Contents, Contexts, and Basics of Contextuality

Ehtibar N. Dzhafarov

Purdue University

Abstract

This is a non-technical introduction into theory of contextuality. More precisely, it presents the basics of a theory of contextuality called Contextuality-by-Default (CbD). One of the main tenets of CbD is that the identity of a random variable is determined not only by its content (that which is measured or responded to) but also by contexts, systematically recorded conditions under which the variable is observed; and the variables in different contexts possess no joint distributions. I explain why this principle has no paradoxical consequences, and why it does not support the holistic "everything depends on everything else" view. Contextuality is defined as the difference between two differences: (1) the difference between content-sharing random variables when taken in isolation, and (2) the difference between the same random variables when taken within their contexts. Contextuality thus defined is a special form of context-dependence rather than a synonym for the latter. The theory applies to any empirical situation describable in terms of random variables. Deterministic situations are trivially noncontextual in CbD, but some of them can be described by systems of epistemic random variables, in which random variability is replaced with epistemic uncertainty. Mathematically, such systems are treated as if they were ordinary systems of random variables.

1 Contents, contexts, and random variables

The word *contextuality* is used widely, usually as a synonym of *context-dependence*. Here, however, contextuality is taken to mean a special form of context-dependence, as explained below. Historically, this notion is derived from two independent lines of research: in quantum physics, from studies of existence or nonexistence of the so-called hidden variable models with context-independent mapping [1–10],[1] and in psychology, from studies of the so-called selective influences [11–18]. The two lines of research merged relatively recently, in the 2010's [19–24], to form an abstract mathematical theory, Contextuality-by-Default (CbD), with multidisciplinary applications [25–57].[2]

The example I will use to introduce the notion of contextuality reflects the fact that even as I write these lines the world is being ravaged by the Covid-19 pandemic, forcing lockdowns and curtailing travel.

Suppose we ask a randomly chosen person two questions:

q_1 : would you like to take an overseas vacation this summer?
q_2 : are you wary of contracting Covid-19?

[1] Here, I mix together the early studies of nonlocality and those of contextuality in the narrow sense, related to the Kochen-Specker theorem [3]. Both are special cases of contextuality.

[2] The theory has been revised in two ways since 2016, the changes being presented in Refs. [39, 42].

Suppose also we ask these questions in two orders:

c^1 : first q_1 then q_2
c^2 : first q_2 then q_1

To each of the two questions, the person can respond in one of two ways: Yes or No. And since we are choosing people to ask our questions randomly, we cannot determine the answer in advance. We assume therefore that the answers can be represented by *random variables*. A random variable is characterized by its *identity* (as explained shortly) and its *distribution*: in this case, the distribution means responses Yes and No together with their probabilities of occurrence.[3]

One can summarize this imaginary experiment in the form of the following *system of random variables*:

R_1^1	R_2^1	$c^1 = q_1 \to q_2$
R_1^2	R_2^2	$c^2 = q_2 \to q_1$
q_1 = "vacation?"	q_2 = "Covid-19?"	system $\mathcal{C}_{2(a)}$

. (1)

This is the simplest system that can exhibit contextuality (as defined below). The random variables representing responses to questions are denoted by R with subscripts and superscripts determining its identity. The subscript of a random variable in the system refers to the question this random variable answers: e.g., R_1^1 and R_1^2 both answer the question q_1. The superscript refers to the *context* of the random variable, the circumstances under which it is recorded. In the example the context is the order in which the two questions are being asked. Thus, R_2^1 answers question q_2 when this question is asked second, whereas R_2^2 answers the same question when it is is asked first.

The question a random variable answers is generically referred to as this variable's *content*. Contents can always be thought of as having the logical function of questions, but in many cases other than in our example they are not questions in the colloquial meaning. Thus, a q may be one's choice of a physical object to measure, say, a stone to weigh, in which case the stone will be the content of the random variable R_q^c representing the outcome of weighing it (in some context c). Of course, logically, this R_q^c answers the question of how heavy the stone is, and q can be taken to stand for this question.

Returning to our example, each variable R_q^c in our set of four variables is identified by its content ($q = q_1$ or $q = q_2$) and by its context ($c = c^1$ or $c = c^2$). It is this double-identification that imposes a structure on this set, rendering it a *system* (specifically, a *content-context system*) of random variables. There may be other variable circumstances under which our questions are asked, such as when and where the questions were asked, in what tone of voice, or how high the solar activity was when they were asked. However, it is a legitimate choice not to take such concomitant circumstances into account, to ignore them. If we do not, which is a legitimate choice too, our contexts will have to be redefined, yielding a different system, with more than just four random variables. The legitimacy of ignoring all but a select set of contexts is an important aspect of contextuality analysis, as we will see later.

The reason I denote our system $\mathcal{C}_{2(a)}$ is that it is a specific example (the specificity being indicated by index a) of a *cyclic system of rank* 2, denoted $\mathcal{C}_2$. More generally, cyclic systems of rank n, denoted $\mathcal{C}_n$, are characterized by the arrangement of n contents, n contexts, and $2n$ random variables shown in Figure 1.

[3] I set aside the intriguing issue of whether responses Yes and No may be indeterministic but not assignable probabilities.

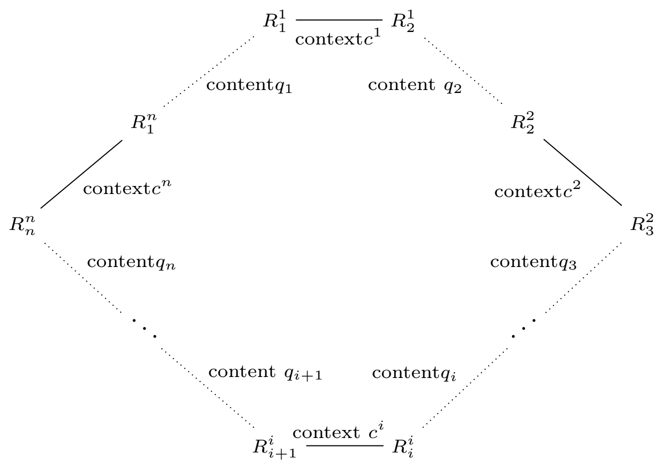
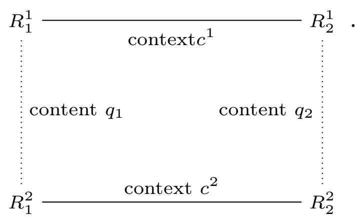


Figure 1: A cyclic system of rank n.

A system of the $\mathcal{C}_2$-type is the smallest such system (not counting the degenerate system consisting of R_1^1 alone):



What else do we know of our random variables? First of all, the two variables within a context, $\left(R_1^1, R_2^1\right)$, or $\left(R_1^2, R_2^2\right)$, are *jointly distributed*. By the virtue of being responses of one and same person, the values of these random variables come in pairs. So it is meaningful to ask what the probabilities are for each of the joint events

$$\begin{array}{lcl} R_1^1 = +1 & \text{and} & R_2^1 = +1, \\ R_1^1 = +1 & \text{and} & R_2^1 = -1, \\ R_1^1 = -1 & \text{and} & R_2^1 = +1, \\ R_1^1 = -1 & \text{and} & R_2^1 = -1, \end{array}$$

where $+1$ and -1 encode the answers Yes and No, respectively. One can meaningfully speak of correlations between the variables in the same context, probability that they have the same value, etc.

By contrast, different contexts, in our case the two orders in which the questions are asked, are mutually exclusive. When asked two questions, a given person can only be asked them in one order. Respondents represented by R_1^1 answer question q_1 asked first, before q_2, whereas the respondents represented by R_1^1 answer question q_1 asked second, after q_2. Clearly, these are different sets of respondents, and one would not know how to pair them. It is meaningless to ask, e.g., what the probability of

$$R_1^1 = +1 \quad \text{and} \quad R_1^2 = +1$$

may be. Random variables in different contexts are *stochastically unrelated.*

2 Intuition of (non)contextuality

Having established these basic facts, let us consider now the two random variables with content q_1, and let us make at first the (unrealistic) assumption that their distributions are the same in both contexts, c^1 and c^2:

$$\begin{array}{c|c} \textit{value} & \textit{probability} \\ \hline R_1^1 = +1 & a \\ \hline R_1^1 = -1 & 1-a \end{array} \text{ and } \begin{array}{c|c} \textit{value} & \textit{probability} \\ \hline R_1^2 = +1 & a \\ \hline R_1^2 = -1 & 1-a \end{array}. \tag{2}$$

If we consider the variables R_1^1 and R_1^2 in isolation from their contexts (i.e., disregarding the other two random variables), then we can view them as simply one and the same random variable. In other words, the subsystem

R_1^1	$c^1 = q_1 \rightarrow q_2$
R_1^2	$c^2 = q_2 \rightarrow q_1$
q_1 = "vacation?"	$\mathcal{C}_{2(a)}$/only q_1

appears to be replaceable with just

R_1
q_1 = "vacation?"

