Article

Quantum Theory and Probability Theory: Their Relationship and Origin in Symmetry

Philip Goyal [1,*] **and Kevin H. Knuth** [2]

[1] Department of Physics, University at Albany (SUNY), 1400 Washington Avenue, Albany,
NY 12222, USA

[2] Departments of Physics and Informatics, University at Albany (SUNY), 1400 Washington Avenue,
Albany, NY 12222, USA; E-Mail: drknuth@gmail.com

[*] Author to whom correspondence should be addressed; E-Mail: pgoyal@albany.edu;
Tel.: +1-518-442-2610.

Abstract: Quantum theory is a *probabilistic calculus* that enables the calculation of the probabilities of the possible outcomes of a measurement performed on a physical system. But what is the relationship between this probabilistic calculus and probability theory itself? Is quantum theory compatible with probability theory? If so, does it *extend* or *generalize* probability theory? In this paper, we answer these questions, and precisely determine the relationship between quantum theory and probability theory, by explicitly *deriving* both theories from first principles. In both cases, the derivation depends upon identifying and harnessing the appropriate symmetries that are operative in each domain. We prove, for example, that quantum theory is compatible with probability theory by explicitly deriving quantum theory on the assumption that probability theory is generally valid.

1. Introduction

Quantum theory is an extraordinarily successful theory which, since its creation in the mid-1920s, has provided us with a quantitatively precise understanding of a vast and an ever-growing range of

physical phenomena. In essence, quantum theory is a *probabilistic calculus* that yields a list of the probabilities of the possible outcomes of a measurement performed on a physical system prepared in some specified manner. For example, in Schroedinger's wave equation for a system of N particles, the quantum state, $\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)$, of the system determines the probability, $|\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)|^2 d^3\mathbf{r}_1 \ldots d^3\mathbf{r}_N$, that a measurement of the positions of the particles will localize them within the volume $d^3\mathbf{r}_1 \ldots d^3\mathbf{r}_N$ of configuration space located around $\mathbf{r}_1, \ldots, \mathbf{r}_N$. More generally, in the abstract quantum formalism articulated by von Neumann, the state of a system is given by a (possibly infinite dimensional) complex vector, v, while the probability that a particular measurement yields outcome i when performed on the system is given by $p_i = |\mathsf{v}_i^\dagger \mathsf{v}|^2$ (known as the Born rule), where v_i is the vector that corresponds to the ith outcome of the measurement.

The probabilistic character of quantum theory naturally raises the question: What is the relationship of quantum theory—of the quantum probabilistic calculus—to probability theory itself? Is quantum theory consistent with probability theory? Is it some kind of extension of probability theory, and, if so, what is the nature and conceptual foundation of that extension?

A significant source of difficulty in clearly answering these questions is that, apart from the notion of probability that they both utilize, probability theory and the standard von Neumann formulation of quantum theory share little in the way of language, conceptual foundations or mathematical structure. In 1948, this gap was narrowed by Feynman, who provided an alternative formulation of the standard quantum formalism [1]. Feynman's formulation strips away much of the elaborate mathematical machinery of the standard von Neumann quantum formalism, leaving behind essentially a single key idea: To each *path* that a system can take classically from some initial event, E_i, to some final event, E_f, is associated a complex number, or *amplitude*. Each such *event* is to be understood as being the outcome of a measurement performed on the system. Feynman's rules can then be stated as follows:

(a) *Amplitude Sum Rule*: If a system classically can take $n > 1$ possible paths from E_i to E_f, but the experimental apparatus does not permit one to determine which path was taken, then the total amplitude, z, for the transition from E_i to E_f is given by the *sum* of the amplitudes, z_k, associated with these paths, so that $z = \sum_{k=1}^n z_k$;

(b) *Amplitude Product Rule*: If the transition from E_i to E_f takes place via intermediate event E_m, the total amplitude, u, is given by the *product* of the amplitudes, u' and u'', for the transitions $E_i \to E_m$ and $E_m \to E_f$, respectively, so that $u = u'u''$; and

(c) *Probability-Amplitude Rule*: The *probability*, $p_{E_i \to E_f}$, of the transition from E_i to E_f is equal to the modulus-squared of the total amplitude, z, for the transition, so that $p_{E_i \to E_f} = |z|^2$.

Although these rules apply to measurements performed upon an abstract quantum system, a helpful concrete picture to have in mind when interpreting these rules is that of a particle moving in space-time from some initial point, α (corresponding to E_i), to some final point, β (corresponding to E_f). The upshot of these rules is that, if a particle classically has, say, two available paths from α to β, but the experimental apparatus used by the experimenter does not actually establish that the particle took one path or the other, one is *not* permitted to compute the transition probability, $p_{\alpha \to \beta}$, by simply summing the transition probabilities along each of these two paths. Rather, one is required to sum the *amplitudes* associated

with these two paths, and *then* compute the transition probability by taking the modulus-squared of the resultant amplitude. As a result, the quantum transition probability $p_{\alpha \to \beta}$, is *not* in general the sum of the two respective transition probabilities, but can take values less than or greater than this value, a fact often summarized by saying that it is as if the paths can "interfere" constructively or destructively with one another.

As one can see from these rules, there is a close formal parallel between Feynman's rules and the rules of probability theory. In probability theory, the probability of the joint proposition $\mathbf{A} \vee \mathbf{B}$ given proposition $\mathbf{C}$ is given by the *sum rule* $\Pr(\mathbf{A}|\mathbf{C}) + \Pr(\mathbf{B}|\mathbf{C})$ in the case where $\mathbf{A}$ and $\mathbf{B}$ are mutually exclusive propositions, while the probability of the proposition $\mathbf{A} \wedge \mathbf{B}$ for any $\mathbf{A}$, $\mathbf{B}$ given proposition $\mathbf{C}$ is given by the *product rule* $\Pr(\mathbf{A}|\mathbf{B} \wedge \mathbf{C}) \Pr(\mathbf{B}|\mathbf{C})$. Formally, these rules are closely paralleled by Feynman's amplitude sum and product rules, respectively.

However, although Feynman's formulation narrows the gulf, it does not close it. Feynman's rules are a curious admixture of the language of physics on the one hand and the event language of probability theory on the other. In particular, while the rules speak of initial and final events (language appropriate to Kolmogorov's probability theory), the notion of a classical path available to a system is a physical one and presupposes the framework of classical physics. Moreover, the manner in which Feynman's rules (particularly the amplitude sum rule) are stated suggest that probability theory is inapplicable in certain situations, and that, in those situations, one should instead use Feynman's rules. On this view, probability theory carries the imprint of the assumptions embodied in classical physics, rendering it inapplicable when one is dealing with quantum phenomena. Indeed, this view is not uncommon, and is bolstered by such commonly-used phrases as "(classical) paths can interfere constructively or destructively with one another", which we mentioned above, or Feynman's evocative image that it is as if the particle sniffs out all of the classically-available paths.

In this paper, we seek to close the gulf between probability theory and quantum theory, and thereby to precisely establish their relationship. For example, we show that the apparent inapplicability of probability theory to quantum systems is due to the failure to identify assumptions external to probability theory which are rooted in classical physics, and we prove that Feynman's rules are compatible with probability theory by explicitly *deriving* Feynman's rules *on the assumption that probability theory is generally valid.*

Our approach is inspired by Cox's pioneering derivation of probability theory [2,3]. The first modern formulation of probability theory was due to Kolmogorov in 1933 [4]. In this formulation based on set theory, propositions are represented by sets, and probabilities by measures on sets. The key components of Kolmogorov's formulation are what we recognize as the sum and product rules of probability theory stated above, which, at the time of Kolmogorov's formulation, were regarded as ultimately justified by recourse to the frequency interpretation of probability. However, in 1946, Cox showed that it was possible to *derive* these rules from much more primitive ideas, and to understand probability in a more general way. As a result of Cox's development, the probability calculus can be regarded a *systematic generalization* of the Boolean logic of propositions. In a nutshell, his line of thinking runs as follows.

In Boolean logic, existing logical propositions (well-formed statements which are objectively true or false) can be used to generate new logical propositions by using the unary negation (or complementation) operator and the binary operators AND and OR [5]. The logic is solely concerned with propositions

that are true or false, and formalizes the process of *deductive* reasoning. Cox showed that it was possible to systematically generalize Boole's logic by quantifying over the space of propositions in such a way as to remain faithful to the symmetries of the logic, thereby formalizing the process of *inductive* reasoning (that is, reasoning on the basis of incomplete information). In particular, to each pair, $\mathbf{A}, \mathbf{D}$, of propositions, he associates a real number, $p(\mathbf{A} \,|\, \mathbf{D})$, which is interpreted as quantifying the *degree* to which an agent believes proposition $\mathbf{A}$ is true given that the agent believes proposition $\mathbf{D}$ is true. Cox then requires that the quantification be consistent with the symmetries of the Boolean logic. For example, due to the *associativity* of the logical AND operator $\wedge$, for any propositions $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$, one has that $\mathbf{A} \wedge (\mathbf{B} \wedge \mathbf{C}) = (\mathbf{A} \wedge \mathbf{B}) \wedge \mathbf{C}$, which leads to the constraint $p(\mathbf{A} \wedge (\mathbf{B} \wedge \mathbf{C}) \,|\, \mathbf{D}) = p((\mathbf{A} \wedge \mathbf{B}) \wedge \mathbf{C} \,|\, \mathbf{D})$. These constraints on p yield a set of functional equations whose solution yields the standard sum and product rules of probability theory. Thus, Cox showed that probability theory can be understood as a calculus that systematically generalizes the Boolean logic of propositions, and that probability could be interpreted as an agent's degree of belief in a proposition on some given evidence. Very importantly, this view of probability recognizes that, from the outset, all probability statements are *conditional* in nature—one always speaks of the probability of a proposition *given* some other proposition—which greatly encourages explicit statement of the assumptions that, in application of Kolmogorov's formulation, are oftentimes left implicit.

Apart from the importance of Cox's work in establishing a new mathematical and conceptual foundation for probability theory, his work also offered a *methodological innovation,* namely to show how one can systematically generalize a *logic* to a *calculus* in a manner that respects the symmetries that characterize the logic. In recent years, Cox's example has been expanded into a general methodology [6] that has been used to yield insights not only into existing areas such as measure theory [7] but also to aid in the construction of new calculi, such as a calculus of questions [7,8]. It is this methodology which we employ here to derive Feynman's rules of quantum theory.

In the following, we proceed in three stages. First, defining probability as a real-valued quantification of the degree to which one logical proposition implies another, we show how probability theory can be derived as the unique calculus for manipulating these probabilities which is consistent with the underlying Boolean logic of propositions. Our presentation is based on the work of Cox [2,3], but reflects the substantial conceptual and mathematical refinement due to Knuth [8,9] and Knuth and Skilling [10]. We thereby establish that probability theory is free from physical assumptions that would be deemed objectionable from the standpoint of quantum theory.

Second, we analyze the double-slit interference experiment using probability theory. We show that the naive application of probability theory to the situation tacitly introduces an assumption rooted in classical physics. However, if the situation is analyzed carefully, taking due account of the conditional nature of probability assignments, the classical assumption is clearly visible. We show that, if this classical assumption is not made, probability theory is no longer able to provide predictions that conflict with Feynman's rules.

Third, we derive Feynman's rules of quantum theory in a manner closely analogous to our derivation of probability theory. The derivation we describe is based on that presented in [11], but offers an alternative line of argument which has the benefit of allowing us to much more clearly exhibit the relationship of Feynman's rules to the rules of probability theory. The derivation has four main phases:

1. *Operational Framework*: First, we establish a fully operational framework in which to describe measurements performed upon physical systems. The framework allows the results of an experiment to be described in purely operational terms by simply stating which sequence of measurements was performed and what were their results. In particular, any metaphysical speculations or physical pictures about how a system behaves *between* measurements (such as imagining "classical paths" of a "particle" between initial and final position measurements as envisaged in Feynman's rules) is eschewed.

2. *Experimental Logic*: Second, we identify an *experimental logic* in which *parallel* and *series* operators can be used to combine together *sequences of measurement outcomes* obtained in experiments. These measurement sequences list the outcomes obtained when a sequence of measurements are performed on a physical system. The action of applying the logical parallel and series operators allows us to formally relate the results of different experiments. The logic itself is characterized by five symmetries that are induced by the operational definition of these operators.

3. *Process Calculus*: Third, we *represent* these measurement sequences with pairs of real numbers, this choice of representation being inspired by the principle of complementarity articulated by Bohr [12]. This representation induces a *pair-valued calculus* characterized by a set of functional equations, which are then solved to yield the possible forms of the two *pair operators* which correspond to the parallel and series sequence operators.

4. *Connection with Probability Theory*: Fourth, and finally, we associate a logical proposition with each measurement sequence, and postulate that the pair associated with each sequence determines the probability of this proposition. We further require that (a) the calculus be consistent with probability theory when applied to series-combined sequences, and (b) when applied to parallel-combined sequences, the maximum and minimum values of the probabilistic predictions of the calculus are placed symmetrically about what one would predict using probability theory on the assumption that these sequences are *probabilistically independent* (an assumption which follows from classical physics). The resulting calculus—which we refer to as the *process calculus*—coincides with Feynman's rules of quantum theory.

Our derivation explicitly demonstrates that Feynman's rules are fully compatible with probability theory. In particular, a vital part of our derivation involves *requiring* that the process calculus agrees with probability theory wherever probability theory is able to yield predictions. We are thereby able to see that the role of the process calculus is to allow us to interrelate the probabilities associated with different experiments in certain situations when probability theory is by itself unable to provide any interrelation. However, the process calculus cannot be viewed as a *generalization* of probability theory as the former is *specialized* to the particular purpose of relating together the results of experiments on physical systems, and it thus concerned with *particular kinds* of propositions, while the latter is concerned with logical propositions *in general*.

The remainder of this paper is organized as follows. First, in Section 2, we give an overview of the derivation of probability theory. Then, in Section 3, we analyze the double-slit experiment using probability theory and Feynman's rules. In Section 4, we present the derivation of Feynman's rules.

Finally, in Section 5, we summarize the main findings, and, in Section 6, conclude with a summary of the key points.

2. Symmetries in Probability Theory

In this section, we review how probability theory arises as a quantification of implication amongst logical propositions constrained by the symmetries of Boolean algebra. These symmetries are outlined in Table 1.

Table 1. Boolean Algebra.

Unary Operation	
Complementation	$\text{NOT} \equiv \neg$
Complementation 1	$\mathbf{A} \wedge \neg\mathbf{A} = \bot$
Complementation 2	$\mathbf{A} \vee \neg\mathbf{A} = \top$
Idempotency	$\mathbf{A} = \neg\neg\mathbf{A}$

Binary Operations	
Disjunction	$\text{OR} \equiv \vee$
Conjunction	$\text{AND} \equiv \wedge$
Idempotency	$\mathbf{A} \vee \mathbf{A} = \mathbf{A}$
	$\mathbf{A} \wedge \mathbf{A} = \mathbf{A}$
Commutativity	$\mathbf{A} \vee \mathbf{B} = \mathbf{B} \vee \mathbf{A}$
	$\mathbf{A} \wedge \mathbf{B} = \mathbf{B} \wedge \mathbf{A}$
Associativity	$\mathbf{A} \vee (\mathbf{B} \vee \mathbf{C}) = (\mathbf{A} \vee \mathbf{B}) \vee \mathbf{C}$
	$\mathbf{A} \wedge (\mathbf{B} \wedge \mathbf{C}) = (\mathbf{A} \wedge \mathbf{B}) \wedge \mathbf{C}$
Absorption	$\mathbf{A} \vee (\mathbf{A} \wedge \mathbf{B}) = \mathbf{A} \wedge (\mathbf{A} \vee \mathbf{B}) = \mathbf{A}$
Distributivity	$\mathbf{A} \wedge (\mathbf{B} \vee \mathbf{C}) = (\mathbf{A} \wedge \mathbf{B}) \vee (\mathbf{A} \wedge \mathbf{C})$
	$\mathbf{A} \vee (\mathbf{B} \wedge \mathbf{C}) = (\mathbf{A} \vee \mathbf{B}) \wedge (\mathbf{A} \vee \mathbf{C})$
De Morgan 1	$\neg\mathbf{A} \wedge \neg\mathbf{B} = \neg(\mathbf{A} \vee \mathbf{B})$
De Morgan 2	$\neg\mathbf{A} \vee \neg\mathbf{B} = \neg(\mathbf{A} \wedge \mathbf{B})$

Consistency	
$\mathbf{A} \rightarrow \mathbf{B} \quad \Leftrightarrow \quad \mathbf{A} \wedge \mathbf{B} = \mathbf{A} \quad \Leftrightarrow \quad \mathbf{A} \vee \mathbf{B} = \mathbf{B}$	

While the first derivation relying on symmetries was originally performed by Cox [2,3], the derivation we present follows Knuth [9] and Knuth and Skilling [10], which relies on the more general class of distributive algebras (all the operations in Table 1 except those involving the unary complementation operation) and more closely mirrors the steps involved in the present derivation of Feynman's rules. These symmetries associated with the logical AND and OR operations are augmented with the symmetries associated with joining independent spaces of statements and the symmetries associated with combining inferences to obtain a probability calculus. This is accomplished by quantifying the

degree to which one logical proposition implies another by introducing a function called a *bivaluation* that takes a pair of logical propositions to a real number (scalar). A scalar representation suffices since the aim is to rank the propositions based on a single ordering relation: implication. These symmetries lead to constraint equations in the bivaluation assignments, which are the sum and product rules of probability theory.

