# Conversations on Contextuality

Ehtibar N. Dzhafarov
Purdue University

Janne V. Kujala
University of Jyväskylä

**Dramatis personæ:**

EXPOSITOR, trying to present and clarify the main points of a certain view of contextuality.

INTERLOCUTOR, skeptical but constructive.

AUTHORS, supportive of Expositor but sympathetic to Interlocutor (remain off-stage except for occasionally inserting footnotes).

## Conversation 1

EXPOSITOR: My dear Interlocutor, as we have agreed, we will discuss a certain approach to probabilistic contextuality. Its authors call it, perhaps not too descriptively, the Contextuality-by-Default theory (CbD).[1] I think I should begin by giving you an informal overview of the main ideas.

INTERLOCUTOR: My dear Expositor, I always find an informal presentation of ideas a dubious exercise. If I do not understand the presentation clearly (which happens often), it is never clear to me whether this is because it was dumbed down so much as to become deficient of information, or because the ideas themselves are deficient. Nevertheless I should let you proceed.

EXP: Let me try. *Objects* (or things, or properties — choose what you like) are measured under varying conditions, called *contexts*. The measurements are generally random variables, and their identity is defined by *what* is measured (object) and *under what conditions* it is measured (context). As a result, the same object measured in different contexts is represented by different random variables: it is meaningless to ask *why* they are different (hence the designation "contextuality-by-default"). Moreover, measurements made in different contexts, whether of the same object or of different objects, *do not*

[1] AUT: The term indeed may not have been optimally chosen. It may suggest that all systems of measurements are contextual until proven otherwise. This is not true. The "by-default" in the name of the theory refers to the identification of the measurements as random variables: their identity always ("by default") depends on context.

*have a joint distribution.* One cannot, e.g., speak of their correlation or of the probability with which they have the same value. All measurements made within one and the same context, however, are *jointly distributed.* The overall picture we have therefore is one of stochastically unrelated to each other islands of jointly distributed measurements ("*bunches* of random variables"). Is this sufficiently clear?

INT: I am not sure. How does one define "objects" and "contexts"?

EXP: Primitives of a theory cannot be explained conceptually except in their relations to other primitives of the theory, and their operational meaning may be outside the theory. The "objects" and "contexts" are such primitives: formally, they are no more than labels defining the identity of a measurement (so that each measurement is defined by two labels, one for "what" and another for "in what context").

INT: Perhaps we could clarify this with examples.

EXP: Here is an example. Suppose we pose two Yes/No questions to randomly chosen people and record their responses. The questions can be asked verbally or presented in writing. Intuitively, a question asked is the "object" being measured, the presentation mode (verbal or written) is context, and the response to a given question by a randomly chosen person is measurement, a random variable labeled by the question asked and by its presentation mode.

INT: So we have four random variables, if I understood you correctly: response to question $A$ presented verbally, response to question $A$ presented in writing, and analogously for the second question, $B$.

EXP: Yes. Let me suggest notation for these four random variables:

$$R_A^V, R_A^W, R_B^V, R_B^W,$$

or, better still,

$$\left(R_A^V, R_B^V\right), \left(R_A^W, R_B^W\right).$$

$R$ stands for response (whose values can be Yes or No), the subscript shows the object (question), and the superscript shows the context (presentation mode). We record $R_A^V$ and $R_B^V$ together because the two questions are posed to a same person. For the same reason, $R_A^W$ and $R_B^W$ are recorded together. Therefore $\left(R_A^V, R_B^V\right)$ are jointly distributed, and so are $\left(R_A^W, R_B^W\right)$. This accords with the general rule: random variables recorded in the same context are jointly distributed. Now, we assume that a person is asked either two written questions or two verbal questions. So, $R_A^V$ is never recorded together with (which means here, never obtained from the same person as) $R_A^W$ or $R_B^W$. Therefore such pairs as $R_A^V, R_A^W$ or $R_A^V, R_B^W$ have no joint distribution.

INT: What does this mean exactly, not to have a joint distribution?

EXP: It is clear that for each of our four variables we have well-defined probabilities with which their value is Yes: this is the probability with which a randomly chosen person will respond Yes to the corresponding question in the corresponding context. Say, the probability of the event $\left[R_A^V = \text{Yes}\right]$ is 0.4, the probability of $\left[R_B^V = \text{Yes}\right]$ is 0.5, and for $\left[R_B^W = \text{Yes}\right]$ the probability is, say, 0.7. All these probabilities are well-defined theoretically and can be estimated empirically. Since we ask $A$ and $B$ in the verbal mode together (from a same person), we can also define and estimate the joint probability of $\left[R_A^V = \text{Yes and } R_B^V = \text{Yes}\right]$. For instance, if it equals $0.4 \times 0.5$, then the two random variables are independent; if it equals 0.4, they are maximally positively correlated, etc.

INT: I see. And, I understand, the situation is different with $R_A^V, R_B^W$ or $R_B^V, R_B^W$, because there is no meaning in which one could define "and" in, say, $\left[R_B^V = \text{Yes and } R_B^W = \text{Yes}\right]$.

EXP: Precisely. The probability of this conjunction is undefined and cannot be estimated empirically. So the general rule is: no two random variables recorded in different (mutually exclusive) contexts possess a joint distribution. Let's agree to call such variables *stochastically unrelated* (to each other).

INT: What if I modified the design of the survey, and asked only one question per person, in writing or verbally? Wouldn't then even $R_A^V$ and $R_B^V$ be stochastically unrelated? And wouldn't this contradict our general rules?

EXP: No combinations of objects and contexts can contradict our general rules because these combinations have to be chosen in accordance with these rules. In your modified set-up, if we continue to view the questions $A$ and $B$ as our sole and distinct objects, then the contexts involve not only the mode of presentation but also the identity of the questions themselves: $(A, V)$, $(B, W)$, etc. So the random variables we record are

$$R_A^{(A,V)}, R_A^{(A,W)}, R_B^{(B,V)}, R_B^{(B,W)}.$$

Since no two of them share a context, they are pairwise stochastically unrelated.

INT: But how do I know which of the representations to use, $R_A^V$ or $R_A^{(A,V)}$?

