* are keyed in *Bernoulli 1975*.

types of arguments as:

Type 1.	A necessary,	A-implies-C contingent
Type 2.	A contingent,	A-implies-C necessary,
Type 3.	A contingent,	A-implies-C contingent.

Note that Bernoulli is allowing contingent argumentations as well as contingent argument-premises. We shall later on encounter writers who, unaware of Bernoulli's analysis, restricted themselves to necessary argumentations, i.e., to Bernoulli's Type 2. This would exclude, for example, argumentations with insufficient premises.

Probability, Bernoulli's synonym for degree of certainty, is brought into the discussion (*1713*, 218–19):

> It is clear from what has been said thus far that the force of proof by which any given argument avails depends on the large number of cases whereby it can exist or not exist, imply or not imply, or even imply the contrary of the thing. Indeed, the degree of certainty or probability which the argument generates can be computed from these cases by the doctrine of the first part [of this book], just as the fates of gamblers in games of chance are usually investigated. In order to show this, we assume b is the number of cases where a given argument exists, c is the number where it does not exist, and $a = b+c$ is the number of both together. Similarly, we assume β is the number of cases where it implies, γ is the number where it does not imply or else implies the contrary of the thing and $\alpha = \beta + \gamma$ is the number of both together. Moreover, I suppose that all the cases are equally possible, or can happen with equal ease. Otherwise, discretion must be applied and in the place of any case that happens more easily than the others one must count as many cases as it happens more easily. For example, in place of a case that happens three times more easily than the others, I count three cases which can happen equally as easily as the others.

We see that Bernoulli is going to provide a numerical measure for what he refers to as the 'force of proof by which any given argument avails' or as the 'degree of certainty or probability which the arguments generates', and that he will do this by use of the doctrine of the first part. This doctrine is Huygens' theory, which has as its basic concept the value of a gamble in games of chance. In particular he will be using the following result (*1713*, 7):

> PROPOSITION III. If the number of cases which result in my getting g is p and, moreover if the number which results in my getting

l is q; (then) taking all cases to be of equal proclivity, my expectation is worth

$$\frac{gp + lq}{p + q}. \tag{1}$$

When rewritten in the form

$$g \cdot \frac{p}{p+q} + l \cdot \frac{q}{p+q}$$

it would appear—from the contemporary point of view—that (1) is the expected value of a random variable whose two values are g and l with respective associated probabilities $p/(p+q)$ and $q/(p+q)$. Bernoulli will be using (1) for the case of $g = 1$ and $l = 0$. Calling those cases for which $g = 1$ 'favorable' and those for which $l = 0$ 'unfavorable', then $p/(p+q)$ is the ratio of the number of favorable cases to the sum of the number of favorable and unfavorable cases. Thus the expectation which Bernoulli computes would be a probability in the 'classical' definition sense—but only if the cases are not only of 'equal proclivity' but also are mutually exclusive and exhaustive.

Bernoulli's '[argument] implies [conclusion]' we shall abbreviate to '$A \to C$', not thereby necessarily attributing to him the use of 'implies' in a truth-functional sense. When Bernoulli refers to an argument as being necessary we shall take it to mean that $A \to C$ is true in all cases. Although it will not be clear until we see how he computes it, what Bernoulli uses as his measure of the 'degree of certainty or probability generated by an argument' evidently depends on the cases of A and of $A \to C$.

In addition to the tripartite classification of types of argument mentioned earlier, Bernoulli also distinguishes arguments as to being pure or mixed (*1713*, 218):

> I call those arguments *pure* which prove [*probant*] a thing in certain cases in such a way that they prove nothing positively in other cases; I call those *mixed* which prove the thing in some cases in such a way that they prove the contrary in the remaining cases.

Here he is talking about an argument which proves rather than, as has been the situation up to now, of implying a thing in a case. We take this to mean that both A and $A \to C$ are true for the case. He elucidates the notions of pure and mixed by an example (*1713*, 218):

> A certain man has been stabbed with a sword in the midst of a rowdy mob, and it is established by the testimony of trustworthy men who were standing at a distance that the crime was committed by a man in a black cloak. If it is found that Gracchus and three others in the crowd were wearing tunics of that color, this tunic is something of an argument that the murder was committed by Gracchus, but is mixed: for in one case it proves his guilt, in three cases his innocence, according to whether the murder was perpetrated by himself or by one of the remaining three; for it is not possible that one of these perpetrated it without Gracchus being thereby supposed innocent. But if indeed in a subsequent hearing Gracchus paled, this pallor of face is a pure argument; for it proves Gracchus' guilt if it arises from a guilty conscience, but does not, on the other hand, prove his innocence if it arises otherwise; for it could be that Gracchus pales from a different cause yet is still the murderer.

The distinction may be illustrated by the following diagrams in which inclusion of regions corresponds to implication (Figures 1.21 and 1.22). In this situation 'case' refers only to A since in Bernoulli's example the $A \to C$ part is necessary, i.e., true in any case. In Figure 1.21 A is resolvable into four mutually exclusive cases, the argument (A, $A \to C$) proving C in case 1 and proving $\neg C$ in cases 2, 3, 4. In Figure 1.22 for the argument (B, $B \to C$) there are two cases; when the first case holds the argument (B, $B \to C$) proves C, but in the second case the argument proves 'nothing'. This latter case corresponds, in Bernoulli's example, to Gracchus's pallor, which is compatible with his being guilty or not guilty.

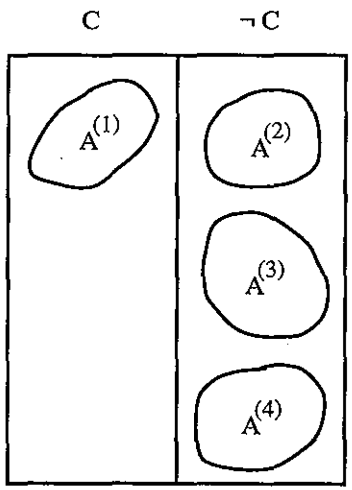

FIGURE 1.21: MIXED

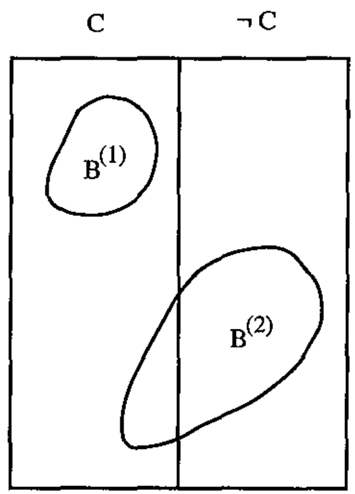

FIGURE 1.22: PURE

Exactly what are Bernoulli's cases? In parts I and III of *Ars conjectandi*, which deals with expectations in connection with games of chance, a case corresponds to a possible (chance) outcome. These can all be explicitly listed and there is one, and only one, outcome in a given case. But when Bernoulli carries the notion over from games of chance to propositions we are no longer sure of what this notion means. There are three aspects in which this new usage differs from that in games of chance:

1. The element of chance (as an outcome in a dice roll, or in a blind drawing from an urn) is absent. But this is not a material difference since Bernoulli speaks of cases which are "equally possible or can happen with equal ease".

2. It isn't clear that there is only one set of possible cases associated with a given argument (A, $A \to C$) as there would be for a given game. Bernoulli allows for a set of cases for A (a in number) and another set of cases for $A \to C$ (α in number). That the cases in each set are mutually exclusive is tacitly assumed, but nothing is indicated concerning their being exhaustive of all possibilities. Moreover, as we shall see, he tacitly assumes that the cases of the two sets are independent of each other.

3. For a pure argument a case for $A \to C$ need not determine a truth value for $A \to C$ (e.g., $B^{(2)}$ in Figure 2). In a game of chance, however, if one outcome is the case no other is, i.e., all outcomes are uniquely determined as to happening or not.

Despite these differences Bernoulli believes the theory of part I (which "determines the fates of gamblers") applies to computing the degrees of certainly of an argument. We discuss his results in our next section and, unlike *Hacking 1974* and *Shafer 1978*, offer explanations of what Bernoulli's results mean in terms of standard probability theory.