We begin by considering two mutually exclusive propositions **A** and **B** such that $\mathbf{A} \wedge \mathbf{B} = \bot$, where $\bot$ represents the logical absurdity, which is always false. We also consider a third proposition, **Z**, such that $\mathbf{A} \wedge \mathbf{Z} \neq \bot$ and $\mathbf{B} \wedge \mathbf{Z} \neq \bot$. The degree of implication is introduced by defining a bivaluation, ϕ, that takes two propositions to a real number. For example, the quantity $\phi(\mathbf{A}|\mathbf{Z})$ represents the degree to which the proposition **Z** implies the proposition **A**. Furthermore, the bivaluation encodes the rank of the statements so that whenever $\mathbf{A} \to \mathbf{B}$ and $\mathbf{A} \neq \mathbf{B}$, we have that $\phi(\mathbf{A}|\mathbf{Z}) < \phi(\mathbf{B}|\mathbf{Z})$.

We now consider the composite proposition $\mathbf{A} \vee \mathbf{B}$ and the degree to which it is implied by **Z**, $\phi(\mathbf{A} \vee \mathbf{B}|\mathbf{Z})$. For this calculus to encode the underlying algebra, the degree $\phi(\mathbf{A} \vee \mathbf{B}|\mathbf{Z})$ must be a function of the degrees $\phi(\mathbf{A}|\mathbf{Z})$ and $\phi(\mathbf{B}|\mathbf{Z})$, so that

$$\phi(\mathbf{A} \vee \mathbf{B}|\mathbf{Z}) = \phi(\mathbf{A}|\mathbf{Z}) \oplus \phi(\mathbf{B}|\mathbf{Z}) \tag{1}$$

where $\oplus$ is a binary operator to be determined.

Consider another logical proposition **C** where $\mathbf{A} \wedge \mathbf{C} = \bot$, $\mathbf{B} \wedge \mathbf{C} = \bot$, $\mathbf{C} \wedge \mathbf{Z} \neq \bot$, and form the element $(\mathbf{A} \vee \mathbf{B}) \vee \mathbf{C}$. We can use associativity of $\vee$ to write this element two ways

$$(\mathbf{A} \vee \mathbf{B}) \vee \mathbf{C} = \mathbf{A} \vee (\mathbf{B} \vee \mathbf{C}) \tag{2}$$

Applying Equation (1), we obtain

$$\phi(\mathbf{A} \vee \mathbf{B}|\mathbf{Z}) \oplus \phi(\mathbf{C}|\mathbf{Z}) = \phi(\mathbf{A}|\mathbf{Z}) \oplus \phi(\mathbf{B} \vee \mathbf{C}|\mathbf{Z}) \tag{3}$$

Applying Equation (1) again to the arguments $\phi(\mathbf{A} \vee \mathbf{B}|\mathbf{Z})$ and $\phi(\mathbf{B} \vee \mathbf{C}|\mathbf{Z})$ above, we get

$$(\phi(\mathbf{A}|\mathbf{Z}) \oplus \phi(\mathbf{B}|\mathbf{Z})) \oplus \phi(\mathbf{C}|\mathbf{Z}) = \phi(\mathbf{A}|\mathbf{Z}) \oplus (\phi(\mathbf{B}|\mathbf{Z}) \oplus \phi(\mathbf{C}|\mathbf{Z})) \tag{4}$$

Defining $u = \phi(\mathbf{A}|\mathbf{Z})$, $v = \phi(\mathbf{B}|\mathbf{Z})$, and $w = \phi(\mathbf{C}|\mathbf{Z})$, the above expression can be written as the functional equation

$$(u \oplus v) \oplus w = u \oplus (v \oplus w) \tag{5}$$

known as the *Associativity Equation* [13], whose general solution [13] is

$$u \oplus v = f^{-1}\left(f(u) + f(v)\right) \tag{6}$$

where f is an arbitrary invertible function. This means that there exists a function $f : \mathbb{R} \to \mathbb{R}$ that re-maps u and v to a representation in which

$$f(u \oplus v) = f(u) + f(v) \tag{7}$$

Writing this in terms of the original expressions, and defining $\mathrm{Pr} \equiv f \circ \phi$ we have

$$\mathrm{Pr}(\mathbf{A} \vee \mathbf{B} \,|\, \mathbf{Z}) = \mathrm{Pr}(\mathbf{A} \,|\, \mathbf{Z}) + \mathrm{Pr}(\mathbf{B} \,|\, \mathbf{Z}). \tag{8}$$

We can consider more general propositions $\mathbf{X}$ and $\mathbf{Y}$ formed from mutually disjoint propositions $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ by $\mathbf{X} = \mathbf{A} \vee \mathbf{B}$ and $\mathbf{Y} = \mathbf{B} \vee \mathbf{C}$. Since $\mathbf{A}$ and $\mathbf{B}$ are disjoint

$$\Pr(\mathbf{A} \vee \mathbf{B} \,|\, \mathbf{Z}) = \Pr(\mathbf{A} \,|\, \mathbf{Z}) + \Pr(\mathbf{B} \,|\, \mathbf{Z}) \tag{9}$$

Similarly, since $\mathbf{A}$ and $\mathbf{Y}$ are disjoint

$$\Pr(\mathbf{A} \vee \mathbf{Y} \,|\, \mathbf{Z}) = \Pr(\mathbf{A} \,|\, \mathbf{Z}) + \Pr(\mathbf{Y} \,|\, \mathbf{Z}) \tag{10}$$

Solving both equations for $\Pr(\mathbf{A} \,|\, \mathbf{Z})$, we find

$$\Pr(\mathbf{A} \vee \mathbf{Y} \,|\, \mathbf{Z}) - \Pr(\mathbf{Y} \,|\, \mathbf{Z}) = \Pr(\mathbf{A} \vee \mathbf{B} \,|\, \mathbf{Z}) - \Pr(\mathbf{B} \,|\, \mathbf{Z}) \tag{11}$$

Noting that $\mathbf{X} \vee \mathbf{Y} = \mathbf{A} \vee \mathbf{Y}$ and $\mathbf{X} \wedge \mathbf{Y} = \mathbf{B}$, and recalling that $\mathbf{X} = \mathbf{A} \vee \mathbf{B}$, we have

$$\Pr(\mathbf{X} \vee \mathbf{Y} \,|\, \mathbf{Z}) - \Pr(\mathbf{Y} \,|\, \mathbf{Z}) = \Pr(\mathbf{X} \,|\, \mathbf{Z}) - \Pr(\mathbf{X} \wedge \mathbf{Y} \,|\, \mathbf{Z}) \tag{12}$$

which is the familiar *sum rule*, or inclusion-exclusion relation

$$\Pr(\mathbf{X} \vee \mathbf{Y} \,|\, \mathbf{Z}) = \Pr(\mathbf{Y} \,|\, \mathbf{Z}) + \Pr(\mathbf{X} \,|\, \mathbf{Z}) - \Pr(\mathbf{X} \wedge \mathbf{Y} \,|\, \mathbf{Z}) \tag{13}$$

It should be noted that this result is symmetric with respect to interchange of the logical AND and OR operation. The operations are dual to one another and associativity of one operation guarantees associativity of the other. The result is that the symmetry of associativity of both the logical OR and logical AND operations results in a constraint equation for the bivaluation, which is the sum rule of probability theory. More generally, the sum rule ensures the symmetry of associativity of the binary operations of the distributive algebra [7].

Another important symmetry represents the fact that one should be able to append statements that have absolutely nothing to do with the problem at hand without affecting one's inferences. For example, the inferential relationship between the statement "it is cloudy" and the statement "it is raining" cannot be affected if we append the fact that "eggplants are purple" to each. More specifically, the degree to which the statement $\mathbf{C} = $ "it is cloudy" implies the statement $\mathbf{R} = $ "it is raining", $\Pr(\mathbf{R} \,|\, \mathbf{C})$, should be equal to the degree to which the statement $(\mathbf{C}, \mathbf{E}) = $ "it is cloudy, and eggplants are purple" implies the statement $(\mathbf{R}, \mathbf{E}) = $ "it is raining, and eggplants are purple", denoted $\Pr\big((\mathbf{R}, \mathbf{E}) \,|\, (\mathbf{C}, \mathbf{E})\big)$.

Consider the direct (Cartesian) product of two conceptually independent spaces of logical propositions. The direct product is associative since

$$(\mathbf{A}, (\mathbf{B}, \mathbf{C})) = ((\mathbf{A}, \mathbf{B}), \mathbf{C}) \tag{14}$$

which allows us to drop the internal parentheses and write $(\mathbf{A}, \mathbf{B}, \mathbf{C})$. Furthermore, it is distributive over logical OR

$$(\mathbf{A_1}, \mathbf{B}) \vee (\mathbf{A_2}, \mathbf{B}) = ((\mathbf{A_1} \vee \mathbf{A_2}), \mathbf{B}) \tag{15}$$

Given bivaluations $\Pr(\mathbf{A} \,|\, \mathbf{Z_A})$ and $\Pr(\mathbf{B} \,|\, \mathbf{Z_B})$ defined in the two independent spaces, we aim to define the bivaluation $\Pr\big((\mathbf{A}, \mathbf{B}) \,|\, (\mathbf{Z_A}, \mathbf{Z_B})\big)$ in the joint space. These bivaluations must be consistent with the symmetries associated with combining the two spaces above. In addition, the bivaluations

defined on the joint space must obey associativity of the direct product, which as we saw above requires that the bivaluations must be additive

$$g\Big(\operatorname{Pr}\big((\mathbf{A},\mathbf{B})\,|\,(\mathbf{Z_A},\mathbf{Z_B})\big)\Big) = g\Big(\operatorname{Pr}(\mathbf{A}\,|\,\mathbf{Z_A})\Big) + g\Big(\operatorname{Pr}(\mathbf{B}\,|\,\mathbf{Z_B})\Big) \tag{16}$$

where g is some invertible function.

In the case where $\mathbf{A_1}$ and $\mathbf{A_2}$ are mutually exclusive, the joint propositions $(\mathbf{A_1},\mathbf{B})$ and $(\mathbf{A_2},\mathbf{B})$ are also mutually exclusive, and we can write the bivaluation quantifying the degree to which the joint proposition $(\mathbf{Z_A},\mathbf{Z_B})$ implies the joint proposition $(\mathbf{A_1}\vee\mathbf{A_2},\mathbf{B})$ as

$$\operatorname{Pr}\Big((\mathbf{A_1}\vee\mathbf{A_2},\mathbf{B})\,|\,(\mathbf{Z_A},\mathbf{Z_B})\Big) = \operatorname{Pr}\Big((\mathbf{A_1},\mathbf{B})\,|\,(\mathbf{Z_A},\mathbf{Z_B})\Big) + \operatorname{Pr}\Big((\mathbf{A_2},\mathbf{B})\,|\,(\mathbf{Z_A},\mathbf{Z_B})\Big) \tag{17}$$

Additivity requires that

$$\operatorname{Pr}\Big((\mathbf{A_1},\mathbf{B})\,|\,(\mathbf{Z_A},\mathbf{Z_B})\Big) = g^{-1}\Big(g\Big(\operatorname{Pr}(\mathbf{A_1}\,|\,\mathbf{Z_A})\Big) + g\Big(\operatorname{Pr}(\mathbf{B}\,|\,\mathbf{Z_B})\Big)\Big)$$

$$\operatorname{Pr}\Big((\mathbf{A_2},\mathbf{B})\,|\,(\mathbf{Z_A},\mathbf{Z_B})\Big) = g^{-1}\Big(g\Big(\operatorname{Pr}(\mathbf{A_2}\,|\,\mathbf{Z_A})\Big) + g\Big(\operatorname{Pr}(\mathbf{B}\,|\,\mathbf{Z_B})\Big)\Big)$$

$$\operatorname{Pr}\Big((\mathbf{A_1}\vee\mathbf{A_2},\mathbf{B})\,|\,(\mathbf{Z_A},\mathbf{Z_B})\Big) = g^{-1}\Big(g\Big(\operatorname{Pr}(\mathbf{A_1}\vee\mathbf{A_2}\,|\,\mathbf{Z_A})\Big) + g\Big(\operatorname{Pr}(\mathbf{B}\,|\,\mathbf{Z_B})\Big)\Big)$$

Writing

$$x = g\Big(\operatorname{Pr}(\mathbf{A_1}\,|\,\mathbf{Z_A})\Big)$$
$$y = g\Big(\operatorname{Pr}(\mathbf{A_2}\,|\,\mathbf{Z_A})\Big)$$
$$z = g\Big(\operatorname{Pr}(\mathbf{B}\,|\,\mathbf{Z_B})\Big)$$
$$k(x,y) = g\Big(\operatorname{Pr}(\mathbf{A_1}\vee\mathbf{A_2}\,|\,\mathbf{Z_A})\Big)$$

and writing $h = g^{-1}$, we have the *product equation*

$$h(k(x,y) + z) = h(x + z) + h(y + z) \tag{18}$$

which encodes the symmetry that the direct product is distributive over the logical OR. Solving this functional equation one finds that g is the logarithm function [10], so that

$$\operatorname{Pr}\Big((\mathbf{A},\mathbf{B})\,|\,(\mathbf{Z_A},\mathbf{Z_B})\Big) = C\operatorname{Pr}(\mathbf{A}\,|\,\mathbf{Z_A})\operatorname{Pr}(\mathbf{B}\,|\,\mathbf{Z_B}) \tag{19}$$

where C is an arbitrary positive constant, which can be set equal to unity without loss of generality. This results in the *direct product rule*

$$\operatorname{Pr}\Big((\mathbf{A},\mathbf{B})\,|\,(\mathbf{Z_A},\mathbf{Z_B})\Big) = \operatorname{Pr}(\mathbf{A}\,|\,\mathbf{Z_A})\operatorname{Pr}(\mathbf{B}\,|\,\mathbf{Z_B}) \tag{20}$$

Last, we consider the symmetry involved in chaining our inferences together. We consider three logical statements where $\mathbf{X}\to\mathbf{Y}$ and $\mathbf{Y}\to\mathbf{Z}$, so that by transitivity $\mathbf{X}\to\mathbf{Z}$. This symmetry can be viewed in terms of the logical AND operation, or dually the logical OR operation, since $\mathbf{X}\to\mathbf{Z}$ implies that $\mathbf{X}\wedge\mathbf{Z}=\mathbf{X}$ and $\mathbf{X}\vee\mathbf{Z}=\mathbf{Z}$ as listed in Table 1. Transitivity of implication can be viewed as associativity of the logical AND when applied to the implicate since

$$((\mathbf{X}\wedge\mathbf{Y})\wedge\mathbf{Z}) = (\mathbf{X}\wedge(\mathbf{Y}\wedge\mathbf{Z})) \tag{21}$$

which, in the case where $\mathbf{X} \to \mathbf{Y}$ and $\mathbf{Y} \to \mathbf{Z}$, can be simplified to

$$
\begin{align}
(\mathbf{X} \wedge \mathbf{Z}) &= (\mathbf{X} \wedge \mathbf{Y}) \tag{22} \\
&= \mathbf{X} \tag{23}
\end{align}
$$

which implies that $\mathbf{X} \to \mathbf{Z}$. The transitivity implies that, when $\mathbf{X} \to \mathbf{Y} \to \mathbf{Z}$, the bivaluation $\mathrm{Pr}(\mathbf{X} \,|\, \mathbf{Z})$ must be a function of $\mathrm{Pr}(\mathbf{X} \,|\, \mathbf{Y})$ and $\mathrm{Pr}(\mathbf{Y} \,|\, \mathbf{Z})$.

