Contextuality and Random Variables

Ehtibar Dzhafarov

Abstract Contextuality is a property of systems of random variables. The identity of a random variable in a system is determined by its joint distribution with all other random variables in the same context. When context changes, a variable measuring some property is instantly replaced by another random variable measuring the same property, or instantly disappears if this property is not measured in the new context. This replacement/disappearance requires no action, signaling, or disturbance, although it does not exclude them. The difference between two random variables measuring the same property in different contexts is measured by their maximal coupling, and the system is noncontextual if one of its overall couplings has these maximal couplings as its marginals.

1 Preamble

Quantum physicists like telling people (perhaps even themselves) that their field is strange and counterintuitive. Contextuality, especially when it takes the form of nonlocality, is one of its strange and counterintuitive notions. Indications of contextuality, such as violations of the Bell-type inequalities, are sometimes referred to as paradoxes. Like everything else in quantum physics, contextuality involves probabilities, hence random variables. And these, unlike the physical issues described by them, are usually taken to be clear and well-known: nothing strange or counterintuitive about random variables, they are merely mathematical tools, on a par with derivatives and integrals.

However, if I had any propensity to mystify my readers, I would argue that random variables are very strange objects. A random variable is a pure potentiality until it is realized, i.e., until it "collapses" into a single value being observed. Why is it less intriguing that the measurement problem in quantum physics (the wonderment

Ehtibar Dzhafarov
Purdue University, West Lafayette, IN, USA e-mail: ehtibar@purdue.edu

at why the Schrödinger wave, which is essentially a special way of describing a random variable with a spatiotemporal distribution, collapses into a specific value being observed)? The textbook definition says that a random variable is a function from a probability space to a measurable space, but one would look in vain for any utilization of this fact in the quantum physical literature. What are these probability spaces on which the random variables are defined? How does one know that two observations belong to a single random variable rather than two different variables? Can one speak of values of a random variable counterfactually, in terms of what its value might have been, had it not been what it was observed to be? Questions and wonderments like this can be multiplied. I am not, however, into mystifying my readers. All these questions have clear answers, but this clarity is not of an evident variety, one cannot achieve it without nontrivial conceptual work. And once one achieves clarity about random variables, I will argue, this clarity is imparted on the substantive issues they describe, contextuality including.

One researcher who forcefully argued that contextuality and nonlocality are primarily matters of probability theory rather than physics was Andrei Khrennikov [1,2]. He does not seem to maintain this position currently, and his arguments when he maintained it were different from those presented in this paper. Nevertheless, I think my views are close to Khrennikov's former views in spirit.

2 Random variables within a system

In 1989 David Mermin published a popular-level discussion of the nonlocal form of contextuality [3] (based on his 1981 work, added to [3] as an appendix). I will present Mermin's reasoning in a modified form. Consider the well-known Alice-Bob scenario, in which Alice chooses between two settings, denoted 1 and 2, and Bob chooses between his two settings, also denoted 1 and 2. The outcomes of Alice's measurements at either setting can be $+1$ or -1, and the same is true for Bob's measurements. Because of the way the experiment is set up (e.g., with Alice's and Bob's measurements being spacelike separated), Alice's choice of a setting cannot influence Bob's measurements, and vice versa. Let us use the term "context ij" to describe the situation in which Alice chooses setting $i \in \{1,2\}$ and Bob chooses setting $j \in \{1,2\}$. Suppose that the outcomes of the two measurements in the four contexts have the following probabilities:

context 22	Bob's outcome $= +1$	Bob's outcome $= -1$	
Alice's outcome $= +1$	0	1/2	1/2
Alice's outcome $= -1$	1/2	0	1/2
	1/2	1/2	

contexts 11, 12, 21	Bob's outcome $= +1$	Bob's outcome $= -1$	
Alice's outcome $= +1$	1/2	0	1/2
Alice's outcome $= -1$	0	1/2	1/2
	1/2	1/2	

. (1)

This describes what is commonly referred to as a *PR-box* [4], a highly contextual system by all reasonable measures [5], and one that violates the CHSH inequalities [6] to the maximal extent algebraically possible (I will explain this in Section 3). Mermin's reasoning is aimed at showing that there is something paradoxical ("extremely perplexing," he says) about such a system of random variables.[1]

We begin with context 11, and denote the two random variables representing the outcomes of Alice's and Bob's measurements as follows:

$$\text{context 11} \left[\begin{array}{r|l} \text{Alice's setting 1} & \text{random variable } A_1 \\ & \\ \text{Bob's setting 1} & \text{random variable } B_1 \end{array} \right. \quad \Pr[A_1 = B_1] = 1. \tag{2}$$

That the probability of $A_1 = B_1$ is 1 follows from the joint distribution (1) for context 11.

Now, Mermin proposes what he calls the *Strong Baseball Principle* (SBP),[2] that I will formulate as follows:

(SBP) if Alice's random variable at her setting i can in no way be influenced by Bob's choice of his setting, then the same random variable can represent her measurement outcomes in both context $i1$ and context $i2$; analogously, Bob's measurement outcomes at his setting j can be represented by the same random variable in both context $1j$ and context $2j$.

In particular, in context 12, Alice's measurement outcomes can be represented by the same A_1 as they are in context 11, and in context 21, Bob's measurement outcomes can be represented by the same B_1 as they are in context 11:

---

[1] Mermin does mention the distributions (1), with perfect correlations and anticorrelations, but for a detailed reasoning he uses the distributions in which the joint probabilities 1/2 and 0 are replaced with $(1+\cos\frac{\pi}{4})/4$ and $(1-\cos\frac{\pi}{4})/4$, respectively. The difference is not significant for my presentation.

[2] So dubbed because of the simile with the belief that watching a baseball game on one's TV cannot affect the game's outcome.

$$\text{context 12}\left[\begin{array}{r|l}\text{Bob's setting 2} & ?\\ \text{Alice's setting 1} & A_1\end{array}\right.$$

$$\text{context 11}\left[\begin{array}{r|l}\text{Alice's setting 1} & A_1\\ & \\ \text{Bob's setting 1} & B_1\end{array}\right. \quad \Pr[A_1 = B_1] = 1\,. \tag{3}$$

$$\text{context 21}\left[\begin{array}{r|l}\text{Bob's setting 1} & B_1\\ \text{Alice's setting 2} & ?\end{array}\right.$$

We can easily fill in the places held by the question marks. We know from (1) that the two random variables in context 12 are perfectly correlated, so once we have determined one of them, the other must copy it (and the same holds for context 21):

$$\text{context 12}\left[\begin{array}{r|l}\text{Bob's setting 2} & B_2\\ & \\ \text{Alice's setting 1} & A_1\end{array}\right. \quad \Pr[A_1 = B_2] = 1$$

$$\text{context 11}\left[\begin{array}{r|l}\text{Alice's setting 1} & A_1\\ & \\ \text{Bob's setting 1} & B_1\end{array}\right. \quad \Pr[A_1 = B_1] = 1\ . \tag{4}$$

$$\text{context 21}\left[\begin{array}{r|l}\text{Bob's setting 1} & B_1\\ & \\ \text{Alice's setting 2} & A_2\end{array}\right. \quad \Pr[A_2 = B_1] = 1$$

Then we apply SBP once again, and conclude that in context 22 the measurements by Alice and Bob can be represented by the same random variables A_2 and B_2 as they are in contexts 21 and 12, respectively:

$$\text{context 22}\left[\begin{array}{r|l}\text{Alice's setting 2} & A_2\\ & \\ \text{Bob's setting 2} & B_2\end{array}\right. \quad \Pr[A_2 = B_2] = 1. \tag{5}$$

That the probability of $A_2 = B_2$ is 1 follows from the chain

$$\Pr[A_2 = B_1] = 1, \Pr[B_1 = A_1] = 1, \Pr[A_1 = B_2] = 1. \tag{6}$$

But we know from (1) for context 22 that the probability of $A_2 = B_2$ is zero, not 1. We have run into a contradiction.