§1.3. Degree of certainty of an argument

Bernoulli's computations for the degree of certainty of an argument is apparently the earliest application of probabilistic notions to logic. He carries these computations out for each of the three types of arguments, and for the pure and mixed kind in the following manner (*1713*, 218-19):

> 1. So first let the argument *exist contingently and imply necessarily.* By what has just been said, there will be b cases where the argument exists and thus proves the thing (or 1), and c cases where it does not exist and thus proves nothing. By Corollary I of Proposition III of part I, this is worth
>
> $$\frac{b \cdot 1 + c \cdot 0}{a} = \frac{b}{a},$$
>
> so that such an argument establishes b/a of the thing, or of the certainty of the thing.
>
> 2. Next let the argument *exist necessarily and imply contingently.* By hypothesis, there will be β cases where it implies the thing, and γ cases where it does not imply or implies the contrary; this now gives a force of argument for proving the thing of
>
> $$\frac{\beta \cdot 1 + \gamma \cdot 0}{\alpha} = \frac{\beta}{\alpha}.$$
>
> Therefore an argument of this kind establishes β/α of the thing; and moreover, if it is mixed it establishes (as is clear in the same way)
>
> $$\frac{\gamma \cdot 1 + \beta \cdot 0}{\alpha} = \frac{\gamma}{\alpha}$$
>
> of the certainty of the contrary.
>
> 3. If some argument *exists contingently and implies contingently,* I suppose first that it exists, in which case it is judged in the manner just shown to prove β/α of the thing and moreover, if it is mixed, γ/α of the contrary. Hence, since there are b cases where it exists and c cases where it does not exist and hence cannot prove anything, this argument is worth
>
> $$\frac{b \cdot \frac{\beta}{\alpha} + c \cdot 0}{a} = \frac{b\beta}{a\alpha}$$
>
> for proving the thing, and if it is mixed, is worth
>
> $$\frac{b \cdot \frac{\gamma}{\alpha} + c \cdot 0}{a} = \frac{b\gamma}{a\alpha}$$
>
> for proving the contrary.

In items 1. and 2. Bernoulli is computing the value of a gamble where the pay-off is 1 (certainty) if the "thing is proved" and 0 otherwise. He uses the corollary to Huygens' rule, cited above, in which $g = 1$ and $l = 0$. In item 3. the b cases for which the argument exists are valued at β/α rather that 1, and the remainder at 0. Despite his numerous references to the "degree of certainty of the thing" it is important to note that what Bernoulli computes is not the probability of the conclusion but the "force of proof of the argument", i.e., the probability that the conclusion is proved by the argument.

To see this we compare his results with a modern treatment of the problem, namely computing the expected value of an appropriate random variable with a given probability distribution. As such a random variable we choose X, where

$$X = \begin{cases} 1, & \text{if } A \text{ proves } C \\ 0, & \text{otherwise.} \end{cases}$$

Its expected value is

$$\begin{aligned} E(X) &= 1 \cdot P(X=1) + 0 \cdot P(X=0) \\ &= P(X=1) \\ &= P(A \text{ proves } C). \end{aligned}$$

Assume, first of all, that the argument under consideration is of the mixed kind (Fig. 1.21). Then 'A proves C' is equivalent to 'A holds and A implies C', i.e., to the conjunction $A(A \to C)$, supposing '$\to$' to be the usual truth-functional conditional. Thus

$$E(X) = P(A(A \to C)). \qquad (1)$$

Now for the probability distribution. As Bernoulli is assuming a set of equi-possible cases which determine when the argument exists (A holds), and also a set of equi-possible cases which determine when it implies the thing, one can have probabilities defined for each of these events using the 'classical' definition. Bernoulli's assumptions then allow us to write

$$P_H(A) = b/a, \qquad P_I(A \to C) = \beta/\alpha, \qquad (2)$$

where P_H and P_I are the respective probability functions connected with the cases for A and for $A \to C$. But for $P(A(A \to C))$ to have a meaning we need to have a conjoint probability distribution for A and $A \to C$. This is something not considered by Bernoulli and, in effect, he assumes that it is the product distribution, i.e., that A and $A \to C$ are stochastically independent, with P_H and P_I the marginal probabilities.[4] Granted this independence, so that

$$P(A(A \to C)) = P_H(A)P_I(A \to C),$$

4. In §5.3 below we show, in the comment following Theorem 5.34, that A and $A \to C$ (with $\to$ taken as the 2-valued conditional) are stochastically independent if and only if either $P(A) = 1$ or $P(A \to C) = 1$, i.e., if and only if one or the other is necessary.

and noting that $b = a$ if A is necessary, and that $\beta = \alpha$ if $A \to C$ is necessary, the results Bernoulli obtains for the degree of certainty [of proving] the thing for the three types of arguments 1.–3. then follow (for the mixed kind). As for proving the contrary (end of item 2.), since it is being assumed that the argument is of the mixed kind, $A \to \neg C$ is equivalent to $\neg(A \to C)$. Hence (assuming independence)

$$\begin{aligned} P(A \text{ proves } \neg C) &= P(A(A \to \neg C)) \\ &= P(A\neg(A \to C) \\ &= P(A)P(\neg(A \to C)) \\ &= \frac{b}{a}(1 - \frac{\beta}{\alpha}) = \frac{b\gamma}{a\alpha}. \end{aligned}$$

We next consider the argument to be of the pure kind (Fig. 1.22). Here 'A proves C' has a new meaning occasioned by the occurrence of indecisive cases (overlapping both C and $\neg C$ region in Fig. 1.22). These cases "prove nothing", hence the β count in β/α treats such cases as though they implied the contrary. Let '$A \overset{*}{\to} C$' designate the implication resulting when the indecisive cases are made into ones implying the contrary. Then, assuming independence of A and $A \overset{*}{\to} C$,

$$\begin{aligned} P(A \text{ proves } C) &= P(A(A \overset{*}{\to} C)) \\ &= P(A)P(A \overset{*}{\to} C). \end{aligned}$$

And since Bernoulli uses β/α for $P(A \overset{*}{\to} C)$ as well as for $P(A \to C)$ the same expression, namely, $b\beta/a\alpha$, results for the pure as for the mixed argument kind. Nevertheless there is a difference. For

$$P(A \text{ proves } C) = \begin{cases} P(A(A \to C)) = P(AC), & \text{if mixed} \\ P(A(A \overset{*}{\to} C)) \leq P(AC), & \text{if pure.} \end{cases}$$

Shafer (*1978*) argues for a different interpretation of Bernoulli's probabilistic analysis of contingent arguments. His contention is that Bernoulli has an epistemic notion of probability which is non-additive (i.e., probabilities for and against need not add up to 1). Admittedly there are difficulties in understanding Bernoulli's ideas, but in view of his evident reliance on games of chance and gambling as his paradigm and the absence of any indication of a conceptual change, we find Shafer's position not convincing. Additionally, as Shafer acknowledges, his explanation encounters difficulties in accounting for some aspects of Bernoulli's treatment (e.g., the matters discussed in *Shafer 1978*, 332 and 334). As we have just seen, Bernoulli gives the value $b\gamma/a\alpha$ for an argument proving the contrary of the thing in the mixed type case. Concerning this Shafer says (*1978*, 332):

> So in the case of a mixed type argument which exists only contingently, we do indeed have a positive probability p for a thing and a positive probability q for its contrary such that $p+q<1$.
> For
> $$\frac{b\beta}{a\alpha}+\frac{b\gamma}{a\alpha}=\frac{b}{a},$$
> which is less than 1 if the argument really exists only contingently.

Shafer is considering Bernoulli's worth of an argument for, and against, a conclusion as being complementary probabilities which in standard probability theory should add up to 1. Since they don't he concludes that non-additive probability must be involved. But from our point of view Bernoulli is computing expected values of proving the thing, and its contrary—in the one case this is

$$P(A \text{ proves } C) = P(A(A \to C)) = P(AC)$$

and in the other

$$P(A \text{ proves } \neg C) = P(A(A \to \neg C)) = P(A\neg C).$$

The sum of these is $P(A)$, the b/a which Shafer computes, and there is no special need to account for its value being less than 1.

§1.4. On combining two or more arguments

After obtaining the results just discussed (i.e., items 1., 2., and 3.) Bernoulli then goes on to the question of computing the force of proof [for a conclusion] when there are two or more arguments for the same thing. In this connection he has the table:

arguments	*1st*	*2nd*	*3rd*	*4th*	*5th*	etc.
total number of cases	a	d	g	p	s	etc.
proving	b	e	h	q	t	etc.
non-proving or else proving the contrary	c	f	i	r	u	etc.