EXP: First you have to decide (outside the CbD theory) on the empirical meaning of "co-occurrence" or "occurrence together" in your study. In our examples you consider questions posed to one and the same person (and responses obtained from one and the same person) as co-occurring. You also know the rules, so you always use different contexts, whatever your choice of the labels for them, for the measurements that do not co-occur. Suppose I simplify your design by forgetting about the presentation modes. I ask one of two questions, $A$ or $B$, of randomly chosen people. The objects being measured then are $A$ and $B$ again, and we know that the measurements of these objects never co-occur, hence they are stochastically unrelated. Therefore the questions themselves (or any two labels corresponding to them one-to-one) are the contexts here,

$$R_A^A, R_B^B.$$

Int: I think I now understand the notion of stochastic unrelatedness and your notation. But couldn't we also say in all such cases that the two random variables, e.g., $R_A^A, R_B^B$, are stochastically independent?

Exp: This would be a common way of thinking of this situation. But it is incorrect. One could only say, with some caution, that they can always be treated *as if* they were independent. We will get to this later, when we consider the notion of a coupling. You can, however, appreciate the difference between the situation when a randomly chosen person is being asked one of two questions, $A$ or $B$, and the situation when a randomly chosen person is being asked both these questions, $A$ and $B$. The two responses in the latter case are random variables

$$\left(R_A^{(A,B)}, R_B^{(A,B)}\right),$$

jointly distributed: one can define and estimate the probability of the joint event $\left[R_A^{(A,B)} = \text{Yes and } R_B^{(A,B)} = \text{Yes}\right]$. Suppose we find out that this probability equals the product of the probability of $\left[R_A^{(A,B)} = \text{Yes}\right]$ and the probability of $\left[R_B^{(A,B)} = \text{Yes}\right]$ taken separately. Then we say that the random variables $R_A^{(A,B)}$ and $R_B^{(A,B)}$ are stochastically independent. This is fundamentally different from the situation with $R_A^A$ and $R_B^B$, when the joint event $\left[R_A^A = \text{Yes and } R_B^B = \text{Yes}\right]$ is simply undefined.[2]

Int: I see the difference. I still have questions about the objects and contexts, but I think I should allow you to continue your presentation of CbD. The double-notation, I understand, is only a departure point.

Exp: Yes, it is. And we will indeed continue to address your misgivings as we proceed. By now we have established a certain picture of a system of measurements: it consists of stochastically unrelated to each other islands (or bunches) of jointly distributed random variables. The main idea of CbD is that these isolated bunches can be characterized by exploring *all possible couplings thereof (*or *all possible joint distributions imposable on them) under well-chosen constraints.*

Int: What is a *coupling*?

Exp: A coupling for stochastically unrelated random variables $X, Y, \ldots, Z$ is a jointly distributed $\left(\widetilde{X}, \widetilde{Y}, \ldots, \widetilde{Z}\right)$ in which $\widetilde{X}$ has the same distribution as $X$, $\widetilde{Y}$ has the same distribution as $Y$, and so on.[3] For instance, in our

[2] Aut: For a detailed discussion, see Dzhafarov and Kujala (2014c) and Dzhafarov (2015).

[3] Aut: See Thorisson (2000). A traditional definition of a coupling does not require that the random variables being coupled be stochastically unrelated, but in the present context it is the only application.

example with $R_A^A$ and $R_B^B$, let the distributions of these random variables be defined by (with Pr standing for probability)

$$\Pr\left[R_A^A = \text{Yes}\right] = 0.4, \Pr\left[R_B^B = \text{Yes}\right] = 0.7.$$

The measurements $R_A^A$ and $R_B^B$ are stochastically unrelated, so $\Pr\left[R_A^A = \text{Yes and } R_B^B = \text{Yes}\right]$ is undefined. To couple them means to create a new pair of random variables, $\widetilde{R}_A^A$ and $\widetilde{R}_B^B$, that are distributional copies of, respectively, $R_A^A$ and $R_B^B$ but are jointly distributed. That is,

$$\Pr\left[\widetilde{R}_A^A = \text{Yes}\right] = 0.4, \Pr\left[\widetilde{R}_B^B = \text{Yes}\right] = 0.7,$$

but, unlike the "originals" $R_A^A$ and $R_B^B$, their distributional copies have a well-defined joint probability,

$$\Pr\left[\widetilde{R}_A^A = \text{Yes and } \widetilde{R}_B^B = \text{Yes}\right].$$

It can be shown that this probability can have any value from 0.1 (maximally negative relation) to 0.4 (maximally positive relation). The independent coupling, with this probability equal to $0.4 \times 0.7$, is within this range.

INT: And any of these values will define a pair $\left(\widetilde{R}_A^A, \widetilde{R}_B^B\right)$ that is a coupling for $R_A^A$ and $R_B^B$?

EXP: Yes. The number of all possible couplings for a given pair of random variables is typically infinite. For binary $X, Y$ (say, Yes/No ones) the only exceptions are the pairs with $\Pr[X = \text{Yes}]$ or $\Pr[Y = \text{Yes}]$ having the values 0 or 1. For such a pair only one coupling is possible.

INT: Let me try an example with continuous distributions. Take random variables $R$ and $S$ that are standard normally distributed. Then any bivariate normally distributed $\left(\widetilde{R}, \widetilde{S}\right)$ with standard normal marginals is a coupling of $R$ and $S$.

EXP: Yes, and there are other couplings for these $R$ and $S$ too: it can be any $\left(\widetilde{R}, \widetilde{S}\right)$ whose joint distribution is well-defined, and whose individual (marginal) distributions are standard normal.

INT: Just so that we don't focus on pairs of random variables exclusively, what would be a coupling for the four random variables $R_A^V, R_A^W, R_B^V, R_B^W$ in our original example? There we had two pairs (you called them "bunches") with jointly distributed components, $\left(R_A^V, R_B^V\right)$ and $\left(R_A^W, R_B^W\right)$.