,

with contexts being superfluous.

Analogously, if the distributions of the two random variables with content q_2 are assumed to be the same,

$$\begin{array}{c|c|c} \textit{value} & R_2^1 = +1 & R_2^1 = -1 \\ \hline \textit{probability} & b & 1-b \end{array} \text{ and } \begin{array}{c|c|c} \textit{value} & R_2^2 = +1 & R_2^2 = -1 \\ \hline \textit{probability} & b & 1-b \end{array}, \tag{3}$$

and if we consider them in isolation from their contexts, the subsystem

R_2^1	$c^1 = q_1 \rightarrow q_2$
R_2^2	$c^2 = q_2 \rightarrow q_1$
q_2 = "Covid-19?"	$\mathcal{C}_{2(a)}$/only q_2

appears to be replaceable with

R_2
q_2 = "Covid-19?"

.

It is tempting now to say: we have only two random variables, R_1 and R_2, whatever their contexts. But a given pair of random variables can only have one *joint distribution*, this distribution cannot be somehow different in different contexts. We should predict therefore, that if the probabilities in system $\mathcal{C}_{2(a)}$ are

$$\Pr\left[R_1^1 = +1, R_2^1 = +1\right] = r_1 \text{ and } \Pr\left[R_1^2 = +1, R_2^2 = +1\right] = r_2,$$

then

$$r_1 = r_2.$$

Suppose, however, that this is shown to be empirically false, that in fact $r_1 > r_2$. For instance, assuming $0 < a < b$, suppose that the joint distributions in the two contexts of system $\mathcal{C}_{2(a)}$ are

context c^1	$R_2^1 = +1$	$R_2^1 = -1$	
$R_1^1 = +1$	$r_1 = a$	0	a
$R_1^1 = -1$	$b - a$	$1 - b$	$1 - a$
	b	$1 - b$	

(4)

and

context c^2	$R_2^2 = +1$	$R_2^2 = -1$	
$R_1^2 = +1$	$r_2 = 0$	a	a
$R_1^2 = -1$	b	$1 - a - b$	$1 - a$
	b	$1 - b$	

. (5)

Clearly, we have then a *reductio ad absurdum* proof that the assumption we have made is wrong, the assumption being that we can drop contexts in R_1^1 and R_1^2 (as well as in R_2^1 and R_2^2), and that we can therefore treat them as one and the same random variable R_1 (respectively, R_2). This is the simplest case when we can say that a system of random variables, here, the system $\mathcal{C}_{2(a)}$, is *contextual.*

This understanding of contextuality can be extended to more complex systems. However, is far from being general enough. It only applies to *consistently connected* systems, those in which any two variables with the same content are identically distributed.[4] This assumption is often unrealistic. Specifically, it is a well-established empirical fact that the individual distributions of the responses to two questions do depend on their order [58]. Besides, this is highly intuitive in our example. If one is asked about an overseas vacation first, the probability of saying "Yes, I would like to take an overseas vacation" may be higher than when this question is asked second, after the respondent has been reminded about the dangers of the pandemic.

In order to generalize the notion of contextuality to arbitrary systems, we need to develop answers to the following two questions:

A: For any two random variables sharing a content, how different are they when taken in isolation from their contexts?

B: Can these differences be preserved when all pairs of content-sharing variables are taken within their contexts (i.e., taking into account their joint distributions with other random variables in their contexts)?

For our system $\mathcal{C}_{2(a)}$ with the within-context joint distributions given by (4) and (5), our informal answer to question **A** was that two random variables with the same content (i.e., R_1^1 and R_1^2 or R_2^1 and R_2^2) are not different at all when taken in isolation. The informal answer to question **B**, however, was that in these two pairs (or at least in one of them) the random variables are not the same when taken in relation to other random variables in their respective contexts. One can say therefore that

[4] The term "consistent connectedness" is due to the fact that in CbD the content-sharing random variables are said to form *connections* (between contexts). In quantum physics consistent connectedness is referred to by such terms as lack of signaling, lack of disturbance, parameter invariance, etc.

the contexts make R_1^1 and R_1^2 (and/or R_2^1 and R_2^2) more dissimilar than when they are taken without their contexts.

This is the intuition we will use to construct a general definition of contextuality.

3 Making it rigorous: Couplings

First, we have to agree on how to measure the difference between two random variables that are not jointly distributed, like R_1^1 and R_1^2. Denote these random variables X and Y, both dichotomous (± 1), with

$$\Pr[X = +1] = u \text{ and } \Pr[Y = +1] = v.$$

Consider all possible pairs of *jointly distributed* variables (X', Y') such that

$$X' \overset{dist}{=} X, Y' \overset{dist}{=} Y,$$

where $\overset{dist}{=}$ stands for "has the same distribution as." Any such pair (X', Y') is called a *coupling* of X and Y. For obvious reasons, two couplings of X and Y having the same joint distribution are not distinguished.

Now, for each coupling (X', Y') one can compute the probability with which $X' \neq Y'$ (recall that the probability of $X \neq Y$ is undefined, we do need couplings to make this inequality a meaningful event). It is easy to see that among the couplings (X', Y') there is one and only one for which this probability is minimal. This coupling is defined by the joint distribution

	$Y' = +1$	$Y' = -1$	
$X' = +1$	$\min(u, v)$	$u - \min(u, v)$	u
$X' = -1$	$v - \min(u, v)$	$\min(1 - u, 1 - v)$	$1 - u$
	v	$1 - v$	

, (6)

and the minimal probability in question is obtained as

$$(u - \min(u, v)) + (v - \min(u, v)) = |u - v|.$$

This probability is a natural measure of difference between the random variables X and Y:[5]

$$\delta(X, Y) = \min_{\substack{\text{all couplings} \\ (X', Y') \text{ of } X \text{ and } Y}} \Pr[X' \neq Y'] = |u - v|. \tag{7}$$

If X and Y are identically distributed, i.e. $u = v$, the joint distribution of X' and Y' can be chosen as

context c^1	$Y = +1$	$Y = -1$	
$X = +1$	u	0	u
$X = -1$	0	$1 - u$	$1 - u$
	u	$1 - u$	

,

[5] It is a special case of the so-called *total variation distance*, except that it is usually defined between two probability distributions, while I use it here as a measure of difference (formally, a *pseudometric*) between two stochastically unrelated random variables.

yielding

$$\delta(X,Y) = \min_{\substack{\text{all couplings} \\ (X',Y') \text{ of } X \text{ and } Y}} \Pr[X' \neq Y'] = 0.$$

Let us apply this to our example, in order to formalize the intuition behind our saying earlier that two identically distributed random variables, taken in isolation, can be viewed as being "the same." For R_1^1 and R_1^2 in (2),

$$\delta\left(R_1^1, R_1^2\right) = \min_{\substack{\text{all couplings} \\ \left(S_1^1, S_1^2\right) \text{ of } R_1^1 \text{ and } R_1^2}} \Pr\left[S_1^1 \neq S_1^2\right] = 0,$$

and, analogously, for R_2^1 and R_2^2 in (3),

$$\delta\left(R_2^1, R_2^2\right) = \min_{\substack{\text{all couplings} \\ \left(S_2^1, S_2^2\right) \text{ of } R_2^1 \text{ and } R_2^2}} \Pr\left[S_2^1 \neq S_2^2\right] = 0.$$

4 Making it rigorous: Contextuality

What is then the rigorous way of establishing that these differences cannot both be zero when considered within their contexts? For this, we need to extend the notion of a coupling to an entire system. A coupling of our system $\mathcal{C}_{2(a)}$ is a set of *corresponding jointly distributed* random variables

$$\begin{array}{|c|c|} \hline S_1^1 & S_2^1 \\ \hline S_1^2 & S_2^2 \\ \hline \end{array} \tag{8}$$

such that

$$\left(S_1^1, S_2^1\right) \overset{dist}{=} \left(R_1^1, R_2^1\right), \left(S_1^2, S_2^2\right) \overset{dist}{=} \left(R_1^2, R_2^2\right). \tag{9}$$

