Algorithmic Information Theory and the Hidden Variable Question

Christopher Fuchs
Dept. of Physics & Astronomy, University of North Carolina
Chapel Hill, NC 27599

Abstract

This note explores, via information theory, the admissibility of certain nonlocal hidden-variable theories. Consider a pair of Stern-Gerlach devices with fixed nonparallel orientations that periodically perform spin measurements on identically prepared pairs of electrons in the singlet spin state. Suppose the outcomes are recorded as binary strings l and r (with l_n and r_n denoting their n-length prefixes). The hidden-variable theories considered here require that there exists a recursive function which may be used to transform l_n into r_n for any n. This note demonstrates that such a theory cannot reproduce all the statistical predictions of quantum mechanics. Specifically, consider an ensemble of outcome pairs (l,r). From the associated probability measure, the Shannon entropies H_n and $\bar{H}_n$ for strings l_n and pairs (l_n,r_n) may be formed. It is shown that such a theory requires that $|\bar{H}_n - H_n|$ be bounded – contrasting the quantum mechanical prediction that it grow with n.

I. Introduction

The class of inequalities initiated by Bell[1] do not absolutely exclude the possibility of hidden variables underlying the phenomena statistically described by quantum mechanics. Hidden-variable theories of the so-called *nonlocal* variety are not constrained by Bell's theorem. Although there is no pressing theoretical reason for taking the existence of such a theory seriously, it is clear that one can only truly begin to understand quantum mechanics when one first understands what it is *not*. This note will attempt to make a contribution to this end. Here a seemingly not- *a priori* unreasonable class of nonlocal hidden-variable theories called the "computable hidden-variable theories" (CHV's) will first be defined for a particular thought experiment and then shown to be inconsistent with certain statistical requirements of quantum mechanics. The reason for this procedure is to make explicit, through the language of *algorithmic information theory*,[2,3] an aspect of quantum theory hitherto seldom discussed and then demonstrate the practical use of this aspect in answering foundational questions. This aspect is that the data obtained from identical measurements performed on identically prepared systems is generally "algorithmically incompressible."

II. The Thought Experiment and the Result

The thought experiment described is a modification of the standard one used for discussions of Bell's theorem. Consider a pair of distantly separated Stern-Gerlach (SG) devices situated so to (flawlessly) measure the spins of a pair of correlated electrons. These are called the left and right devices, respectively. For definiteness, suppose that the correlated electrons are in the singlet spin state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\Big(|\uparrow\rangle_L|\downarrow\rangle_R - |\downarrow\rangle_L|\uparrow\rangle_R\Big). \qquad (1)$$

Assume that the left and right SG devices, respectively, are oriented so that they invariably measure spin along $\hat{z}$ and an axis that differs from $\hat{z}$ by a computable angle θ. "θ computable" simply means that there is an algorithm for generating the decimal expansion of θ. E.g., $\theta = \pi/6$ is clearly computable. Suppose that at periodic time intervals these devices are supplied with identically prepared pairs of correlated electrons. (This would allow the measurement outcomes to serve as a "window" into the "hidden" dynamics of the devices, if such a dynamics did indeed exist.) Finally, imagine that each SG device is endowed with the capability of recording its measurement outcomes as a string of binary digits – 0 and 1 denoting down and up outcomes, respectively. Denote the left and right strings, respectively, by l and r, and their n-length prefixes by l_n and r_n. E.g., a typical run of the devices might give $l = 01101011\ldots$ and $r = 10110100\ldots$; the length-4 prefixes for these strings are $l_4 = 0110$ and $r_4 = 1011$. To cap off the description of the thought experiment, assume that there is in fact an ensemble of such devices: each macroscopically identical to the next, each with its own supply of electrons, and each performing the operation described. Associated with this ensemble will be an ensemble of ordered pairs (l,r) and consequently ensembles of pairs (l_n,r_n).