EXP: Because of the latter, this example too can be presented as involving just pairs: $\left(R_A^V, R_B^V\right)$ is a random variable too, unless we use the term very narrowly. But this is not critical: it is the same thing to seek a coupling for $\left(R_A^V, R_B^V\right)$ and $\left(R_A^W, R_B^W\right)$, each with a known distribution, and to seek a coupling for $R_A^V, R_A^W, R_B^V, R_B^W$ in which you know the distributions of $\left(R_A^V, R_B^V\right)$

and $\left(R_A^W, R_B^W\right)$. The joint distribution of $\left(R_A^V, R_B^V\right)$ is determined by four probabilities

$$\begin{array}{rcccc} \left(R_A^V, R_B^V\right): & (\text{Yes}, \text{Yes}) & (\text{Yes}, \text{No}) & (\text{No}, \text{Yes}) & (\text{No}, \text{No}) \\ \text{probability}: & p_{YY} & p_{YN} & p_{NY} & p_{NN} \end{array},$$

with the probabilities summing to 1, of course. And the joint distribution of $\left(R_A^W, R_B^W\right)$ is determined analogously,

$$\begin{array}{rcccc} \left(R_A^W, R_B^W\right): & (\text{Yes}, \text{Yes}) & (\text{Yes}, \text{No}) & (\text{No}, \text{Yes}) & (\text{No}, \text{No}) \\ \text{probability}: & q_{YY} & q_{YN} & q_{NY} & q_{NN} \end{array}.$$

To couple these pairs (equivalently, to couple all four random variables $R_A^V, R_A^W, R_B^V, R_B^W$) means to create a quadruple $\left(\widetilde{R}_A^V, \widetilde{R}_A^W, \widetilde{R}_B^V, \widetilde{R}_B^W\right)$ with jointly distributed components such that the distributions of the pairs $\left(\widetilde{R}_A^V, \widetilde{R}_B^V\right)$ and $\left(\widetilde{R}_A^W, \widetilde{R}_B^W\right)$ are the same as those of $\left(R_A^V, R_B^V\right)$ and $\left(R_A^W, R_B^W\right)$, respectively. For instance,

$$\Pr\left[\widetilde{R}_A^V = \text{Yes and } \widetilde{R}_B^V = \text{No}\right] = p_{YN},\ \Pr\left[\widetilde{R}_A^W = \text{No and } \widetilde{R}_B^W = \text{No}\right] = q_{NN},\ \text{etc.}$$

INT: And to create a quadruple $\left(\widetilde{R}_A^V, \widetilde{R}_A^W, \widetilde{R}_B^V, \widetilde{R}_B^W\right)$ with jointly distributed components means ...

EXP: It means to assign to each of the 16 possible quadruples of values a probability value:

$$\begin{array}{rcccc} (\widetilde{R}_A^V, \widetilde{R}_A^W, \widetilde{R}_B^V, \widetilde{R}_B^W): & (\text{Yes}, \text{Yes}, \text{Yes}, \text{Yes}) & (\text{Yes}, \text{Yes}, \text{Yes}, \text{No}) & \ldots & (\text{No}, \text{No}, \text{No}, \text{No}) \\ \text{probability}: & r_{YYYY} & r_{YYYN} & \ldots & r_{NNNN} \end{array}.$$

To form a coupling, these probabilities should agree with the observed $p$ and $q$ values: e.g.,

$$\sum_{i,j \in \{\text{Yes},\text{No}\}} r_{YiNj} = p_{YN}, \quad \sum_{i,j \in \{\text{Yes},\text{No}\}} r_{iNjN} = q_{NN}.$$

INT: And, of course, this can generally be done in an infinite number of ways. I think it is clear. You said earlier that you wanted to characterize the isolated bunches of random variables by exploring *all possible couplings of these bunches under well-chosen constraints.* Tell me now what you mean by the "well-chosen constraints" for couplings.

EXP: The "well-chosen constraints" depend on (and determine) the aspect of the system of unrelated to each other bunches you want to characterize. If we are interested in contextuality, the idea (arguably, the most original idea in the CbD approach) is to look for *couplings in which the measurements of one and the same object under different conditions are equal to each other with* *as high a probability as possible.* Contextuality is determined by computing these highest probabilities for each object in isolation and then determining if they are compatible with the observed bunches of measurements.

INT: I assume you want to elaborate.

EXP: I will. But I think we will relegate this to our next conversation.

## Conversation 2

EXP: My dear Interlocutor, you asked me to elaborate what I said about the measurements of an object in different contexts being equal to each other with as high a probability as possible. First of all, let me emphasize that the probability of being equal to each other applies to the *couplings* rather than the coupled random variables themselves. When I find a bivariate-normally distributed coupling $\left(\widetilde{R}, \widetilde{S}\right)$ for standard normally distributed $R$ and $S$, I do not make $R$ and $S$ jointly distributed, I merely create jointly distributed "copies" of $R$ and $S$. And there are generally an infinite set of such couplings. Among them there is one coupling, with the correlation between $\widetilde{R}$ and $\widetilde{S}$ equal to 1 (defining a degenerate bivariate-normal distribution), in which $\Pr\left[\widetilde{R} = \widetilde{S}\right]$ has the highest possible value (in this case, 1). It is called a *maximal coupling*. If $R$ and $S$ are normally distributed but with different means and/or variances, then the maximal couplings still exist, but not among bivariate normally distributed $\left(\widetilde{R}, \widetilde{S}\right)$, and the highest possible probability for $\left[\widetilde{R} = \widetilde{S}\right]$ is less than 1.

INT: Let me switch back to our original example to understand this. In $R_A^V, R_A^W, R_B^V, R_B^W$ we have $R_A^V$ and $R_A^W$ measuring the same object $A$ in two contexts. We take them and look for a coupling $\left(\widehat{R}_A^V, \widehat{R}_A^W\right)$ for them, forgetting for the time being all about the remaining two variables. We ask the question: what is the maximal possible probability for the event $\left[\widehat{R}_A^V = \widehat{R}_A^W\right]$? We know that, by definition of a coupling,

$$\Pr\left[\widehat{R}_A^V = 1\right] = \Pr\left[R_A^V = 1\right] = p_A^V,$$

$$\Pr\left[\widehat{R}_A^W = 1\right] = \Pr\left[R_A^W = 1\right] = p_A^W.$$

I assume this allows me to compute the maximal possible value for $\Pr\left[\widehat{R}_A^V = \widehat{R}_A^W\right]$?