In other words, the distributions within contexts, (4) and (5), remain intact when we replace the R-variables with the corresponding S-variables,

	$S_2^1 = +1$	$S_2^1 = -1$	
$S_1^1 = +1$	a	0	a
$S_1^1 = -1$	$b - a$	$1 - b$	$1 - a$
	b	$1 - b$	

and

	$S_2^2 = +1$	$S_2^2 = -1$	
$S_1^2 = +1$	0	a	a
$S_1^2 = -1$	b	$1 - a - b$	$1 - a$
	b	$1 - b$	

. (10)

Such couplings always exist, not only for our example, but for any other system of random variables. Generally, there is an infinity of couplings for a given system.[6] Thus, to construct a

[6] One need not have separate definitions of couplings for pairs of random variables and for systems. In general, given any set of random variables $\mathfrak{R}$, its coupling is a set of random variables S, in a one-to-one correspondence with $\mathfrak{R}$, such that the corresponding variables in $\mathfrak{R}$ and S have the same distribution, and all variables in S are jointly distributed. To apply this definition to $\mathfrak{R}$ representing a system of random variables one considers all variables within a given context as a single element of $\mathfrak{R}$. In our example, (8) is a coupling of two stochastically unrelated random variables, $\left(R_1^1, R_2^1\right)$ and $\left(R_1^2, R_2^2\right)$.

coupling for system $\mathcal{C}_{2(a)}$, one has to assign probabilities to all quadruples of joint events,

S_1^1	S_2^1	S_1^2	S_2^2	*probability*
$+1$	$+1$	$+1$	$+1$	p_{++++}
$+1$	$+1$	$+1$	-1	p_{+++-}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
-1	-1	-1	-1	p_{----}

so that the appropriately chosen subsets of these probabilities sum to the joint probabilities shown in (10):

$$\begin{array}{c}
p_{++++}+p_{+++-}+p_{++-+}+p_{++--}=\Pr\left[S_1^1=+1,S_2^1=+1\right]=a,\\
p_{+-++}+p_{+-+-}+p_{+--+}+p_{+---}=\Pr\left[S_1^1=+1,S_2^1=-1\right]=0,\\
p_{++++}+p_{+-++}+p_{-+++}+p_{--++}=\Pr\left[S_1^2=+1,S_2^2=+1\right]=0,\\
\textit{etc.}
\end{array}$$

This is a system of seven independent linear equations with 16 unknown p-probabilities, subject to the additional constraint that all probabilities must be nonnegative. It can be shown that this linear programming problem always has solutions, and infinitely many of them at that, unless one of the probabilities a and b equals 1 or 0 (in which case the solution is unique).

Unlike in system $\mathcal{C}_{2(a)}$ itself, in any coupling (8) of this system the random variables have joint distributions across the contexts. In particular, $\left(S_1^1,S_1^2\right)$ is a jointly distributed pair. Since from (9) we know that

$$S_1^1\overset{dist}{=}R_1^1 \text{ and } S_1^2\overset{dist}{=}R_1^2,$$

$\left(S_1^1,S_1^2\right)$ is a coupling of R_1^1 and R_2^1. Similarly, $\left(S_2^1,S_2^2\right)$ is a coupling of R_2^1 and R_2^2. We ask now: what are the possible values of

$$\Pr\left[S_1^1\neq S_1^2\right] \text{ and } \Pr\left[S_2^1\neq S_2^2\right]$$

across all possible couplings (8) of the entire system $\mathcal{C}_{2(a)}$? Consider two cases.

Case 1. In some of the couplings (8),

$$\Pr\left[S_1^1\neq S_1^2\right]=0 \text{ and } \Pr\left[S_2^1\neq S_2^2\right]=0.$$

We can say then that both $\delta\left(R_1^1,R_1^2\right)$ and $\delta\left(R_1^1,R_1^2\right)$ preserve their individual (in-isolation) values when considered within the system. The system $\mathcal{C}_{2(a)}$ is then considered *noncontextual*.

Case 2. In all couplings (8), at least one of the values

$$\Pr\left[S_1^1\neq S_1^2\right] \text{ and } \Pr\left[S_2^1\neq S_2^2\right]$$

is greater than zero. That is, when considered within the system, $\delta\left(R_1^1,R_1^2\right)$ and $\delta\left(R_1^1,R_1^2\right)$ cannot both be zero. Intuitively, the contexts "force" either R_1^1 and R_1^2 or R_2^1 and R_2^2 (or both) to be more dissimilar than when taken in isolation. The system $\mathcal{C}_{2(a)}$ is then considered *contextual*.

We can quantify the degree of contextuality in the system in the following way. We know that

$$\begin{array}{l}
\delta\left(R_1^1,R_1^2\right)+\delta\left(R_2^1,R_2^2\right)\\
\\
=\displaystyle\min_{\substack{\text{all couplings}\\ \left(S_1^1,S_1^2\right)\text{ of }R_1^1\text{ and }R_1^2}}\left(\Pr\left[S_1^1\neq S_1^2\right]\right)+\min_{\substack{\text{all couplings}\\ \left(S_2^1,S_2^2\right)\text{ of }R_2^1\text{ and }R_2^2}}\left(\Pr\left[S_2^1\neq S_2^2\right]\right)=0.
\end{array}$$

This quantity is compared to

$$\delta\left(\left(R_1^1,R_1^2\right),\left(R_2^1,R_2^2\right)\right) = \min_{\substack{\text{all couplings}\\ \left(S_1^1,S_2^1,S_1^2,S_2^2\right)\text{ of system }\mathcal{C}_{2(a)}}} \left(\Pr\left[S_1^1\neq S_1^2\right]+\Pr\left[S_2^1\neq S_2^2\right]\right),$$

which can be interpreted as the total of the pairwise differences between same-content variables within the system. The system is contextual if this quantity is greater than zero, and this quantity can be taken as a measure of the degree of contextuality. This is by far not the only possible measure, but it is arguably the simplest one within the conceptual framework of CbD.

5 Generalizing to arbitrary systems

Consider now a realistic version of our example, when

$$\begin{array}{c}\Pr\left[R_1^1=+1\right]=a_1,\Pr\left[R_2^1=+1\right]=b_1,\\ \Pr\left[R_1^2=+1\right]=a_2,\Pr\left[R_2^2=+1\right]=b_2,\end{array}$$

with a_1 allowed to be different from a_2, and b_1 from b_2. The within-context joint distributions then generally look like this:

context c^1	$R_2^1=+1$	$R_2^1=-1$	
$R_1^1=+1$	r_1	a_1-r_1	a_1
$R_1^1=-1$	b_1-r_1	$1-a_1-b_1+r_1$	$1-a_1$
	b_1	$1-b_1$	

(11)

and

context c^2	$R_2^2=+1$	$R_2^2=-1$	
$R_1^2=+1$	r_2	a_2-r_2	a_2
$R_1^2=-1$	b_2-r_2	$1-a_2-b_2+r_2$	$1-a_2$
	b_2	$1-b_2$	

. (12)

Let us call the system in (1) with these within-context distributions $C_{2(b)}$. We clearly have context-dependence now (unless the two joint distributions are identical), but can we also say that the system is contextual? If we follow the logic of the definition of contextuality as it was presented above, for consistently connected systems, the answer cannot automatically be affirmative. The logic in question requires that we answer the questions **A** and **B** formulated in Section 2. By now we have all necessary conceptual tools for this.