Considering first pure arguments, for which the last line in the table refers to the number of non-proving cases, Bernoulli obtains for the force of proof in combining the first and second arguments,

$$\frac{e\cdot 1+f\cdot\frac{b}{a}}{d} \quad \left(=\frac{e}{d}+\frac{f}{d}\cdot\frac{b}{a}\right). \tag{1}$$

His reason is that for the second argument there are e of the d cases that prove the thing (with weight value 1) and of the remaining f cases there is a weight of b/a of proving the thing by virtue of the first argument.

Introducing 'A proves C' and 'B proves C' for the two arguments, and P_I and P_{II} as the (classical) probability notions pertinent to the two sets of cases, we can write in place of (1)

$$P_{II}(B \text{ proves } C) + P_{II}(B \text{ does not prove } C)\cdot(P_I(A \text{ proves } C). \tag{2}$$

This can also be written as

$$P((B \text{ proves } C) \text{ or } (A \text{ proves } C)), \tag{3}$$

where P is the product probability function of which P_I and P_{II} are its *independent* marginals. In obtaining (1) Bernoulli implicitly assumes independence of the cases, that is that there are no inter-relations. Moreover, the arguments being pure, 'proves' has a special meaning, namely 'A proves C' means '$A(A \overset{*}{\to} C)$'. Note that elements of the product probability space, which are ordered pairs of cases one from each of the two sets of cases, are to be considered equally possible. Moreover, to get Bernoulli's result one needs to consider a pair having at least one proving case to be a proving case for the combination. Without a proper clarification of 'cases' Bernoulli's result, i.e., (1), has dubious reliability.

A different approach is needed for the combining of mixed type arguments, in which case the last line in Bernoulli's table refers to the number of cases proving the contrary. Here he says (*1713*, 220–21):

> 5. Next let all the arguments be mixed. Since the number of proving cases in the first argument is b, in the second e, in the third h, etc., and the number proving the contrary, c, f, i, etc., the probability of the thing to the probability of the contrary is as b is to c on the strength of the first argument alone, as e is to f on the strength of the second alone, and as h is to i on the strength of the third alone, etc. Hence it is evident enough that the total force of proof resulting from the assemblage of all the arguments should be composed of the forces of all the arguments taken singly, i.e.,, that the probability of the thing to the probability of its contrary should be in the ratio of

$beh\ldots$ to $cfi\ldots$.Hence the absolute [i.e., not relative] probability of the thing is $\frac{beh}{beh+cfi}$, and the absolute probability of the contrary is $\frac{cfi}{beh+cfi}$.

A present day argument for this result would run somewhat as follows. For a mixed type argument if in any given case the argument does not prove then the contrary is proved. Hence for multiple mixed type arguments for the same thing all arguments (in any given case) must prove, or all arguments must prove the contrary—otherwise the case leads to an impossibility. The number of cases in which all prove is (for three arguments) beh, and in which all prove the contrary is cfi. (Note the tacit use of independence.) Hence Bernoulli's result

$$\frac{beh}{beh+cfi}, \quad \frac{cfi}{beh+cfi}$$

for the probabilities for, and against, the thing being proved. These are, of course, conditional probabilities, the conditions being that all three arguments prove the thing or all three prove the contrary. In terms of the product probability space for the three sets of cases, only those ordered triples are counted in which all three cases prove, or all three cases prove the contrary, since the other triples represent impossible situations. Next Bernoulli treats the situation when there are both pure and mixed arguments for a conclusion. He supposes, as an example, that there are five arguments of which the first three are pure and the last two are mixed. From his result on combining pure arguments he had that the first three (pure) arguments provide $(adg-cfi)/adg$ of a certainty of the thing, so that only cfi/adg of the certainty remains, there being $adg-cfi$ cases that prove and cfi cases that do not prove. The two mixed arguments (taken by themselves) provide a weight of $qt/(qt+ru)$, hence resulting in an expected value of

$$\frac{(adg-cfi)\cdot 1 + cfi\cdot\frac{qt}{qt+ru}}{adg}$$

that is

$$\frac{adg-cfi}{adg}+\frac{cfi}{adg}\cdot\frac{qt}{qt+ru}, \tag{4}$$

which is the probability of the three pure arguments proving or, if not, then the two mixed proving. In addition to Bernoulli's usual tacit assumption of (stochastic) independence of the individual arguments of each kind proving, there is also the tacit use of independence of the pure arguments from the mixed ones.

A half century later Lambert (*1764*, 402) declared that this result of Bernoulli's (i.e., (4)) must be erroneous since if $q=0$, or $t=0$, (no case of a mixed argument proving, i.e., all cases proving the contrary) the formula doesn't reduce to 0, as it ought to, but to $1-cfi/adg$. R. Haussner, translator of *Ars conjectandi* into German, comes to Bernoulli's defense. He remarks, in an end-note to this passage (*Bernoulli 1899*, 153) that Lambert's objection is not admissible for the following reason. If $q=0$, or $t=0$, i.e., if the contrary is absolutely certain by virtue of a mixed argument, then the pure arguments can't change anything—are made powerless (*entkräftet*) by the mixed arguments—and therefore should not be taken into consideration. Accordingly it is the formula for combining mixed arguments alone which should be used, and this gives the value 0; when using (4) the values $q=0$ and $t=0$ must, according to Haussner, be excluded.

To discuss the issue we point out that when combining pure arguments (for the same thing) no impossible combinations of cases can arise since all cases (of any of the arguments) are either proving or non-proving, i.e., there are no cases proving the contrary. But impossible combinations do arise when combining mixed arguments. As we have seen, Bernoulli takes care of this by using, in effect, conditional probabilities. However, his (4) is defective in not taking into account the possibility of impossible combinations between cases for pure and mixed. To illustrate the matter we use a simple form of Lambert's example in which there is one argument of each kind, as shown in Figure 1.41.

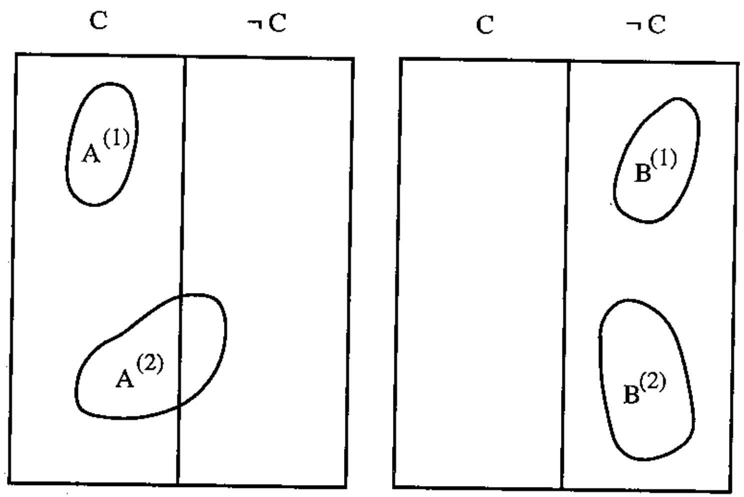

FIGURE 1.41

For this situation, with P_I and P_{II} the probability functions defined for the two (separate) arguments, Bernoulli's result gives

$$P_I(A \text{ proves } C) + P_I(A \text{ does not prove } C) \cdot P_{II}(B \text{ proves } C) = \frac{1}{2} + \frac{1}{2} \cdot 0 = \frac{1}{2}. \quad (5)$$

But in order to correctly assess the chances of both proving we have to use a probability space allowing for all possible combinations of the two sets of cases. Here there are four such combinations, namely

	$B^{(1)}$	$B^{(2)}$
$A^{(1)}$	$(1,0)$	$(1,0)$
$A^{(2)}$	$(?,0)$	$(?,0)$

where in place of having an ordered pair ($A^{(i)}$, $B^{(j)}$) we use the character '1' to indicate a proving case (based on the one argument), '0' to indicate a proving-the-contrary case, and '?' to indicate a non-proving case. For '(1, 0)' we associate no probability since it represents an impossibility; for (?, 0) we associate, in keeping with Haussner's view, the probability 0. Thus we have the probability distribution

	$B^{(1)}$	$B^{(2)}$
$A^{(1)}$	0/0	0/0
$A^{(2)}$	0	0

where '0/0' means not a possible occurrence. Letting P be its probability function we compute

$$\begin{aligned} P((A \text{ proves } C) \text{ or } (B \text{ proves } C)) \\ &= P(A \text{ proves } C) + P(A \text{ does not prove } C) \cdot P(B \text{ proves } C) \\ &= \frac{0}{2} + \frac{2}{2} \cdot \frac{0}{2} \\ &= 0, \end{aligned}$$

in agreement with Lambert and Haussner's interpretation. Bernoulli's result would be correct—as Haussner intimates—if there were no impossible combinations and if, also, the marginal probabilities were independent. This latter condition is not mentioned by either Bernoulli or Haussner.