EXP: Yes. This maximal value is $1 - \left|p_A^V - p_A^W\right|$. It is very easy to prove.[4]

[4] AUT: See, e.g., Dzhafarov *et al.* (2015a).

INT: I will take you word for it. So we have

$$\max_{\substack{\text{all couplings}\\ \left(\widehat{R}_A^V, \widehat{R}_A^W\right)}} \Pr\left[\widehat{R}_A^V = \widehat{R}_A^W\right] = 1 - \left|p_A^V - p_A^W\right|.$$

After that I forget all about $R_A^V$ and $R_A^W$ and focus on $R_B^V$ and $R_B^W$, the other two measurements of one and the same object in two contexts. By analogy,

$$\max_{\substack{\text{all couplings}\\ \left(\widehat{R}_B^V, \widehat{R}_B^W\right)}} \Pr\left[\widehat{R}_B^V = \widehat{R}_B^W\right] = 1 - \left|p_B^V - p_B^W\right|.$$

How does one proceed from here?

EXP: Now we have to take all four our random variables and construct a coupling $\left(\widetilde{R}_A^V, \widetilde{R}_A^W, \widetilde{R}_B^V, \widetilde{R}_B^W\right)$ for them. We have already discussed how we do this. Except in special cases, there is an infinity of such couplings. What we are now interested in is whether among all these couplings there is at least one in which

$$\Pr\left[\widetilde{R}_A^V = \widetilde{R}_A^W\right] = 1 - \left|p_A^V - p_A^W\right|$$
and
$$\Pr\left[\widetilde{R}_B^V = \widetilde{R}_B^W\right] = 1 - \left|p_B^V - p_B^W\right|.$$

If the answer to this question is affirmative, then we say that the system of measurements, in this case comprised of $\left(R_A^V, R_B^V\right)$ and $\left(R_A^W, R_B^W\right)$, is *noncontextual*. If there is no such couplings, then the system is *contextual*.

INT: Let me first see how this works if

$$p_A^V = p_A^W \text{ and } p_B^V = p_B^W.$$

The maximal probabilities of $\left[\widehat{R}_A^V = \widehat{R}_A^W\right]$ and of $\left[\widehat{R}_B^V = \widehat{R}_B^W\right]$ then are equal to 1, for both $A$ and $B$. So, if I can find a coupling $\left(\widetilde{R}_A^V, \widetilde{R}_A^W, \widetilde{R}_B^V, \widetilde{R}_B^W\right)$ in which $\widetilde{R}_A^V$ is always the same as $\widetilde{R}_A^W$ and $\widetilde{R}_B^V$ is always the same as $\widetilde{R}_B^W$, then the system is noncontextual.

EXP: This is in fact the traditional understanding of contextuality (expressed in the language of CbD).[5] If $\widetilde{R}_A^V$ and $\widetilde{R}_A^W$ are always the same, one can say that the measurement of $A$ does not depend on what the context is, $V$ or $W$; and analogously for $\widetilde{R}_B^V$ and $\widetilde{R}_B^W$.

INT: The adjective "noncontextual" here seems intuitive to me. Let me now consider the case when $p_A^V \neq p_A^W$, i.e., the maximal probability $1 - \left|p_A^V - p_A^W\right| < 1$. So we begin by computing ... But wait: the measurements of $A$ here have different distributions in context $V$ and in context $W$, so these measurement

[5] AUT: See Dzhafarov and Kujala (2014a,c).

are context-dependent. Why don't we declare this system of measurements contextual, without computing anything else?

EXP: You are touching on a subtle conceptual and terminological issue. Nothing prevents one from calling this system contextual, but in the terminology of CbD it is called *inconsistently connected* (and if $p_A^V = p_A^W$ and $p_B^V = p_B^W$, the system is *consistently connected*). If you insist on using the word "contextuality" whenever a system has $p_A^V \neq p_A^W$ or $p_B^V \neq p_B^W$, call this "contextuality-1." Or use another qualifier, but distinguish this form of contextuality from the contextuality in the sense of CbD (call it "contextuality-2" if you like). It is the form of contextuality that may exist on top of the "contextuality-1."

INT: I still have misgivings, but we can return to this later. Let me resume my attempt to understand how this "contextuality-2" works in the case $p_A^V \neq p_A^W$ and/or $p_B^V \neq p_B^W$. We compute the maximal probability of $\left[\widehat{R}_A^V = \widehat{R}_A^W\right]$ across all couplings $\left(\widehat{R}_A^V, \widehat{R}_A^W\right)$ for $R_A^V$ and $R_A^W$; and, separately, we compute the maximal probability of $\left[\widehat{R}_B^V = \widehat{R}_B^W\right]$ across all couplings $\left(\widehat{R}_B^V, \widehat{R}_B^W\right)$ for $R_B^V$ and $R_B^W$. These probabilities, you tell me, are $1 - \left|p_A^V - p_A^W\right|$ and $1 - \left|p_B^V - p_B^W\right|$, respectively. Then we look at all possible couplings $\left(\widetilde{R}_A^V, \widetilde{R}_A^W, \widetilde{R}_B^V, \widetilde{R}_B^W\right)$ for all four of our random variables, and for each of them we compute the probabilities of $\left[\widetilde{R}_A^V = \widetilde{R}_B^V\right]$ and of $\left[\widetilde{R}_A^W = \widetilde{R}_B^W\right]$.

EXP: It is clear that these probabilities cannot exceed the values $1 - \left|p_A^V - p_A^W\right|$ and $1 - \left|p_B^V - p_B^W\right|$, respectively — because every sub-coupling $\left(\widetilde{R}_A^V, \widetilde{R}_A^W\right)$ of $\left(\widetilde{R}_A^V, \widetilde{R}_A^W, \widetilde{R}_B^V, \widetilde{R}_B^W\right)$ is also one of the possible couplings $\left(\widehat{R}_A^V, \widehat{R}_A^W\right)$ for $R_A^V$ and $R_A^W$ taken separately; and analogously for $\left(\widetilde{R}_B^V, \widetilde{R}_B^W\right)$ and $\left(\widehat{R}_B^V, \widehat{R}_B^W\right)$.