Since this functional relationship holds in general, it must hold in special cases. We consider a special case that completely constrains this relationship, which therefore results in the general solution. Consider the following relations obtained by applying the direct product rule

$$
\begin{align}
\mathrm{Pr}\Big((\mathbf{A_1} \vee \mathbf{A_2}, \mathbf{B_1}) \,|\, (\mathbf{A_1} \vee \mathbf{A_2}, \mathbf{B_1} \vee \mathbf{B_2})\Big) &= \mathrm{Pr}(\mathbf{A_1} \vee \mathbf{A_2} \,|\, \mathbf{A_1} \vee \mathbf{A_2}) \, \mathrm{Pr}(\mathbf{B_1} \,|\, \mathbf{B_1} \vee \mathbf{B_2}) \\
&= \mathrm{Pr}(\mathbf{B_1} \,|\, \mathbf{B_1} \vee \mathbf{B_2}) \tag{24}
\end{align}
$$

$$
\begin{align}
\mathrm{Pr}\Big((\mathbf{A_1}, \mathbf{B_1}) \,|\, (\mathbf{A_1} \vee \mathbf{A_2}, \mathbf{B_1})\Big) &= \mathrm{Pr}(\mathbf{A_1} \,|\, \mathbf{A_1} \vee \mathbf{A_2}) \, \mathrm{Pr}(\mathbf{B_1} \,|\, \mathbf{B_1}) \\
&= \mathrm{Pr}(\mathbf{A_1} \,|\, \mathbf{A_1} \vee \mathbf{A_2}) \tag{25}
\end{align}
$$

and

$$
\mathrm{Pr}\Big((\mathbf{A_1}, \mathbf{B_1}) \,|\, (\mathbf{A_1} \vee \mathbf{A_2}, \mathbf{B_1} \vee \mathbf{B_2})\Big) = \mathrm{Pr}(\mathbf{A_1} \,|\, \mathbf{A_1} \vee \mathbf{A_2}) \, \mathrm{Pr}(\mathbf{B_1} \,|\, \mathbf{B_1} \vee \mathbf{B_2}) \tag{26}
$$

Setting $\mathbf{X} = (\mathbf{A_1}, \mathbf{B_1})$, $\mathbf{Y} = (\mathbf{A_1} \vee \mathbf{A_2}, \mathbf{B_1})$, $\mathbf{Z} = (\mathbf{A_1} \vee \mathbf{A_2}, \mathbf{B_1} \vee \mathbf{B_2})$ so that $\mathbf{X} \to \mathbf{Y} \to \mathbf{Z}$ and substituting Equation (24) and Equation (25) into Equation (26) we find

$$
\mathrm{Pr}(\mathbf{X} \,|\, \mathbf{Z}) = \mathrm{Pr}(\mathbf{X} \,|\, \mathbf{Y}) \, \mathrm{Pr}(\mathbf{Y} \,|\, \mathbf{Z}) \tag{27}
$$

This is the *chain product rule*.

This can be extended to accommodate arbitrary statements $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$. Consider the following application of the sum rule where the implicate is $\mathbf{X}$

$$
\mathrm{Pr}(\mathbf{X} \,|\, \mathbf{X}) + \mathrm{Pr}(\mathbf{Y} \,|\, \mathbf{X}) = \mathrm{Pr}(\mathbf{X} \wedge \mathbf{Y} \,|\, \mathbf{X}) + \mathrm{Pr}(\mathbf{X} \vee \mathbf{Y} \,|\, \mathbf{X}) \tag{28}
$$

Since $\mathbf{X} \to \mathbf{X} \vee \mathbf{Y}$, we have that $\mathrm{Pr}(\mathbf{X} \,|\, \mathbf{X}) = \mathrm{Pr}(\mathbf{X} \vee \mathbf{Y} \,|\, \mathbf{X})$, which implies that

$$
\mathrm{Pr}(\mathbf{Y} \,|\, \mathbf{X}) = \mathrm{Pr}(\mathbf{X} \wedge \mathbf{Y} \,|\, \mathbf{X}) \tag{29}
$$

Setting $\mathbf{X} = \mathbf{A} \wedge \mathbf{B} \wedge \mathbf{Z}$, $\mathbf{Y} = \mathbf{B} \wedge \mathbf{Z}$ in Equation (27) and applying the identity above, we have the general form of the *product rule* for probability,

$$
\mathrm{Pr}(\mathbf{A} \wedge \mathbf{B} \,|\, \mathbf{Z}) = \mathrm{Pr}(\mathbf{A} \,|\, \mathbf{B} \wedge \mathbf{Z}) \, \mathrm{Pr}(\mathbf{B} \,|\, \mathbf{Z}) \tag{30}
$$

The result of the considerations above is that the symmetries of Boolean algebra place constraints on bivaluation assignments used to quantify degrees of implication among logical statements. These constraints take the form of constraint equations, which are the familiar sum and product rules of probability theory

$$
\mathrm{Pr}(\mathbf{X} \vee \mathbf{Y} \,|\, \mathbf{Z}) = \mathrm{Pr}(\mathbf{Y} \,|\, \mathbf{Z}) + \mathrm{Pr}(\mathbf{X} \,|\, \mathbf{Z}) - \mathrm{Pr}(\mathbf{X} \wedge \mathbf{Y} \,|\, \mathbf{Z}) \qquad \text{Sum Rule} \tag{31}
$$

$$
\mathrm{Pr}(\mathbf{A} \wedge \mathbf{B} \,|\, \mathbf{Z}) = \mathrm{Pr}(\mathbf{A} \,|\, \mathbf{B} \wedge \mathbf{Z}) \, \mathrm{Pr}(\mathbf{B} \,|\, \mathbf{Z}) \qquad \text{Product Rule} \tag{32}
$$

$$
\mathrm{Pr}\Big((\mathbf{A}, \mathbf{B}) \,|\, (\mathbf{Z_A}, \mathbf{Z_B})\Big) = \mathrm{Pr}(\mathbf{A} \,|\, \mathbf{Z_A}) \, \mathrm{Pr}(\mathbf{B} \,|\, \mathbf{Z_B}) \qquad \text{Direct Product Rule} \tag{33}
$$

The remainder of this section focuses on demonstrating how the rules above result in bivaluations that agree with the standard definitions and results for probability measures, such as in Kolmogorov [4]. Specifically, we will use the sum and product rules to derive the range of values that the bivaluations may take, and show that, in the extreme cases of falsity and truth, the bivaluations take values of zero and unity, respectively; and in general range from zero to unity. We will also show that the definition of a bivaluation leads directly to the standard definition of conditional probability. These results are derived below in a sequence that lends itself to an efficient derivation, but will be summarized at the end of the section in a logically-unified fashion.

The falsity, or absurdity, $\perp$, is defined by taking the logical conjunction (AND) of two mutually exclusive (disjoint) statements $\mathbf{X} \wedge \mathbf{Y} = \perp$. As such, it is always false. Previously, we found that for any two disjoint statements, $\Pr(\mathbf{X} \vee \mathbf{Y} \mid \mathbf{Z}) = \Pr(\mathbf{X} \mid \mathbf{Z}) + \Pr(\mathbf{Y} \mid \mathbf{Z})$. But the sum rule above applies to all logical statements, so we can write

$$
\begin{aligned}
\Pr(\perp \mid \mathbf{Z}) = \Pr(\mathbf{X} \wedge \mathbf{Y} \mid \mathbf{Z}) &= \Pr(\mathbf{X} \mid \mathbf{Z}) + \Pr(\mathbf{Y} \mid \mathbf{Z}) - \Pr(\mathbf{X} \vee \mathbf{Y} \mid \mathbf{Z}) \\
&= \Pr(\mathbf{X} \mid \mathbf{Z}) + \Pr(\mathbf{Y} \mid \mathbf{Z}) - (\Pr(\mathbf{X} \mid \mathbf{Z}) + \Pr(\mathbf{Y} \mid \mathbf{Z})) \\
&= 0
\end{aligned}
$$

so that, for all $\mathbf{Z}$

$$
\Pr(\perp \mid \mathbf{Z}) = 0 \tag{34}
$$

Of particular note is the fact that $\Pr(\perp \mid \perp) = 0$.

All statements, excluding the bottom, imply themselves with certainty. To see this, set $\mathbf{A} = \mathbf{B} = \mathbf{C}$ in the product rule

$$
\Pr(\mathbf{A} \wedge \mathbf{A} \mid \mathbf{A}) = \Pr(\mathbf{A} \mid \mathbf{A} \wedge \mathbf{A}) \Pr(\mathbf{A} \mid \mathbf{A}) \tag{35}
$$

so that

$$
\Pr(\mathbf{A} \mid \mathbf{A}) = \Pr(\mathbf{A} \mid \mathbf{A}) \Pr(\mathbf{A} \mid \mathbf{A}) \tag{36}
$$

which implies that either $\Pr(\mathbf{A} \mid \mathbf{A}) = 0$ or that $\Pr(\mathbf{A} \mid \mathbf{A}) = 1$. We first consider the possibility that $\Pr(\mathbf{A} \mid \mathbf{A}) = 0$ and consider the special case of the degree to which the truism implies itself, $\Pr(\top \mid \top)$. As described above, the bivaluation is meant to encode the rank of the statements so that whenever $\mathbf{A} \to \mathbf{B}$ and $\mathbf{A} \neq \mathbf{B}$ we have that $\Pr(\mathbf{A} \mid \mathbf{Z}) < \Pr(\mathbf{B} \mid \mathbf{Z})$. If we let $A = \perp$ and $B = Z = \top$, we have that $\Pr(\perp \mid \top) < \Pr(\top \mid \top)$, which implies that $\Pr(\top \mid \top) > 0$. Therefore, it must be that the solution to Equation (36) is $\Pr(\mathbf{A} \mid \mathbf{A}) = 1$. This numerically encodes the fact that all non-absurd statements imply themselves with certainty. This particular value of unity is determined by our choice of setting the arbitrary constant $C = 1$ in Equation (19), which, as we will show, defines the scale for the maximum value that can be assigned to the bivaluation.

The result above suggests that for all statements $\mathbf{X} \to \mathbf{Y}$ we should have that $\Pr(\mathbf{Y} \mid \mathbf{X}) = 1$. This can be shown by noting that $\mathbf{X} \wedge \mathbf{Y} = \mathbf{X}$ when $\mathbf{X} \to \mathbf{Y}$. This enables one to write the product rule $\Pr(\mathbf{X} \wedge \mathbf{Y} \mid \mathbf{X}) = \Pr(\mathbf{X} \mid \mathbf{X} \wedge \mathbf{Y}) \Pr(\mathbf{Y} \mid \mathbf{X})$ as

$$
\Pr(\mathbf{X} \mid \mathbf{X}) = \Pr(\mathbf{X} \mid \mathbf{X}) \Pr(\mathbf{Y} \mid \mathbf{X}) \tag{37}
$$

which implies that $\Pr(\mathbf{Y} \mid \mathbf{X}) = 1$ whenever $\mathbf{X} \to \mathbf{Y}$ and $\mathbf{X} \neq \perp$. Since, the truism, $\top$, which is the top element of the lattice, trivially satisfies $\mathbf{X} \to \top$ for all $\mathbf{X}$, we have that $\Pr(\top \mid \mathbf{X}) = 1$ for all $\mathbf{X}$.

We have examined the extreme cases, and have seen that, for all $\mathbf{Z}$, $\Pr(\perp \mid \mathbf{Z}) = 0$ and $\Pr(\top \mid \mathbf{Z}) = 1$, while $\Pr(\mathbf{Y} \mid \mathbf{X}) = 1$ whenever $\mathbf{X} \to \mathbf{Y}$. We now consider the intermediate cases. By considering non-disjoint $\mathbf{X}$ and $\mathbf{Y}$ so that $\mathbf{X} \wedge \mathbf{Y} \neq \perp$, Boolean logic implies that $\perp \to (\mathbf{X} \wedge \mathbf{Y}) \to \mathbf{Y}$. In the case where $\mathbf{Y}$ does not imply $\mathbf{X}$ so that $\mathbf{X} \wedge \mathbf{Y} \neq \mathbf{Y}$, the fact that bivaluations rank the statements implies that

$$\Pr(\perp \mid \top) < \Pr(\mathbf{X} \wedge \mathbf{Y} \mid \top) < \Pr(\mathbf{Y} \mid \top) \tag{38}$$

which can be simplified to

$$0 < \Pr(\mathbf{X} \wedge \mathbf{Y} \mid \top) < \Pr(\mathbf{Y} \mid \top) \tag{39}$$

We can apply the product rule to $\Pr(\mathbf{X} \wedge \mathbf{Y} \mid \top)$ and write

$$0 < \Pr(\mathbf{X} \mid \mathbf{Y} \wedge \top) \Pr(\mathbf{Y} \mid \top) < \Pr(\mathbf{Y} \mid \top) \tag{40}$$

Since $\Pr(\mathbf{X} \mid \mathbf{Y} \wedge \top) > 0$, we have that $\Pr(\mathbf{Y} \mid \top) > 0$. Dividing through by $\Pr(\mathbf{Y} \mid \top)$, and recognizing that $\mathbf{Y} = \mathbf{Y} \wedge \top$

$$0 < \Pr(\mathbf{X} \mid \mathbf{Y}) < 1 \tag{41}$$

indicating the normalized range of values over which the bivaluation $\Pr(\cdot)$ is defined.

The definition of bivaluation suggests that the $\Pr(\mathbf{X} \mid \mathbf{Y})$ is to be identified with the conditional probability of $\mathbf{X}$ given $\mathbf{Y}$. We now show that this is indeed the case. We consider the product rule

$$\Pr(\mathbf{X} \wedge \mathbf{Y} \mid \top) = \Pr(\mathbf{X} \mid \mathbf{Y} \wedge \top) \Pr(\mathbf{Y} \mid \top) \tag{42}$$

for the case where $\mathbf{Y} \neq \perp$. Recalling that $\mathbf{Y} = \mathbf{Y} \wedge \top$, we can divide by $\Pr(\mathbf{Y} \mid \top)$ and solve for $\Pr(\mathbf{X} \mid \mathbf{Y})$ to obtain

$$\Pr(\mathbf{X} \mid \mathbf{Y}) = \frac{\Pr(\mathbf{X} \wedge \mathbf{Y} \mid \top)}{\Pr(\mathbf{Y} \mid \top)} \tag{43}$$

which is Kolmogorov's definition of conditional probability, arrived at here from an entirely different perspective. As a final check, by considering the case where $\mathbf{X} \wedge \mathbf{Y} = \perp$, we have that $\Pr(\mathbf{X} \wedge \mathbf{Y} \mid \top) = \Pr(\perp \mid \top) = 0$, which, by Equation (43), implies that $\Pr(\mathbf{X} \mid \mathbf{Y}) = 0$ whenever $\mathbf{X} \wedge \mathbf{Y} = \perp$.