A contradiction always means that some of the assumptions made in the process of reasoning, perhaps unawares, are wrong. The first impulse one might have is to declare that the random variables with the distributions (1) are impossible, but this can easily be dismissed. One way to do this is to note that the same reasoning with the same contradiction at the end can be obtained with distributions that are empirically observed and codified by a well-established theory. This is the argument chosen by Mermin, who uses distributions that are predicted by quantum mechanics for a certain choice of the four settings (directions in which spins are measured) in the standard EPR/Bohm experiment with spin-$1/2$ particles [6] (see footnote 1). Of course, one can always challenge the validity of quantum mechanics and the veracity of the experiments corroborating its predictions (which one would have to do if the contradiction we arrived at could not be dissolved by any other means). A much better argument therefore would be to simply note that the random variables with distributions (1) exist mathematically, as appropriately chosen functions on certain probability spaces. We will get to this later, however.

Assuming we are satisfied there is nothing wrong with our distributions, where else can one seek the cause of the contradiction we derived? One might try to deny the possibility that Bob's setting have no influence on Alice's measurements (which probably remains the most commonly held interpretation of nonlocality among non-physicists). However, this would be wrong ("disquieting," Mermin says) in view of what physics says about information propagation: e.g., it is ruled out if Alice's and Bob's measurements are spacelike separated. Note that Alice has no means to infer Bob's setting because in contexts $i1$ and $i2$ the random variables representing the outcomes of her measurements have indistinguishable distributions ($+1$ and -1 occurring with equal probabilities). Therefore the hypothetical ways in which Bob's setting would influence Alice's measurements would have to be contrived to remain hidden, in addition to contradicting physical theory.

Mermin too dismisses the "hidden action at a distance" resolution of the contradiction he derives, and he suggests that the culprit is SBP.

> Many people want to conclude from this [the contradiction – E.D.] that what happens at A does depend on how the switch is set at B, which is disquieting in view of the absence of any connections between the detectors. The conclusion can be avoided, if one renounces the Strong Baseball Principle, maintaining that indeed what happens at A does not depend on how the switch is set at B, but that this is only to be understood in its statistical sense, and most emphatically cannot be applied to individual runs of the experiment. To me this alternative conclusion is every bit as wonderful as the assertion of mysterious actions at a distance. I find it quite exquisite that, setting quantum metaphysics entirely aside, one can demonstrate directly from the data and the assumption that there are no mysterious actions at a distance, that there is no conceivable way consistently to apply the Baseball Principle to individual events (p. 49).

While it is quite obvious to me that SBP is wrong, I do not think Mermin's explanation is sufficiently transparent. Let us try to understand it. Mermin says "individual runs" because in his exposition he does not even mention random variables, speaking instead of very long sequences of realizations thereof. If treated informally, this only obfuscates analysis, and if treated rigorously, complicates it. A sequence of realizations of a random variables is a random process, an indexed set of identically distributed random variables. I suggest therefore that Mermin's transplantation of a sequence of realizations from one context to another should simply be understood as placing in these contexts one and the same random variable, the way we have done this in (4) and (5). And this is what must not be done, Mermin says based on the contradiction this led us into, and I think there is no reasonable way to disagree with this prohibition. Consider any single context, say, 11 in (2). The realizations of A_1 and B_1 there come in pairs, they have therefore a well-defined joint distribution. In particular, one has an opportunity to decide, by looking at a long enough sequence of the paired realizations, whether they are perfectly correlated. By contrast, in (4), if one looks at A_1 in context 11 and A_1 in context 12, their realizations cannot co-occur, because context 11 and 12 are mutually exclusive. One has no non-arbitrary way of pairing a value of A_1 in context 11 with a value of A_1 in context 12. There is no meaningful sense of asking whether they are correlated, perfectly or otherwise. But then it means that we have made a mistake by denoting them by the same symbol: A_1 is a single random variable, even if mentioned many times, and $A_1 = A_1$ holds with probability 1. In contexts 11 and 12 therefore we have two different random variables with one and the same distribution. This is, I suggest, how one could understand Mermin's assertion "that indeed what happens at A does not depend on how the switch is set at B, but that this is only to be understood in its statistical sense."

How should we amend the representations (4) and (5) to avoid contradiction? The answer is simple: we should use different symbols for A_1 in context 11 and A_1 in context 12, e.g., denote them by A_1^{11} and A_1^{12}, respectively (and analogously for other pairs of random variables transplanted from one context to another by SBP).[3] One can write $A_1^{11} \sim A_1^{12}$, where $\sim$ means "is distributed as", but statements like $A_1^{11} = A_1^{12}$, $A_1^{11} \neq A_1^{12}$, $A_1^{11} + A_1^{12} = 2$, etc. are all void of meaning because A_1^{11} and A_1^{12} posses no joint distribution. The system of random variables in our example can now be presented as follows:

$$
\begin{tikzcd}
 & A_1^{11} \arrow[r, dash, "\Pr\left[A_1^{11}=B_1^{11}\right]=1"'] & B_1^{11} \arrow[dr, dotted, dash, "B_1^{11}\sim B_1^{21}"] & \\
A_1^{12} \arrow[ur, dotted, dash, "A_1^{12}\sim A_1^{11}"] \arrow[d, dash, "\Pr\left[A_1^{12}=B_2^{12}\right]=1"'] & & & B_1^{21} \arrow[d, dash, "\Pr\left[A_2^{21}=B_1^{21}\right]=1"] \\
B_2^{12} \arrow[dr, dotted, dash, "B_2^{22}\sim B_2^{12}"'] & & & A_2^{21} \\
 & B_2^{22} \arrow[r, dash, "\Pr\left[A_2^{22}=B_2^{22}\right]=0"] & A_2^{22} \arrow[ur, dotted, dash, "A_2^{21}\sim A_2^{22}"'] &
\end{tikzcd}
\tag{7}
$$

We are no longer driven into contradiction: the value of $\Pr\left[A_2^{22} = B_2^{22}\right]$ can in no way be inferred from other components of this diagram, because none of them contains the variables A_2^{22} and B_2^{22}. If my interpretation of Mermin's conclusion is deemed

[3] This notation for the random variables has redundancy in it: the subscripts can be recovered from the superscripts (contexts) and the symbol (A or B) used for the random variable. Thus, A_1^{11} could simply be written as A^{11}, and B_2^{12} as B^{12}. I leave the notation redundant by a deliberate choice, however, to make the structure of the random variables maximally transparent.

reasonable, his was a valuable observation for the 1980s, although I doubt it could be well understood the way it was formulated.

To generalize, a random variable within a system of random variables is identified not only by what it measures (e.g., A_1^{11} and A_1^{12} both measure the same property, say, the spin of a particle in Alice's direction 1) but also by the context in which it is recorded (here, by the directions chosen by both Alice and Bob for their measurements). Let me dispel two possible objections to this general statement.