INT: Yes, I see this. Now, if in some of the couplings $\left(\widetilde{R}_A^V, \widetilde{R}_A^W, \widetilde{R}_B^V, \widetilde{R}_B^W\right)$ both these probabilities are achieved, then we say the system is noncontextual (lacks "contextuality-2"). Otherwise it is contextual.

EXP: Yes. The intuition here is that there is something in the relationship between $\left(R_A^V, R_B^V\right)$ and $\left(R_A^W, R_B^W\right)$ that cannot be reduced to the separate effects of the context changes on the responses to $A$ and on the responses to $B$.

INT: Wait, wait. Couldn't we then simply reformulate the problem by taking $(A, B)$ as a single object? Then we would have two stochastically unrelated random variables

$$R_{(A,B)}^V, R_{(A,B)}^W,$$

each with four possible values, (Yes-Yes), (Yes-No), etc. We will have contextuality if their distributions are different. This would be "contextuality-1," of course.

Exp: This is definitely a possible approach. I have mentioned already that objects and contexts are primitives of the CbD theory, which means that the theory does not dictate their choice. The only constraint imposed by the theory is that the random variables measured in the same context have a joint distribution, while random variables in different contexts do not. If you choose a single object in two contexts, as you have proposed, you will simply be dealing with a different problem. The system comprised of $R^V_{(A,B)}$ and $R^W_{(A,B)}$ may or may not be inconsistently connected (I suggest we stick to this term instead of "contextuality-1" and the like), but irrespective of this, it is noncontextual. If it is inconsistently connected, then the system we considered previously, $\left(R^V_A, R^V_B\right)$ and $\left(R^W_A, R^W_B\right)$, may or may not be consistently connected, and in either case it can be contextual or noncontextual.

Int: Then I was wrong: inconsistent connectedness in $\left\{R^V_{(A,B)}, R^W_{(A,B)}\right\}$ does not predict or account for the contextuality in $\left\{\left(R^V_A, R^V_B\right), \left(R^W_A, R^W_B\right)\right\}$.

Exp: No, it does not. The system $\left\{\left(R^V_A, R^V_B\right), \left(R^W_A, R^W_B\right)\right\}$ can be shown to be noncontextual if and only if

$$\left|\left\langle R^V_A R^V_B\right\rangle - \left\langle R^W_A R^W_B\right\rangle\right| \leq \left|\left\langle R^V_A\right\rangle - \left\langle R^W_A\right\rangle\right| + \left|\left\langle R^V_B\right\rangle - \left\langle R^W_B\right\rangle\right|,$$

where $\langle\cdot\rangle$ is expected value.[6] You can easily verify that this inequality may hold or fail with the distributions of $R^V_{(A,B)}$ and $R^W_{(A,B)}$ being different.

Int: I wonder: even if we get a completely different system by doing this, is it always possible to get rid of contextuality in the sense of CbD by redefining the objects?

Exp: The answer to this is yes, but generally not by grouping the objects together, as in $R^V_{(A,B)}, R^W_{(A,B)}$. Consider, e.g., a system involving three objects $q, q', q''$ and three contexts $c, c', c''$, combined in the measurements as follows:[7]

$$\left(R^c_q, R^c_{q'}\right), \left(R^{c'}_{q'}, R^{c'}_{q''}\right), \left(R^{c''}_{q''}, R^{c''}_q\right).$$

As always, the pairs of random variables labeled by the same context are jointly distributed, and different pairs are stochastically unrelated. One cannot now put all three objects together as a single object. Instead one can do something universally applicable: taking a measurement's context as part of the identity of the measurement's object. In our case this means replacing $q$ in $R^c_q$ with $(q, c)$, replacing $q$ in $R^{c''}_q$ with $(q, c'')$, etc. We will get then in place of the system above a new system

$$\left(R^c_{(q,c)}, R^c_{(q',c)}\right), \left(R^{c'}_{(q',c')}, R^{c'}_{(q'',c')}\right), \left(R^{c''}_{(q'',c'')}, R^{c''}_{(q,c'')}\right).$$

[6] Aut: This is an example of a cyclic system with $n = 2$, as defined in Conversation 3.

[7] Aut: This is an example of a cyclic system with $n = 3$.

In this system no two measurements share their object, and the system is readily seen as trivially noncontextual (and also consistently connected, again in the trivial sense).

INT: This universal trick then consists in declaring any object in a new context to be a new object. For instance, a question presented verbally and the same (in content) question presented in writing are different questions.

EXP: Yes, and there is nothing incorrect about this, at least not from the point of view of CbD. It just makes the issue of contextuality uninteresting.

INT: So we will not use this trick for the sake of keeping our discussion interesting. However, philosophically speaking, it may very well be the case that finding a system of measurements contextual means that "the same" objects in different contexts are not really the same.

EXP: Perhaps. But I find such philosophical formulations unsatisfactory. We need a language rich enough to lead to interesting classifications and quantifications of contextuality. The language of CbD is rich enough. Trivial renaming of all objects into object-contexts is not.

INT: I agree. I think I understand the definition of contextuality in CbD. However, I may need more persuasion to accept it. Let me return to my misgivings about "contextuality-1" and "contextuality-2." Why do we need the latter?

EXP: Let me remind to you that "contextuality-1" is inconsistent connectedness: for some objects, their measurements have different distributions in different contexts. We may very well have a consistently connected system, however, without "contextuality-1." Will we simply declare it noncontextual since all contextuality is "contextuality-1"? Again, one can say this if one so wishes, but this terminology will not change the fact that there is an important distinction within the class of consistently connected systems of measurements.

INT: Please remind me what this distinction is.