With this as a scaffold, the CHV notion can be formalized. Simply put, a CHV is said to be responsible for the measurement outcomes if for every pair (l,r) in the ensemble, there is at least one of a finite set of computer programs (more formally recursive functions) that, for any n, produces the string r_n as output whenever given l_n as input. Note that each string l can have as its origin any process whatsoever:

deterministic or indeterministic. This definition only requires that a rigid, mechanistic relation between l and r be maintained. Furthermore, such a theory is inherently nonlocal; the program provides the "medium" for the instantaneous action of the one string on the other. The finiteness of the set of programs is meant to allow for the possibility that the SG devices in the ensemble might have differing microscopic initial conditions.

The main result may now be stated. Because a CHV provides a compression scheme for the measurement data, it must contradict the statistical predictions of quantum mechanics. Suppose the probability distributions for the ensembles of strings l_n and pairs (l_n, r_n) are $p_n(l_n)$ and $q_n(l_n, r_n)$, respectively. The Shannon entropies for these distributions are:

$$H_n = -\sum_{l_n} p_n(l_n) \, log[p_n(l_n)]$$
$$\text{and} \tag{2}$$
$$\overline{H}_n = -\sum_{l_n} \sum_{r_n} q_n(l_n, r_n) \, log[q_n(l_n, r_n)],$$

where *log* denotes the base-2 logarithm. Consider the quantity $|\overline{H}_n - H_n|$. Standard quantum theory requires that this be proportional to n. For CHV's, however, this quantity is necessarily bound by a constant independent of n. The remainder of this section will be devoted to justifying the quantum mechanical result; the corresponding result for a CHV will be derived in the next two sections.

Suppose that standard quantum theory does indeed hold in the thought experiment. In that case, the required Shannon entropies are straightforward to derive. The essential ingredient in this derivation is simply noted: quantum theory declares that the only condition determining the measurement outcomes is the probability distribution derivable from (1). Hence, the probability of a 0 or a 1 occurring in the k'th place of a string l_n must be independent of k. Furthermore, this probability is independent of which left-hand SG device in the ensemble produced l_n. Analogous results hold for any string r_n and for the correlation probabilities between the k'th places of l_n and r_n. With these considerations, it is a simple exercise in quantum mechanics to show that

$$|\overline{H}_n - H_n| = -f(\theta) n, \text{ where}$$

$$f(\theta) = \left(sin^2\frac{\theta}{2} \, log[sin^2\frac{\theta}{2}] + cos^2\frac{\theta}{2} \, log[cos^2\frac{\theta}{2}] \right).$$

Therefore, for any θ other than $\theta = 0$ or $\theta = \pi$,

$$|\overline{H}_n - H_n| \propto n. \tag{3}$$

IV. Algorithmic Information Theory

This section introduces enough of the apparatus of algorithmic information theory that the main result can be proven. It does not purport to be a general introduction to the subject. The notion of a "recursive function" is taken as primitive. For the most part, this section follows the development of algorithmic information theory found in Ref. 3.

Notation and Definitions:

Let $\mathbf{X} = \{\Lambda, 0, 1, 00, 01, \ldots\}$ be the set of finite binary strings in lexicographic order, where Λ is the empty string. Elements of $\mathbf{X}$ may be thought of dually as strings and natural numbers. Let $\mathbf{X}_n$ be the set of n-length strings. $\mathcal{O}(1)$ denotes a bounded function. The variables s, t, v, x, and y denote elements of $\mathbf{X}$. The length, n-length prefix, and k'th digit of s are denoted by $|s|$, s_n, and $s(k)$, respectively. A set $\mathbf{S} \subset \mathbf{X}$ is called an *instantaneous code* if for any x, y in $\mathbf{S}$, neither x nor y is a prefix of the other. Elements of $\mathbf{S}$ (denoted generally by r) are called *programs*. A *computer* C is a recursive function $C: \mathbf{S} \times \mathbf{X} \to \mathbf{X}$. A computer U is said to be *universal* iff for each computer C there is a constant k_C such that: if $C(r, v)$ is defined, then there exists a program $r' \in \mathbf{S}$ such that $U(r', v) = C(r, v)$ and $|r'| \leq |r| + k_C$. Let a particular countably infinite instantaneous code $\mathbf{S}$ and universal computer U be chosen as standard. Finally, let $\langle , \rangle : \mathbf{X} \times \mathbf{X} \to \mathbf{X}$ be a recursive bijection with the property that if $|s| = |t|$, then $\langle s, t \rangle = s(1)t(1)s(2)t(2)\ldots$. E.g., if $s = 011$ and $t = 101$, then $\langle s, t \rangle = 011011$.