We would be remiss in our account of Bernoulli's ideas on combining arguments if we left the impression that he was naive about the need for considering the inter-relationships of arguments. At the end of his chapter 3 he cautions against carelessly applying the rules for combining arguments without heed to the nature of the arguments. Some arguments, he says,

which appear to be different may really not be so; vice versa, arguments appearing to be different may be the same; or arguments may be such as to make the contrary impossible, and so forth. He illustrates these matters with story-like examples. Such informal cautions are needed since Bernoulli's formal theory for combining arguments does not include a representation of the logical structure of the sentences involved in premises and conclusions, and hence no way of indicating inter-relationships.

Bernoulli also discusses having pure arguments on both sides of a question, i.e., for C and for $\neg C$, remarking that the resulting probabilities on combination could for each side considerably exceed $\frac{1}{2}$. This astonishes Shafer who says (*1978*, 336): "Not only does Bernoulli allow the probability of a thing and its contrary to add to less than one, he also allows it to add to more than one!". Bernoulli's actual words are (*1713*, 210):

> Here it should be noted that if the arguments adduced on each side are strong enough, it may happen that the absolute probability of each side significantly exceeds half of certainty, i.e., that both of the contraries are rendered probable, though relatively speaking one is less probable than the other. So it is possible that one thing should have $\frac{2}{3}$ of certainty while its contrary will have $\frac{3}{4}$; in this way both contraries will be probable, yet the first less probable than its contrary, in the ratio $\frac{2}{3}$ to $\frac{3}{4}$, or 8 to 9.

It would appear from this quotation that Bernoulli is content with a situation in which 'super-additive' probabilities arise. However at the end of his chapter he presents some examples (artificially constructed) to illustrate the need for care in applying the rules. One of these contains super-additive probabilities. It is instructive to see what he says about it. We quote in full the 'story' and his discussion (*1713*, 222–23):

> With regard to a written contract doubts are raised as to whether the appended date is fraudulent (predated). An argument against this could be that the document bares the signature of a notary, i.e., an official who takes an oath of office, who would be unlikely to commit fraud since he couldn't do this without greatly endangering his honor and position; so that of 50 notaries scarcely one is to be found who would venture to be so base. And yet arguments for the affirmative [that it was fraudulent] could be that the notary has a bad reputation, that he could expect to profit greatly by fraud and, especially, that he had attested to what had no probability—as for example that someone had loaned another 10,000 gold coins at a time when by every estimation he could scarcely have had a 100 to his name. If we here consider, by itself, the argument from the office and position of

the attestor, we can estimate the probability of the authenticity of the document at $\frac{49}{50}$ of certainty. But when we evaluate the arguments for the contrary, we must grant that it could hardly be unfalsified and therefore that fraud was committed is morally certain, i.e., has $\frac{999}{1000}$ of certainty. But there is no need to conclude from this that the probability of authenticity to the probability of fraud is (by §7) in the ratio of $\frac{49}{50}$ to $\frac{999}{1000}$, i.e., that they are about equal [and both near 1]. Certainly when we suppose that the notary has a bad reputation we suppose also that he is not among the 49 righteous notaries who abhor fraud, but that he is the 50th whose conscience doesn't bother him when he is faithless in office. But then that argument which could otherwise prove the authenticity of the document loses all force and is valueless.

So Bernoulli doesn't consider super-additivity as acceptable: it arises when there is a failure to analyze data sufficiently. One shouldn't just count cases but also look at the contents of the arguments for possible interrelationships. When they exist a different set of cases could be appropriate. Expressed in present-day terms the argument for the conclusion, and that for its contrary, were based on different conditions, i.e., what is involved are the conditional probabilities

$$P(C \mid H_1) \quad \text{and} \quad P(\neg C \mid H_2);$$

and if H_1 and H_2 can differ there is no reason why the two probabilities couldn't both be near 1. But with a common condition, e.g., H_1H_2 (assuming H_1H_2 consistent) then $P(C \mid H_1H_2)+P(\neg C \mid H_1H_2) = 1$. Moreover, it is hard to reconcile acceptance of non-additivity with Bernoulli's belief, based on the law of large numbers which he was so proud of, that probabilities could be obtained *a posteriori* from frequencies. For relative frequencies for and against an event necessarily add to 1. We think a conscious use of conditional probability in non-games-of-chance situations clarifies the matter.

As a side observation it is interesting to note the two different ways in which Bernoulli estimates his probabilities. For the honesty of the notary it is statistical: "... of 50 notaries scarcely one is to be found ..."; while for the dishonesty it is moral certainty, to which (in his chapter 1, part IV) he has assigned the ratio $\frac{999}{1000}$.

§1.5. J. H. Lambert—probabilistic syllogisms

The material to be discussed in this section comes from Lambert's *Neues Organon oder Gedanken über die Erforschung und Bezeichnung des Wahren und dessen Unterscheidung von Irrtum und Schein (1764)*. It is an extensive two-volume work of which only the fifth chapter in volume 2, entitled 'Von dem Wahrscheinlichen', will be of interest to us. And even then we shall be selective, extracting from it those portions deemed to be relevant to probability logic.

It is clear that Lambert was thoroughly familiar with chapter 3, part IV of Bernoulli's *Ars conjectandi*. For example in §239 of his *Neues Organon* we find him discussing Bernoulli's result on combining pure and mixed arguments and, as we mentioned in our preceding section, disputing its correctness. The connection between logic and probability begun by Bernoulli becomes more visible and strengthened in Lambert. Where Bernoulli refers to the probability of a thing (*res*), Lambert refers to the probability of a proposition (*Satz*). Moreover, probability is a notion in its own right, not derivative from expectation. The logical structure of simple propositions is symbolically rendered, the basic form considered being 'A is B' (e.g., Man is mortal). To such propositions Lambert assigns a 'natural' degree of probability, this being the number of A's which are B's, to the total number which are or are not, as they are encountered in experience without selection. In addition to the traditional syllogistic forms Lambert also introduces—long before De Morgan did—numerically quantified forms such as '$\frac{3}{4}A$ are B' meaning that three quarters of the A's are B's. But, in addition, such sentences are endowed with a probabilistic significance which we shall be discussing presently.

Preliminary to the first of Lambert's probabilistic results there is the following description of a necessary inference form (*1764*, vol. 2, 336):

> The premises
>
> B is C, D, E, F, etc. [i.e., B is C, and B is D, and etc.]
> A is C, D, E, F, etc.
>
> do not yield 'A is B' unless the predicates C, D, E, F, etc. are, singly or in combination [*zusammengenommen*], a characterizing attribute [*eigenes Merkmal*] of B. If there is one such then, calling it M, one has an identity 'B is M' or, by simple conversion, 'M is B'. The second premise gives 'A is M', and hence the conclusion 'A is B'. For each such M there is a valid argument establishing 'A is B'.