In summary, we have shown that bivaluations fall within the following range

$$0 \leq \Pr(\mathbf{X} \mid \mathbf{Y}) \leq 1 \tag{44}$$

where, in the extreme cases of truth and falsity, we have that, for all $\mathbf{Y}$

$$\Pr(\top \mid \mathbf{Y}) = 1 \tag{45}$$
$$\Pr(\perp \mid \mathbf{Y}) = 0 \tag{46}$$

In general, for $\mathbf{X}, \mathbf{Y} \neq \perp$, we have shown that bivaluations representing degrees of implication obey [14]

$$\Pr(\mathbf{X} \mid \mathbf{Y}) = 1 \quad \text{when} \quad \mathbf{Y} \to \mathbf{X} \tag{47}$$
$$\Pr(\mathbf{X} \mid \mathbf{Y}) = 0 \quad \text{when} \quad \mathbf{X} \wedge \mathbf{Y} = \perp \tag{48}$$
$$0 < \Pr(\mathbf{X} \mid \mathbf{Y}) < 1 \qquad \text{otherwise} \tag{49}$$

These results for the values of the bivaluation $\Pr(\cdot)$, together with the sum and product rules above, encompass standard probability theory [2–4]. The theory describes how to reason consistently in the sense of agreement with Boolean logic.

We observe that the constraint equations do not totally constrain the bivaluation (probability) assignments. In particular, consider the set of *atomic* propositions, $\mathbf{A}_1, \ldots, \mathbf{A}_N$, namely those propositions which cannot be obtained via the union of other propositions, and write the truism as $\top = \mathbf{A}_1 \vee \mathbf{A}_2 \cdots \vee \mathbf{A}_N$. Then the N probabilities $\Pr(\mathbf{A}_i | \top)$ are freely assignable up to the normalization condition $\sum_i \Pr(\mathbf{A}_i | \top) = \Pr(\top | \top) = 1$. This means that there is freedom for the probability assignments to be problem-dependent. We will see that, in quantum theory, the set of constraints on the quantification of logical statements identified above are coupled to a set of constraints governing the quantification of measurement sequences, so that the probabilities of statements about a quantum system will depend on the quantum amplitudes.

3. Feynman's Rules of Quantum Theory

The first general quantum theories of matter were formulated in 1925–1926 by Schroedinger [15] and Heisenberg [16]. From these specific theories, the general-purpose quantum formalism—which provides a general framework for building quantum theories—was shortly thereafter abstracted by Dirac [17] and put in precise mathematical form by von Neumann [18]. As mentioned earlier, in 1948, Feynman abstracted a set of rules (Feynman's rules) from the von Neumann formalism [1] which do away with most of the elaborate mathematical machinery of the standard abstract quantum formalism, and which can be put in close formal correspondence with sum and product rules of probability theory. Furthermore, the formal similarity of Feynman's rules to the rules of probability theory suggests that it might be possible to *derive* Feynman's rules in a manner analogous to that used to derive probability theory described above. We shall indeed show this to be the case.

In this section, we shall analyze the double-slit experiment (see Figure 1) using probability theory and Feynman's rules.

On the left, a heated electrical filament, A, emits electrons (all assumed to have the same energy), that pass through a wire-loop detector which registers the time of their emission. These electrons then impinge on screen B in which there are two slits, B_1 and B_2. Some of the electrons will subsequently emerge, pass through a wire-loop detector which registers their time of emergence, and finally reach screen C, which is covered by a grid of highly-sensitive detectors, each of which is capable of detecting an individual electron [19]. When the experiment is run, if the temperature of the electrical element is reduced sufficiently, one finds that one and only one of the screen detectors fires within the resolution time of the detectors [20]. For the purposes of visualization, let us suppose that the outputs of the detectors is wired into a grid of light-emitting diodes (LEDs) on the backside of C. So, watching the experiment, one sees a sequence of flashes, each emanating from a particular cell in the LED grid. After the accumulation of many such flashes, one builds up an *intensity pattern* over the screen, which allows us to estimate the probability that a given electron will strike a particular detector on the screen on a given run of the experiment.

Figure 1. Sketch of the double-slit experiment. On the left, a heated electrical filament serves as an electron source. Electrons emerging from the filament are collimated, and are detected as they pass through a wire-loop detector. The electrons then encounter a screen, B, containing two slits, and the electrons that pass B are registered by another wire-loop detector. Finally, the electrons that pass through B will be detected on the screen on the right-hand side.

electron source

collimator

screen with slits

detection screen
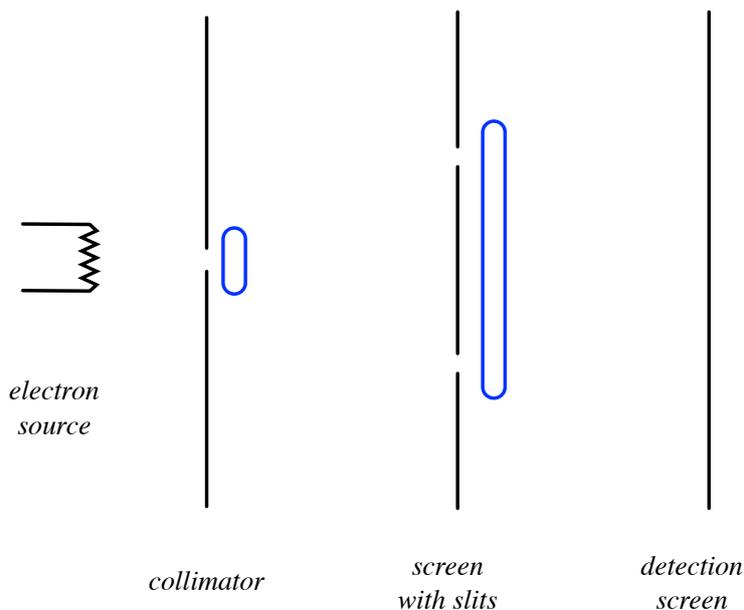
Note that it is the atomicity of these detections—only one detector fires at a particular time—that leads us to visualize electrons as entities that fly through space in highly-localized bundles of energy. However, this model extrapolates too far from the observed facts, and leads to manifestly incorrect predictions. To see this, let us compute the probability that, in a given run of the experiment, a detection is obtained at some given location on C at time t_3 and passes B at time $t_2 < t_3$ given that an electron is emitted from the filament, A, at time $t_1 < t_2$. We shall suppose that screen B is sufficiently large that these holes provide the only avenue through which the electrons can reach C. According to probability theory

$$\Pr(\mathbf{C}|\mathbf{A}) = \Pr(\mathbf{C}, \mathbf{B}|\mathbf{A}) \tag{50}$$

where we have defined the propositions

$\mathbf{A} = $ "Electron is emitted from A at time t_1"

$\mathbf{B} = $ "Electron passes through the union of the space occupied by slits B_1 and B_2 at time t_2"

$\mathbf{C} = $ "Electron is detected at given cell at C at time t_3"

Note that we are inferring the truth of proposition $\mathbf{B}$ from the truth of $\mathbf{A}$ and $\mathbf{C}$, which one might legitimately question if one wishes to avoid interpolating beyond the observed data. However, if we wish, we could employ a detector (consisting of an induction loop, as shown in the figure) which can

register the passage of an electron through the slits, without localizing its passage through one slit or the other. In that case, one would find that this detector fires whenever $\mathbf{A}$ and $\mathbf{C}$ are both true, thereby providing empirical justification for Equation (50).

Suppose, now, we admit the classical model of electrons as *highly localized entities*. Then, in the above experiment, given detection at C and emission at A, one would infer that the electron must have travelled through B via either slit B_1 or B_2. That is, one would infer the statement

$$\mathbf{B} = \mathbf{B}_1 \vee \mathbf{B}_2 \tag{51}$$

where

$\qquad \mathbf{B}_1 = $ "Electron passes through slit B_1 at time t_2"

$\qquad \mathbf{B}_2 = $ "Electron passes through slit B_2 at time t_2"

Under this assumption, Equation (50) becomes

$$\begin{aligned}
\Pr(\mathbf{C}|\mathbf{A}) &= \Pr(\mathbf{C}, \mathbf{B}_1 \vee \mathbf{B}_2|\mathbf{A}) \\
&= \Pr(\mathbf{C}, \mathbf{B}_1|\mathbf{A}) + \Pr(\mathbf{C}, \mathbf{B}_2|\mathbf{A})
\end{aligned} \tag{52}$$

where the sum rule of probability theory, Equation (31), has been used to arrive at the second line.

Now, Equation (52) asserts a relationship that can readily be tested. If slit B_2 is closed and the experiment is run, the only electrons that strike C must pass through slit B_1 [21]. On the assumption that probability $\Pr(\mathbf{C}, \mathbf{B}_1|\mathbf{A})$ is unaffected by the closure of slit B_2, which assumption follows naturally from the classical model of the electron, the probability $\Pr(\mathbf{C}, \mathbf{B}_1|\mathbf{A})$ can be estimated. Similarly, if slit B_1 is closed but B_2 left open, one can estimate $\Pr(\mathbf{C}, \mathbf{B}_2|\mathbf{A})$. One then finds *experimentally* that Equation (52) does *not* hold true. The conclusion seems inescapable: the model of electrons as highly localized entities must be false [22].

The intensity pattern over screen C obtained when both slits are open in fact resembles the pattern that would be expected if classical waves (such as water waves or electromagnetic waves) were being emitted from the electrical filament. In particular, one observes constructive and destructive interference, the hallmark of wave phenomena. Indeed, the intensity pattern can be *quantitatively* reasonably well described using a classical wave model [23]. Yet, this model manifestly conflicts with the atomicity of the detections at C. Thus, we find ourselves in a somewhat perplexing situation where we require both the classical particle model and the classical wave model to account for the observations, but where each appears to be in conflict with the other, a situation known as the "wave-particle duality" of the electron.

The failure of Equation (51) naturally raises the question: how can the probability $\Pr(\mathbf{C}, \mathbf{B}|\mathbf{A})$ be related to the probabilities $\Pr(\mathbf{C}, \mathbf{B}_1|\mathbf{A})$ and $\Pr(\mathbf{C}, \mathbf{B}_2|\mathbf{A})$? From the experimental findings mentioned above, it can be seen that, although $\Pr(\mathbf{C}, \mathbf{B}|\mathbf{A})$ is not *determined* by $\Pr(\mathbf{C}, \mathbf{B}_1|\mathbf{A})$ and $\Pr(\mathbf{C}, \mathbf{B}_2|\mathbf{A})$, it is not *independent* of them either. So, presumably there is a looser relation between these three probabilities, involving additional degrees of freedom. But, if so, what is that relationship?

According to Feynman's rules, the probability $\Pr(\mathbf{C}, \mathbf{B}|\mathbf{A})$ is equal to $|z|^2$, where z is a complex number, referred to as the *amplitude,* associated with the process leading from the detection at A at time t_1 via detection at B at time t_2 to the detection at C at time t_3. This amplitude is, in turn, the sum

of the amplitudes z_1 and z_2, where z_i ($i = 1, 2$) is the amplitude associated with the process leading from the detection at A at time t_1 to detection at C at time t_3 via slit B_i at time t_2. Thus, in place of Equation (52), one has the *amplitude sum rule*

$$z = z_1 + z_2 \tag{53}$$

while the relationship between the amplitude, u, associated with a process and its probability, p, is given by the *amplitude-probability rule* as $p = |u|^2$, so that

$$\begin{aligned}
\Pr(\mathbf{C}, \mathbf{B}|\mathbf{A}) &= |z|^2 \\
\Pr(\mathbf{C}, \mathbf{B}_1|\mathbf{A}) &= |z_1|^2 \\
\Pr(\mathbf{C}, \mathbf{B}_2|\mathbf{A}) &= |z_2|^2
\end{aligned} \tag{54}$$

As a consequence of these relations, in place of Equation (52), one obtains

$$\Pr(\mathbf{C}, \mathbf{B}|\mathbf{A}) = p_1 + p_2 + 2\sqrt{p_1 p_2} \cos \phi \tag{55}$$

where, for brevity, we have written $p_i = \Pr(\mathbf{C}, \mathbf{B}_i|\mathbf{A})$ and $\phi = \arg(z_1^* z_2)$. Hence, as anticipated, there does exist a non-trivial relationship between the three probabilities, but this relationship involves an additional degree of freedom, ϕ.

In summary, we see that the conflict of predictions is not between probability theory and Feynman's rules *per se,* but rather between (a) the *union* of probability theory and an assumption whose origin lies in classical physics and (b) Feynman's rules. The conflict disappears once the classical assumption is dropped.

We conclude by noting that, in addition to the two Feynman rules listed above (the amplitude sum rule and the amplitude-probability rule), there is also a third rule, the *amplitude product rule,* which states that, if a process (with amplitude u) can be broken into two sub-processes (with amplitudes u' and u'') concatenated in series, then

$$u = u'u'' \tag{56}$$

For example, the process leading from the detection at A at time t_1 to the detection at C at time t_3 via B_1 at time t_2 can be broken into (i) the detection at A at time t_1 to B_1 at time t_2, and (ii) B_1 at time t_2 to the detection at C at time t_3, so that $z_1 = z_1' z_1''$, where z_1' and z_1'' are the amplitudes of the sub-processes.

We note that, in this example, the amplitude product rule implies that

$$\Pr(\mathbf{C}, \mathbf{B}_1|\mathbf{A}) = \Pr(\mathbf{C}|\mathbf{B}_1) \Pr(\mathbf{B}_1|\mathbf{A}) \tag{57}$$

In contrast, the application of the product rule of probability theory, Equation (32), implies that

$$\Pr(\mathbf{C}, \mathbf{B}_1|\mathbf{A}) = \Pr(\mathbf{C}|\mathbf{B}_1, \mathbf{A}) \Pr(\mathbf{B}_1|\mathbf{A}) \tag{58}$$

which is the same provided that $\Pr(\mathbf{C}|\mathbf{B}_1, \mathbf{A}) = \Pr(\mathbf{C}|\mathbf{B}_1)$, which is true provided that the second sub-process is Markov (probabilistically independent) with respect to the first, which is indeed experimentally valid. Hence, there is a close formal relationship between Feynman's amplitude sum and product rules on the one hand, and the sum and product rules of probability theory on the other.

4. Derivation of Feynman's Rules

As we have seen above, Feynman's rules express the content of the quantum formalism with a minimum of formal means, and does so in a manner which establishes a close formal parallel to the rules of probability theory. The latter observation raises the question of whether Feynman's rules may be derivable in a manner analogous to that described in Section 2, namely by quantifying over a suitably-defined logic. In this section, we shall outline such a derivation.

4.1. Operational Experimental Framework

We begin by establishing a fully operational framework for describing experimental set-ups consisting of sequences of measurements and interactions on a physical system. The motivation for such a framework is twofold. First, as described above, the application of Feynman's rules to an experiment involving electrons requires that one considers the various *classical* paths that an electron (modeled as a particle) could take from some initial point to some final point in spacetime. To appeal to a classical model in the statement of rules for a theory that is inconsistent with that same classical model can (and indeed does) lead to confusion. Hence, it is highly desirable to formulate a way to describe experiments which is sufficiently precise as to obviate the need for such an appeal. Second, although the primitive terms "measurement", "outcome", and "interaction" may seem very simple and transparent, it turns out that they require very careful formal specification in order that they can be consistently used in a derivation of Feynman's rules. The experimental framework provides such a specification. The experimental framework is described in [11], to which the reader is referred for full details. Below, for completeness, we shall recount the main points.

The key primitives of an experimental set-up are as follows. A *source* is a black box which issues physical systems which behave identically as far as a set of given measurements and interactions are concerned. Measurements are black boxes which take a physical system from the source as input, yield one of a finite number of possible *repeatable outcomes*, and then output the physical system. A repeatable outcome is a *macroscopically stable output* (such as the illumination of an LED) of the measurement device which is such that, if the output is obtained when a measurement is performed on a system, then the same output is again obtained with certainty if the measurement is immediately repeated on the system. We refer to measurements all of whose outcomes are repeatable as *repeatable measurements*. Finally, an *interaction* is anything which happens to a system in between measurements which is itself not a measurement, but which non-trivially influences the outcome probabilities of subsequent measurements.