One is that we still can write A_1 for both A_1^{11} and A_1^{12} but keep in mind that we deal with two different sequences of realizations of A_1. Indeed, one might argue, there is nothing wrong in saying that one records a random variable today, and then the same random variable is recorded tomorrow. The response to this argument is that it is acceptable only if one is allowed to be informal, hoping this will not lead to confusion. A rigorous treatment of random variables requires that whenever one attaches different contexts to them (in this example, day of the measurement, today or tomorrow), one deals with different random variables. After all, to say "A_1 today" and "A_1 tomorrow" means to denote them differently, albeit sloppily. They are (perhaps) identically distributed, but they are distinct and have no joint distribution. Recall that the rigorous definition of a sequence of realizations of a random variable R (a sample of its values) is the sequence of different random variables, $R_1, R_2, \ldots$, each of which is distributed as R. The fact that in most applications they are also assumed to be independent is more subtle, and its mathematical meaning is captured through the notion of couplings that we will discuss later.

Another, often heard objection is that by saying that A_1^{11} and A_1^{12} are different random variables, one somehow admits that something in the contexts *causes* A_1^{11} to *transform* into A_1^{12} as context 11 is replaced with context 12. So in the EPR/Bohm scenario, one might argue, we still have some kind of an action at a distance. This objection is merely a play on the words "causes" and "transforms." If A_1^{11} and A_1^{12} are identically distributed, Alice has no means to distinguish them. Which means that no information, no action is transferred from Bob's setting to Alice's measurements. A_1^{11} and A_1^{12} are different random variables only for someone who, like Mermin, gets information from both Alice and Bob, both about their settings chosen and outcomes obtained. What changes when context 11 is replaced with context 12 is the relation between Alice's measurements and Bob's measurements, and this, because of the fundamentally relational nature of random variables (as explained below), means that Mermin knows that in contexts 11 and 12 Alice deals with two different random variables.

Having dealt with these objections, can we say that with my interpretation of Mermin's analysis we, at least informally (because a more formal treatment is to follow), have explained the strangeness of contextuality in terms of random variables? The answer is, we have not. In fact, unexpected as this might come, Mermin's conclusion in my interpretation equally applies to any and all systems of random variables, contextual or not. Consider, e.g., the following modification of the distributions (1):

contexts 11, 12, 21, 22	Bob's outcome $= +1$	Bob's outcome $= -1$	
Alice's outcome $= +1$	$1/2$	0	$1/2$
Alice's outcome $= -1$	0	$1/2$	$1/2$
	$1/2$	$1/2$	

. (8)

This system is clearly noncontextual. It can be viewed as describing a single pair of perfectly correlated random variables, with setting choices being fake. It is still true, however, that the contexts 11 and 12 are mutually exclusive, and insofar as these settings are not ignored, Alice's measurements in these contexts must be represented by different (though identically distributed) random variables that have no joint distribution. The diagram representing this situation is identical to (7), except $\Pr\left[A_2^{22} = B_2^{22}\right]$ is now 1 rather than zero. This makes no difference for how one treats the random variables because in both cases $\Pr\left[A_2^{22} = B_2^{22}\right]$ is completely unrelated to other elements of the diagrams.

$$
\begin{tikzcd}
 & A_1^{11} \arrow[r, dash, "{\Pr\left[A_1^{11}=B_1^{11}\right]=1}"'] & B_1^{11} \arrow[dr, dotted, dash, "B_1^{11}\sim B_1^{21}"] & \\
A_1^{12} \arrow[ur, dotted, dash, "A_1^{12}\sim A_1^{11}"] \arrow[d, dash, "{\Pr\left[A_1^{12}=B_2^{12}\right]=1}"'] & & & B_1^{21} \arrow[d, dash, "{\Pr\left[A_2^{21}=B_1^{21}\right]=1}"] \\
B_2^{12} \arrow[dr, dotted, dash, "B_2^{22}\sim B_2^{12}"'] & & & A_2^{21} \arrow[dl, dotted, dash, "A_2^{21}\sim A_2^{22}"] \\
 & B_2^{22} \arrow[r, dash, "{\Pr\left[A_2^{22}=B_2^{22}\right]=1}"] & A_2^{22} &
\end{tikzcd} \tag{9}
$$

Somewhat paradoxically, therefore, having resolved the contradiction brought in by SBP, Mermin (or at least my interpretation of his analysis) loses the distinction between contextual and noncontextual systems. There is no way, e.g., to derive Bell-type inequalities for systems like (7) and (9), because the joint distributions involved of the four pairs of random variables

$$\left(A_1^{11}, B_1^{11}\right), \left(A_2^{21}, B_1^{21}\right), \left(A_2^{22}, B_2^{22}\right), \left(A_1^{12}, B_2^{12}\right), \tag{10}$$

are logically unrelated to each other.

Some researchers derive from this that the notion of contextuality is flawed. In particular, Bell-type inequalities, according to this view, are simply invalid, based on the mistake of following SBP. It seems that this is also Andrei Khrennikov's view [7], although his implementation of the context-dependence of random variables is different from the one presented here [8].

While one is free not to make distinctions one finds uninteresting, I find this position less than constructive. It is true that the context-indexing of random variables precludes the naive notion of contextuality, but what one better get rid of is the naivety rather than the notion. As it turns out, there is a conceptual and mathematical tool that enables us to readily distinguish systems like (7) from those like (9). As a bonus, this tool, while incompatible with SBP, justifies and formalizes the counterfactual reasoning on which SBP is based. The questions like "what would the outcome of one's measurement be if it were made in a context other than the one in which it is made" translate into rigorous and non-controversial mathematical problems.

3 Sample spaces and couplings

The mathematical tool in question is (probabilistic) *coupling*. Before introducing it, however, let us make sure we understand why two random variables in different contexts do not have a joint distribution (from which it follows also that they can never be one and the same random variable).

All random variables in this paper are assumed to be *dichotomous*,[4] because of which a random variable is defined as a function $X : S \to \{-1, 1\}$, with the following properties:

1. S belongs to a probability space (S, Σ, μ), where Σ is a sigma algebra of subsets of S, and μ a probability measure $\Sigma \to [0, 1]$;
2. $X^{-1}(\{1\}) \in \Sigma$, and $\Pr[X = 1] = \mu\left(X^{-1}(\{1\})\right)$; $X^{-1}(\{-1\}) \in \Sigma$, and $\Pr[X = -1] = \mu\left(X^{-1}(\{-1\})\right)$;
3. $\Pr[X = 1] + \Pr[X = -1] = 1$.

The set S is often called a *sample space*, but I prefer to use this term for the probability space (S, Σ, μ). Random variables X and Y are jointly distributed if and only if they are functions on the same sample space. If they are, then

$$\Pr[X = 1, Y = 1] = \mu\left(X^{-1}(\{1\}) \cap Y^{-1}(\{1\})\right). \tag{11}$$

Realizations of X and Y are then defined in pairs. If they are not on the same sample space, $\Pr[X = 1, Y = 1]$ is undefined, and no pairing scheme for their realizations exists.