To answer **A** we look at all possible couplings $\left(S_1^1,S_1^2\right)$ and $\left(S_2^1,S_2^2\right)$ of the content-sharing pairs $\left\{R_1^1,R_1^2\right\}$ and $\left\{R_2^1,R_2^2\right\}$, respectively, and determine

$$\delta\left(R_1^1,R_1^2\right) = \min_{\substack{\text{all couplings}\\ \left(S_1^1,S_1^2\right)\text{ of }\left\{R_1^1,R_1^2\right\}}} \Pr\left[S_1^1\neq S_1^2\right],$$

and

$$\delta\left(R_2^1,R_2^2\right) = \min_{\substack{\text{all couplings}\\ \left(S_2^1,S_2^2\right)\text{ of }\left\{R_2^1,R_2^2\right\}}} \Pr\left[S_2^1\neq S_2^2\right].$$

To answer **B**, we look at all possible couplings

$$\begin{array}{|c|c|}\hline S_1^1 & S_2^1 \\ \hline S_1^2 & S_2^2 \\ \hline\end{array}$$

of the entire system $\mathcal{C}_{2(b)}$, and determine if we can find couplings in which

$$\Pr\left[S_1^1 \neq S_1^2\right] = \delta\left(R_1^1, R_1^2\right)$$

and

$$\Pr\left[S_2^1 \neq S_2^2\right] = \delta\left(R_2^1, R_2^2\right).$$

If such couplings exist, we say that the system is noncontextual, even if it exhibits context-dependence in the form of *inconsistent connectedness.*

Recall that consistently connected systems are those in which any two variables with the same content are identically distributed, as it was in our initial (unrealistic) example. For such systems $\delta\left(R_1^1, R_1^2\right) = 0$ and $\delta\left(R_2^1, R_2^2\right) = 0$. However, if

$$R_1^1 \overset{dist}{\neq} R_1^2,$$

then $\delta\left(R_1^1, R_1^2\right) > 0$, and analogously for $\delta\left(R_2^1, R_2^2\right)$. In fact, we know from (6) and (7) that if the within-context distributions in the system are as in (11) and (12), then

$$\delta\left(R_1^1, R_1^2\right) = |a_1 - a_2|, \delta\left(R_2^1, R_2^2\right) = |b_1 - b_2|.$$

This means that system $\mathcal{C}_{2(b)}$ is contextual if and only if

$$\delta\left(\left(R_1^1, R_1^2\right), \left(R_2^1, R_2^2\right)\right) = \min_{\substack{\text{all couplings} \\ \left(S_1^1, S_2^1, S_1^2, S_2^2\right) \\ \text{of system } \mathcal{C}_{2(b)}}} \left(\Pr\left[S_1^1 \neq S_1^2\right] + \Pr\left[S_2^1 \neq S_2^2\right]\right)$$

$$> |a_1 - a_2| + |b_1 - b_2|.$$

Indeed, this inequality indicates that in all couplings either

$$\Pr\left[S_1^1 \neq S_1^2\right] > \delta\left(R_1^1, R_1^2\right),$$

or

$$\Pr\left[S_2^1 \neq S_2^2\right] > \delta\left(R_2^1, R_2^2\right),$$

or both. The intuition remains the same as above: the contexts "force" the same-content variables to be more dissimilar than they are in isolation. The difference

$$\delta\left(\left(R_1^1, R_1^2\right), \left(R_2^1, R_2^2\right)\right) - \delta\left(R_1^1, R_1^2\right) - \delta\left(R_2^1, R_2^2\right)$$

is a natural (although by far not the only) measure of the degree of contextuality.[7]

[7] For other measures of contextuality, see Refs. [50, 53–55]

6 Other examples

The system $\mathcal{C}_{2(b)}$ of the previous section, with the within-context distributions (11) and (12), is not a toy example, despite its simplicity. Except for the specific choice of the questions, it describes an empirical situation one sees in polls of public opinion, with two questions asked in one order of a large group of participants, and the same two questions asked in the other order of another large group of participants [58, 59].

In quantum physics, system of the $\mathcal{C}_2$-type can describe the outcomes of successive measurements of two spins along two directions, encoded 1 and 2, in the same spin-$1/2$ particle (e.g., electron). Without getting into details, in such an experiment the spin-$1/2$ particles are *prepared* in one and the same *quantum state*, and then subjected to two measurements in one of the two orders. Each measurement results in one of two outcomes, spin up ($+1$) or spin down (-1).

R_1^1	R_2^1	$c^1 = q_1 \rightarrow q_2$
R_1^2	R_2^2	$c^2 = q_2 \rightarrow q_1$
$q_1 =$ "is spin in direction 1 up?"	$q_2 =$ "is spin in direction 2 up?"	system $\mathcal{C}_{2(c)}$

. (13)

The computations in accordance with the standard quantum-mechanical rules yield the following two results [30]. First, the system is inconsistently connected, i.e. generally the probability of spin-up in a given direction depends on whether it is measured first or second,

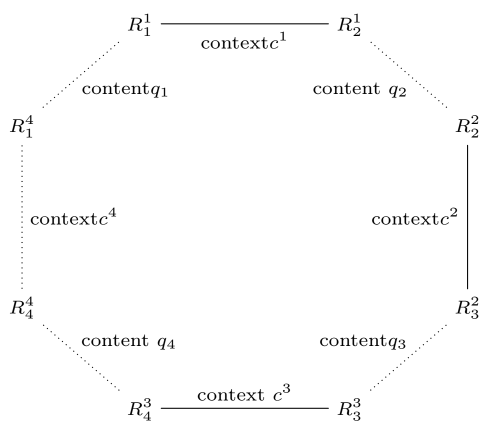
$$\Pr\left[R_1^1 = +1\right] \neq \Pr\left[R_1^2 = +1\right] \text{ and } \Pr\left[R_2^1 = +1\right] \neq \Pr\left[R_2^2 = +1\right].$$

Second, the system is noncontextual,[8] i.e., it is always the case that

$$\delta\left(\left(R_1^1, R_1^2\right), \left(R_2^1, R_2^2\right)\right) \leq \delta\left(R_1^1, R_1^2\right) + \delta\left(R_2^1, R_2^2\right).$$

As we see, systems of the $\mathcal{C}_2$-type may be of interest in both physics and behavioral studies.

However, in both these fields, the origins of the research of what we now call contextuality are dated back to another cyclic system, in which the arrangement shown in Figure 1 specializes to



.

[8] For those familiar with CbD, this follows from the fact the expected values $\langle R_1^1 R_2^1 \rangle$ and $\langle R_1^2 R_2^2 \rangle$ are always equal to each other, whereas the criterion for contextuality of a cyclic system [36], when specialized to $n = 2$, is $\left|\langle R_1^1 R_2^1 \rangle - \langle R_1^2 R_2^2 \rangle\right| > \left|\langle R_1^1 \rangle - \langle R_1^2 \rangle\right| + \left|\langle R_2^1 \rangle - \langle R_2^2 \rangle\right|$.

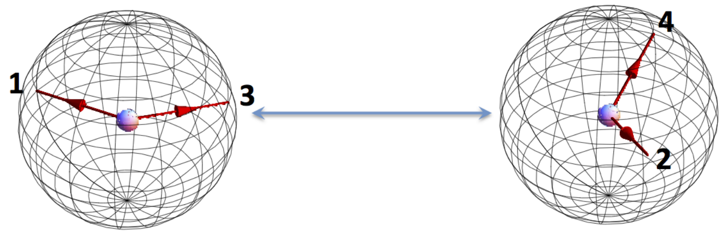


Figure 2: A schematic representation of the EPR/Bohm experimental set up. Explanations in the text.

Figure 2 illustrates the empirical situation described by this system, and the first for which contextuality was mathematically established [2, 4–6, 60]. Two spin-$1/2$ particles are prepared in a special quantum state making them *entangled*, and they move away from each other. The "left" particle's spin is measured along one of the two directions (encoded 1 and 3) by someone we will call Zora, and simultaneously the "right" particle's spin is measured along one of the two directions (encoded 2 and 4) by a Nico.[9] The outcomes of the measurements are spin-up or spin-down, and each random variable R_i^j answers the question

$$q_i : \text{is the spin in direction } i \text{ up? } (i = 1, 2, 3, 4).$$

In the form of a content-context matrix the system can be presented as

$$\begin{array}{|c|c|c|c||c|}
\hline
R_1^1 & R_2^1 & & & c^1 \\
\hline
 & R_2^2 & R_3^2 & & c^2 \\
\hline
 & & R_3^3 & R_4^3 & c^3 \\
\hline
R_1^4 & & & R_4^4 & c^4 \\
\hline
\hline
\begin{array}{c} q_1 \\ \text{(Zora's 1)} \end{array} & \begin{array}{c} q_2 \\ \text{(Nico's 2)} \end{array} & \begin{array}{c} q_3 \\ \text{(Zora's 3)} \end{array} & \begin{array}{c} q_4 \\ \text{(Nico's 4)} \end{array} & \mathcal{C}_{4(a)} \\
\hline
\end{array}\,. \tag{14}$$

The measurements by Zora and Nico are made simultaneously, or at least close enough in time so that no signal about Zora's choice of a direction can reach Nico before he makes his measurement, and vice versa. Because of this, the system is consistently connected,

$$R_i^j \overset{dist}{=} R_i^{j'}$$

for any content q_i and two contexts c^j and $c^{j'}$ in which q_i is measured. Following the logic of contextuality analysis, we first establish that (because of the consistent connectedness)

$$\delta\left(R_1^1, R_1^4\right) = \delta\left(R_2^1, R_2^2\right) = \delta\left(R_3^2, R_3^3\right) = \delta\left(R_4^3, R_4^4\right) = 0.$$

[9] For no deep reason, I decided to deviate from the established tradition to call the imaginary performers of the measurements in this task Alice and Bob.