An experimental set-up is defined by specifying a source, a sequence of measurements, and the interactions occur that during the experiment. In a run of an experiment, a physical system from the source passes through a sequence of measurements $M_1, M_2, \ldots$, which respectively yield outcomes $m_1, m_2, \ldots$ at times $t_1, t_2, \ldots$. These outcomes are summarized in the *measurement sequence* $[m_1, m_2, \ldots]$. For notational brevity, the measurements that yield these outcomes, and the times at which these outcomes occur, are left implicit. In between these measurements, the system may undergo interactions with the environment. For example, as illustrated in Figure 2, in particular run of an

experiment involving a sequence of three measurements, each of whose outcomes are labeled 1 and 2, the sequence $[1, 1, 1]$ is obtained.

> **Figure 2.** An experimental set-up consisting of three successive measurements, each of which has two possible outcomes. In a particular run of the experiment, the measurement outcome sequence $[1, 1, 1]$ is obtained.
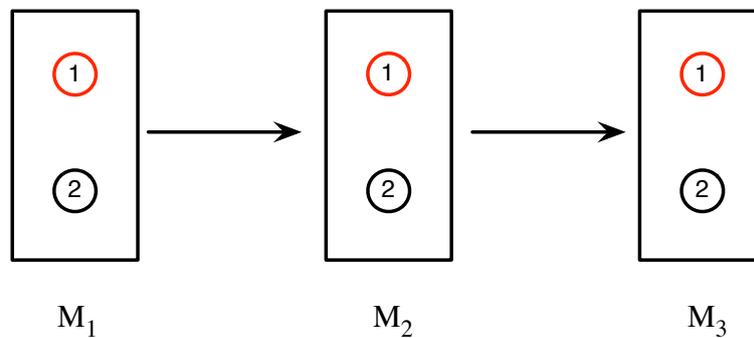
In this case, the system may, for instance, be a silver atom, upon which Stern-Gerlach measurements (which, in this case, would each have two possible outcomes) are performed.

Over many runs of the experiment, the experimenter will observe the frequencies of the various possible measurement sequences, from which one can (using Bayes' rule) estimate the probability associated with each sequence. We define the probability $P(A)$ associated with sequence $A = [m_1, m_2, \ldots, m_n]$ as the probability of obtaining outcomes $m_2, \ldots, m_n$ conditional upon obtaining m_1

$$P(A) = \Pr(m_n, m_{n-1}, \ldots, m_2 \,|\, m_1) \tag{59}$$

The conditionalization on outcome m_1 ensures that $P(A)$ is independent of the history of the system prior to the first measurement, and so is fully under experimental control.

A particular outcome of a measurement is either *atomic* or *coarse-grained*. An atomic outcome cannot be more finely divided in the sense that the detector whose output corresponds to the outcome cannot be sub-divided into smaller detectors whose outputs correspond to two or more outcomes. A coarse-grained outcome is one that does not differentiate between two or more outcomes. For example, in the experiment in Figure 3, the second measurement, $\widetilde{M}_2$, has a single outcome which is a coarse-graining of the outcomes labeled 1 and 2 of measurement M_2. Accordingly, the outcome of $\widetilde{M}_2$ is labeled $(1, 2)$, and we apply this notational convention to coarse-grained outcomes generally. The measurement sequence obtained in this case is $[1, (1, 2), 1]$.

In general, if all of the possible outcomes of a measurement are atomic, we shall call the measurement itself atomic. Otherwise, we shall refer to it as a coarse-grained measurement, and sometimes symbolize this as $\widetilde{M}$ if we wish to indicate that it is obtained by coarse-graining over some of the outcomes of the atomic measurement M.

It is important that all of the measurements, $M_1, M_2, \ldots$ that are employed in an experimental set-up come from the same *measurement set*, $\mathcal{M}$, or are coarsened versions of measurements in this set. The set consists only of atomic, repeatable measurements, and satisfies the *closure* condition that, if any pair of measurements, M, N are selected from $\mathcal{M}$, then, if an experiment is performed where M

is used to prepare a system (namely, selecting those systems that yield a particular outcome of **M**) and measurement **N** is performed immediately afterwards, then the outcome probabilities of **N** are independent of interactions with the system prior to **M**. Interactions *between* measurements are likewise selected from a set, $\mathcal{I}$, of possible interactions, which are such that they preserve closure when performed between any pair of measurements from $\mathcal{M}$. These rather intricate requirements are necessary to ensure that the behavior of the system under study is independent of the history of the system prior to the start of the experiment, and that all of the measurements performed on the system are probing the same aspect of the system. We shall restrict consideration throughout to sequences whose initial and final outcomes are atomic.

Figure 3.   An experimental set-up consisting of three measurements in which the second measurement is coarse-grained.   In a particular run of the experiment, the sequence $[1, (1, 2), 1]$ is obtained.
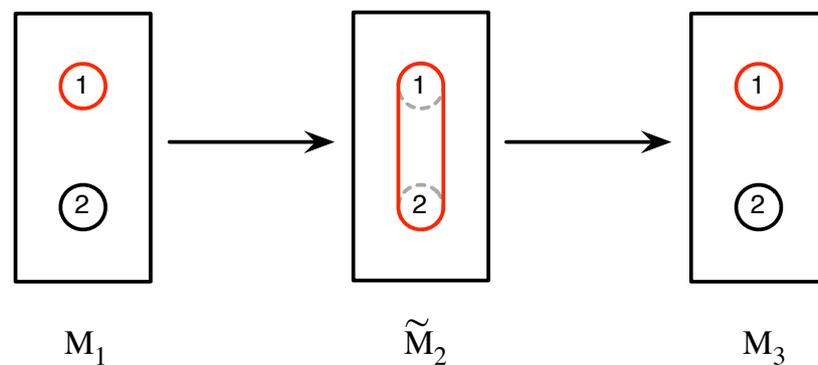
4.2.   Sequence Combination Operators

We wish to develop a calculus—which we shall refer to henceforth as the **process calculus**—that is capable of establishing a relation between the probabilities observed in experimental set-ups such as those in Figure 2 and 3.   For example, in a run of the first experiment, one might observe the sequences $A = [1, 1, 1]$ or $B = [1, 2, 1]$, while, in a run of the second experiment, one might observe $C = [1, (1, 2), 2]$.   The calculus should provide a relationship between the probabilities $P(A)$, $P(B)$ and $P(C)$ associated with these sequences.

As the discussion of the double-slit experiment above makes clear, probability theory cannot by itself establish a relationship between these probabilities unless an additional assumption drawn from classical physics is made.   That is, if one assumes that, in Figure 3, when the large detector of $\widetilde{M}_2$ fires, the system *in fact* went via the top or bottom of the detector (that is, via outcome 1 or 2) *even though it was not measured doing so,* one is led by a simple application of probability theory to the relationship $P(C) = P(A) + P(B)$, which is in manifest conflict with experimental data.   If, however, we refrain from making this classical assumption, probability theory provides no relationship between $P(A)$, $P(B)$ and $P(C)$ whatsoever. The calculus we seek to develop is designed to fill this void.

Recognizing that the probabilities associated with sequences A, B and C cannot be simply related, we seek a deeper theoretical description of the sequences where the description of C is *determined* by

the descriptions of A and B, and where the description of each sequence yields the probability associated with that sequence. That is, we seek to introduce a level of theoretical description which is one level lower (more fundamental) than the probability-level of description.

To carry out this program, we begin by formalizing the relationship between A, B and C by introducing the *sequence parallel composition operator,* $\vee$, so that we write $C = A \vee B$. Formally, if two sequences can be obtained from the same experimental set-up, agree in the first and last outcomes, but differ in precisely one outcome, then they can be combined in parallel. That is, if we have sequences $[m_1, \ldots, m_i, \ldots, m_n]$ and $[m_1, \ldots, m_i', \ldots, m_n]$ obtained from the same experimental set-up which differ only in the ith outcome ($m_i \neq m_i'$), but neither the first nor the last, then

$$[m_1, \ldots, (m_i, m_i'), \ldots, m_n] = [m_1, \ldots, m_i, \ldots, m_n] \vee [m_1, \ldots, m_i', \ldots, m_n] \tag{60}$$

where (m_i, m_i') symbolizes a coarse-grained outcome.

In order to reflect the idea of series concatenation of experimental set-ups, we also introduce a *series composition operator*, symbolized as $\cdot$. For example, the sequence $[m_1, m_2, m_3]$ can be viewed as the series composition of the shorter sequences $[m_1, m_2]$ and $[m_2, m_3]$, so that

$$[m_1, m_2, m_3] = [m_1, m_2] \cdot [m_2, m_3] \tag{61}$$

More generally, if two sequences are obtained in two different experimental set-ups that immediately follow one another in time, and the sequences are such that the last measurement and outcome of one sequence are the same as the first measurement and outcome of the other sequence, then they can be combined together in series.

From these definitions, it follows that the two binary operators have the following five symmetries:

$$A \vee B = B \vee A \tag{62}$$

$$(A \vee B) \vee C = A \vee (B \vee C) \tag{63}$$

$$(A \cdot B) \cdot C = A \cdot (B \cdot C) \tag{64}$$

$$(A \vee B) \cdot C = (A \cdot C) \vee (B \cdot C) \tag{65}$$

$$C \cdot (A \vee B) = (C \cdot A) \vee (C \cdot B) \tag{66}$$

namely commutativity and associativity of $\vee$ (Equations (62) and (63)), associativity of $\cdot$ (Equation (64)), and right- and left- distributivity of $\cdot$ over $\vee$ (Equations (65) and (66)).

4.3. Pair Representation of Sequences

We now introduce the desired theoretical level of description of the sequences. In particular, we represent each sequence, A, with a *real number pair* [24], $(a_1, a_2)^\mathsf{T}$, and require that this representation is consistent with the five symmetries identified above. For example, if pairs $\mathbf{a}, \mathbf{b}$ represent the sequences A, B, respectively, then the pair $\mathbf{c}$ that represents $C = A \vee B$ must be determined by $\mathbf{a}$, $\mathbf{b}$ through the relation

$$\mathbf{c} = \mathbf{a} \oplus \mathbf{b} \tag{67}$$

where $\oplus$ is a pair-valued binary operator, assumed continuous, to be determined. Similarly, if the sequences A, B and C are related by $C = A \cdot B$, then the pair $\mathbf{c}$ that represents C must be determined by $\mathbf{a}, \mathbf{b}$ through the relation

$$\mathbf{c} = \mathbf{a} \odot \mathbf{b} \tag{68}$$

where $\odot$ is another pair-valued binary operator, assumed continuous, also to be determined.

From these definitions of $\oplus$ and $\odot$, the five symmetries of the sequence combination operators immediately imply that

$$\mathbf{a} \oplus \mathbf{b} = \mathbf{b} \oplus \mathbf{a} \tag{S1}$$

$$(\mathbf{a} \oplus \mathbf{b}) \oplus \mathbf{c} = \mathbf{a} \oplus (\mathbf{b} \oplus \mathbf{c}) \tag{S2}$$

$$(\mathbf{a} \odot \mathbf{b}) \odot \mathbf{c} = \mathbf{a} \odot (\mathbf{b} \odot \mathbf{c}) \tag{S3}$$

$$(\mathbf{a} \oplus \mathbf{b}) \odot \mathbf{c} = (\mathbf{a} \odot \mathbf{c}) \oplus (\mathbf{b} \odot \mathbf{c}) \tag{S4}$$

$$\mathbf{a} \odot (\mathbf{b} \oplus \mathbf{c}) = (\mathbf{a} \odot \mathbf{b}) \oplus (\mathbf{a} \odot \mathbf{c}) \tag{S5}$$

In [11], we show that these symmetry conditions impose strong restrictions on the possible form of the operators $\oplus$ and $\odot$. In particular, commutativity and associativity of $\oplus$, subject to some minor auxiliary mathematical conditions, imply that, without loss of generality, one can take

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \oplus \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix} \tag{69}$$

which we refer to as the *sum rule*. The distributivity conditions (S4) and (S5) then imply that $\mathbf{a} \odot \mathbf{b}$ has a bilinear multiplicative form

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \gamma_1 a_1 b_1 + \gamma_2 a_1 b_2 + \gamma_3 a_2 b_1 + \gamma_4 a_2 b_2 \\ \gamma_5 a_1 b_1 + \gamma_6 a_1 b_2 + \gamma_7 a_2 b_1 + \gamma_8 a_2 b_2 \end{pmatrix} \tag{70}$$

where the γ_i are real-valued quantities to be determined. Finally, the associativity of $\odot$ implies that $\mathbf{a} \odot \mathbf{b}$ has one of five possible standard forms, namely

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 - a_2 b_2 \\ a_1 b_2 + a_2 b_1 \end{pmatrix} \tag{C1}$$

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_1 b_2 + a_2 b_1 \end{pmatrix} \tag{C2}$$

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_2 b_2 \end{pmatrix} \tag{C3}$$

and

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_1 b_2 \end{pmatrix} \tag{N1}$$

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 \\ a_2 b_1 \end{pmatrix} \tag{N2}$$

We recognize possibility (C1) as complex multiplication, while (C2) and indeed also (C3) are variations thereof known respectively as dual numbers and split-complex numbers. Finally, the last two possibilities (N1 and N2) are non-commutative multiplication rules.
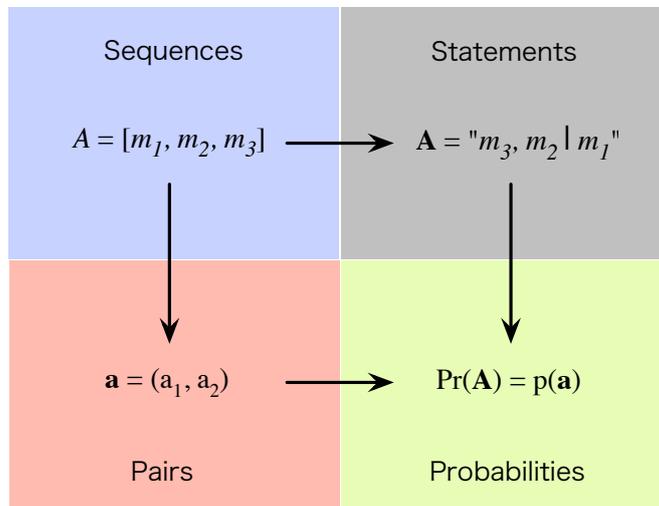
4.4. Probabilities, and the Probability Product Equation

We defined the probability $P(A)$ associated with sequence $A = [m_1, m_2, \ldots, m_n]$ in Equation (59) as $P(A) = \Pr(m_n, m_{n-1}, \ldots, m_2 \mid m_1)$. We now create a link between the theoretical representation, $\mathbf{a}$, of a sequence, A, and the probability, $P(A)$, associated with the sequence by requiring that the former determine the latter, so that

$$P(A) = p(\mathbf{a}) \tag{71}$$

where $p(\cdot)$ is a continuous real-valued function that depends non-trivially on both real components of its argument. The overall structure that this link establishes is shown in Figure 4.

Figure 4. Illustration of the overall logical structure of the process calculus. On the top left is the space of *sequences*. To each sequence, A, corresponds a conditional statement, $\mathbf{A}$ (top right). Each sequence is represented by a *pair,* $\mathbf{a}$ (bottom left), while each conditional statement is represented by a *probability*, $\Pr(\mathbf{A})$ (bottom right). The link between these representations, $\Pr(\mathbf{A}) = p(\mathbf{a})$, is given on the bottom right.

Now, once a link between pairs and probabilities has been postulated, the process calculus provides the means to relate the probabilities associated with different sequences. In certain cases, such as for the sequences $A = [1, 1, 1]$, $B = [1, 2, 1]$, and $C = [1, (1, 2), 2]$ mentioned in Section 4.2 above, probability theory alone can establish no relation, which is precisely the void the process calculus is designed to fill. However, there are situations where both the process calculus *and* probability theory can be applied, and in these circumstances they must agree, on pain of the process calculus being inconsistent with probability theory. This consistency requirement sharply delimits the form that the function $p(\cdot)$ can take.