The *algorithmic complexities* are defined by:

$$\begin{aligned} K_C(s/t) &= \min\{|r| : C(r, t) = s\} \\ K_C(s) &= K_C(s/\Lambda) \\ K_C(s, t) &= K_C(\langle s, t \rangle) \\ K(s/t) &= K_U(s/t) \end{aligned}$$

The *canonical program* s^* for s is defined by $s^* = \min\{r : U(r, \Lambda) = s\}$. Clearly $|s^*| = K(s)$.

Now let $p: \mathbf{X} \to [0, 1]$ and let p_n denote the restriction of p to $\mathbf{X}_n$. p is said to be a *probability measure for a stochastic process* if it satisfies:

$$\sum_{|x| = n} p_n(x) = 1 \quad \& \quad p_{n-1}(y) = p_n(y0) + p_n(y1)$$

for any n and any $y \in \mathbf{X}_{n-1}$. If p is recursive, then p is said to be a *computable* measure. In this section, only computable measures are considered. The *Shannon entropy* H_n for p over n-length strings is:

$$H_n = -\sum_{|x| = n} p_n(x) \, log[p_n(x)].$$

Finally, with the measure p, the *average complexities* $\langle K/y \rangle_p^n$ and $\langle K \rangle_p^n$ for n-length strings are defined by:

$$\langle K/y \rangle_p^n = \sum_{|x| = n} p_n(x) K(x/y) \quad \& \quad \langle K \rangle_p^n = \langle K/\Lambda \rangle_p^n.$$

Theorems:

Theorems (a)-(f), from Ref. 3, are listed so that the present treatment will be self-contained. Theorem (i), from Ref. 4, provides the link relating complexity to entropy. Theorems (g), (h), and (j) are simple results due to the author. When crucial, rather than relegating bounded terms to a term written as $\mathcal{O}(1)$, a constant written in the form $D_{\dots}$, where the ellipsis symbolizes a set of subscripts, will be used so that all dependencies are clear. (E.g., D_p denotes a constant that depends only on the measure p.)

(a) For any computer C, $K(s/t) \leq K_C(s/t) + k_C$.

(b) $K(s) \leq K(s,t) + \mathcal{O}(1)$.

(c) $K(s/t) \leq K(s) + \mathcal{O}(1)$.

(d) $K(s,t) = K(t,s) + \mathcal{O}(1)$.

(e) $K(s,t) = K(s) + K(t/s^*) + \mathcal{O}(1)$. (To make the $\mathcal{O}(1)$ term's dependence on U explicit, this can be written as $|\, K(s,t) - K(s) - K(t/s^*) \,| \leq D_U$.)

(f) If $f: \mathbb{N} \to \mathbb{N}$ is a recursive function and $\sum \left(\frac{1}{2}\right)^{f(n)}$ converges, then $K(n) \leq f(n) + \mathcal{O}(1)$.

(g) $K(t/s^*) \leq K(t/s) + \mathcal{O}(1)$.

Proof: Consider a computer C such that $C(r,v) = U(r, U(v, \Lambda))$. Then $C(r, s^*) = U(r,s)$. Hence it must be the case that $K_C(t/s^*) = K(t/s)$ for any t. By (a) then, $K(t/s^*) \leq K(t/s) + \mathcal{O}(1)$. □

(h) $-D_U \leq K(s) - K(s/\,|\,s\,|^*) \leq 2log\,|\,s\,| + D_U$.