Then we compute

$$\delta\left(\left(R_1^1,R_1^4\right),\left(R_2^1,R_2^2\right),\left(R_3^2,R_3^3\right),\left(R_4^3,R_4^4\right)\right)$$

$$= \min_{\substack{\text{all couplings}\\ \left(S_1^1,S_4^1,S_1^2,S_2^2,S_2^3,S_3^3,S_3^4,S_4^4\right)\\ \text{of system } \mathcal{C}_{4(a)}}} \left(\Pr\left[S_1^1\neq S_1^4\right]+\Pr\left[S_2^1\neq S_2^2\right]+\Pr\left[S_3^2\neq S_3^3\right]+\Pr\left[S_4^3\neq S_4^4\right]\right).$$

The system is noncontextual if and only if this quantity is zero. As it turns out (and this is what was established by John Bell in his celebrated papers in the 1960s, [1,2]), the directions $1,2,3,4$ can be chosen so that, by the laws of quantum mechanics, this quantity is greater than zero, making the system contextual.

In psychology, systems of the same $\mathcal{C}_4$-type have been of interest as representing the following empirical situation [11–13, 15–18, 23]. Consider two variables having two values each, that can be manipulated in an experiment. Think, e.g., of a briefly presented visual object that can have one of two colors (red or green) and one of two shapes (square or oval), combined in the 2×2 ways. In the experiment, an observer responds to the object by answering two Yes-No questions: "is the object red?" and "is the object square?". If we simply identify these questions with contents, the resulting system of random variables looks like this:

R_1^1	R_2^1	c^1 : red and oval
R_1^2	R_2^2	c^2 : green and oval
R_1^3	R_2^3	c^3 : red and square
R_1^4	R_2^4	c^4 : green and square
q_1 : red?	q_2 : square?	$\mathcal{R}$

, (15)

with the contexts describing the object being presented, and the contents the questions asked.

Although possible, this is not, however, an especially interesting way of conceptualizing the situation. It is more informative to describe the contents of the random variables as color and shape responses to the color and shape of the visual stimuli, respectively:

q_1 : does this red object appear red?
q_2 : does this square object appear square?
q_3 : does this green object appear red?
q_4 : does this oval object appear square?

With the contexts remaining as they are in system (15), the experiment is now represented by a system of the $\mathcal{C}_4$-type:

R_1^1	R_2^1			c^1
	R_2^2	R_3^2		c^2
		R_3^3	R_4^3	c^3
R_1^4			R_4^4	c^4
q_1 (red)	q_2 (square)	q_3 (green)	q_4 (oval)	$\mathcal{C}_{4(b)}$

Compared to system $\mathcal{C}_{4(c)}$ in (14), the physical situation described by $\mathcal{C}_{4(b)}$ is, of course, very different: e.g., instead of R_1^j and R_3^j being outcomes of spin measurements by Zora along two different directions, these random variables represent now responses to the color question when the color is red and when it is green, respectively. However, the logic of the contextuality analysis does not change. If this system turns out to be consistently connected and noncontextual, the interpretation of this in psychology is that the judgment of color is *selectively influenced by object's color* (irrespective of its shape), and the judgment of shape is *selectively influenced by object's shape* (irrespective of its color). Deviations from this pattern of selective influences, whether in the form of inconsistent connectedness or contextuality, or both,[10] provide an interesting way of classifying (and quantifying) the ways object's color may influence one's judgment of its shape and vice versa.

7 What if the system is deterministic?

A *deterministic quantity* r is a special case of a random variable: it is a random variable R that attains the value r with probability 1:

$$\Pr[R = r] = 1.$$

It is convenient to present this as

$$R \equiv r.$$

A *deterministic system* is one containing only deterministic variables. For instance,

$$\begin{array}{|c|c|c|c|c||c|}
\hline
r_1^1 & r_2^1 & & r_4^1 & & c^1 \\
\hline
r_1^2 & & r_3^2 & & & c^2 \\
\hline
 & r_2^3 & r_3^3 & r_4^3 & r_5^3 & c^3 \\
\hline
 & & r_3^4 & & r_5^4 & c^4 \\
\hline
\hline
q_1 & q_2 & q_3 & q_4 & q_5 & \mathcal{D} \\
\hline
\end{array} \tag{16}$$

is a deterministic systems in which r_i^j represents a random variable $R_i^j \equiv r_i^j$. The system can be consistently connected (if the value of r_i^j does not depend on j) or inconsistently connected (otherwise).

It is easy to see, however, that a deterministic system is always noncontextual.[11] Indeed, any two content-sharing $R_i^j \equiv r_i^j$ and $R_i^{j'} \equiv r_i^{j'}$ in this system have a single coupling $(S_i^j \equiv r_i^j, S_i^j \equiv r_i^j)$,

[10] System $\mathcal{C}_{4(d)}$ is almost certainly inconsistently connected (guessing of an imaginary experiment based on the results of many real ones).

[11] This fact was first mentioned to me years ago by Matt Jones of the University of Colorado.

consisting of the same deterministic quantities but considered jointly distributed.[12] It follows that

$$\delta\left(r_i^j, r_i^{j'}\right) = \begin{cases} 1 & \text{if} \quad r_i^j \neq r_i^{j'} \\ 0 & \text{if} \quad r_i^j = r_i^{j'} \end{cases}.$$

The entire deterministic system in (16) also has a single coupling, one containing the same deterministic quantities as the system itself, but considered jointly distributed. Clearly, the subcoupling $\left(S_i^j \equiv r_i^j, S_i^{j'} \equiv r_i^{j'}\right)$ extracted from this coupling is precisely the same as the coupling of $R_i^j \equiv r_i^j$ and $R_i^{j'} \equiv r_i^{j'}$ taken in isolation, and

$$\delta\left(\left\{\left(r_i^j, r_i^{j'}\right) : \text{all such pairs}\right\}\right) = \sum_{\text{all such pairs}} \delta\left(r_i^j, r_i^{j'}\right).$$

One might conclude that deterministic systems are of no interest for contextuality analysis. This is not always true, however. There are cases when we know that a system is deterministic, but we do not know which of a set of possible deterministic systems it is, because it can be any of them. Let us look at this in detail, using as examples systems consisting of logical truth values of various statements.

Consider first the following $\mathcal{C}_4$-type system:

$R_1^1 \equiv +1$	$R_2^1 \equiv -1$			c^1
	$R_2^2 \equiv +1$	$R_3^2 \equiv -1$		c^2
		$R_3^3 \equiv +1$	$R_4^3 \equiv -1$	c^3
$R_1^4 \equiv -1$			$R_4^4 \equiv +1$	c^4
q_1	q_2	q_3	q_4	$\mathcal{C}_{4(c)}$

, (17)

where +1 and −1 encode truth values (true and false), and the contents are the statements

q_1 : "my name is Zora"	q_2 : "my name is Nico"
q_3 : "my name is Max"	q_4 : "my name is Alex"

.

Equivalently, the contents could also be formulated as questions, "is my name Zora?" and "is my name Nico?", in which case +1 and −1 would encode answers Yes and No. In the following, however, I will refer to the q's as statements, and the values of the variables as truth values. The contexts justifying the truth values in (17) are

c^1 : the statements are made by Zora	c^2 : the statements are made by Nico
c^3 : the statements are made by Max	c^4 : the statements are made by Alex

.

[12]There is a subtlety here, first pointed out to me by Janne Kujala of Turku University. If $R_i^j \equiv r_i^j$ and $R_i^{j'} \equiv r_i^{j'}$, one may be tempted to say that the joint event $\left(R_i^j \equiv r_i^j, R_i^{j'} \equiv r_i^{j'}\right)$ has the probability one, and this would create an exception from the principle that random variables in different contexts are not jointly distributed. This is wrong, however, because $\left(R_i^j \equiv r_i^j, R_i^{j'} \equiv r_i^{j'}\right)$ can only be thought of counterfactually, as it involves mutually exclusive contexts. In fact, the only justification (or, better put, excuse) for the intuition that $\left(R_i^j \equiv r_i^j, R_i^{j'} \equiv r_i^{j'}\right)$ is a meaningful joint event is that $R_i^j \equiv r_i^j$ and $R_i^{j'} \equiv r_i^{j'}$ have a single coupling, and in this coupling $\Pr\left[S_i^j \equiv r_i^j, S_i^j \equiv r_i^j\right] = 1$. More generally, use of couplings is a rigorous way of dealing with counterfactuals [49].