EXP: If the distribution of the measurements of a given object is the same across all contexts involving this object, then it is possible to couple these measurements so that their copies in the coupling are equal to each other with probability 1. Let's call this the *identity coupling*. Now, there are two possibilities. There can be a consistently connected system that has a coupling in which this probability 1 is achieved for all objects; in other words the identity couplings for different objects can all be put together so that they are compatible with the observed distributions of the bunches of measurements in each of the contexts. And there can be systems in which such couplings do not exist: the observed bunches of measurements are not compatible with the identity couplings for all the objects. This is an important distinction, and it is captured by calling the systems of the latter kind contextual. In quantum physics this distinction is related to such questions as the (non)existence of hidden variables of which all observed random variables in an experiment are functions.[8]

Int: Yes, I agree this distinction is important. But it seems to me in quantum physics the case of consistent connectedness is the main if not the only case to consider.[9] If we retain the traditional definition of contextuality for such systems (perhaps in the CbD formulation), do we need to extend it to inconsistent connectedness? You said the contextuality in the sense of CbD exists "on top of" inconsistent connectedness. Couldn't we, however, simply ignore it? In other words, couldn't we have a single notion of contextuality, which coincides with "contextuality-2" for consistently connected systems and with "contextuality-1" otherwise? I have almost asked this question before, but we digressed (or at least I see it now as a digression) into the discussion of how one can define objects and contexts.

Exp: Let me think of how to respond to this, and we will return to this in our next conversation.

# Conversation 3

Exp: My dear Interlocutor, 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, but I may be unable to discuss them with you comprehensively.

Int: Let us try intuitive plausibility.

Exp: One argument I find persuasive is appealing to "small" inconsistencies added to consistently connected systems with "large" contextuality. Contextuality in CbD can be rigorously quantified,[10] but I will only need intuitive guidance to present the argument. Consider our system $\left\{\left(R_A^V, R_B^V\right), \left(R_A^W, R_B^W\right)\right\}$. You may recall that the criterion (necessary and sufficient condition) for noncontextuality here is given by the inequality

$$\left|\left\langle R_A^V R_B^V\right\rangle - \left\langle R_A^W R_B^W\right\rangle\right| \leq \left|\left\langle R_A^V\right\rangle - \left\langle R_A^W\right\rangle\right| + \left|\left\langle R_B^V\right\rangle - \left\langle R_B^W\right\rangle\right|.$$

[8]Aut: This is, e.g., how the problem was formulated by Bell (1964). The possibility of reformulating this in terms of the (non)existence of certain couplings (without using this concept explicitly) was realized later, in Suppes and Zanotti (1981) and Fine (1982).

[9]Aut: It may be a prevalent case but definitely not the only one: see, e.g., Bacciagaluppi (2015).

[10]Aut: See Dzhafarov *et al.* (2015a); Kujala *et al.* (2015); Kujala and Dzhafarov (2015); de Barros *et al.* (2015).

Your proposal is to accept this formula only for consistently connected systems, when

$$\left|\left\langle R_A^V\right\rangle - \left\langle R_A^W\right\rangle\right| + \left|\left\langle R_B^V\right\rangle - \left\langle R_B^W\right\rangle\right| = 0.$$

In this case the system is contextual if and only if

$$\left|\left\langle R_A^V R_B^V\right\rangle - \left\langle R_A^W R_B^W\right\rangle\right| > 0.$$

Now, the largest possible value of the last expression is 2, and, I think you would agree, it is reasonable to say that the system $\left\{\left(R_A^V, R_B^V\right), \left(R_A^W, R_B^W\right)\right\}$ with this value equal to 2 exhibits the greatest possible degree of contextuality, given the consistent connectedness.

INT: Is this maximum value 2 compatible with consistent connectedness?

EXP: Yes, it is. See these two distributions:

| | $R_B^V = +1$ | $R_B^V = -1$ | |
|---|---|---|---|
| $R_A^V = +1$ | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ |
| $R_A^V = -1$ | 0 | $\frac{1}{2}$ | $\frac{1}{2}$ |
| | $\frac{1}{2}$ | $\frac{1}{2}$ | |

| | $R_B^W = +1$ | $R_B^W = -1$ | |
|---|---|---|---|
| $R_A^W = +1$ | 0 | $\frac{1}{2}$ | $\frac{1}{2}$ |
| $R_A^W = -1$ | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ |
| | $\frac{1}{2}$ | $\frac{1}{2}$ | |

The computations yield all the expected values $\left\langle R_A^V\right\rangle, \left\langle R_A^W\right\rangle, \left\langle R_B^V\right\rangle, \left\langle R_B^W\right\rangle$ equal to zero, $\left\langle R_A^V R_B^V\right\rangle = 1$ and $\left\langle R_A^W R_B^W\right\rangle = -1$.

INT: I see. Please continue with your example.

EXP: Let us now introduce a minuscule inconsistency, say,

| | $R_B^V = +1$ | $R_B^V = -1$ | |
|---|---|---|---|
| $R_A^V = +1$ | $\frac{1}{2}$ | $\varepsilon$ | $\frac{1}{2} + \varepsilon$ |
| $R_A^V = -1$ | 0 | $\frac{1}{2} - \varepsilon$ | $\frac{1}{2} - \varepsilon$ |
| | $\frac{1}{2}$ | $\frac{1}{2}$ | |

| | $R_B^W = +1$ | $R_B^W = -1$ | |
|---|---|---|---|
| $R_A^W = +1$ | 0 | $\frac{1}{2}$ | $\frac{1}{2}$ |
| $R_A^W = -1$ | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ |
| | $\frac{1}{2}$ | $\frac{1}{2}$ | |

The only difference of this system from the previous one is that now $\langle R_A^V \rangle = 2\varepsilon$ rather than 0 and $\langle R_A^V R_B^V \rangle = 1 - 2\varepsilon$ rather than 1; but $\varepsilon$ can be chosen arbitrarily small. The system therefore can be made arbitrarily close to the previous one. If I continue to follow your proposal, however, I should abandon the joint expectations altogether and focus on the marginals only: the system is contextual now simply because it is inconsistently connected,

$$\left|\langle R_A^V \rangle - \langle R_A^W \rangle\right| + \left|\langle R_B^V \rangle - \langle R_B^W \rangle\right| = 2\varepsilon > 0.$$

But you would probably agree that if $\varepsilon$ is minuscule, the degree of contextuality in the system is minuscule too, wouldn't you?

INT: Indeed I would. And I can guess the rest of your argument too. When $\varepsilon$ is very small but nonzero, the system has a small degree of contextuality (which will be "contextuality-1"). As you make $\varepsilon$ smaller and smaller, the contextuality gets smaller and smaller. But as soon as $\varepsilon$ reaches zero, the contextuality jumps from the limiting zero value to the maximal possible value (because now it is "contextuality-2"). It is a strange behavior, I should admit.