Proof: A similar result is derived in Ref. 4. The left-hand inequality is a consequence of (c). By (f), there is a constant D_U such that $K(n) \leq 2\,log(n) + D_U$ for all n. The right-hand inequality then follows from successive applications of (b), (d), and (e). □

(i) There is a constant $D_{U,p}$ such that, for all n, $0 \leq \langle K/n^* \rangle_p^n - H_n \leq D_{U,p}$.

(j) There are constants D_U and $D_{U,p}$ such that, for all n, $-D_U \leq \langle K \rangle_p^n - H_n \leq 2log\,n + D_{U,p}$.

Proof: This is a simple consequence of (h) and (i). □

V. Computable Hidden-Variable Theories

Armed with the last section's tools, a precise definition of a CHV can now be formed. Let $\mathcal{H}$ denote the set of all possible pairs (l,r) in the ensemble of strings produced by the thought experiment.

Def: A CHV V is said to be responsible for the measurement outcomes if there is a finite subset $\mathbf{V} \subset \mathbf{S}$ such that for each $(l,r) \in \mathcal{H}$ there exists a $v \in \mathbf{V}$ for which it is the case that $U(v, l_n) = r_n$ for every n.

Notice immediately that if V is responsible for the outcome strings, then for each $(l,r) \in \mathcal{H}$ it follows that $K(r_n/l_n) \leq \max\{\,|\,v\,| : v \in \mathbf{V}\}$ for all n. But then by (g), $K(r_n/l_n^*) \leq D_{U,V}$ for all n. This, coupled with (e), leads to the following conclusion.

Thm 1: If V is responsible for the measurement outcomes, then for each $(l,r) \in \mathcal{H}$, it follows that $|\, K(l_n, r_n) - K(l_n) \,| \leq D_{U,V}$ for all n.

Now consider the probability measures for the outcome strings using the notation introduced in Section II. It is assumed that these measures are computable. (This will be the case if standard quantum theory is valid since θ is required to be computable. If it were not the case here, by being noncomputable, p and q would trivially differ from the values predicted by quantum mechanics and there would be no need for further discussion.) For these measures: $\sum_{r_n} q_n(l_n, r_n) = p_n(l_n)$ for all n. An important fact to note is that $q_n(s_n, t_n)$ vanishes iff $(s,t) \notin \mathcal{H}$. Hence from Theorem 1 it follows that, if V generates the measurement outcomes, for all n, the quantity

$$\left|\, \sum_{l_n, r_n} q_n(l_n, r_n)\, K(l_n, r_n) - \sum_{l_n} p_n(l_n)\, K(l_n) \right|$$

will be bounded by a constant $D_{U,V}$. Now because of the form of the bijection $\langle l_n, r_n \rangle$, the double sum in this expression may be construed as a single sum over strings of length $2n$. This fact leads to the following:

Thm 2: If V is responsible for the outcome strings, $|\, \langle K \rangle_q^{2n} - \langle K \rangle_p^n \,| \leq D_{U,V}$ for all n.

Combining Theorems 2 and (j), the following emerges:

Thm 3: If V is responsible for the outcome strings, $|\, \bar{H}_n - H_n \,| \leq D_{U,V,p,q}$ for all n.

This is the sought after identity; for, although $D_{U,V,p,q}$ depends on the the CHV explicitly (through V and possibly p and q), it is independent of n.

I wish to thank D. Brown, G. Comer, M. Fuchs, J. Kinsella, and J. W. York, Jr. for helpful discussions and encouragement. This work was supported by NSF Grant PHY-8908741.

References

[1] For a more recent effort see S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **61**, 662 (1988).

[2] G. J. Chaitin, *Information, Randomness & Incompleteness* (World Scientific, Singapore, 1987); W. H. Zurek, Phys. Rev. A **40**, 4731 (1989).

[3] G. J. Chaitin, J. ACM **22**, 329 (1975).

[4] S. K. Leung-Yan-Cheong and T. M. Cover, IEEE Trans. Inf. Theory **IT-24**, 331 (1978).