This is a situation when the truth values are determined uniquely, the system is deterministic, and consequently it is noncontextual (even though context-dependence in it is salient in the form of inconsistent connectedness).

Consider next another system of the C_4-type,

R_1^1	R_2^1			c^1
	R_2^2	R_3^2		c^2
		R_3^3	R_4^3	c^3
R_1^4			R_4^4	c^4
q_1 : "q_2 is true"	q_2 : "q_3 is true"	q_3 : "q_4 is true"	q_4 : "q_1 is false"	$\mathcal{C}_{4(d)}$

,

with contents/statements of a very different kind, and the contexts which here (at least provisionally) can simply be defined by which statements they include: c^1 includes (q_1, q_2), c^2 includes (q_2, q_3), etc.

One can recognize here a formalization of the quadripartite version of the Liar antinomy: one can begin with any statement, say q_3, assume it is true, conclude that then q_4 is true, then q_1 is false, then q_2 is false, and then q_3 is false; and if one assumes that q_3 is false, then by the analogous chain of assignments one arrives to q_3 being true. There is no consistent assignment of truth values in this system. In the language of CbD, the truth values of the statements in $\mathcal{C}_{4(d)}$ can only be described by an inconsistently connected deterministic system.

We come to the main issue now: $\mathcal{C}_{4(d)}$ is certainly a deterministic system (because truth values of statements within a context are fixed), but which deterministic system is it? There are 16 possible ways of filling this system with truth values:

$+1$	$+1$		
	$+1$	$+1$	
		$+1$	$+1$
-1			$+1$

,

$+1$	$+1$		
	$+1$	$+1$	
		-1	-1
$+1$			-1

,

$+1$	$+1$		
	-1	-1	
		$+1$	$+1$
-1			$+1$

,

$+1$	$+1$		
	$+1$	$+1$	
		-1	-1
$+1$			-1

,

etc.

The only constraint in generating these systems is that

1. in the first three contexts (rows) the truth values of the two variables coincide (because the first statement in them says that the second one is true, and the second one does not refer to the first one);

2. in context c^4 (the last row) the truth values of the two variables are opposite (because q_4 says that q_1 is false, and q_1 does not refer to q_4).

We see that although random variability in $\mathcal{C}_{4(d)}$ is absent, we have in its place *epistemic uncertainty.* This opens the possibility of attaching epistemic (Bayesian) probabilities to the 16 possible deterministic variants of $\mathcal{C}_{4(d)}$, and obtaining as a result a *system of epistemic random variables.* Mathematically, such a variable is treated in precisely the same way as an ordinary ("frequentist") random variable. For instance, we can say that an epistemic variable R can have values $+1$ and -1 with Bayesian probabilities p and $1-p$. This means that R in fact is a deterministic quantity that can be either $+1$ or -1, and the degree of rational belief that R is $+1$ (given what we know of it) is p. In all computational respects, however, R is treated as if it was a variable that sometimes can be $+1$ and sometimes -1.

If we choose equal weights for all 16 deterministic variants of $\mathcal{C}_{4(d)}$ (simply because we have no rational grounds for preferring some of them to others), the resulting system will have the following Bayesian distributions:

$$\begin{array}{c|c|c|c}
\text{context } c^i, \; i=1,2,3 & R^i_{i+1}=+1 & R^i_{i+1}=-1 & \\
\hline
R^i_i=+1 & 1/2 & 0 & 1/2 \\
\hline
R^i_i=-1 & 0 & 1/2 & 1/2 \\
\hline
 & 1/2 & 1/2 &
\end{array} \tag{18}$$

and

$$\begin{array}{c|c|c|c}
\text{context } c^4 & R^4_1=+1 & R^4_1=-1 & \\
\hline
R^4_4=+1 & 0 & 1/2 & 1/2 \\
\hline
R^4_4=-1 & 1/2 & 0 & 1/2 \\
\hline
 & 1/2 & 1/2 &
\end{array}. \tag{19}$$

This system is clearly contextual. Indeed, since it is consistently connected,

$$\delta\left(R^1_1, R^4_1\right) = \delta\left(R^1_2, R^2_2\right) = \delta\left(R^2_3, R^3_3\right) = \delta\left(R^3_4, R^4_4\right) = 0. \tag{20}$$

At the same time,

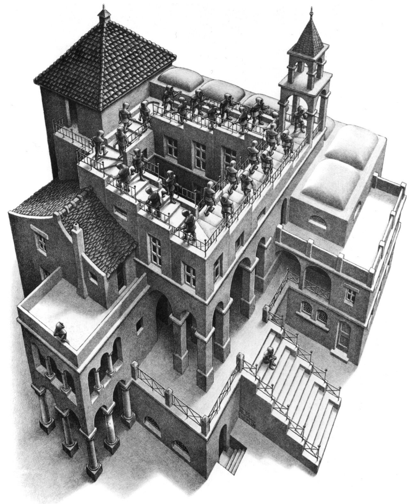
$$\begin{gathered}
\delta\left(\left(R^1_1, R^4_1\right), \left(R^1_2, R^2_2\right), \left(R^2_3, R^3_3\right), \left(R^3_4, R^4_4\right)\right) \\
= \min_{\substack{\text{all couplings} \\ \left(S^1_1, S^1_4, S^2_1, S^2_2, S^3_2, S^3_3, S^4_3, S^4_4\right) \\ \text{of system } \mathcal{C}_{4(d)}}} \left(\Pr\left[S^1_1 \neq S^4_1\right] + \Pr\left[S^1_2 \neq S^2_2\right] + \Pr\left[S^2_3 \neq S^3_3\right] + \Pr\left[S^3_4 \neq S^4_4\right]\right) > 0.
\end{gathered} \tag{21}$$

This is easy to see. This quantity could be zero only if, in some coupling of $\mathcal{C}_{4(a)}$, the equalities in the first row below all held with probability 1:

$$\begin{tikzcd}
S^4_1 = S^1_1 \arrow[d] & S^1_2 = S^2_2 \arrow[d] & S^2_3 = S^3_3 \arrow[d] & S^3_4 = S^4_4 \arrow[d] \\
S^1_1 = S^1_2 \arrow[ur] & S^2_2 = S^2_3 \arrow[ur] & S^3_3 = S^3_4 \arrow[ur] & S^4_4 \neq S^4_1
\end{tikzcd}.$$

But in any coupling of $\mathcal{C}_{4(a)}$, the equalities in the second row also hold with probability 1, because they copy (18) and (19). Reading now all the equalities above from left to right along the arrows as a chain

$$S^4_1 = S^1_1 = S^1_2 = S^2_2 = \ldots,$$



R_1^1	R_2^1			c^1
	R_2^2	R_3^2		c^2
		R_3^3	R_4^3	c^3
R_1^4			R_4^4	c^4
R_1^5	R_1^5	R_1^5	R_1^5	c^5
q_1	q_2	q_3	q_4	$\mathcal{E}$

Figure 3: "Ascending-Descending" by M. C. Escher. The four flights of stairs are enumerated q_1, q_2, q_3, q_4. The epistemic random variables have values *ascending* and *descending*, and in each of the the first four contexts they are perfectly correlated. The fifth context is a mixture of the quadruples of values precisely two of which are ascending (so that travelers always end up in the same place from where they started). The resulting epistemic system is contextual [55].

one arrives at a contradiction

$$S_1^4 \neq S_1^4.$$

In essence, this is the same reasoning as that establishing the unremovable contraction in the Liar antinomy. However, this time it merely serves the purpose of establishing that our system is contextual. In fact, the degree of contextuality here, computed as the difference between (21) and the (zero) sum of the deltas in (20), is maximal among all possible systems of the $\mathcal{C}_4$-type.