To exhibit this consistency requirement, consider the two sequences $A = [m_1, m_2]$ and $B = [m_2, m_3]$ of atomic outcomes, with the sequences represented by pairs $\mathbf{a}$ and $\mathbf{b}$. Since outcome m_2 is the same in

each, $C = A \cdot B$ is given by $C = [m_1, m_2, m_3]$, represented by the pair $\mathbf{a} \odot \mathbf{b}$. The probability, $P(C)$, associated with sequence C is given by

$$P(C) = \Pr(m_3, m_2 \,|\, m_1)$$

which, by the product rule of probability theory, can be rewritten as

$$P(C) = \Pr(m_3 \,|\, m_2, m_1) \Pr(m_2 \,|\, m_1)$$

Since m_2 is atomic, measurement $\mathbf{M}_2$ (with outcome m_2) establishes closure with respect to $\mathbf{M}_3$ (with outcome m_3). Therefore, the probability of outcome m_3 is independent of m_1, and the above equation simplifies to

$$P(C) = \Pr(m_3 \,|\, m_2) \Pr(m_2 \,|\, m_1)$$
$$= P(B)\,P(A)$$

Since $P(A) = p(\mathbf{a})$, $P(B) = p(\mathbf{b})$, and $P(C) = p(\mathbf{a} \odot \mathbf{b})$, the consistency of the process calculus with probability theory requires that, for any $\mathbf{a}$, $\mathbf{b}$, the function $p(\cdot)$ must satisfy the equation

$$p(\mathbf{a} \odot \mathbf{b}) = p(\mathbf{a})\,p(\mathbf{b}) \tag{72}$$

As shown in [11], if one solves for the function $p(\cdot)$ that satisfies this equation in each of the five forms of $\odot$ given above, one obtains

Case C1: $p(\mathbf{a}) = (a_1^2 + a_2^2)^{\alpha/2}$;

Case C2: $p(\mathbf{a}) = |a_1|^\alpha e^{\beta a_2/a_1}$;

Case C3: $p(\mathbf{a}) = |a_1|^\alpha |a_2|^\beta$;

Case N1: $p(\mathbf{a}) = |a_1|^\alpha$;

Case N2: $p(\mathbf{a}) = |a_1|^\alpha$;

with α, β real constants.

We now note that, while in the three commutative forms (C1), (C2) and (C3), the function $p(\cdot)$ depends upon *both* components of its pair argument, in the two non-commutative forms (N1) and (N2), it depends only on the first component of its argument. Consequently, in the latter two cases, the process calculus reduces to a scalar calculus insofar as predictions as concerned.

To see this, consider case (N1): pairs $\mathbf{a} = (a_1, a_2)^\mathsf{T}$ and $\mathbf{b} = (b_1, b_2)^\mathsf{T}$ can be combined using $\oplus$ and $\odot$ to yield $\mathbf{c} = \mathbf{a} \oplus \mathbf{b} = (a_1 + b_1, a_2 + b_2)^\mathsf{T}$ and $\mathbf{d} = \mathbf{a} \odot \mathbf{b} = (a_1 b_1, a_1 b_2)^\mathsf{T}$, respectively. The associated probabilities are $p(\mathbf{a}) = |a_1|^\alpha$, $p(\mathbf{b}) = |b_1|^\alpha$, $p(\mathbf{c}) = |a_1 + b_1|^\alpha$ and $p(\mathbf{d}) = |a_1 b_1|^\alpha$, all of which are independent of the second components of pairs $\mathbf{a}$ and $\mathbf{b}$. Hence, the second component of each pair can be dropped without in any way affecting the probabilistic predictions made by the calculus. Hence, the calculi in cases (N1) and (N2) are effectively scalar calculi, contrary to our design desideratum. Accordingly, we reject these two cases. Of the five possible forms of $\odot$, we are therefore left with three: (C1), (C2) and (C3).

In Reference [11], additional arguments were mounted which eliminated cases (C2) and (C3), and picked out case (C1) with $\alpha = 2$. In the remainder of this paper, we shall present a novel line of argument which leads to the same conclusion.

4.5. Pair Symmetry

When representing a sequence with a pair, we did not distinguish the role played by each component of pair. That is to say, of the resulting process calculi consistent with the constraints imposed thus far, there should be at least one whose predictions are invariant under the operation which swaps the components of every pair used to represent a sequence. This *pair symmetry* requirement implies that $p(\cdot)$ must be invariant under this swap operation, namely that, for any $\mathbf{a} = (a_1, a_2)$,

$$p\left((a_1, a_2)\right) = p\left((a_2, a_1)\right) \tag{73}$$

In case (C1), $p(\mathbf{a}) = (a_1^2 + a_2^2)^{\alpha/2}$, which already satisfies this symmetry for any α. However, in case (C2), $p(\mathbf{a}) = |a_1|^\alpha e^{\beta a_2/a_1}$, which is not symmetric under the swap operation for any α, β apart from the trivial $\alpha = \beta = 0$. Therefore, case (C2) must be eliminated. Finally, in case (C3), $p(\mathbf{a}) = |a_1|^\alpha |a_2|^\beta$, which satisfies this symmetry provided that $\alpha = \beta$.

In summary, after having imposed the pair symmetry condition, we are left with the case (C1) and case (C3) with $\alpha = \beta$.

4.6. Independent Parallel Processes

Consider an experimental set-up consisting of three measurements, $\mathbf{M}_1, \mathbf{M}_2$ and $\mathbf{M}_3$ performed in succession. On one run, this generates sequence $A = [m_1, m_2, m_3]$ and, on another run, sequence $B = [m_1, m_2', m_3]$, with $m_2 \neq m_2'$. Then consider a second set-up, identical to the first except that the intermediate measurement $\widetilde{\mathbf{M}}_2$ coarsens outcomes m_2 and m_2' of $\mathbf{M}_2$, and suppose that this generates the sequence $C = [m_1, (m_2, m_2'), m_3]$. If sequences A and B are represented by pairs $\mathbf{a}, \mathbf{b}$, respectively, then sequence C is represented by the pair $\mathbf{c} = \mathbf{a} + \mathbf{b}$.

Now, as we have discussed above in the context of the double-slit experiment, if one were to assume according to classical physics that, in the second experiment, the system went through either the portion of the large detector corresponding to outcome m_2 or through the portion corresponding to outcome m_2' *even though neither was explicitly measured,* it would follow that

$$\Pr(m_3, (m_2, m_2')|\, m_1) = \Pr(m_3, m_2 \vee m_2'|\, m_1) \tag{74}$$

Using the sum rule of probability theory, the right hand side can be written as

$$\begin{aligned} \Pr(m_3, m_2 \vee m_2'|\, m_1) &= \Pr(m_3, m_2|m_1) + \Pr(m_3, m_2'|m_1) \\ &= p(\mathbf{a}) + p(\mathbf{b}) \end{aligned} \tag{75}$$

Thus, in the special case where the classical assumption is valid—that is to say, the two processes represented by sequences A and B are independent (do not interfere)—it follows that

$$p(\mathbf{a}) + p(\mathbf{b}) = p(\mathbf{a} + \mathbf{b}) \tag{76}$$

It seems very reasonable to require that the pair-calculus that we are developing should include the possibility of independent (non-interfering) processes as a special case. Accordingly, we shall require that the pair-calculus satisfy the:

Additivity Condition: *For any given probabilities p_1 and p_2 for which $p_1 + p_2 \leq 1$, there exist pairs* **a** *and* **b** *satisfying $p(\mathbf{a}) = p_1, p(\mathbf{b}) = p_2$ such that Equation* (76) *(which we shall henceforth refer to as the* additivity equation*) holds true whenever $p(\mathbf{a} + \mathbf{b}) \leq 1$.*

We shall now investigate the constraints that this condition imposes in cases (C1) and (C3).

4.7. Case (C1)*, with $p(\mathbf{x}) = (x_1^2 + x_2^2)^{\alpha/2}$*

Setting $\mathbf{a} = (r_1 \cos\theta_1, r_1 \sin\theta_1)$ and $\mathbf{b} = (r_2 \cos\theta_2, r_2 \sin\theta_2)$, the additivity Equation (76) becomes

$$r_1^\alpha + r_2^\alpha = \left[r_1^2 + r_2^2 + 2r_1 r_2 \cos\theta \right]^{\alpha/2} \tag{77}$$

where $\theta = \theta_1 - \theta_2$. By the additivity condition, for any r_1, r_2 for which $p(\mathbf{a} + \mathbf{b}) \leq 1$, there must exist some value of θ which satisfies this equation. Since $-1 \leq \cos\theta \leq 1$

$$|r_1 - r_2|^\alpha \leq \left[r_1^2 + r_2^2 + 2r_1 r_2 \cos\theta \right]^{\alpha/2} \leq (r_1 + r_2)^\alpha \tag{78}$$

so that (as illustrated in Figure 5) a value of θ can be found that satisfies Equation (77) provided that

$$|r_1 - r_2| \leq (r_1^\alpha + r_2^\alpha)^{1/\alpha} \leq r_1 + r_2 \tag{79}$$

The inequality $(r_1^\alpha + r_2^\alpha)^{1/\alpha} \leq r_1 + r_2$ is satisfied provided that $\alpha \geq 1$, which also satisfies $|r_1 - r_2| \leq (r_1^\alpha + r_2^\alpha)^{1/\alpha}$.

Therefore, the additivity condition is satisfied by case (C1) for any $\alpha \geq 1$.

Case (C3), with $p(\mathbf{x}) = |x_1 x_2|^\alpha$

In this case, the additivity Equation (76) reads

$$|a_1 a_2|^\alpha + |b_1 b_2|^\alpha = |a_1 a_2 + b_1 b_2 + (a_1 b_2 + a_2 b_1)|^\alpha \tag{80}$$

Parameterizing **a**, **b** as

$$\mathbf{a} = (\gamma_1 \, p(\mathbf{a})^{1/\alpha}, 1/\gamma_1) \tag{81}$$

$$\mathbf{b} = (\gamma_2 \, p(\mathbf{b})^{1/\alpha}, 1/\gamma_2) \tag{82}$$

where γ_1, γ_2 are real non-zero parameters, we obtain

$$p(\mathbf{a}) + p(\mathbf{b}) = \left| p(\mathbf{a})^{1/\alpha} (1 + \delta) + p(\mathbf{b})^{1/\alpha} \left(1 + \delta^{-1} \right) \right|^\alpha \tag{83}$$

where $\delta = \gamma_1/\gamma_2$. Extremising with respect to δ, one finds that

$$0 \leq p(\mathbf{a}) + p(\mathbf{b}) \leq \left[p(\mathbf{a})^{1/2\alpha} + p(\mathbf{b})^{1/2\alpha} \right]^{2\alpha} \tag{84}$$

with the lower and upper limits obtained, respectively, when $\delta = -1$ and $\delta = (p(\mathbf{b})/p(\mathbf{a}))^{1/2\alpha}$ (see Figure 6).

The inequality $p(\mathbf{a}) + p(\mathbf{b}) \leq \left[p(\mathbf{a})^{1/2\alpha} + p(\mathbf{b})^{1/2\alpha} \right]^{2\alpha}$ is satisfied by any $\alpha \geq 1/2$. Therefore, the additivity condition is satisfied by case (C3) for any $\alpha \geq 1/2$.

Figure 5. For $p(\mathbf{a}) = p(\mathbf{b}) = 1/8$, this graph shows, as a function of α, (i) the extreme values of $p(\mathbf{a} + \mathbf{b})$, and (ii) the value of $p(\mathbf{a}) + p(\mathbf{b}) = 1/4$. For $\alpha < 1$, the maximum of $p(\mathbf{a} + \mathbf{b})$ is less than $p(\mathbf{a}) + p(\mathbf{b})$.
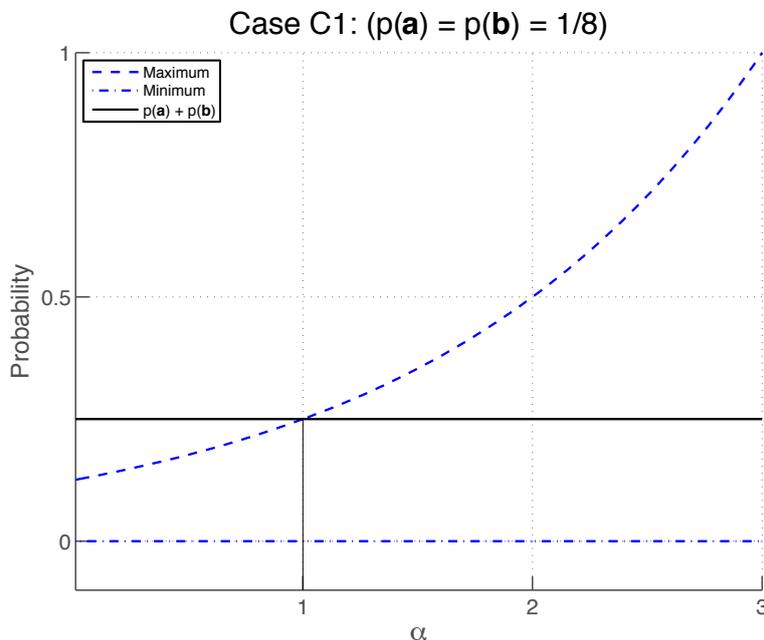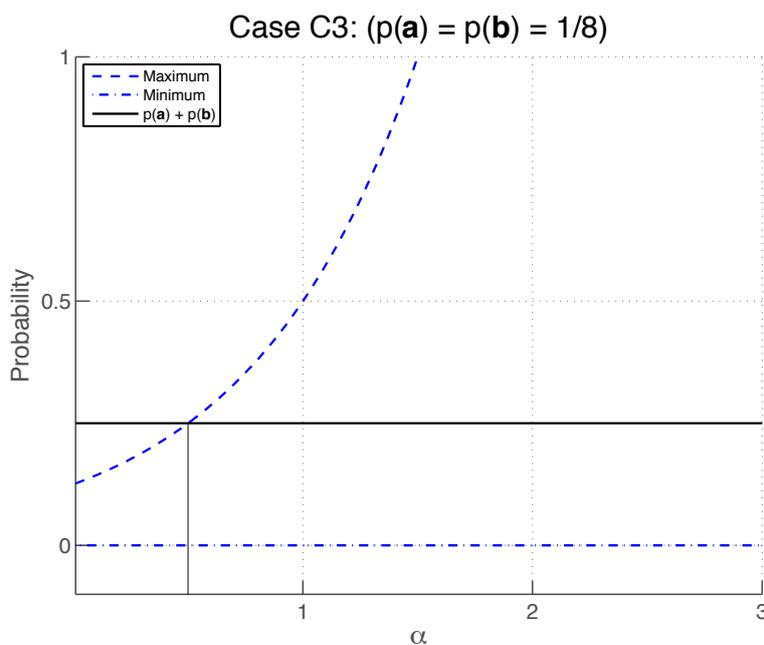

Figure 6. For $p(\mathbf{a}) = p(\mathbf{b}) = 1/8$, this graph shows, as a function of α, (i) the extreme values of $p(\mathbf{a} + \mathbf{b})$, and (ii) the value of $p(\mathbf{a}) + p(\mathbf{b}) = 1/4$. For $\alpha < 1/2$, the maximum of $p(\mathbf{a} + \mathbf{b})$ is less than $p(\mathbf{a}) + p(\mathbf{b})$.

4.8. Symmetric Bias

Although the additivity condition imposes constraints on the value of α in cases (C1) and (C3), it is by itself insufficient to pick out either of these cases and to pick out a unique value of α. In this section, we strengthen the additivity condition in such a way that uniquely picks out case (C1) with $\alpha = 2$.

Consider again the three sequences A, B and $C = A \vee B$ of Section 4.6. In general, for fixed values of $p(\mathbf{a})$ and $p(\mathbf{b})$, it will not be true that $p(\mathbf{a} + \mathbf{b}) = p(\mathbf{a}) + p(\mathbf{b})$. Instead, the possible values of $p(\mathbf{a} + \mathbf{b})$ will span a range whose endpoints will (in general) be a function of $p(\mathbf{a})$ and $p(\mathbf{b})$.