We see that, by definition, to construct a set of jointly distributed random variables means to specify a sample space and define these random variables as functions on this sample space. What usually remains unclear to a student of these textbook definitions is the nature of a sample space. What is it and how can it be (re)constructed? The answer to this question is so simple that it can be surprising. Let us discuss this answer in detail, using a system of the same format as above, but

[4] In my favorite theory of contextuality, confining the consideration to dichotomous variables is not a loss of generality, because all random variables within a system have to be replaced by sets of jointly distributed dichotomous variables before the system can be subjected to contextuality analysis [9, 10].

this time with more arbitrary joint distributions in the four contexts: for $i \in \{1,2\}$ and $j \in \{1,2\}$,

$$\begin{array}{c|c|c|c}
\text{context } ij & B_j^{ij}=+1 & B_j^{ij}=-1 & \\
\hline
A_i^{ij}=+1 & r_{ij} & p_i - r_{ij} & p_i \\
\hline
A_i^{ij}=-1 & q_j - r_{ij} & 1-p_i-q_j+r_{ij} & 1-p_i \\
\hline
 & q_j & 1-q_j &
\end{array}. \tag{12}$$

The only constraint imposed by my notation here is that the distribution of A_i^{ij} is the same for $j=1$ and $j=2$ (in both cases $\Pr\left[A_i^{ij}=1\right]=p_i$), and analogously for the distribution of B_j^{ij}. This property of a system of random variables is called *consistent connectedness*, or *no-disturbance*. This is precisely the property that guarantees that Alice has no way of distinguishing A_i^{i1} and A_i^{i2}, the difference being only available to someone who receives information about both Alice's and Bob's settings and outcomes.

We assume, of course, that the distribution in (12) is well-defined, that is, all probabilities shown are numbers between 0 and 1. As it turns out, this is all one needs to say that the random variables A_i^{ij} and B_j^{ij} with this joint distribution exist as mathematical objects. Indeed, consider the following probability space (S, Σ, μ_{ij}) for context ij: $S=\{a,b,c,d\}$, $\Sigma = 2^S$, and μ_{ij} is defined by the probability mass function

$$\begin{array}{r||c|c|c|c}
x= & a & b & c & d \\
\hline
\mu_{ij}(\{x\})= & r_{ij} & p_i - r_{ij} & q_j - r_{ij} & 1-p_i-q_j+r_{ij}
\end{array}. \tag{13}$$

The random variables are now defined as the functions

$$A_i^{ij}(x) = \begin{cases} +1 \text{ if } x \in \{a,b\} \\ -1 \text{ if } x \in \{c,d\} \end{cases}, B_j^{ij}(x) = \begin{cases} +1 \text{ if } x \in \{a,c\} \\ -1 \text{ if } x \in \{b,d\} \end{cases}. \tag{14}$$

The resulting system of random variables can be presented in the form of the following *content-context matrix*:

A_1^{11}		B_1^{11}		context 11
A_1^{12}			B_2^{12}	context 12
	A_2^{21}	B_1^{21}		context 21
	A_2^{22}		B_2^{22}	context 22
Alice's 1	Alice's 2	Bob's 1	Bob's 2	Alice-Bob system

. (15)

The term *content* (of a random variable) refers to that which the random variable is measuring, or settings from which the measured property can be deduced. In the matrix above, the contents are listed at the bottom.

Clearly, up to the labeling of the values of S, this construction is unique. Moreover, any other sample space $\left(S'_{ij}, \Sigma'_{ij}, \mu'_{ij}\right)$ on which A_i^{ij} and B_j^{ij} can be defined is reducible to this (S, Σ, μ_{ij}), in the following sense: denoting by X_a the pre-image of $\left[A_i^{ij}=+1,B_j^{ij}=+1\right]$ in S'_{ij}, by X_b the pre-image of $\left[A_i^{ij}=+1,B_j^{ij}=-1\right]$, etc., one can map X_a into a, X_b into b, and so on, to define A_i^{ij} and B_j^{ij} as functions on (S,Σ,μ_{ij}). The latter therefore is the most economic sample space possible. The general logic of the construction should be clear. Whenever a joint distribution of hypothetical random variables is well-defined, these random variables exist as functions defined on a sample space, and the most economic version of the latter can be uniquely constructed from the joint distribution. There is never a situation in which one can say that random variables with a given joint distribution do not exist (provided no *a priori* constraints are imposed on their sample space).

It is also clear from this construction why A^{i1} and A^{i2} are distinct random variables even if they are identically distributed. They are defined on different sample spaces: even if one chooses to denote the elements of the sample set in the same way, $\{a,b,c,d\}$, the respective measures μ_{i1} and μ_{i2} are as distinct as are the joint distributions in contexts $i1$ and $i2$ in (12).

In the previous section I mentioned "the fundamentally relational nature of random variables," because of which the identity of random variables cannot be the same in different contexts. One can see now that this expression has a precise mathematical meaning: the sample space on which a given random variable is defined is determined by its joint distribution with all other random variables in the same context. A reasonable analogy is provided by a set of points in a metric space without coordinates. Each point is characterized by its distances to the rest of the points, so moving even one of the latter changes the point in question instantly. No spooky transfer of information is involved, these are changes that occur by definition. Of course, like all analogies, this one also has its drawbacks. In particular, it is possible to say that "this point" (one pointed at) changes its identity when other points change their positions. In a system of random variables one can only say that a random variable in one context is different from a random variable that measures the same thing in a different context. However, it seems to me that the analogy with distances does a very good job in dispelling remnants of mystery in the term "nonlocality."

Let me now introduce the conceptual tool that would allow us to speak of contextual and noncontextual systems. A *coupling* of several random variables $X_1,\ldots,X_n$ is a set of jointly distributed variables $Y_1,\ldots,Y_n$ such that $Y_i \sim X_i$ for $i=1,\ldots,n$. Note that $X_1,\ldots,X_n$ need not be jointly distributed, and in fact in all applications we are interested in, they are not. In other words, each of X_i is defined on its own sample space, whereas all Y's are defined on yet another sample space. Using the term "s*tochastically unrelated*" for random variables no two of which possess a joint distribution, $X_1,\ldots,X_n$, and $(Y_1,\ldots,Y_n)$ viewed as a single random variable, are stochastically unrelated. In relation to our discussion of Mermin's SBP, couplings can be thought of as answers to the counterfactual question "How could these random variables be jointly distributed if they were jointly distributed?"

Any set $X_1,\ldots,X_n$ of random variables has a coupling, and generally it has an infinity of couplings, i.e., infinity of $(Y_1,\ldots,Y_n)$ with different joint distributions (but the same marginal distributions, because by definition, $Y_i \sim X_i$ for $i = 1, \ldots, n$).[5] Therefore, to use couplings as a means for categorizing the sets $X_1, \ldots, X_n$, one should only be interested in whether a set $X_1, \ldots, X_n$ has a coupling *subject to some specified constraints* [11], as we will see shortly .