We could use other multipartite versions of the Liar paradox, with three or five or any number of statements, all leading to the same outcome. A special mention is needed of the bipartite version. In this system it is no longer possible to define the contexts simply by the contents of the variables they include. Instead we once again need to use the order of the contents, this time interpreted as the direction of inference: $q \to q'$ means that we assign truth values to q and infer the corresponding truth values for q'.[13] The resulting system is

R_1^1	R_2^1	$c^1 : q_1 \to q_2$
R_1^2	R_2^2	$c^2 : q_2 \to q_1$
$q_1 =$ "q_2 is true"	q_2 : "q_1 is false"	$\mathcal{C}_{2(d)}$

,

[13]The interpretation of contexts in terms of the direction of inference is the right one also in systems with larger number of statements. It is merely a coincidence that for $n > 2$ in the systems depicting the n-partite Liar paradox the direction of inference in a context is uniquely determined by the pairs of contents involved in this context.

with four possible deterministic variants:

$$\begin{array}{|c|c|}\hline +1 & +1 \\ \hline -1 & +1 \\ \hline \end{array}\,, \quad \begin{array}{|c|c|}\hline +1 & +1 \\ \hline +1 & -1 \\ \hline \end{array}\,,$$

$$\begin{array}{|c|c|}\hline -1 & -1 \\ \hline -1 & +1 \\ \hline \end{array}\,, \quad \begin{array}{|c|c|}\hline +1 & +1 \\ \hline +1 & -1 \\ \hline \end{array}\,.$$

Mixing them with equal epistemic probabilities creates a consistently connected and highly contextual system (maximally contextual among all cyclic systems of rank 2).

Logical paradoxes are not, of course, the only application of contextuality analysis with epistemic random variables. It seems that many "strange" or "paradoxical" situations can be converted into contextual epistemic systems [55, 57]. Among other applications are such objects as the Penroses' "impossible figures" and M. C. Escher pictures (as in Figure 3).

8 The right to ignore (or not to)

I will mention now some aspects of the Contextuality-by-Default theory (CbD) that seem to pose difficulties for understanding. Questions about them are being asked often and in spite of having been repeatedly addressed in published literature.

The most basic aspect of CbD is double indexation of the random variables. The response to a given question q is a random variable R_q^c whose identity is determined not only by q but also by the context c in which q is responded to. This looks innocuous enough, but it puzzles some when a system being analyzed is consistently connected, i.e. when changing c in R_q^c does not change the distribution. And the puzzlement may increase when our knowledge tells us there is no possible way in which different contexts c can differently influence the random variables R_q^c.

Consider again the system $\mathcal{C}_{4(a)}$ in (14), from which we date contextuality studies. I reproduce it here for reader's convenience:

R_1^1	R_2^1			c^1
	R_2^2	R_3^2		c^2
		R_3^3	R_4^3	c^3
R_1^4			R_4^4	c^4
q_1 (Zora's 1)	q_2 (Nico's 2)	q_3 (Zora's 3)	q_4 (Nico's 4)	$\mathcal{C}_{4(a)}$

In this system, Nico's choice between directions 2 and 4 can in no ways affect Zora's measurements of spin along direction 1. Nevertheless, when Nico switches from direction 2 to 4, the random variable describing the outcome of Zora's measurement of spin along direction 1 ceases to be R_1^1 and becomes R_1^4. It looks like Nico has influenced Zora's measurements after all. Isn't it an example of what Albert Einstein famously called a "spooky action at a distance"?

The answer is, it is not. Nico's choices are undetectable by Zora. Whether he chooses direction 2 or direction 4, Zora can see no changes in the statistical properties of what she observes when she measures spins along direction 1. "Action" means information transmitted, and no information is transmitted from Nico to Zora (and vice versa). The fact that in at least one of the pairs

$$\left\{R_1^1, R_1^4\right\}, \left\{R_2^1, R_2^2\right\}, \left\{R_3^3, R_3^4\right\}, \left\{R_4^3, R_4^4\right\}$$

the two random variables cannot be viewed as being the same can be established by neither Zora nor Nico. It can only be established by a Max who receives the choice of directions and outcomes of measurements from both Zora and Nico and computes the joint distributions in contexts c^1, c^2, c^3, c^4.

An important point here is that compared to Max, Zora does not misunderstand or miss anything when she sees no difference between R_1^1 and R_1^4 or between R_3^3 and R_3^4. Her understanding is no less complete or less correct. Zora and Max simply deal with different systems of random variables. In the same way Max's understanding is no less complete or less correct than that of an Alex who, in addition to knowing what Max knows, observes whether solar activity during the measurements is high or low. In Alex's system, each context of system $\mathcal{C}_{4(a)}$ is split into two contexts, e.g., c^1 is replaced with

$R_1^{1,high}$	$R_2^{1,high}$			$c^{1,high}$
$R_1^{1,low}$	$R_2^{1,low}$			$c^{1,low}$
q_1	q_2	q_3	q_4	$\mathcal{C}_{4(a)}/c^1$ only

.

In studying a system of random variable one always can ignore any of the circumstances that do not affect the distributions of the variables.[14] Or one can choose not to ignore such circumstances, to systematically record them and make them part of the contexts. If a circumstance is irrelevant (as it may be in the case of Alex's recording of solar activity), one will find this out by considering couplings of the system. Thus, one may establish that the contextuality analysis of the system does not change if all couplings are constrained by

$$\Pr\left[S_i^{j,high} = S_i^{j,low}\right] = 1,$$

for any R_i^j in the original system $\mathcal{C}_{4(a)}$. This would mean that $R_i^{j,high}$ and $R_i^{j,low}$ can be viewed as being one and the same random variable (assuming, of course, that solar activity is indeed irrelevant).

This reasoning fully applies to the issue often raised by those who enjoy shallow paradoxes. If one records values of a random variable R in, say, chronological order, and simultaneously records the ordinal positions of these values in the sequence (as part of their contexts),

r_1	r_2	$\dots$	r_n	$\dots$
1	2	$\dots$	n	$\dots$

,

would not this transform all these realizations of a single random variable into pairwise stochastically unrelated random variables

$$R^1, R^2, \dots, R^n, \dots$$

with a single realization each? The answer is yes, if one so wishes (one may also choose to ignore the ordinal positions of the observations altogether), but then a standard view is immediately restored when one considers couplings of these random variables. For instance, the *iid coupling* (corresponding to the standard statistical concept of *independent identically distributed variables*)

[14] This statement can even be extended to ignoring circumstances when distributions do change (inconsistent connectedness). However, this issue has more complex ramifications, and we will set it aside.

has the structure

$$\begin{array}{cccccc}
 & R^1 & R^2 & \ldots & R^n & \ldots \\
S^1 & \boxed{r_1^1 = r_1} & r_2^1 & \ldots & r_n^1 & \ldots \\
S^2 & r_1^2 & \boxed{r_2^2 = r_2} & \ldots & r_n^2 & \ldots, \\
\vdots & \vdots & \vdots & \ddots & \vdots & \ddots \\
S^n & r_1^n & r_2^n & \ldots & \boxed{r_n^n = r_n} & \ldots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \ddots
\end{array}$$

where the boxed values are those factually observed, whereas all other values are independently sampled from the distribution of R. More details are available in Refs. [37, 61].

Finally, does the double-indexation in CbD lend any support to the holistic view of the universe, the view that "everything depends on everything else"? The answer is that the opposite is true, CbD supports a radically analytic view. First, as we have established, unless distributions of two given content-sharing variables are found to be different (which is ubiquitous but not universal) one can ignore the difference between their contexts, i.e., disregard all other variables in these contexts. This will redefine the system, but will not be wrong. Second, the difference in the identity of two content-sharing variables in different contexts (whether their distributions are the same or not) involves no change in the colloquial meaning of the word. The notion of a change implies that something that preserves its identity (e.g., a moving body) changes some of its properties (e.g., position in space). However, R_2^1 and R_2^2 (having the same content in different contexts) are simply different random variables, stochastically unrelated because they occur in mutually exclusive contexts. The difference between them is precisely the same as that between R_2^1 and R_1^1 (different contents in the same context). By choosing a different question to ask, one switches to considering another random variable rather than "changes" the previous one. The same happens when one chooses a different context: one simply switches to considering a different random variable. If I see Max and then see Alex, it does not mean that Max has changed into Alex.

The core of these and other problems with understanding CbD, it seems to me, is in the tendency to view random variables as empirical objects. They are not. Random variables are our descriptions of empirical objects. They are part of our knowledge of the world, and the same as any other knowledge, they can appear, disappear, and be revised as soon as we adopt a new point of view or gain new evidence.

References

[1] Bell, J. (1964). On the Einstein-Podolsky-Rosen paradox. Physics 1, 195-200.

[2] Bell, J. (1966). On the problem of hidden variables in quantum mechanics. Review of Modern Physic 38:447-453.

[3] Kochen, S., Specker, E. P. (1967). The problem of hidden variables in quantum mechanics. Journal of Mathematics and Mechanics 17:59–87.

[4] 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.

[5] Clauser, J.F., Horne, M.A. (1974). Experimental consequences of objective local theories. Physical Review D 10:526-535.

[6] Fine, A. (1982). Joint distributions, quantum correlations, and commuting observables. Journal of Mathematical Physics 23:1306-1310.