From this necessary type of inference a transition is made to a probabilistic one (*1764*, vol. 2, 338–39):

§169. If the predicates C, D, E, F, etc. are not characterizing attributes of B, then they apply to additional subjects [beyond B]. Accordingly if one takes into account, as considered in §154, *et seq.*, the enumeration of cases in which they apply to B and in which they do not apply to B, or finds on other grounds the ratio between the two, then the degree of probability that accrues from them [towards a proof] is calculable. We content ourselves with reducing this calculation to the theory of games of chance. One imagines as many piles of tickets as there are arguments. In each pile let this number of valid or marked tickets to the number of unmarked tickets be in exactly the same ratio as that of the cases in which the argument [associated with that pile] is valid to that in which it is invalid. One supposes then that Caius blindly takes a ticket from each pile; the question is, how probable is it that among these selected tickets there is no valid one? It would be as probable or as improbable that all the arguments found on behalf of the proposition do not prove it. The theory of games of chance specifies the following rule for this calculation. *One multiplies together the number of tickets in each pile, and likewise, one multiplies together the number of non-valid or unmarked tickets in each pile; then dividing the latter product by the former yields the degree of probability that the arguments do not prove. And if this degree, which is necessarily a fraction, is subtracted from 1, then the remainder is the degree of probability that the arguments prove.*

Lambert does not explicitly say what the arguments that he is working with are, but from the context we infer that they are arguments whose conclusion is 'A is B' and whose two premises are of the form 'M is B' and 'A is M', where M is a predicate formed by combination (i.e., conjunction) from one or more of the predicates C, D, E, F, etc. and both of 'B is C, D, E, F, etc.' and 'A is C, D, E, F, etc' are necessary. Then for any M as described, 'A is M' is necessary but 'M is B' is only contingent (has a "degree of probability"). If there are n of the predicates C, D, E, F, etc. then there are $2^n - 1$ ways of forming a combination of one or more of the predicates. Not all of these lead to independent arguments—for example, if 'M_a is M_b' then 'M_b is B' implies 'M_a is B'. Lambert is aware of this and assumes that he has only independent arguments. Designating these combinations which lead to independent arguments as $M_1, \ldots, M_m$, we formulate his question as follows:

Given the independent arguments α_i, where α_i is the conjunction

$$(M_i \text{ is } B) \text{ and } (A \text{ is } M_i) \qquad (i = 1, \ldots, m) \tag{1}$$

with the second component 'A is M_i' being necessary and the first component 'M_i is B' contingent with probability

$$p_i = P(\alpha_i) = P(M_i \text{ is } B),$$

what is the probability that 'A is B' is proved by the arguments?

As each α_i logically implies 'A is B' the statement 'α_i proves (A is B)', i.e., '$\alpha_i(\alpha_i \to (A \text{ is } B))$' is equivalent to α_i. Thus to ask if an α_i proves 'A is B' is to ask if $\alpha_1 \vee \cdots \vee \alpha_m$ holds, and the probability of this is $P(\alpha_1 \vee \cdots \vee \alpha_m)$. The answer Lambert gives,

$$1 - (1 - p_1)(1 - p_2) \ldots (1 - p_m), \tag{2}$$

is indeed, by virtue of the assumed independence, equal to $P(\alpha_1 \vee \cdots \vee \alpha_m)$. (This is not the probability of 'A is B', for although $\alpha_1 \vee \cdots \vee \alpha_m$ implies 'A is B' it need not be equivalent to it—its probability is lower bound for that of 'A is B'.) Note that Lambert uses the metaphor of urns with lottery tickets to enable him to speak of the probability of compound propositions, since initially only the simple (universal) categorical proposition has a 'natural' probability assigned. To a present-day probabilist he is constructing an m-fold product space of the m independent arguments.

The inference form just discussed is contrasted with another type of probabilistic inference (*1764*, vol. 2, 355–56):

§185. On the other hand it is quite otherwise when the probability of a conclusion [*Schlußsatz*] is to be determined from the probability of the premises [*Vordersätze*]. For the premises cannot be viewed as arguments which are separate and independent of each other, since the conclusion depends on both conjointly; the conclusion then holds when all premises do. This being presupposed, calculation of the probability of the conclusion for an entire sorites [*Schlußkette*] can also equally well be reduced to the theory of games of chance. To this end we will again take piles of tickets and, indeed, as many as there are premises in the sorites. In each pile let the number of valid to nonvalid be in the same ratio as the cases in which the premises is true to those in which it is not. One then supposes that Caius blindly takes a ticket from each pile. The question is: How probable is that among the drawn tickets no nonvalid [one] occurs, or that all are valid? This is the degree of probability which the conclusion of the given sorites would have. To calculate this the theory of games gives the following rule: *One multiplies together the number of tickets in each pile; and likewise one multiplies together the number of valid tickets in each pile; the division of the latter by the former specifies the degree of probability of the conclusion.*

We note, first of all, that Lambert is considering only necessary inferences (and, of course, syllogistic ones since no others were recognized). When he refers to the (collection of) premises of the sorites we assume that he means with deletion of those which are consequences of others. The answer he gives for the probability of the conclusion, namely

$$P(\pi_1)(P(\pi_2)\dots P(\pi_m),$$

where $\pi_1, \dots, \pi_m$ are the premises in the sorites, is the probability of the conjunction of these premises only if one assumes their stochastic independence. Moreover $P(\pi_1\pi_2\dots\pi_m)$, or $P(\pi_1)P(\pi_2)\dots P(\pi_m)$ if one assumes independence, is not necessarily the probability of the conclusion since the conclusion is only implied by this conjunction, the value $P(\pi_1\pi_2\dots\pi_m)$ is just a lower bound.

It is interesting to note that Lambert believed that a contingent (probable) proposition could follow from non-contingent premises. He gives the example, for 'C ein Individuum' (*1764*, vol. 2, 359):

$$\begin{array}{rl} & \tfrac{3}{4}\,A \text{ sind } B \\ & C \text{ ist } A \\ \textit{folglich} & C\,\tfrac{3}{4} \text{ ist } B. \end{array} \qquad (3)$$

That is to say, translating his special symbolism, three quarters of the A's are B's, C is an A, therefore with probability $\frac{3}{4}$, C is a B. His justification for the inference form is the following (pp. 358–59):

> §189. One has, then, the two propositions
>
> $$\tfrac{3}{4}A \text{ are } B$$
> $$C \text{ is } A,$$
>
> and the question is: what kind of conclusion, since they have the common middle term A, can be drawn? We suppose that both propositions are true and definite; namely, that with regard to the upper one, one is assured that neither more nor less that $\frac{3}{4}$ of the A's have predicate B; and that with regard to the lower one C is an individual and that it is an A. If one knows no more than this it remains absolutely undetermined whether C is among the $\frac{3}{4}$ of the A's which are B's or among the $\frac{1}{4}A$'s which are not B's. Were this to be determined one could forthwith conclude whether C is a B or not, and there would be complete certainty. But as we are supposing that we know of C only that it is an A, we can determine no conclusion other than: it is more likely [*vermuthlicher*] that B applies to C than it doesn't. Inasmuch as among four A's there are always three which have predicate B, and since with regard to C no selection is allowed, it is then three times more likely that C is among the A's which are B's than among those which are not. Accordingly the conclusion that C is a B is not fully certain but deviates by $\frac{1}{4}$ from certainty, that is to say its probability is $\frac{3}{4}$.

To us this demonstration of Lambert involves a (tacit) assumption of a uniform probability distribution, namely the assumption that any $\frac{3}{4}$ of the A's is equally likely of being in B as any other. Then the 'classical' definition of probability applies and we may compute the probability that any preselected A (e.g., the individual C) is in B. A simple combinatorial calculation shows this to be $\frac{3}{4}$.

We contrast Lambert's 'syllogistic' inference form (3) with the following probabilistic one which we believe reproduces the essence of it:

$$\begin{array}{rr} \text{For all } x, & P(x \in B \mid x \in A) = \tfrac{3}{4} \\ & P(C \in A) = 1 \\ \text{therefore} & P(C \in B) = \tfrac{3}{4}. \end{array}$$

This inference is valid since, on instantiating in the first premise the x to C, we have

$$\frac{P(C \in B \wedge C \in A)}{P(C \in A)} = \frac{3}{4},$$

from which the conclusion follows since $P(C \in A) = 1$ and hence $P(C \in B \wedge C \in A) = P(C \in B)$.

Lambert goes on to consider probabilistic inference for a syllogism when there is a limitation on the middle term in the minor premise (*1764*, vol. 2, 360–61):

> §191. In all these cases [so far considered] the degree which probability [theory?] determines [for the conclusion] comes to the conclusion from the major premise. We will now invert the situation and show also how it stems from the minor premise. Let $MNPQ$ be the attributes of the concept B which fill up its extension [*die seinen Umfang ausfüllen*], it being left undetermined whether there is present a characterizing attribute of B. One then has the two propositions
>
> $$MNPQ \text{ is } B$$
> $$C \text{ is } MNP,$$
>
> which again allows only a probable conclusion
>
> $$C \text{ is } B,$$

since we are supposing it left undecided as to whether C also has the attribute Q. Here the degree of probability is assigned in proportion as the magnitude [*grösse*] and the number of attributes MPQ which one already has found in C is to the magnitude and number of those which are yet to be found. One sets, e.g.,

$$MNPQ = A$$
$$MNP = \tfrac{2}{3}A$$

and then has the inference

$$\begin{array}{c} \text{all } A \text{ are } B \\ C \text{ is } \tfrac{2}{3}A \\ \text{therefore} \quad C\,\tfrac{2}{3} \text{ is } B. \end{array}$$

The inference still holds [not only for singular C but also] for C general, or particular, or with a definite degree of particularity.