In the Contextuality-by-Default (CbD) theory, the notion of a coupling is applied to a system of random variables in two ways. We first construct couplings for all pairs of the same-content random variables. We have four of them in system (15):

$$\left\{A_1^{11}, A_1^{12}\right\}, \left\{A_2^{21}, A_2^{22}\right\}, \left\{B_1^{11}, B_1^{21}\right\}, \left\{B_2^{12}, B_2^{22}\right\}. \tag{16}$$

Denoting a coupling of $\left\{A_1^{11}, A_1^{12}\right\}$ by $\left(\tilde{A}_1^{11}, \tilde{A}_1^{12}\right)$, we look among these couplings for one in which the probability of $\tilde{A}_1^{11} = \tilde{A}_1^{12}$ is as large as possible, given the individual distribution of $\tilde{A}_1^{11} \sim A_1^{11}$ and $\tilde{A}_1^{12} \sim A_1^{12}$. The reason one is interested in such *maximal couplings* is that the maximal probability in question is a natural measure of similarity between A_1^{11} and A_1^{12}, when they are taken in isolation from their respective contexts. Put counterfactually, "if they were jointly distributed and no other random variables existed," they could coincide as often as this maximal probability. Suppose we have computed these maximal probabilities, and

$$\begin{array}{l} \max \Pr\left[\tilde{A}_1^{11} = \tilde{A}_1^{12}\right] = \omega_1, \max \Pr\left[\tilde{A}_2^{21} = \tilde{A}_2^{22}\right] = \omega_2, \\ \max \Pr\left[\tilde{B}_1^{11} = \tilde{B}_1^{21}\right] = \omega_3, \max \Pr\left[\tilde{B}_2^{12} = \tilde{B}_2^{22}\right] = \omega_4, \end{array} \tag{17}$$

where each of the maxima is taken over all possible couplings of the corresponding pair in (15).

We next construct a coupling for the entire system (15), or more precisely, a coupling of the four stochastically unrelated random variables

$$X_1 = \left(A_1^{11}, B_1^{11}\right), X_2 = \left(A_1^{12}, B_2^{12}\right), X_3 = \left(A_2^{21}, B_1^{21}\right), X_4 = \left(A_2^{22}, B_2^{22}\right). \tag{18}$$

Such a coupling consists of the jointly distributed random variables

$$Y_1 = \left(\bar{A}_1^{11}, \bar{B}_1^{11}\right), Y_2 = \left(\bar{A}_1^{12}, \bar{B}_2^{12}\right), Y_3 = \left(\bar{A}_2^{21}, \bar{B}_1^{21}\right), Y_4 = \left(\bar{A}_2^{22}, \bar{B}_2^{22}\right), \tag{19}$$

with $Y_i \sim X_i$ $(i = 1, \ldots, 4)$. There are generally an infinity of such couplings, and in each of them we can compute

$$\begin{array}{l} \Pr\left[\bar{A}_1^{11} = \bar{A}_1^{12}\right] = \omega_1', \Pr\left[\bar{A}_2^{21} = \bar{A}_2^{22}\right] = \omega_2', \\ \Pr\left[\bar{B}_1^{11} = \bar{B}_1^{21}\right] = \omega_3', \Pr\left[\bar{B}_2^{12} = \bar{B}_2^{22}\right] = \omega_4'. \end{array} \tag{20}$$

Obviously,

$$\omega_1' \leq \omega_1, \omega_2' \leq \omega_2, \omega_3' \leq \omega_3, \omega_4' \leq \omega_4. \tag{21}$$

What we need to determine is whether there is a coupling in which all these inequalities become equalities, i.e.,

[5] One could, obviously, create many copies of $(Y_1, \ldots, Y_n)$, identically distributed but defined on different sample spaces. We should agree therefore that we make no distinction between them: a coupling is entirely identified by its distribution.

$$\begin{array}{l}\Pr\left[\bar{A}_1^{11}=\bar{A}_1^{12}\right]=\omega_1,\ \Pr\left[\bar{A}_2^{21}=\bar{A}_2^{22}\right]=\omega_2,\\ \Pr\left[\bar{B}_1^{11}=\bar{B}_1^{21}\right]=\omega_3,\ \Pr\left[\bar{B}_2^{12}=\bar{B}_2^{22}\right]=\omega_4.\end{array} \tag{22}$$

In other words, we seek couplings of system (15) that preserve both the distributions within the contexts (as they should by the definition of a coupling) and the similarity values (17) between the content-sharing variables. If no such couplings exist, one can say that the contexts make the content-sharing random variables to be more dissimilar than they are when they were taken in isolation. Such a system is called *contextual*. Otherwise, if such a coupling exists (generally not uniquely), the system is *noncontextual*.

This simple and, I would argue, highly intuitive logic of (non)contextuality is sufficient to restore all the contextuality results obtained in the literature and then go much further, doing so without compromising the rigorous mathematics of random variables and, in particular, without falling into the SBP trap. For a consistently connected system, as in (12), the elements of each pair in (16) are identically distributed, and it is easy to see that in this case

$$\omega_1=\omega_2=\omega_3=\omega_4=1. \tag{23}$$

Indeed, the maximal coupling $\left(\tilde{A}_1^{11},\tilde{A}_1^{12}\right)$ of $\left\{A_1^{11},A_1^{12}\right\}$, e.g., has the distribution

content 1 of Alice	$\tilde{A}_1^{12}=+1$	$\tilde{A}_1^{12}=-1$	
$\tilde{A}_1^{11}=+1$	p_1	0	p_1
$\tilde{A}_1^{11},=-1$	0	$1-p_1$	$1-p_1$
	p_1	$1-p_1$	

. (24)

It follows that in this case we seek couplings (Y_1,Y_2,Y_3,Y_4) in (19) such that $\Pr\left[\bar{A}_1^{11}=\bar{A}_1^{12}\right]=1$, $\Pr\left[\bar{B}_1^{11}=\bar{B}_1^{21}\right]=1$, and so on. Equivalently, we seek a *reduced coupling* [12] in which the variables $\bar{A}_1^{11}$and $\bar{A}_1^{12}$ can be replaced with a single $\bar{A}_1$, the variables $\bar{B}_1^{11}$and $\bar{B}_1^{21}$ can be replaced with a single $\bar{B}_1$, etc.

With this reformulation, one comes as close to the intuition underlying SBP as it is possible without committing the logical error of the original SBP. Thus, the system in our opening example, with distributions (1), is contextual because otherwise it would have to have a coupling in which

$$
\begin{tikzcd}
 & \bar{A}_1^{11} \arrow[r, dash, "{\Pr[\bar{A}_1^{11}=\bar{B}_1^{11}]=1}"'] & \bar{B}_1^{11} \arrow[dr, dash, dotted, "{\Pr[\bar{B}_1^{11}=\bar{B}_1^{21}]=1}"] & \\
\bar{A}_1^{12} \arrow[ur, dash, dotted, "{\Pr[\bar{A}_1^{12}=\bar{A}_1^{11}]=1}"] \arrow[d, dash, "{\Pr[\bar{A}_1^{12}=\bar{B}_2^{12}]=1}"] & & & \bar{B}_1^{21} \arrow[d, dash, "{\Pr[\bar{A}_2^{21}=\bar{B}_1^{21}]=1}"'] \\
\bar{B}_2^{12} \arrow[dr, dash, dotted, "{\Pr[\bar{B}_2^{22}=\bar{B}_2^{12}]=1}"'] & & & \bar{A}_2^{21} \arrow[dl, dash, dotted, "{\Pr[\bar{A}_2^{21}=\bar{A}_2^{22}]=1}"] \\
 & \bar{B}_2^{22} \arrow[r, dash, "{\Pr[\bar{A}_2^{22}=\bar{B}_2^{22}]=0}"] & \bar{A}_2^{22} &
\end{tikzcd}
\tag{25}
$$

A chain of equalities it contains,

$$\Pr\left[\bar{A}_1^{11}=\bar{B}_1^{11}\right]=1,\ldots,\Pr\left[\bar{A}_2^{22}=\bar{B}_2^{22}\right]=0,\ldots,\Pr\left[\bar{A}_1^{12}=\bar{A}_1^{11}\right]=1, \tag{26}$$

is obviously impossible.