EXP: Precisely. I conclude that your concept of contextuality is not well-formed. If we distinguish inconsistent connectedness from contextuality in accordance with CbD, however, the problem disappears. The degree of contextuality in the second system, if $\varepsilon$ is very small, is only slightly smaller than the degree of contextuality in the first system. The inequality

$$2 = \left|\langle R_A^V R_B^V \rangle - \langle R_A^W R_B^W \rangle\right| > \left|\langle R_A^V \rangle - \langle R_A^W \rangle\right| + \left|\langle R_B^V \rangle - \langle R_B^W \rangle\right| = 0$$

changes into

$$2 - 2\varepsilon = \left|\langle R_A^V R_B^V \rangle - \langle R_A^W R_B^W \rangle\right| > \left|\langle R_A^V \rangle - \langle R_A^W \rangle\right| + \left|\langle R_B^V \rangle - \langle R_B^W \rangle\right| = 2\varepsilon.$$

INT: I agree this feature speaks in favor of the CbD concept. Does this mean, however, that inconsistent connectedness and contextuality (or "contextuality-1" and "contextuality-2," even if you don't like this terminology) have fundamentally different ontologies?

EXP: It is a question to which I do not have a definitive answer. Inconsistent connectedness in most, if not all cases have trivial and well-understood causes: conditions under which measurements are made affect these measurements, either through physical interference or through context-dependent measurement biases. In the example with written and verbal questions, reading a question invokes very different psychological processes than hearing it asked: there is nothing remarkable in the distributions of responses in the two cases being different. Or consider a formally identical but empirically different example: replace the presentation mode with order in which the two questions are asked, so that instead of $V$ we have context $A \to B$ and instead of $W$ the context $B \to A$. In this case it is natural to expect that the first question affects a person's response to the second question.[11] In physics, it is common that different contexts correspond to different experimental set-ups, so that one and the same object in different contexts is simply measured differently.[12]

INT: And you claim that the causes for contextuality are different?

EXP: At least they may be different, and they are definitely different in some cases. Take the famous Alice-Bob experiment (more formally known as the EPR/Bohm or EPR/Bell paradigm), in which Alice measures spins in a particle 1 and Bob measures spins in a particle 2. The two particles are *entangled*, meaning that they were created in a special, singular state that makes Alice's and Bob's spins that are measured in the opposite directions perfectly correlated. Alice and Bob make their measurements simultaneously from some Charlie's point of view, and this precludes any information traveling from Alice to Bob or vice versa. Nevertheless, we have a clear case of contextuality ("contextuality-2") in this case.[13] If we modify this experiment so that Alice and Bob (from Charlie's point of view) make their measurement with an interval between them that allows for signaling, and if we assume that some form of signaling is indeed effected, then we may have distributions of Alice's measurement depending on Bob's settings and/or vice versa. This would be inconsistent connectedness. But contextuality may still be measurable "on top of" this inconsistency.[14]

INT: But you don't know if contextuality is always so different in nature from inconsistent connectedness, do you?

EXP: No, which is why I said I did not have a definitive answer.

INT: Do we know if contextuality of the kind we find in quantum physics also exists in non-physical systems? Perhaps in human behavior?

EXP: No, we don't. A recent analysis of available experimental data in psychology seems to suggest that the answer is negative.[15] But we can't know for sure, because in psychology we lack a theory analogous to quantum mechanics and have to grope in darkness trying now this and then that.

INT: So it is possible that contextuality only exists in quantum physics? This would be disappointing, wouldn't it?

EXP: Not necessarily. Lack of contextuality, if it can be formulated as a general principle in some domain, allows us to predict outcomes of experiments, or at least predict what outcomes are not possible. In psychology this hypothetical principle would, in a sense, create a general theory that we otherwise lack.

[11] AUT: See Moore (2002); Wang *et al.* (2014); Dzhafarov *et al.* (2015c).
[12] AUT: See the discussion of an experiment by Lapkiewicz *et al.* (2011) in Kujala *et al.* (2015).
[13] AUT: See, e.g., Dzhafarov and Kujala (2014b).
[14] AUT: See the chapter by Kujala and Dzhafarov in this volume.
[15] AUT: See Dzhafarov *et al.* (2015c).

INT: I see. Well, we have covered a lot of ground in our conversations. Would you like to summarize?

EXP: I could summarize the definition of contextuality in a more formal way. I recall you did not like informal presentations.

INT: Please do.

EXP: The primitive concepts of the theory are[16]

1. set $Q$ of "objects being measured,"
2. set $C$ of "contexts of measurements,"
3. relation "object $q$ is measured in context $c$," $q \prec c$, and
4. set of "measurements," random variables $R = \left\{R_q^c : q \prec c\right\}_{q \in Q}^{c \in C}$.

Two postulates of the theory are

1. for a given context $c \in C$, the set of measurements $R^c = \left\{R_q^c : q \prec c\right\}_{q \in Q}$ is a random variable (which means the measurements are jointly distributed);
2. any two measurements $R_q^c, R_{q'}^{c'}$ with $c \neq c'$ are stochastically unrelated (whether $q = q'$ or not).

We call the set $R_q = \left\{R_q^c : q \prec c\right\}^{c \in C}$ a *connection* for object $q$. The elements of a connection are pairwise stochastically unrelated. Let $T_q = \left\{T_q^c : q \prec c\right\}^{c \in C}$ be a coupling for connection $R_q$, i.e., $T_q^c$ has the same distribution as $R_q^c$ for all $c$ such that $q \prec c$. This coupling is called *maximal* if $\Pr\left[\text{for any } c, c' \in C,\ T_q^c = T_q^{c'}\right]$ is maximal across all possible couplings for $R_q$. Such a coupling exist, and the *maximal probability* in question, $p_q$, is uniquely defined.