Now, in general, the maximum and minimum values of $p(\mathbf{a} + \mathbf{b})$ will *not* be symmetrically placed about $p(\mathbf{a}) + p(\mathbf{b})$. However, it seems natural to suppose that the process calculus should allow two processes to interfere constructively and destructively to an equal degree. That is, if we define the *biases*

$$\beta_+ = \max_{\mathbf{a},\mathbf{b}} p(\mathbf{a} + \mathbf{b}) - \big[p(\mathbf{a}) + p(\mathbf{b})\big] \tag{85}$$

$$\beta_- = \big[p(\mathbf{a}) + p(\mathbf{b})\big] - \min_{\mathbf{a},\mathbf{b}} p(\mathbf{a} + \mathbf{b}) \tag{86}$$

where the maximizations and minimizations are constrained by given values of $p(\mathbf{a})$ and $p(\mathbf{b})$, it seems natural to suppose that $\beta_+ = \beta_-$. Accordingly, we now require that the process calculus satisfy the:

Symmetric Bias Condition [25]: *For any given probabilities p_1 and p_2 for which $p_1 + p_2 \leq 1$, there exist pairs $\mathbf{a}$ and $\mathbf{b}$ satisfying $p(\mathbf{a}) = p_1, p(\mathbf{b}) = p_2$ such that $\beta_+ = \beta_-$ holds true whenever $p(\mathbf{a} + \mathbf{b}) \leq 1$.*

We shall now investigate the constraints that this condition imposes in cases (C1) and (C3).

Case (C1), with $p(\mathbf{x}) = (x_1^2 + x_2^2)^{\alpha/2}$

If we rewrite Equation (79) in terms of $p(\mathbf{a})$ and $p(\mathbf{b})$, we obtain

$$\big|p(\mathbf{a})^{1/\alpha} - p(\mathbf{b})^{1/\alpha}\big|^\alpha \leq p(\mathbf{a} + \mathbf{b}) \leq \big[p(\mathbf{a})^{1/\alpha} + p(\mathbf{b})^{1/\alpha}\big]^\alpha \tag{87}$$

from which the biases are

$$\beta_+ = \big[p(\mathbf{a})^{1/\alpha} + p(\mathbf{b})^{1/\alpha}\big]^\alpha - \big[p(\mathbf{a}) + p(\mathbf{b})\big] \tag{88}$$

$$\beta_- = \big[p(\mathbf{a}) + p(\mathbf{b})\big] - \big|p(\mathbf{a})^{1/\alpha} - p(\mathbf{b})^{1/\alpha}\big|^\alpha \tag{89}$$

Let us consider two special cases. For $\alpha = 1$, one finds $\beta_+ = 0$ and $\beta_- = 2\max(p(\mathbf{a}), p(\mathbf{b}))$, so that although no constructive interference is possible, destructive interference is always possible, which is a highly asymmetric situation. In contrast, for $\alpha = 2$, one finds that $\beta_+ = \beta_- = 2\sqrt{p(\mathbf{a})p(\mathbf{b})}$, which means that one has complete symmetry between constructive and destructive interference, so that the symmetric bias condition is satisfied. We now show that $\alpha = 2$ is the *only* value of α that satisfies this condition.

The symmetric bias condition requires that $\beta_+ = \beta_-$ for any given $p(\mathbf{a}), p(\mathbf{b})$ provided that $p(\mathbf{a}) + p(\mathbf{b}) \leq 1$ and $p(\mathbf{a} + \mathbf{b}) \leq 1$ (see Figure 7). This must hold in the special case where $p(\mathbf{b}) \ll p(\mathbf{a})$, in which case, since $\alpha \geq 1$

$$\big[p(\mathbf{a})^{1/\alpha} \pm p(\mathbf{b})^{1/\alpha}\big]^\alpha = p(\mathbf{a})\left[1 \pm \alpha\left[\frac{p(\mathbf{b})}{p(\mathbf{a})}\right]^{1/\alpha} + \frac{\alpha(\alpha - 1)}{2}\left[\frac{p(\mathbf{b})}{p(\mathbf{a})}\right]^{2/\alpha} \pm \dots\right] \tag{90}$$

and, to second order in $\left[p(\mathbf{b})/p(\mathbf{a})\right]^{1/\alpha}$

$$\beta_+ = \alpha p(\mathbf{a}) \left[\frac{p(\mathbf{b})}{p(\mathbf{a})}\right]^{1/\alpha} + \frac{\alpha(\alpha-1)}{2} p(\mathbf{a}) \left[\frac{p(\mathbf{b})}{p(\mathbf{a})}\right]^{2/\alpha} - p(\mathbf{b}) \tag{91}$$

$$\beta_- = \alpha p(\mathbf{a}) \left[\frac{p(\mathbf{b})}{p(\mathbf{a})}\right]^{1/\alpha} - \frac{\alpha(\alpha-1)}{2} p(\mathbf{a}) \left[\frac{p(\mathbf{b})}{p(\mathbf{a})}\right]^{2/\alpha} + p(\mathbf{b}) \tag{92}$$

The symmetric bias condition $\beta_+ = \beta_-$ then implies that

$$\frac{\alpha(\alpha-1)}{2} \left[\frac{p(\mathbf{b})}{p(\mathbf{a})}\right]^{-1+2/\alpha} = 1 \tag{93}$$

which must hold for any ratio $p(\mathbf{b})/p(\mathbf{a})$, which implies that $\alpha = 2$. More generally, when $\alpha = 2$, Equations (88) and (89) imply that $\beta_+ = \beta_- = 2\sqrt{p(\mathbf{a})p(\mathbf{b})}$.

Therefore, the symmetric bias condition is satisfied by case (C1) only if $\alpha = 2$.

Figure 7. Graphs (a) and (b) show, as a function of α for the indicated values of $p(\mathbf{a})$ and $p(\mathbf{b})$, (i) the extreme values of $p(\mathbf{a}+\mathbf{b})$, (ii) the average of these extrema, and (iii) the value of $p(\mathbf{a}) + p(\mathbf{b})$. In both cases, the average of the extrema coincides with the value of $p(\mathbf{a}) + p(\mathbf{b})$ only when $\alpha = 2$.

(a)

(b)
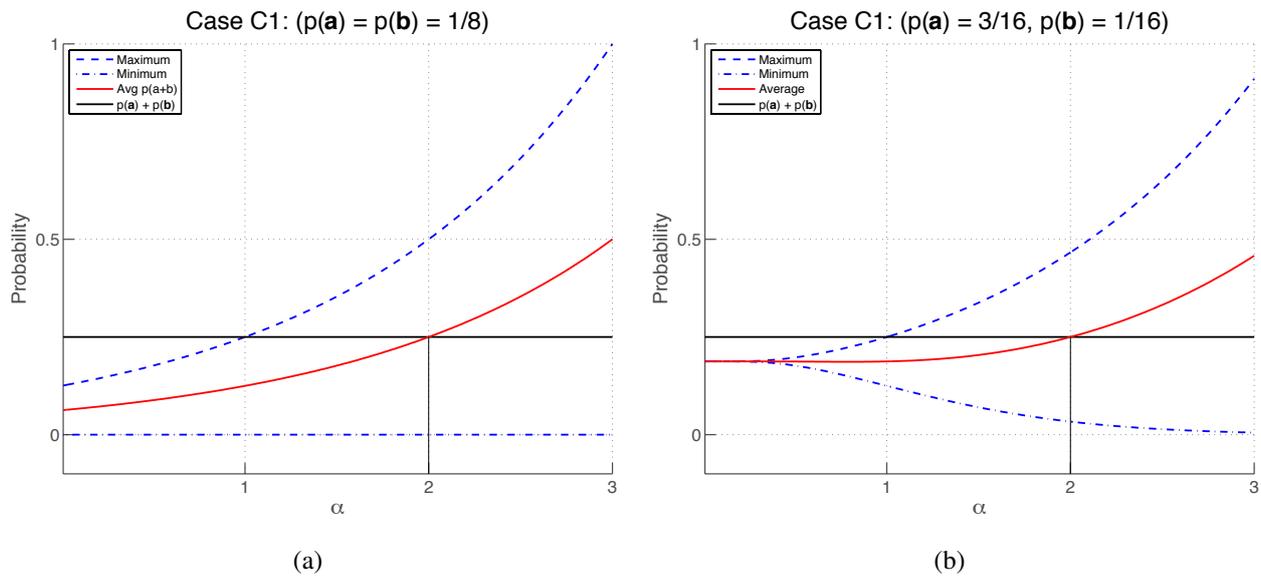
Case (C3), with $p(\mathbf{x}) = |x_1 x_2|^\alpha$

From Equation (84), the biases are

$$\beta_+ = \left[p(\mathbf{a})^{1/2\alpha} + p(\mathbf{b})^{1/2\alpha}\right]^{2\alpha} - \left[p(\mathbf{a}) + p(\mathbf{b})\right] \tag{94}$$

$$\beta_- = p(\mathbf{a}) + p(\mathbf{b}) \tag{95}$$

The symmetric bias condition requires that $\beta_+ = \beta_-$ for any $p(\mathbf{a})$, $p(\mathbf{b})$ provided that $p(\mathbf{a}) + p(\mathbf{b}) \leq 1$ and $p(\mathbf{a}+\mathbf{b}) \leq 1$. In particular, if $p(\mathbf{a}) = p(\mathbf{b}) = p$, it follows from $\beta_+ = \beta_-$ that

$$2^{2\alpha}p = 4p \tag{96}$$

which implies $\alpha = 1$. With this setting, using Equations (94) and (95), $\beta_+ = \beta_-$ becomes

$$2\sqrt{p(\mathbf{a})p(\mathbf{b})} = p(\mathbf{a}) + p(\mathbf{b}) \tag{97}$$

which, as illustrated in Figure 8, cannot hold for all $p(\mathbf{a}), p(\mathbf{b})$. Therefore, the symmetric bias condition cannot be satisfied by case (C3) for any value of α.

> **Figure 8.** Graphs (a) and (b) show, as a function of α for the indicated values of $p(\mathbf{a})$ and $p(\mathbf{b})$, (i) the extreme values of $p(\mathbf{a} + \mathbf{b})$, (ii) the average of these extrema, and (iii) the value of $p(\mathbf{a}) + p(\mathbf{b})$. The average of the extrema coincides with the value of $p(\mathbf{a}) + p(\mathbf{b})$ at *different* values of α in the two graphs.
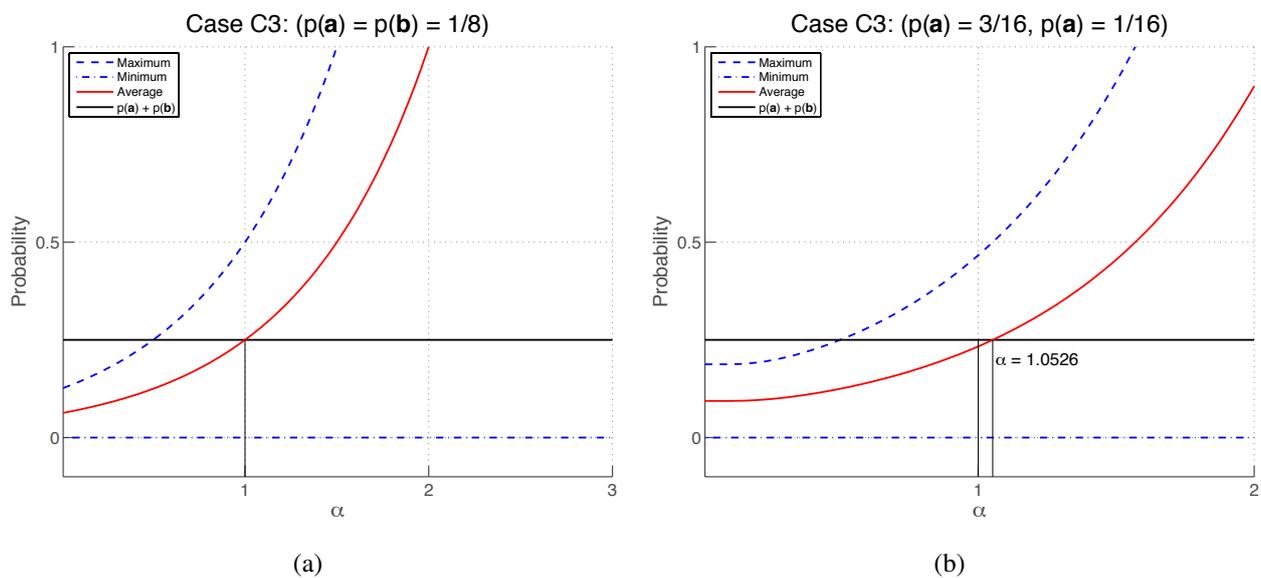
5. Summary

In Section 4, we have shown that Feynman's rules of quantum theory can be derived from an experimental logic through a pair-valued representation. In particular, we have shown the following. First, to combine two sequences in parallel, one combines the pairs $\mathbf{a}$ and $\mathbf{b}$ that represent these sequences using the sum rule of Equation (69)

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \oplus \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix}$$

which we recognize as complex addition of the pairs. In order to combine two sequences in series, we have shown that one must use (C1), so that, for pairs $\mathbf{a}$ and $\mathbf{b}$

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \odot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 b_1 - a_2 b_2 \\ a_1 b_2 + a_2 b_1 \end{pmatrix}$$

which we recognize as complex multiplication. Hence the number pairs $\mathbf{a}, \mathbf{b}, \ldots$ combine according to the rules of complex arithmetic. Finally, the probability associated with a sequence is given by form (C1) with $\alpha = 2$, so that

$$p(\mathbf{x}) = x_1^2 + x_2^2$$

These are Feynman's rules of quantum theory.

In Figure 9, we summarize the relationship of the *space of sequences* (and their complex-valued representation) and the corresponding *space of conditional logical statements* (and their probabilities) for sequences A and B combined in parallel to yield sequence $C = A \vee B$. The diagram illustrates a number of points of crucial importance in understanding the relationship between quantum theory and probability theory:

1. As shown on the right hand side of the diagram, statements **A**, **B**, and **C** are all atomic; in particular, **C** cannot be obtained from **A** and **B** by means of any Boolean logical operations.

2. In probability theory unfettered by additional constraints, the probabilities of the atomic statements **A**, **B**, and **C** would be freely assignable (see Section 2). However, additional constraints *do* exist, as a result of which these probabilities are *not* freely assignable. More precisely, (i) due to the amplitude sum rule operative in the pair space, the pair representing sequence C is determined by the pairs representing A and B and (ii) due to the postulated connection between the sequence space and the statement space, the probability of **C** is determined by the pairs representing sequences A and B. That is, once z_1 and z_2 are fixed, the probabilities of not only propositions **A**, **B**, but *also* proposition **C**, are determined.

3. In the statement space, the probability of proposition **C** is not *independent* of the probabilities of propositions **A**, **B**, but, on the other hand, is not *determined* by them either. The lee-way that exists in the probability of **C** even after the probabilities of **A** and **B** have been fixed arises because these *three* probabilities are determined through *three* independent degrees of freedom, namely $|z_1|$, $|z_2|$, and $\arg(z_1/z_2)$, in pair space.

4. In the statement space, one can construct the statement **A** $\vee$ **B** from **A** and **B** using the Boolean OR operation. The probability of **A** $\vee$ **B** is determined by the sum rule of probability theory that is operative in the probability space (which, in turn, results from the associative symmetry of the logical OR operation). In particular, statement **A** $\vee$ **B** is *not* the same as **C**.

5. The application of probability theory *alone* does not predict any quantitative relation between the probabilities of **A** $\vee$ **B** and **C**. If one adds the appropriate assumption from classical physics, then these two propositions can be *equated*, which implies that the probability of **C** is given by the probability of **A** $\vee$ **B**. Feynman's rules posit an alternative set of assumptions (which we have explicitly identified in the process of deriving Feynman's rules), which lead to the assignment of *different* probabilities to these two propositions.