As shown in detail in [13], the language of probabilisitic couplings, when applied to consistently connected systems, allows one to formalize both the counterfactual formulation of contextuality and its formulation in terms of the hidden variable models with noncontextual mapping. I will not be repeating this discussion here. The Bell-type inequalities for consistently connected systems are derived in essentially the same way as they are derived traditionally. For instance, system (15) with distributions (12) can be shown to be contextual (i.e., not to have a coupling with the stipulated properties) if and only if

$$\max\left(\pm\left\langle A_1^{11}B_1^{11}\right\rangle \pm\left\langle A_1^{12}B_2^{12}\right\rangle \pm\left\langle A_2^{21}B_1^{21}\right\rangle \pm\left\langle A_2^{22}B_2^{22}\right\rangle\right)>2, \tag{27}$$

where the maximum is taken over all choices between $+$ and $-$ in front of each expected value $\langle\ldots\rangle$ such that the number of the minus signs is odd [14, 15]. This is the well-known CHSH inequality, except it is commonly written as

$$\max\left(\pm\langle A_1B_1\rangle \pm\langle A_1B_2\rangle \pm\langle A_2B_1\rangle \pm\langle A_2B_2\rangle\right)>2. \tag{28}$$

The latter form, however, is logically flawed, as it employs SBP and places the same random variable in different contexts. In fact, this inequality simply cannot be satisfied, because the variables A_1, A_2, B_1, B_2 in it must be jointly distributed (defined on the same sample space) by the following diagram:

$$\begin{tikzcd} A_1 \arrow[r, dash, "\text{same sample space}"] \arrow[d, dash, "\text{same sample space}"'] & B_1 \arrow[d, dash, "\text{same sample space}"] \\ B_2 \arrow[r, dash, "\text{same sample space}"'] & A_2 \end{tikzcd} \tag{29}$$

4 Overt influences vs contextuality (or on the magic of words)

There is one situation in which SBP cannot even be considered. It is the case of *inconsistently connected systems*, or systems with disturbance. Using again the system (15) as an example, suppose that the distributions of the random variables are as follows: for $i \in \{1,2\}$, $j \in \{1,2\}$,

context ij	$B_j^{ij} = +1$	$B_j^{ij} = -1$	
$A_i^{ij} = +1$	r_{ij}	$p_{ij} - r_{ij}$	p_{ij}
$A_i^{ij} = -1$	$q_{ij} - r_{ij}$	$1 - p_{ij} - q_{ij} + r_{ij}$	$1 - p_{ij}$
	q_{ij}	$1 - q_{ij}$	

. (30)

The difference between this and (12) is in the marginal distributions: they are no longer necessarily the same for A_i^{i1} and A_i^{i2} (generally, $p_{i1} \neq p_{i2}$), nor are they necessarily the same for B_j^{1j} and B_j^{2j} (generally, $q_{1j} \neq q_{2j}$).[6] Suppose, e.g., that in the EPR/Bohm experiment, Alice's and Bob's measurements are timelike separated, i.e., transmission of information from settings of one of them to measurement outcomes of another is possible. Say, Bob sends certain π-rays of frequency 1 when he chooses setting 1, and he sends π-rays of frequency 2 when he chooses setting 2, so that the outcomes of Alice's measurements, A_i^{i1} and A_i^{i2}, can be affected by these rays differently. Clearly, we have here dependence of Alice's measurements not only on her choice of a setting i but on the entire context ij. By definition, one can speak of *context-dependence* here.

But does this context-dependence necessarily mean that the system (15) with distributions (30) is contextual? Some researchers think that the answer to this question must be affirmative, unless the distributions are reduced to (12), in which case a system may be contextual or noncontextual. However, unless some unknown to me laws compel the meaning of the word "contextuality" to be derived from the

[6] The use of "not necessarily" here is to indicate that consistently connected systems, with distributions (12), are merely a special case of the inconsistently connected ones, with distributions (30).

way it sounds, or its closeness to "context-dependence," this is not the only possible answer. A more constructive approach would be to consider contextuality as a *form* of context-dependence, and to ask whether it can be separated from and studied together with inconsistent connectedness, viewed as another form of context-dependence, on the level of marginal distributions.

The definition of contextuality given in the previous section is in fact formulated for (generally) inconsistently connected systems. The values of $\omega_1, \omega_2, \omega_3, \omega_4$ defined in (17) generally are not all equal to 1. Thus, the maximal couplings of the content-sharing pair $\{A_1^{11}, A_1^{12}\}$ now has the distribution

content 1 of Alice	$\tilde{A}_1^{12} = +1$	$\tilde{A}_1^{12} = -1$	
$\tilde{A}_1^{11} = +1$	$\min(p_{11}, p_{12})$	$p_{11} - \min(p_{11}, p_{12})$	p_{11}
$\tilde{A}_1^{11}, = -1$	$p_{12} - \min(p_{11}, p_{12})$	$1 - p_{11} - p_{12} + \min(p_{11}, p_{12})$	$1 - p_{11}$
	p_{12}	$1 - p_{12}$	

, (31)

whence

$$\omega_1 = \max \Pr\left[\tilde{A}_1^{11} = \tilde{A}_1^{12}\right] = 1 - |p_{11} - p_{12}|. \tag{32}$$

Similar formulas hold for other content-sharing pairs. In all other respects, however, the logic of contextuality remains precisely as previously described: one seeks an overall coupling of the system subject to the constraints (22), and the system is contextual if and only if no such couplings exist. The interpretation of contextuality also remains the same as it was for consistently connected systems: contextuality means that the content-sharing random variable within their respective contexts (i.e., considered jointly distributed with other variables) are more dissimilar than when they are isolated from their contexts. An inconsistently connected system can be contextual or noncontextual, by precisely the same logic as in the special case when the system is consistently connected.

This approach is more constructive than simply declaring any inconsistently connected system contextual, because it provides greater differentiation among systems of random variables, while properly reducing to special cases when more restricted definitions apply. One can offer specific arguments in favor of our definition of contextuality [5, 11, 16–20], of which I will mention one. First, observe that some contextual systems are more contextual than others, with respect to the following, intuitively plausible way of measuring contextuality. Consider the value

$$\begin{aligned}\omega' &= \max\left(\omega_1' + \omega_2' + \omega_3' + \omega_4'\right) \\ &= \max\left(\Pr\left[\bar{A}_1^{11} = \bar{A}_1^{12}\right] + \Pr\left[\bar{A}_2^{21} = \bar{A}_2^{22}\right] + \Pr\left[\bar{B}_1^{11} = \bar{B}_1^{21}\right] + \Pr\left[\bar{B}_2^{12} = \bar{B}_2^{22}\right]\right)\end{aligned} \tag{33}$$

with the maximum taken over all possible couplings of system (15). This value cannot exceed

$$\omega_1 + \omega_2 + \omega_3 + \omega_4 = 4 - |p_{11} - p_{12}| - |p_{21} - p_{22}| - |q_{11} - q_{21}| - |q_{12} - q_{22}|, \tag{34}$$

because of which the nonnegative quantity

$$\mathsf{CNTX} = \omega_1 + \omega_2 + \omega_3 + \omega_4 - \omega' \tag{35}$$

can be taken for a measure of contextuality. A system is noncontextual if this quantity is zero. In this paper's opening example, system (15) with distributions (1) is consistently connected, so $\omega_1 = \omega_2 = \omega_3 = \omega_4 = 1$. This system is contextual because, as we have seen, it is not possible for a coupling to satisfy the chain of equalities in (25). The value of CNTX for this system can be shown to be 1, and this can be shown to be the highest possible value of CNTX across all systems of format (15) [5, 15, 21].