In our discussion of this passage we take the case of C an individual. In the first place we interpret his requirement, that the attributes M, N, P, Q fill up B's extension, to mean that the intersection of the extensions of M, N, P, Q include that of B. Hence if, as here, Lambert has '$A = MNPQ$' and 'All A are B', then A and B have the same extension. As for the minor premise 'C is $\frac{2}{3}A$', there are fewer attributes for $\frac{2}{3}A$ than for A (MNP versus $MNPQ$) so that $\frac{2}{3}A$ has a *larger extension* than A; let us designate this extension by '$(\frac{2}{3}A)_p$'. In addition to Lambert's second premise meaning '$C \in (\frac{2}{3}A)_p$' there is also the tacit assumption that an artitrarily chosen member of $(\frac{2}{3}A)_p$ has a $\frac{2}{3}$ chance of being an A, i.e., that

$$\text{For all } x,\ P(x \in A \,|\, x \in (\tfrac{2}{3}A)_p) = \tfrac{2}{3}.$$

Consequently, the premise 'C is $\frac{2}{3}A$' implies that $P(C \in A) = \frac{2}{3}$. This, together with the premise that B has the same extension as A, yields the conclusion $P(C \in B) = \frac{2}{3}$.

From this inference form Lambert goes on to the more general form in which there is restriction on the middle term in both premises. He cites as valid

$$\begin{array}{c} \tfrac{3}{4}A \text{ are } B \\ C \text{ is } \tfrac{2}{3}A \\ \text{therefore} \quad C\,\tfrac{1}{2} \text{ is } B. \end{array} \tag{4}$$

He also cites as valid a form with major premise complementary to that of (4) (*1764*, vol. 2, 362):

§193. Replacing the major premise in this inference [i.e., (4)] by its negative, one obtains

$$\begin{array}{c} \tfrac{1}{4}A \text{ are not } B \\ C \text{ is } \tfrac{2}{3}A \\ \text{therefore} \quad C\,\tfrac{1}{6} \text{ is not } B. \end{array} \tag{5}$$

Hence the probability of the denial of the conclusion [C is B] is $\frac{1}{6}$, and on the other hand that it is affirmed is $\frac{1}{2}$. Both probabilities together give $\frac{1}{6} + \frac{1}{2} = \frac{2}{3}$, which is the probability of the minor premise.

Shafer (*1978*, 355 *et seq.*), finds this (and similar results) to be evidence for Lambert's adherence to a non-additive concept of probability. Lambert does in fact assert that the sum of the probabilities for a proposition and its denial add up, in his example, to less than 1. We contend, however, that the basis on which he makes this assertion is faulty. Specifically, his inference forms (4) and (5) are invalid, requiring in place of the predicate B in the conclusion the predicate $A \cap B$, respectively, $A \cap \overline{B}$. For if we take Lambert's major premise to mean

$$\text{For all } x,\ P(x \in B \,|\, x \in A) = \tfrac{3}{4}, \tag{6}$$

so that, instantiating x to C,

$$\frac{P(C \in B \wedge C \in A)}{P(C \in A)} = \tfrac{3}{4}, \tag{7}$$

and the minor premise to mean $P(C \in A) = \frac{2}{3}$, then

$$P(C \in B \wedge C \in A) = \tfrac{2}{3} \cdot \tfrac{3}{4} = \tfrac{1}{2}. \tag{8}$$

Similarly, for the complementary inference form (5)

$$P(c \notin B \wedge C \in A) = \tfrac{1}{4} \cdot \tfrac{1}{2} = \tfrac{1}{6}. \tag{9}$$

Adding (8) and (9) gives $P(C \in A)$ with the value $\frac{1}{2} + \frac{1}{6} = \frac{2}{3}$. This is the probability of the minor premise, and our explanation accounts for the fact that the two probabilities involved do add up to this value and not to 1.

Continuing his investigations of the probabilistic syllogism Lambert introduces a generalization in which individuals can be one of three kinds: those to which a given term applies, those to which it does not, and those to which it is undetermined as to whether it does or does not apply. He illustrates the kind of inference he has in mind with the following example:

$$\begin{aligned} &(\tfrac{2}{3}a + \tfrac{1}{4}e + \tfrac{1}{12}u)A \text{ sind } B \\ &C \text{ ist } (\tfrac{3}{5}a + \tfrac{2}{5}u)\, A \qquad (10) \\ \text{folglich,}\quad &C\,(\tfrac{2}{5}a + \tfrac{3}{20}e + \tfrac{9}{20}u) \text{ ist } B. \end{aligned}$$

The major premise here has the meaning that of the totality of individuals which are A's, $\frac{2}{3}$ are certainly B's, $\frac{1}{4}$ certainly are not, and for the remaining $\frac{1}{12}$ it is indefinite, or not determined, as to whether they are, or are not, B's. In the minor premise the coefficient $(\frac{3}{5}a + \frac{2}{5}u)$ represents the attributes of B divided into those which apply to C, weighted at $\frac{3}{5}$, and a remaining part, weighted at $\frac{2}{5}$, for which it is undetermined whether they do or do not. The conclusion of (10) says "of 20 unselected cases that appear with conclusion of this kind and degree, 8 affirm, 3 deny, and 9 remain undetermined; or rather, in a single case, there are 8 grounds to affirm the conclusion, 3 to deny it, and 9 to conclude nothing, or leave it uncertain."

As with the preceding inference forms, we present a reformulation in contemporary terms. Lambert's new feature, the inclusion of indeterminacies, will be replaced by inequalities. Analogous to our expressing '$\frac{3}{4}A$ are B' by 'For all x, $P(x \in B \mid x \in A) = \frac{3}{4}$' we shall express the major premise of (10) by

$$\text{For all } x, \quad \tfrac{2}{3} \le P(x \in B \mid x \in A) \le \tfrac{9}{12}, \qquad (11)$$

from which, using

$$P(x \in B \mid x \in A) = 1 - P(x \notin B \mid x \in A)$$

and simple algebra, we have

$$\text{For all } x, \quad \tfrac{1}{4} \le P(x \notin B \mid x \in A) \le \tfrac{1}{3}. \qquad (12)$$

Then for arbitrary x the value of $P(x \in B \mid x \in A)$ is at least $\frac{2}{3}$, and might be up to $\frac{1}{12}$ more, while $P(x \notin B \mid x \in A)$ is at least $\frac{1}{4}$, and might be up to $\frac{1}{12}$ more. Thus (11) encapsulates all the information in the major premise. As for the minor premise, just as we rendered 'C is $\frac{2}{3}A$' by '$P(C \in A) = \frac{2}{3}$' so for 'C is $(\frac{3}{5}a + \frac{2}{5}u)\,A$' we shall write

$$\tfrac{3}{5} \le P(C \in A). \qquad (13)$$

To obtain the coefficients of a, e, and u in the conclusion of (10) Lambert 'multiplies' the two expressions $(\frac{2}{3}a+\frac{1}{4}e+\frac{1}{12}u)$ and $(\frac{3}{5}a+\frac{2}{5}u)$, and arranges

the result in this manner:

$$\begin{array}{l} \tfrac{2}{5}aa + \tfrac{3}{20}ae + \tfrac{3}{60}au \\ \qquad\qquad\quad + \tfrac{4}{15}au \\ \qquad\qquad\quad + \tfrac{2}{20}eu \\ \qquad\qquad\quad + \tfrac{2}{60}uu \\ \hline \tfrac{2}{5}a \;+ \tfrac{3}{20}e \;+ \tfrac{9}{20}u \end{array}$$

We compare Lambert's conclusion in (10) with what we can obtain from our reformulated version. From (11) and (12) we have, for arbitrary C,

$$\tfrac{2}{3} \cdot P(C \in A) \le P(C \in B \wedge C \in A)$$
$$\tfrac{1}{4} \cdot P(C \in A) \le P(C \notin B \wedge C \in A)$$

so that, by use of (13), we obtain

$$\begin{aligned} &\tfrac{2}{5} \le P(C \in B \wedge C \in A) \\ \text{and}\quad &\tfrac{3}{20} \le P(C \notin B \wedge C \in A), \qquad (14) \end{aligned}$$

whereas Lambert's conclusion would, in our interpretation, be

$$\begin{aligned} &\tfrac{2}{5} \le P(C \in B) \\ \text{and}\quad &\tfrac{3}{20} \le P(C \notin B). \qquad (15) \end{aligned}$$

Since (14) implies (15), we see that Lambert's conclusion, though correct, is weaker than ours. On the other hand might not Lambert's (15), or its equivalent

$$\tfrac{2}{5} \le P(C \in B) \le \tfrac{17}{20},$$

be the strongest conclusion about $P(C \in B)$ which the premises warrant? Clearly Lambert hasn't addressed himself to this type of question. We shall in our chapter 5; in particular, the result of Theorem 5.48 provides the strongest conclusion one can draw about $P(A_1)$ if given only that

$$p_1 \le P(A_1 \mid A_2) \le p_2 \quad \text{and} \quad q_1 \le P(A_2) \le q_2.$$

In our preceding section we have mentioned Lambert's objection to Bernoulli's result on combining pure and mixed arguments. Here is a further comment.