Let $S = \left\{S_q^c : q \prec c\right\}_{q \in Q}^{c \in C}$ be a coupling for the system $R$, i.e., $S^c = \left\{S_q^c : q \prec c\right\}_{q \in Q}$ is distributed as $R^c = \left\{R_q^c : q \prec c\right\}_{q \in Q}$ for all $q \prec c$. This coupling is called *maximally connected* if

$$\Pr\left[\text{for any } c, c' \in C,\ S_q^c = S_q^{c'}\right] = p_q$$

for every $q \in Q$. If $R$ has a maximally connected coupling it is *noncontextual*. If $R$ does not have a maximally connected coupling it is *contextual*.

INT: Do we know criteria of (non)contextuality analogous to the one you mentioned before, for the system $\left\{\left(R_A^V, R_B^V\right), \left(R_A^W, R_B^W\right)\right\}$?

[16] AUT: For details see Dzhafarov *et al.* (2015a); Kujala *et al.* (2015); Kujala and Dzhafarov (2015); de Barros *et al.* (2015).

Exp: Yes, this was a special (in fact, simplest) example of a broad class of systems for which we have criteria of (non)contextuality. These systems are called *cyclic*, and they are defined as follows: for some $n > 1$,

1. the set of objects is $Q = \{q_1, \ldots, q_n\}$,
2. the set of contexts is $C = \{c_1, \ldots, c_n\}$,
3. for each object $q$ there are precisely two contexts $c, c'$ such that $q \prec c$ and $q \prec c'$,
4. for each context $c$ there are precisely two objects $q, q'$ such that $q \prec c$ and $q' \prec c$,
5. all $R_q^c$ with $q \prec c$ are binary, $\pm 1$.

By appropriate enumeration we can always achieve a cyclic structure: $q_i \prec c_i$ and $q_{i \oplus 1} \prec c_i$ for $i = 1, \ldots, n$ (where $\oplus$ is cyclic increment, with $n \oplus 1 = 1$). The system is noncontextual if and only if

$$\max_{\substack{\text{odd number}\\ \text{of minuses}}} \sum_{i=1}^{n} \left( \pm \left\langle R_i^i R_{i\oplus 1}^i \right\rangle \right) \leq (n-2) + \sum_{i=1}^{n} \left| \left\langle R_{i\oplus 1}^i \right\rangle - \left\langle R_{i\oplus 1}^{i\oplus 1} \right\rangle \right|,$$

where each $\pm \left\langle R_i^i R_{i\oplus 1}^i \right\rangle$ is replaced with $+\left\langle R_i^i R_{i\oplus 1}^i \right\rangle$ or $-\left\langle R_i^i R_{i\oplus 1}^i \right\rangle$ so that the minus is chosen an odd number of times $(1, 3, \ldots)$.

Int: This looks like a good point at which to adjourn, my dear Expositor. I have to think this all over. I am sure I will come up with more questions and misgivings.

Exp: I am sure you will, my dear Interlocutor. Until then, good bye.[17]

## Acknowledgments

This work was supported by NSF grant SES-1155956 and AFOSR grant FA9550-14-1-0318. We thank Lasse Leskelä of Aalto University, Matt Jones of University of Colorado, Victor H. Cervantes of Purdue, and Ru Zhang of Purdue for most helpful discussions.

# References

Bacciagaluppi, G. (2015). Leggett-Garg inequalities, pilot waves and contextuality, *International Journal of Quantum Foundations* **1**, pp. 1–17.

[17] Aut: For a brief overview of CbD, see Dzhafarov *et al.* (2015b).

Bell, J. S. (1964). On the Einstein Podolsky Rosen paradox, *Physics* **1**, pp. 195–200.

de Barros, J. A., Dzhafarov, E., Kujala, J. V. and Oas, G. (2015). Measuring observable quantum contextuality, *arXiv:1406.3088* .

Dzhafarov, E. N. (2015). Stochastic unrelatedness, couplings, and contextuality, *arXiv:1506.08218* .

Dzhafarov, E. N. and Kujala, J. V. (2014a). Contextuality is about identity of random variables. *Physica Scripta* **T163**, p. 014009.

Dzhafarov, E. N. and Kujala, J. V. (2014b). No-forcing and no-matching theorems for classical probability applied to quantum mechanics, *Foundations of Physics* **44**, pp. 248–265.

Dzhafarov, E. N. and Kujala, J. V. (2014c). A qualified Kolmogorovian account of probabilistic contextuality, *Lecture Notes in Computer Science* **8369**, pp. 201–212.

Dzhafarov, E. N., Kujala, J. V. and Larsson, J.-Å. (2015a). Contextuality in three types of quantum-mechanical systems, *Foundations of Physics* doi: 10.1007/s10701-015-9882-9.

Dzhafarov, E. N., Kujala, J. V., Larsson, J.-Å. and Cervantes, V. H. (2015b). Contextuality-by-default: A brief overview of ideas, concepts, and terminology, *arXiv:1504.00530* .

Dzhafarov, E. N., Zhang, R. and Kujala, J. V. (2015c). Is there contextuality in behavioral and social systems? *arXiv:1504.07422* .

Fine, A. (1982). Hidden variables, joint probability, and the bell inequalities, *Physical Review Letters* **48**, pp. 291–295.

Kujala, J. V. and Dzhafarov, E. N. (2015). Proof of a conjecture on contextuality in cyclic systems with binary variables, *arXiv:1503.02181* .

Kujala, J. V., Dzhafarov, E. N. and Larsson, J.-Å. (2015). Necessary and sufficient conditions for maximal contextuality in a broad class of quantum mechanical systems, *arXiv:1412.4724* .

Lapkiewicz, R., Li, P., Schaeff, C., Langford, N. K., Ramelow, S., Wieśniak, M. and Zeilinger, A. (2011). Experimental non-classicality of an indivisible quantum system, *Nature* **474**, pp. 490–93.

Moore, D. W. (2002). Measuring new types of question-order effects, *Public Opinion Quarterly* **66**, pp. 80–91.

Suppes, P. and Zanotti, M. (1981). When are probabilistic explanations possible? *Synthese* **48**, pp. 191–199.

Thorisson, H. (2000). *Coupling, Stationarity, and Regeneration*, Springer, New York.

Wang, Z., Solloway, T., Shiffrin, R. M. and Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments, *Proceedings of the National Academy of Sciences* **111**, pp. 9431–9436.