Now, let us introduce a small disturbance in our example, making the distributions

context 22	$B_2^{22} = +1$	$B_2^{22} = -1$	
$A_2^{22} = +1$	0	$1/2 + \varepsilon$	$1/2 + \varepsilon$
$A_2^{22} = -1$	$1/2 - \varepsilon$	0	$1/2 - \varepsilon$
	$1/2 - \varepsilon$	$1/2 + \varepsilon$	

. (36)

contexts $ij = 11, 12, 21$	$B_j^{ij} = +1$	$B_j^{ij} = -1$	
$A_i^{ij} = +1$	$1/2$	0	$1/2$
$A_i^{ij} = -1$	0	$1/2$	$1/2$
	$1/2$	$1/2$	

Intuition tells us that the degree of contextuality in this system should be only slightly different from the value of CNTX in the previous case, for $\varepsilon = 0$. And indeed, the degree of contextuality here is

$$\mathsf{CNTX} = 1 - 2\varepsilon. \tag{37}$$

We would not have such a smooth change of the degree of contextuality with $\varepsilon \to 0$ if we based the contextuality of the system with $\varepsilon > 0$ on the difference of the marginal probabilities alone.

On the other hand, in our second example, system (15) with distributions (8) is noncontextual, i.e., $\mathsf{CNTX} = 0$. If we introduce the same small perturbation as above, the distributions will be

context 22	$B_2^{22} = +1$	$B_2^{22} = -1$	
$A_2^{22} = +1$	$1/2 + \varepsilon$	0	$1/2 + \varepsilon$
$A_2^{22} = -1$	0	$1/2 - \varepsilon$	$1/2 - \varepsilon$
	$1/2 - \varepsilon$	$1/2 + \varepsilon$	

(38)

contexts $ij = 11, 12, 21$	$B_j^{ij} = +1$	$B_j^{ij} = -1$	
$A_i^{ij} = +1$	$1/2$	0	$1/2$
$A_i^{ij} = -1$	0	$1/2$	$1/2$
	$1/2$	$1/2$	

It can be shown that this system remains noncontextual, $\mathsf{CNTX} = 0$, as ε increases from 0 to $1/2$. Again, this is what one should expect based on the definition of contextuality. Zero CNTX at $\varepsilon = 0$ means that the system has a coupling in which the value of ω' reaches $\omega_1 + \omega_2 + \omega_3 + \omega_4$, which in this case has the maximal possible value, 4. Clearly, this is even easier to achieve if $\omega_1 + \omega_2 + \omega_3 + \omega_4$ has a smaller value, $4 - 2\varepsilon$.

As we have seen, an inconsistently connected system can be contextual or noncontextual, and this lays the ground for a richer classification of systems than the indiscriminate notion of context-dependence. It seems reasonable to maintain that being able to make finer differentiations is always desirable, provided it is done in a principled way. Nevertheless some researchers keep coming up with the revelatory insight that it is possible to present both contextuality and inconsistent connectedness as context-dependence and to refuse to distinguish them. Sometimes this is presented as the only position consistent with the "ontological" (or "ontic") models, in which (continuing to use our example) A_i^{ij} and B_j^{ij} are presented as functions

$$A_i^{ij} = f(i, j, \lambda), B_j^{ij} = g(i, j, \lambda), \tag{39}$$

where λ is some "hidden" variable. The term "ontological/ontic" is supposed to hint at something happening in reality, as opposed to purely mathematical descriptions. However, as a general approach, (39) is purely descriptive rather than explanatory, because it is trivially applicable to any system of random variables. It is in fact nothing more than a mathematically lax version of constructing an unconstrained overall coupling for system (15). We know that this is always possible. Recall, that to make $\left(A_i^{ij}, B_j^{ij}\right)$ for $i, j \in \{1, 2\}$ jointly distributed they have to be presented as functions on the same sample space. The random variable λ is nothing but the identity function on this sample space. More rigorously, of course, one has to write

$$\bar{A}_i^{ij} = f(i, j, \lambda), \bar{B}_j^{ij} = g(i, j, \lambda), \tag{40}$$

or

$$\left(A_i^{ij}, B_j^{ij}\right) \sim \left(f(i, j, \lambda), g(i, j, \lambda)\right), \tag{41}$$

because $\left(A_i^{ij}, B_j^{ij}\right)$ for different i, j are stochastically unrelated.

One source of misunderstanding leading some to considering (39) as an alternative to CbD is the suggestive terminology I and my colleagues coined for inconsistent connectedness: we called it (the manifestation of) *direct influences*, as opposed to contextual influences [22–25]. For instance, the distribution of A_i^{i1} may be different from that of A_i^{i2} because Bob sends his π-rays that affect the outcomes of Alice's measurements. This intuition leads some to point out, as if this were a discovery of a flaw in CbD, that the π-rays can also account for contextuality. One needs nothing but the π-rays, according to this reasoning. It is simply that some effects of the π-rays are overt, and are reflected in the differences of marginal distributions, while other effects of π-rays are hidden, and we call them (mistakenly, according to this criticism) contextuality [26]. This assertion is being justified, not surprisingly, by the very same possibility of representing a system by (39). One can construct various toy examples to demonstrate this, but the fact remains that (39) is applicable universally. As I have mentioned, it is simply a restatement of the possibility to construct an unconstrained overall coupling of any system.

To see that all of this is completely off target, it would suffice to replace the term *direct influence* with *overt influence*. In retrospect, this would have been a better term, and I intend to use it in the future. The criticism in question then would look like this: CbD distinguishes overt effects (observable on the level of marginal distributions) and contextual effects in systems of random variables, while we (the critics) say that some context-dependence in such systems can be overt and some hidden. This is no more than a terminological quibble, provided the hidden influences are to be revealed by means of the CbD-based contextuality analysis. However, the criticism in question seems to lead its proponents to simply lump together all context-dependence for systems that are not known to be consistently connected. If one accepts this position, in physics, contextuality analysis will be reserved to situations when no physical transfer of information from Bob's settings to Alice's measurements (and vice versa) is allowed by laws of physics. If tomorrow the physicists concluded that superluminal transmission is possible after all, the EPR/Bohm contextuality would have to be suspect. In systems like the original Kochen-Specker one [27] or KCBS system [28], where the measurements in each context are made on the same particle, contextuality is inherently suspect, as there it hinges on the fact that the current quantum mechanical accounts of these systems involve no forces or other forms of interference. In CbD, however, contextuality does not depend on the state of substantive theories: e.g., the EPR/Bohm system with certain choices of directions by Alice and Bob is contextual in both contemporary quantum theory and in Bohmian mechanics, where hidden superluminal transmission is built in. All of this is discussed and explained in our earlier publications, e.g. [17]. Quoting from the latter work,

> [...] to defend a definition is a difficult task. A good definition of a term should be intuitively plausible (although sometimes one's intuition itself should be "educated" to make it plausible), it should include as special cases all examples and situations that are traditionally considered to fall within the scope of the term, it should lead to productive development (to allow one to prove nontrivial theorems), and have a growing set of applications. I believe contextuality in the sense of CbD satisfies all these desiderata (p. 14).