Lambert says, first of all, that all of Bernoulli's types of arguments can be encompassed in the one form

$$\text{All } A\,(Ma + Nu + Pe) \text{ are } B$$

(a = affirming, u = indefinite, e = denying). For example, if $P = 0$ then we have the pure type and if $N = 0$ then the mixed type. Combining a pure and a mixed argument by Lambert's scheme requires forming the product

$$(Ma + Nu)(ma + pe) = (Mm)aa + (Nm)au + (Np)ue + (Mp)ae,$$

concerning which Lambert says "We have, however (§237), entirely omitted the ae cases, as they are impossible, and this makes the product here given different from that of Bernoulli."

In concluding this section we mention that there is a discussion, from the viewpoint of inductive logic, of this Bernoulli-Lambert disagreement in *Hacking 1974* but with the epistemological notion of evidence in place of argument(-premise).

§1.6. Thomas Bayes—and a problem "... no less important than curious"

The phrase quoted in our title comes from a letter of Richard Price to the Secretary of the Royal Society in which he transmits an essay "found among the papers of our deceased friend Mr. Bayes". The letter and essay were read to the Society December 23, 1763. Some eleven years later Laplace's memoir of 1774 treated essentially the same problem. It is believed that Laplace was unaware of Bayes' essay (*Stigler 1978*). Over the past two centuries a vast literature has been generated by the novel idea in Bayes' essay and Laplace's memoir. We shall restrict our attention to a small aspect of it relevant to our subject.

The problem in Bayes' essay which excited Price is the following (*Bayes 1763*, 376):

PROBLEM

> *Given* the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.

There are two striking features about this problem aside from its evident relevance to statistical inference:

(i) it is concerned with the chance (= probability) *of a probability* lying between two given values, and

(ii) this chance to be determined supposing that the probability is that of an event of which *it is only known* that in $p+q$ trials it happened p times and failed q times.

Feature (i) involves the notion of a probability distribution of a probability value, a notion not within the scope of this monograph. The second feature concerns the probability of an event on given information, i.e., a conditional probability. Nothing is said about a (prior) probability distribution though, in effect, in the course of his solution Bayes attributes one to the $\frac{p}{p+q}$ values, namely that they are uniformly distributed over $[0,1]$ (see *Stigler 1986b*, 128). Before undertaking the solution of his PROBLEM Bayes develops from basic principles some results on conditional probability. One of these is an item for our historical account of probability logic.

Bayes seems to have been the first to make significant use of the multiplication rule (for the probability of the conjunction of two events) to gain information about the probability of one of the two happening when it is known that the other has happened, provided one also knows the probability of their conjunction. Here is his statement of the rule (*1763*, 378):

PROP. 3

> The probability that two subsequent events will both happen is a ratio compounded of the probability of the 1st, and the probability of the 2d on supposition that the first happens.

Bayes doesn't say what he means by a 'subsequent' event. However at the beginning of the essay he explains "An event is said to be determined when it has either happened or failed." We surmise that Bayes is using 'subsequent' to imply that the event is not yet determined, i.e., that it is a contingent or chance event.[5]

Solving in his PROP. 3 for the quantity of interest, i.e., the conditional probability, Bayes then has the corollary

> Hence if of two subsequent events the probability of the 1st be $\frac{a}{N}$, and the probability of both together $\frac{P}{N}$, then the probability the 2nd on supposition of the first happens is $\frac{P}{a}$.

We contrast this corollary with his

5. *Shafer 1982* has a different view of what Bayes means here by 'subsequent'. On the basis of this view he constructs a mathematical framework for probability which takes the timing of events into account. Such a framework, he believes, helps to explain Bayes' ideas. Shafer's paper includes a useful Appendix on the history of conditional probability (though it omits mention of C. S. Peirce's contribution. See our *1988*, 184–85).

PROP. 5

If there be two subsequent events, the probability of the 2nd being $\frac{b}{N}$ and the probability of both together $\frac{P}{N}$, and it being discovered that the second has happened, from hence I guess that the first has also happened, the probability that I am right is $\frac{P}{b}$.

This PROP. 5 appears to be a restatement of his corollary to PROP. 3, with the roles of 1st and 2nd interchanged and, as such, has puzzled some readers. (See *Dale 1991*, 33, for an account). What exactly Bayes has in mind isn't clear, but his restatement (if that is what it is) involves an epistemological element ("...it being discovered that...") which is absent from the corollary to PROP. 3. On the basis of this PROP. one is obtaining new (probability) knowledge about the happening of an event. In an explanatory footnote Price, a minute reader of the essay, remarks: "what is proved by Mr. Bayes in this and the preceding proposition [PROP. 4] is the same with the answer to the following question. What is the probability that a certain event, when it happens, will be accompanied with another to be determined at the same time?" In other words, in terms of the notion of conditional probability, not then yet codified, the question is What is $P(E_1E_2 \mid E_2)$? When Bayes' proposition gives as the answer $P(E_1E_2)/P(E_2)$ it implies that

$$P(E_1E_2 \mid E_2) = P(E_1 \mid E_2). \tag{1}$$

Historically, (1) is the earliest result in probability logic involving conditional probability. (See Theorem 5.30(e) below.)

The solution of Bayes' PROBLEM—which we shall not go into—consists in finding $P(A \mid B)$ where B is 'an event E has, in $p+q$ independent trials, happened p times and failed q times' and A is '$a \leq P(E) \leq b$' for some $[a, b]$ in $[0, 1]$. In his demonstration Bayes assumes that, to put it in modern language, the probability distribution of $P(E)$ is a uniform one. For a detailed discussion see *Dale 1991*, chapter 2.

The first problem considered in Laplace's memoir *1774* is equivalent to Bayes' PROBLEM. Here is how Laplace puts it:[6]

> If an urn contains an infinity of white and black tickets in an unknown ratio, and we draw $p+q$ tickets from it, of which p are white and q are black, then we require the probability that when we draw a new ticket from the urn, it will be white.

(Having an "infinity" of tickets serves to insure that the ratio of white to black is unchanged as the tickets are drawn.)

6. Our English translation comes from *Stigler 1986*.

Whereas Bayes uses (in effect)

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

in his demonstration and in the course of it elaborates the right-hand side, Laplace does the elaboration initially, formulating it as a general principle:

> If an event can be produced by a number n of different causes, the probabilities of these causes given the event are to each other as the probabilities of the event given the causes, and the probability of the existence of each of these is equal to the probability of the event given that cause, divided by the sum of all the probabilities of the event given each of these causes.

(Laplace's use of the term 'cause' is no longer current.) In modern notation the principle is:

$$If \qquad E \rightarrow C_1 \vee C_2 \vee \cdots \vee C_n$$

then [if the $P(C_i)$ are all equal,] for $i, j = 1, \ldots, n$,

$$\frac{P(C_i \mid E)}{P(C_j \mid E)} = \frac{P(E \mid C_i)}{P(E \mid C_j)};$$

and [if the C_i are mutually exclusive] then for $j = 1, \ldots, n$,

$$P(C_j \mid E) \left[= \frac{P(C_jE)}{P(E)}\right] = \frac{P(E \mid C_j)}{\sum_{i=1}^{n} P(E \mid C_i)},$$

a special case of the Bayes' Rule found in just about every text on probability or statistics. Its derivation from basic probability principles requires only very simple logical properties, e.g.,

$$E \leftrightarrow EC_1 \vee \cdots \vee EC_n \quad \text{if} \quad E \rightarrow C_1 \vee \cdots \vee C_n.$$

Conditional probabilities will be the focus of our attention in chapter 5.