5.1. *Real Quantum Theory*

It is interesting to consider what happens if we seek a real *scalar*-valued representation of the process logic rather than a *pair*-valued representation. In that case, one finds that

$$a \oplus b = a + b \tag{98}$$

$$a \odot b = ab \tag{99}$$

Figure 9. A diagram illustrating the connection between the space of measurement sequences and the space of statements. On the left hand side, the sequences A and B are combined together in parallel to generate sequence $C = A \vee B$. If amplitudes z_1 and z_2 represent sequences A and B, respectively, then, by the amplitude sum rule, amplitude $z_1 + z_2$ represents sequence C. On the right hand side, corresponding to the sequences A, B, and C are the atomic statements **A**, **B** and **C**, with probabilities $|z_1|^2$, $|z_2|^2$, and $|z_1 + z_2|^2$, respectively. Note that the probability associated with **C** is *not* freely assignable due to the postulated connection between the two spaces. Also shown is the statement **A** $\vee$ **B**, which is distinct from **C**, and which has probability $|z_1|^2 + |z_2|^2$ determined by the sum rule of probability theory.
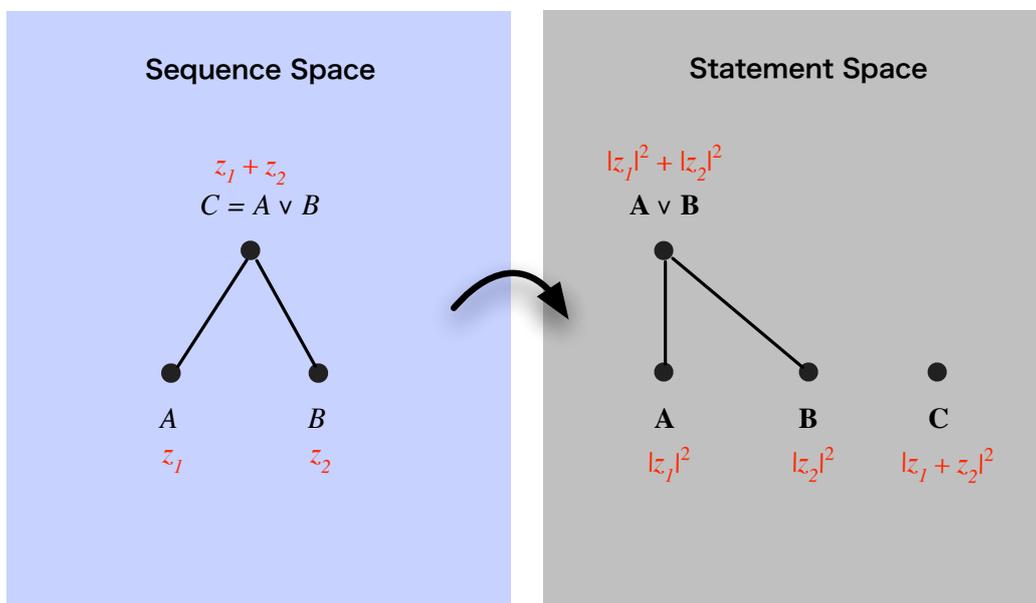
and Equation (71) together with the probability product equation, Equation (72), yield

$$p(a) = |a|^\alpha \tag{100}$$

If one considers sequences A and B with respective scalars a and b, with respective associated probabilities $p_1 = p(a) = |a|^\alpha$ and $p_2 = p(b) = |b|^\alpha$, their parallel combination $A \vee B$ has scalar $a + b$, with associated probability

$$p(a + b) = |a + b|^\alpha \tag{101}$$

$$= \left| p_1^{1/\alpha} + \gamma\, p_2^{1/\alpha} \right|^\alpha \tag{102}$$

where $\gamma = \text{sgn}(a/b)$. The maximum and minimum values of $p(a + b)$ are therefore $\left(p_1^{1/\alpha} + p_2^{1/\alpha} \right)^\alpha$ and $\left| p_1^{1/\alpha} - p_2^{1/\alpha} \right|^\alpha$, respectively, so that, just as in Section 4.8, the symmetric bias condition implies that $\alpha = 2$.

Hence, with a real representation of the experimental logic, we recover Feynman's rules of quantum theory with amplitudes restricted to the real numbers, a formalism referred to as real quantum theory.

In particular, we do *not* recover the predictions made using probability theory within the framework of classical physics.

6. Conclusions

In this paper, we have shown that, by harnessing the symmetries in an experimental logic, and requiring correspondence to probability theory, it is possible to derive the core of the quantum formalism. The key physical inputs in the derivation are the pair-valued representation of the experimental logic, which is inspired by the principle of complementarity, and the symmetric bias condition, which ensures that the predictions of the process calculus are (in some precise sense) symmetrically placed around the predictions that arise from the application of probability theory on the assumption that two processes combined in parallel are probabilistically independent.

Hence, by explicit derivation, we have shown that Feynman's rules can be understood as a probabilistic calculus whose predictions coincide with probability theory whenever probability theory is applicable, but which also yields predictions in certain situations where probability theory alone yields no predictions.

As we mentioned in the Introduction, the view is sometimes expressed that Feynman's rules are inconsistent with probability theory. This misconception, which is unfortunately fairly widespread, stems from the failure to appreciate the fundamental origins of the probability calculus, and a failure to distinguish between *probability theory* on the one hand, and an *assumption that has its roots in classical physics* on the other. As we have shown, probability theory is a precise and controlled generalization of the process of deductive reasoning embodied in Boolean logic, and is not dependent upon assumptions peculiar to classical physics. Furthermore, as we have illustrated in our discussion of the double-slit experiment (Section 3), this classical assumption amounts to assuming that a system (such as an electron) travels through a screen via one of two slits *even though it was not measured doing so*. By deriving Feynman's rules while making use of probability theory in its unmodified form, we have explicitly demonstrated that it is this classical assumption which is at fault, *not* probability theory.

It is interesting to note that, from a conceptual standpoint, the above-mentioned classical assumption is highly speculative in that it asks us to believe something about what happened without the benefit of having made a measurement to verify that it is actually the case. That we are usually willing to grant this assumption is, *as humans,* the result of our long training with macroscopic objects in everyday life, and is, *as physicists,* the result of our long habituation to the assumptions embodied in the framework of classical physics. If one, however, exercises metaphysical caution and refrains from making this assumption, one opens up the possibility of a richer predictive theory. In this sense, quantum theory arises when we stay *closer* to the actual observed phenomena; when we are *more* wary about accepting statements that are not based on empirical data. In that sense, classical physics is *more* metaphysically speculative (and correspondingly *less* well grounded in empirical data) than quantum physics.

Another view that is sometimes expressed is that quantum theory is incompatible with Boolean logic. In this view, which can be traced to Birkhoff and von Neumann [26], the distributive law of Boolean logic fails when one is dealing with certain propositions concerning the properties of quantum systems, and Boolean logic must accordingly be replaced by a so-called quantum logic that is abstracted from the quantum formalism. However, this cannot be the case: as we have demonstrated here, Feynman's

rules are entirely compatible with Boolean logic. Indeed, our derivation of Feynman's rules *depends* upon the validity of the latter—in particular, to each sequence of experiment outcomes is associated a logical proposition that is subject to the usual rules of Boolean logic (as well as the rules of probability theory, which we have shown are a systematic generalization of the Boolean logic). We suspect that this erroneous view arises from a failure similar to that mentioned above in connection with the supposed incompatibility of probability theory and quantum theory, namely a failure to distinguish between Boolean logic on the one hand, and additional assumptions rooted in a particular view of physical reality on the other. Indeed, a significant number of workers in the quantum logic area share the view that quantum logic is not to be regarded as a modification of Boolean logic, but rather is to be regarded as an operational (or experimental) logic rather like the experimental logic we describe above (see [27], for example). It would be interesting to make a comparison of the quantum logic and our experimental logic. It would also be instructive to explicitly pinpoint the misconception that underpins the view that quantum theory is incompatible with Boolean logic. We hope to return to both of these tasks in subsequent papers.

Although our paper has been primarily concerned with delineating the relationship between probability theory and quantum theory, the results presented here have much broader implications. First, our derivation of Feynman's rules does not employ most of the fundamental notions (such as space, mass, energy and momentum) of classical physics. A large part of the input to the derivation is the experimental logic and the consistency conditions emanating from probability theory, and the only other significant input which could be regarded as *physical* is the *pair-valued* representation of sequences. These observations suggest that the quantum formalism is logically prior to most of the concepts we ordinarily regard as fundamental to our description of the physical world, in turn suggesting that the quantum formalism is much closer in its nature to probability theory than was ever previously suspected. Not only do these observations have important implications for the further development of physics (as we elaborate in [11]), but they also suggest that is may be possible to develop a clearer understanding of the meaning of the recent application of the quantum formalism in areas outside physics such as in artificial intelligence, cognition, and psychology (for example, [28,29]).

Second, the methodology we have employed appears to have many possible applications. As we have illustrated, the key idea behind the methodology is to start with a logic which captures the relationships between entities of interest (be they logical propositions, measurement sequences, or something else) and then to derive a *calculus* by *suitable quantification* of this logic. As we have seen, the derivation of quantum theory involves the crucial further step of *connecting* the pair calculus together with the probability calculus, and then using conditions expressed in terms of the latter to constrain the former. The methodology has the great benefit of being transparent and highly systematic. Its efficacy in deriving quantum theory raises the question of whether the same methodology can be applied to better understand other existing physical theories or even to derive new ones. In this connection, one of us is currently exploring whether the application of this methodology to quantify causal sets of events can aid the understanding of space-time structure as an emergent phenomenon [30]. The efficacy of the methodology also raises the question of whether it can be used to derive calculi in other areas of science. This indeed appears to be the case: as mentioned in the Introduction, we have already used this methodology to derive the axioms of measure theory [7] and to develop a new calculus of questions [7,8]. However, we suspect that this work merely scratches the surface of what is possible.

Acknowledgements

The authors would like to thank John Skilling for very helpful discussions and his efforts on the paper that led to the present work. Philip Goyal would like to thank Yiton Fu for very helpful discussions. Kevin Knuth would like to thank Ariel Caticha and Keith Earle for many insightful discussions.

References and Notes

1. Feynman, R.P. Space-time approach to non-relativistic quantum mechanics. *Rev. Mod. Phys.* **1948**, *20*, 367.
2. Cox, R.T. Probability, frequency, and reasonable expectation. *Am. J. Phys.* **1946**, *14*, 1–13.
3. Cox, R.T. *The Algebra of Probable Inference*; The Johns Hopkins Press: Baltimore, MD, USA, 1961.
4. Kolmogorov, A.N. *Foundations of Probability Theory*; Julius Springer: Berlin, Germany, 1933.
5. Boole, G. *An Investigation of the Laws of Thought*; Macmillan: London, UK, 1854.
6. Knuth, K.H. Deriving laws from ordering relations. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, Proceedings of 23rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering; Erickson, G.J., Zhai, Y., Eds.; American Institute of Physics: New York, NY, USA, 2004; pp. 204–235.
7. Knuth, K.H. Measuring on lattices. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, Proceedings of 23rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering; Goggans, P., Chan, C.Y., Eds.; American Institute of Physics: New York, NY, USA, 2004; Volume 707, pp. 132–144.
8. Knuth, K.H. Valuations on lattices and their application to information theory. In *Proceedings of the 2006 IEEE World Congress on Computational Intelligence*, Vancouver, Canada, July 2006.
9. Knuth, K.H. Lattice duality: The origin of probability and entropy. *Neurocomputing* **2005**, *67C*, 245–274.
10. Knuth, K.H.; Skilling, J. The foundations of inference. Available online: http://arxiv.org/abs/1008.4831 (accessed on 27 April 2011).
11. Goyal, P.; Knuth, K.H.; Skilling, J. Origin of complex quantum amplitudes and feynman's rules. *Phys. Rev. A* **2010**, *A81*, 022109:01–022109:12.
12. Bohr, N. Causality and complementarity. *Philos. Sci.* **1937**, *4*, 289–298.
13. Aczél, J. *Lectures on Functional Equations and Their Applications*; Academic Press: New York, NY, USA, 1966.
14. Readers familiar with lattice (order) theory, will observe that bivaluations represent a generalization of the zeta function [31,32], which is an indicator function for the Boolean lattice where

$$\zeta(\mathbf{X}, \mathbf{Y}) = 1 \quad \text{when} \quad \mathbf{X} \to \mathbf{Y}$$
$$\zeta(\mathbf{X}, \mathbf{Y}) = 0 \qquad \qquad \text{otherwise}$$

noting that the arguments of the zeta function are swapped with respect to the definition of a bivaluation. Furthermore, the sum rule is to be identified with the inclusion-exclusion relation.

15. Schroedinger, E. Quantisation as an eigenvalue problem. *Ann. Phys.* **1926**, *79*, 361–376.

16. Heisenberg, W. Quantum-theoretical re-interpretation of kinematic and mechanical relations. *Z. Phys.* **1925**, *33*, 879–893. Translation in [33].

17. Dirac, P. *Principles of Quantum Mechanics*, 4th ed.; Oxford Science Publications: Oxford, UK, 1999.

18. von Neumann, J. *Mathematical Foundations of Quantum Mechanics*; Princeton University Press: Princeton, NJ, USA, 1955.

19. In practice, one would probably use a single Geiger counter to detect the electrons falling on a small patch of the screen, and then move the detector over the screen to build up the intensity pattern over the screen. However, in this thought-experiment, we help ourselves to more sophisticated equipment.

20. The resolution time of a detector is the smallest interval of time between two incident electrons for which two distinct output pulses will be obtained from the detector.

21. Alternatively, we can replace the single wire-loop detector at B with two finer-grained detectors, one placed in front of each of the slits, which are capable of indicating passage through one slit or the other.

22. We remark that, in the deBroglie–Bohm interpretation of quantum theory, the model of an electron has two distinct components: (i) a discrete entity (an "indicator particle"), and (ii) a delocalized wave, which determines the motion of the discrete entity. Since the "electron" does not consist *solely* of a highly localized object in such a hybrid model, one would *not* infer that proposition **B** implies $\mathbf{B}_1 \vee \mathbf{B}_2$, and one would therefore not infer Equation (52). If, instead, one were to redefine the $\mathbf{B}_i$ to refer to the indicator particle alone, one *could* infer (52), but one would not be able to infer that $\mathrm{Pr}(\mathbf{C}, \mathbf{B}_1|\mathbf{A})$ is unaffected by the closure of slit B_2 (as the closure of B_2 would be expected to affect the wave component), and hence one could not subject Equation (52) to the experimental test given above. Thus, in either interpretation of the $\mathbf{B}_i$, the de Broglie–Bohm model of an electron would not be ruled out by the experimental test mentioned above.

23. In this model, the electron waves are taken to move at speed, v, at which speed particle-like electrons would be expected to move, with the wavelength of the waves set equal to the *de Broglie* wavelength of the electrons which, for $v \ll c$, is $\lambda = h/mv$, where h is Planck's constant, and m is the mass of the electron.

24. As mentioned in the Introduction, our choice of representation is inspired by Bohr's principle of complementarity. We are investigating other ways of understanding the origin of the pair representation.

25. This condition is inspired by another condition suggested by J. Skilling in discussion with one of us.

26. Birkhoff, G.; von Neumann, J. The logic of quantum mechanics. *Ann. Math.* **1936**, *37*, 823–843.

27. Gudder, S. *Quantum Probability*; Academic Press: London, UK, 1988.

28. Busemeyer, J.R.; Wang, Z.; Townsend, J.T. Quantum dynamics of human decision making. *J. Math. Psychol.* **2006**, *50*, 220–241.

29. Pothos, E.M.; Busemeyer, J.R. A quantum probability model explanation for violations of "rational" decision theory. *Proc. R. Soc. Lond. B* **2009**, *276*, 2171–2178.

30.   Knuth, K.H.; Bahreyni, N. A derivation of special relativity from causal sets. Available online: http://arxiv.org/abs/1005.4172 (accessed on 27 April 2011).

31.   Rota, G.C. On the foundations of combinatorial theory I. Theory of Möbius functions. *Prob. Theor. and Related Fields* **1964**, *2*, 340–368.

32.   Knuth, K.H. Deriving Laws from Ordering Relations. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering in Science and Engineering*, Proceedings of 23rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering; Erickson, G.J., Zhai, Y., Eds.; American Institute of Physics: New York, NY, USA, 2003; Volume 707, pp. 204–235.

33.   van der Waerden, B.L. *Sources of Quantum Mechanics*; Dover Publications: New York, NY, USA, 1967.