To summarize:

1. (Non)contextuality is a property of systems of random variables. It is a special form of context-dependence, the other form of context-dependence being inconsistent connectedness.
2. Being a purely mathematical property, (non)contextuality of a system does not depend on substantive theories of the empirical situations represented by the system.
3. The identity of a random variable in a system is determined by its joint distribution with all other random variables in the same context. When context changes, a variable measuring some property is instantly replaced by another random variable measuring the same property (in the language of CbD, having the same

content), or it instantly disappears (if the property is not measured in the new context).
4. In particular, if the measurements described by the random variables in each contexts are separated by spacelike intervals, then the disappearance or replacement of a random variable by another random variable with the same content occurs instantly in response to spacelike separated changes in the context. No action at a distance is involved.
5. The difference between two random variables having the same content in different contexts is measured by their maximal coupling, and the system is noncontextual if one of its overall couplings has these maximal couplings as its marginals.
6. A contextual system, by contrast, makes the content-sharing random variables in different contexts more dissimilar than they are in isolation.
7. A system can be contextual or noncontextual irrespective of whether it is consistently connected.

References

1. Khrennikov, A. (2008) Bell-Boole inequality: Nonlocality or probabilistic incompatibility of random variables? *Entropy* 10, 19-32.
2. Khrennikov, A. (2009) Bell's inequality: Physics meets Probability. *Information Science* 179, 492-504.
3. Mermin, N. (1989) Can you help your team tonight by watching on TV? More experimental metaphysics from Einstein, Podolsky, and Rosen. In J. T. Cushin and E. McMullin (Eds.) *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem* (pp. 38-59). University of Notre Dame Press.
4. Popescu, S., Rohrlich, D. (1994) Quantum nonlocality as an axiom. *Foundations of Physics* 24, 379–385.
5. Kujala, J.V., & Dzhafarov, E.N. (2019) Measures of contextuality and noncontextuality. *Philosophical Transactions of the Royal Society A* 377:20190149.
6. Clauser, J.F., Horne, M.A., Shimony, A., Holt, R.A. (1969). Proposed experiment to test local hidden-variable theories. *Physical Review Letters* **23**, 880-884.
7. Avis, D., Fischer, P., Hilbert, A., & Khrennikov, A. (2009) Single, complete, probability spaces consistent with EPR-Bohm-Bell experimental data. In A. Khrennikov (Ed.), *Foundations of Probability and Physics-5, AIP Conference Proceedings* 750 (pp. 294-301). Melville, New York: AIP.
8. Dzhafarov, E.N., & Kujala, J.V. (2014) Embedding quantum into classical: contextualization vs conditionalization. *PLoS One* 9(3): e92818. doi:10.1371/journal.pone.0092818.
9. Dzhafarov, E.N., & Kujala, J.V. (2017) Contextuality-by-Default 2.0: Systems with binary random variables. In J.A. de Barros, B. Coecke, E. Pothos (Eds.) *Lecture Notes in Computer Science* 10106, 16-32.
10. Dzhafarov, E.N., Cervantes, V.H., & Kujala, J.V. (2017). Contextuality in canonical systems of random variables. *Philosophical Transactions of the Royal Society A* 375: 20160389.
11. Dzhafarov, E.N. (2018) Replacing nothing with something special: Contextuality-by-Default and dummy measurements. In A. Khrennikov & T. Bourama (Eds) *Quantum Foundations, Probability and Information*, pp. 39-44. Berlin: Springer.
12. Dzhafarov, E.N., & Kujala, J.V. (2016). Probability, random variables, and selectivity. In W.Batchelder et al. (Eds), pp. 85-150. New Handbook of Mathematical Psychology. Cambridge University Press.

13. Dzhafarov, E.N. (2019) On joint distributions, counterfactual values, and hidden variables in understanding contextuality. *Philosophical Transactions of the Royal Society A* 377:20190144.
14. Dzhafarov, E.N., Kujala, J.V., & Larsson, J.-. (2015) Contextuality in three types of quantum-mechanical systems. *Foundations of Physics* 7, 762-782
15. Kujala, J.V., & Dzhafarov, E.N. (2016) Proof of a conjecture on contextuality in cyclic systems with binary variables. *Foundations of Physics* 46, 282-299.
16. Kujala, J.V., Dzhafarov, E.N., & Larsson, J- (2015) Necessary and sufficient conditions for extended noncontextuality in a broad class of quantum mechanical systems. *Physical Review Letters* 115, 150401.
17. Dzhafarov, E.N., & Kujala, J.V. (2015) Conversations on contextuality. In E.N. Dzhafarov, S. Jordan, R. Zhang, V.H. Cervantes (Eds.) *Contextuality from Quantum Physics to Psychology*, pp. 1-22. New Jersey: World Scientific.
18. Dzhafarov, E.N., & Kujala, J.V. (2016) Context-content systems of random variables: The contextuality-by-default theory. Journal of Mathematical Psychology 74, 11-33.
19. Dzhafarov, E.N., Kujala, J.V. (2018) Contextuality analysis of the double slit experiment (with a glimpse into three slits). Entropy 20, 278; doi:10.3390/e20040278.
20. Dzhafarov, E.N. (2021) Contents, contexts, and basics of contextuality. In Shyam Wuppuluri and Ian Stewart (Eds). *From Electrons to Elephants and Elections: Saga of Content and Context*. To be published by Springer - The Frontiers Collection.
21. Dzhafarov, E.N., Kujala, J.V., & Cervantes, V.H. (2020) Contextuality and noncontextuality measures and generalized Bell inequalities for cyclic systems. *Physical Review A* 101:042119. (available as arXiv:1907.03328.) + Erratum: Contextuality and noncontextuality measures and generalized Bell inequalities for cyclic systems [Phys. Rev. A 101, 042119 (2020)]. Physical Review A 101:069902.
22. Cervantes, V.H., & Dzhafarov, E.N. (2018) Snow Queen is evil and beautiful: Experimental evidence for probabilistic contextuality in human choices. *Decision* 5, 193-204.
23. Cervantes, V.H., Dzhafarov, E.N. (2019) True contextuality in a psychophysical experiment. *Journal of Mathematical Psychology* 91, 119-127.
24. Basieva, I., Cervantes, V.H., Dzhafarov, E.N., Khrennikov, A. (2019) True contextuality beats direct influences in human decision making. *Journal of Experimental Psychology: General* 148, 1925-1937.
25. Dzhafarov, E.N. (2021) The Contextuality-by-Default view of the Sheaf-Theoretic approach to contextuality. To be published in A. Palmigiano and M. Sadrzadeh (Eds.) *Samson Abramsky on Logic and Structure in Computer Science and Beyond (Outstanding Contributions to Logic)*. Springer Nature.
26. Atmanspacher H., Filk T. (2019) Contextuality Revisited: Signaling May Differ From Communicating. In: de Barros J., Montemayor C. (Eds) *Quanta and Mind. Synthese Library (Studies in Epistemology, Logic, Methodology, and Philosophy of Science)*, vol 414. Springer, Cham.
27. Kochen, S., & Specker, E. P. (1967) The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17:59–87.
28. Klyachko, A.A., Can, M.A., Binicioglu, S., & Shumovsky, A.S. (2008) A simple test for hidden variables in spin-1 system. *Physical Review Letters* 101:020403.