§1.7. John Michell on the distribution of the fixed stars

Our topic for this section is a probabilistic inference inconspicuously buried in an astronomical paper, *An Inquiry into the Probable Parallax and Magnitude of the Fixed Stars, from the Quantity of Light which they afford us, and the Particular Circumstance of their Situation* (*Michell 1767*). This paper, appearing in the Transactions of the Royal Society three years after Bayes' essay, is celebrated in the history of astronomy as having provided the first realistic estimate of the distance of the fixed stars, and for establishing the existence of physical (as opposed to optical) double stars by means of a theoretical argument involving probability. The latter argument (*not* its conclusion) was criticized at great length in *Forbes 1850*. Forbes' paper, though prominently appearing in the *Philosophical Magazine* and occasioning considerable discussion, was apparently unknown to the author of the biography of Michell in the *Dictionary of Scientific Biography*, who writes:

> ... The directness of Michell's language [describing his results] leaves something to be desired; but the unimpeachable logic of his arguments gave a convincing theoretical proof of the existence of physical binary stars in the sky long before Herschel (1803) provided a compelling observational proof.

Since Michell's conclusion about binary stars turned out to be true, it could be that the author of this biography was lulled into uncritical acceptance of the argument, for Forbes' criticisms are quite telling. To examine this argument we open with a quote from Michell (*1767*, 428):

> It has always been usual with astronomers to dispose the fixed stars into constellations: this has been done for the sake of remembering and distinguishing them, and therefore it has in general been done merely arbitrarily, and with this view only; nature herself however seems to have distinguished them into groups. What he [Michell] means is, that from the apparent situation, of the stars in the heavens, there is the highest probability, that either by the original act of the Creator, or in consequence of some general law, such perhaps as gravity, they are collected together in great numbers in some parts of space, while in others there are either few or none. The argument he [Michell] intends to make use of, in order to prove this, is of the kind which infers either design, or some general law, from a general analogy, and the greatness of the odds against things having been in the present situation, if it was not owing to some such cause.

On the assumption of a random [actually, uniform] distribution of the

stars over the celestial sphere, each star being in any of the 13,131 subregions of 2° diameter with equal probability, Michell computes the odds that, of the 230 stars comparable in brightness to the double star β Capricorni, no two should fall within that angular distance. He finds the odds to be 80 to 1. When more stars are taken into account, e.g., the six brightest stars of the Pleiades, the odds against such a close grouping amount to 500,000 to 1. Since there are a large number of such groupings Michell's conclusion is that it is "next to certainty" that there is a cause for these groupings and that it is not a matter of chance. To outline the probabilistic element of Michell's argument in contemporary language, let S = present situation of the fixed stars, L = existence of some general law or original act of the Creator. Then since $P(S \mid \neg L) = 1 - P(\neg S \mid \neg L)$,

$$\begin{aligned} P(S \mid \neg L) \approx 0 &\rightarrow P(\neg S \mid \neg L) \approx 1 \\ &\rightarrow P(L \mid S) \approx 1, \end{aligned}$$

and since $P(S \mid \neg L) \approx 0$ (β Capricorni, the Pleiades, etc.), one has then the conclusion $P(L \mid S) \approx 1$. Note, in particular, the inference

$$\begin{aligned} &P(\neg S \mid \neg L) \approx 1 \\ \text{therefore} \quad &P(L \mid S) \approx 1. \end{aligned} \tag{1}$$

We shall only state Forbes' two principal objections:

(i) Michell takes the high improbability of an event's happening, when it is one of a great many possibilities, as that of the event when it is already the case. ("The improbability, for instance, of a given deal producing a given hand at whist is so immense, that were we to assume Mitchell's [sic] principle, we would be compelled to assign to it as the result of an active Cause with far more probability then even found by him for the physical connection of the six stars of the Pleiades.")

(ii) Michell's assumption of a uniform probability distribution for any star of a given magnitude and any subregion of the celestial sphere "leads to conclusions obviously at variance with the idea of random or lawless distribution, and is therefore not the expression of that Idea." Forbes likens it to assuming that any face of a die has an equal chance of coming up before one knows whether the die is loaded or not.

Boole's *1851a*, his first published paper on probability, was occasioned by the appearance of Forbes' paper. It presents a quite different objection to the Michell argument (Boole *1851a* = *1952*, 249–50):

> The proper statement of Mr. Mitchell's problem, as relates to β Capricorni, would therefore, be the following: (1.) Upon the hypothesis that a given number of stars have been distributed over the

heavens according to a law or manner whose consequences we should be altogether unable to foretell, what is the probability that such a star as β Capricorni would nowhere be found? (2.) Such a star as β Capricorni having been found, what is the probability that the law or manner of distribution was not one whose consequences we should be altogether unable to foretell? The first of the above questions certainly admits of a perfectly definite numerical answer. [Forbes denied that this was possible unless one had a specific probability distribution. Michell, in effect, does assume one.] Let the value of the probability in question be p. It has then generally been maintained that the answer to the second question is also p, and against this view Prof. Forbes justly contends. [Forbes, who is not explicit on this point, is being given more credit than is warranted.]

Boole then goes over to an abstract formulation:

Let us state Mr. Mitchell's problem, as we may now do, in the following manner: There is a calculated probability p in favour of the truth in a particular instance of the proposition. If a condition A has prevailed, a consequence B has not occurred. Required the similar probability for the proposition, if a consequence B has occurred, the condition A has not prevailed. Now, the two propositions are logically connected. The one is the "negative conversion" of the other; and hence, it either is *true* universally, the other is so. It seems hence to have been inferred, that if there is a probability p in a special instance in favour of the former, there is the same probability p in a special instance in favour of the latter. But this inference would be quite erroneous. It would be an error of the same kind as to assert that whatever probability there is that a stone arbitrarily selected is a mineral, there is the same probability that a non-mineral arbitrarily selected is a non-stone. But that these probabilities are different will be evident from their fractional expressions, which are—

1. $\dfrac{\text{Number of stones which are minerals}}{\text{Number of stones}}$

2. $\dfrac{\text{Number of non-minerals which are not stones}}{\text{Number of non-minerals}}$

It is true that if either of these fractions rises to 1, the other does also; but otherwise, they will, in general, differ in value.

Note that Boole seems to think that conditional probabilities involve conditional (i.e., if—, then—) sentences since he says "the two propositions are logically connected", one being the "negative conversion" of the other. Nevertheless, as his example shows, he does render the probabilities as conditional probabilities, not as probabilities of conditional sentences. Of special interest is his pointing out that the inference

$$\begin{aligned} &P(\neg B \mid A) = p \\ \text{therefore}\quad &P(\neg A \mid B) = p, \end{aligned} \tag{2}$$

is not valid, except when $p = 1$. Boole either didn't realize, or didn't make clear that it is only this special (valid) case which Michell's argument needs.[7] To see this note that Michell claims, in effect, that for a large number of S_i,

$$P(S_i \mid \neg L) = p_i,$$

but doesn't use 'negative conversion' on these separately but first obtains for a conjunction $S = S_1 S_2 \ldots$ of many of the S_i that, independence of the S_i being (tacitly) assumed,

$$P(S \mid \neg L) = p_1 p_2 \ldots, \tag{3}$$

Since $p_1 p_2 \cdots \approx 0$,

$$P(\neg S \mid \neg L) \approx 1,$$

and now, since $p \approx 1$, 'negative conversion' (i.e., (2)) can be used to obtain

$$P(L \mid S) \approx 1.$$

Clearly it would be worthwhile having a general method such than one could obtain relationships between probabilities such as $P(\neg B \mid A)$ and $P(\neg A \mid B)$, i.e., between probabilities whose arguments are logically related. Boole asserts (in *1851a*) that he has had such a method "for a considerable time." We shall be discussing this method in §§2.4, 2.5 below.

7. We had not appreciated this fact when writing §6.1 of our *1986*. Michell's argument lacks cogency on other grounds. See *Dale 1991*, §4.3.