[7] Cabello, A., Estebaranz, J., & Garcìa-Alcaine, G. (1996). Bell-Kochen-Specker theorem: A proof with 18 vectors. Physics Letters A 212:183-187.

[8] Cabello, A. (2008). Experimentally testable state-independent quantum contextuality. Physical Review Letters 101, 210401.

[9] 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.

[10] Kurzynski, P., Cabello, A., & Kaszlikowski, D. (2014). Fundamental monogamy relation between contextuality and nonlocality. Physical Review Letters 112:100401.

[11] Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), Attention and Performance II. Acta Psychologica 30:276–315.

[12] Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. Journal of Mathematical Psychology, 28, 363–400.

[13] Dzhafarov, E.N. (2003). Selective influence through conditional independence. Psychometrika, 68:7-26.

[14] Dzhafarov, E.N., & Gluhovsky, I. (2006). Notes on selective influence, probabilistic causality, and probabilistic dimensionality. Journal of Mathematical Psychology, 50:390-401.

[15] Kujala, J.V., Dzhafarov, E.N. (2008). Testing for selectivity in the dependence of random variables on external factors. Journal of Mathematical Psychology 52:128–144.

[16] Dzhafarov, E.N., Kujala, J.V. (2010). The joint distribution criterion and the distance tests for selective probabilistic causality. Frontiers in Psychology: Quantitative Psychology and Measurement 1:151 doi: 10.3389/fpsyg.2010.0015.

[17] Dzhafarov, E.N., & Kujala, J.V. (2016). Probability, random variables, and selectivity. In W.Batchelder, H. Colonius, E.N. Dzhafarov, J. Myung (Eds), The New Handbook of Mathematical Psychology (pp. 85-150). Cambridge University Press.

[18] Zhang, R. Dzhafarov, E.N. (2015). Noncontextuality with marginal selectivity in reconstructing mental architectures. Frontiers in Psychology: Cognition 1:12 doi: 10.3389/fpsyg.2015.00735.

[19] Dzhafarov, E.N. & Kujala, J.V. (2012). Selectivity in probabilistic causality: Where psychology runs into quantum physics. Journal of Mathematical Psychology 56, 54-63.

[20] Dzhafarov, E.N., Kujala, J.V. (2012). Quantum entanglement and the issue of selective influences in psychology: An overview. Lecture Notes in Computer Science 7620:184-195.

[21] Dzhafarov, E.N., Kujala, J.V. (2013). All-possible-couplings approach to measuring probabilistic context. PLoS ONE 8(5):e61712. doi:10.1371/journal.pone.0061712.

[22] Dzhafarov, E.N., Kujala, J.V. (2013). Order-distance and other metric-like functions on jointly distributed random variables. Proceedings of the American Mathematical Society 141:3291-3301.

[23] Dzhafarov, E.N., Kujala, J.V. (2014). Selective influences, marginal selectivity, and Bell/CHSH inequalities. Topics in Cognitive Science, 6, 121–128.

[24] Dzhafarov, E.N., Kujala, J.V. (2014). A qualified Kolmogorovian account of probabilistic contextuality. Lecture Notes in Computer Science 8369:201-212.

[25] Dzhafarov, E.N., Kujala, J.V. (2014). Contextuality is about identity of random variables. Physica Scripta T163:014009.

[26] Dzhafarov, E.N., Kujala, J.V. (2015). Conversations on contextuality. In E.N. Dzhafarov, S. Jordan, R. Zhang, V. Cervantes (Eds). Contextuality from Quantum Physics to Psychology (pp. 1-22). New Jersey: World Scientific.

[27] Kujala, J.V., & Dzhafarov, E.N. (2015). Probabilistic Contextuality in EPR/Bohm-type systems with signaling allowed. In E.N. Dzhafarov, S. Jordan, R. Zhang, V. Cervantes (Eds). Contextuality from Quantum Physics to Psychology (pp. 287-308). New Jersey: World Scientific.

[28] Bacciagaluppi, G. (2015). Einsten, Bohm, and Leggett-Garg. In E.N. Dzhafarov, S. Jordan, R. Zhang, V. Cervantes (Eds). Contextuality from Quantum Physics to Psychology (pp. 63-76). New Jersey: World Scientific.

[29] Dzhafarov, E.N., & Kujala, J.V., & Larsson, J.-Å. (2015). Contextuality in three types of quantum-mechanical systems. Foundations of Physics 7, 762-782.

[30] Dzhafarov, E.N., Zhang, R., Kujala, J.V. (2015). Is there contextuality in behavioral and social systems? Philosophical Transactions of the Royal Society A 374:20150099.

[31] Kujala, J.V., Dzhafarov, E.N., Larsson, J-Å (2015). Necessary and sufficient conditions for extended noncontextuality in a broad class of quantum mechanical systems. Physical Review Letters 115:150401.

[32] Dzhafarov, E.N., Kujala, J.V., Cervantes, V.H., Zhang, R., Jones, M. (2016). On contextuality in behavioral data. Philosophical Transactions of the Royal Society A 374:20150234.

[33] Dzhafarov, E.N., Kujala, J.V., & Cervantes, V.H. (2016). Contextuality-by-Default: A brief overview of ideas, concepts, and terminology. Lecture Notes in Computer Science 9535:12-23.

[34] 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.

[35] de Barros, J.A., Dzhafarov, E.N., Kujala, J.V., & Oas, G. (2016). Measuring observable quantum contextuality. Lecture Notes in Computer Science 9535, 36-47.

[36] 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.

[37] Dzhafarov, E.N. (2016). Stochastic unrelatedness, couplings, and contextuality. Journal of Mathematical Psychology 75C:34-41.

[38] Dzhafarov, E.N. and Kujala, J.V. (2017). Probabilistic foundations of contextuality. Fortschritte der Physik 65:1-11.

[39] Dzhafarov, E.N. and Kujala, J.V. (2017). Contextuality-by-Default 2.0: Systems with binary random variables. Lecture Notes Computer Sciences 10106:16-32.

[40] Cervantes, V.H., Dzhafarov, E.N. (2017). Exploration of contextuality in a psychophysical double-detection experiment. Lecture Notes in Computer Science 10106:182-193.

[41] Zhang, R., Dzhafarov, E.N. (2016). Testing contextuality in cyclic psychophysical systems of high ranks. Lecture Notes in Computer Science 10106:151-162.

[42] 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.

[43] Cervantes, V.H., Dzhafarov, E.N. (2017). Advanced analysis of quantum contextuality in a psychophysical double-detection experiment. Journal of Mathematical Psychology 79:77-84.

[44] Dzhafarov, E.N. (2017). 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.

[45] 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.

[46] Dzhafarov, E.N., Kujala, J.V. (2018). Contextuality analysis of the double slit experiment (with a glimpse into three slits). Entropy 20:278.

[47] 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.

[48] Cervantes, V.H., Dzhafarov, E.N. (2019). True contextuality in a psychophysical experiment. Journal of Mathematical Psychology 91, 119-127.

[49] Dzhafarov, E.N. (2019). On joint distributions, counterfactual values, and hidden variables in understanding contextuality. Philosophical Transactions of the Royal Society A 377:20190144.

[50] Kujala, J.V., & Dzhafarov, E.N. (2019). Measures of contextuality and noncontextuality. Philosophical Transactions of the Royal Society A 377:20190149. (available as arXiv:1903.07170.)

[51] Jones, M. (2019). Relating causal and probabilistic approaches to contextuality. Philosophical Transactions of the Royal Society A, 377, 20190133.

[52] Dzhafarov, E.N., & Kujala, J.V. (2020). Systems of random variables and the Free Will Theorem. Physical Review Research 2:043288; doi: 10.1103/PhysRevResearch.2.043288.

[53] 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.

[54] Dzhafarov, E.N., Kujala, J.V., & Cervantes, V.H. (2020). Erratum: Contextuality and noncontextuality measures and generalized Bell inequalities for cyclic systems [Phys. Rev. A 101, 042119 (2020)]. Physical Review A 101:069902.

[55] Cervantes, V.H., & Dzhafarov, E.N. (2020). Contextuality analysis of impossible figures. Entropy 22, 981; doi: 10.3390/e22090981.

[56] Dzhafarov, E.N., Kujala, J.V., & Cervantes, V.H. (in press). Epistemic odds of contextuality in cyclic systems. To be published in European Physics Journal - Special Topics. (available as arXiv:2002.07755.)

[57] Dzhafarov, E.N. (in press). 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, in series Outstanding Contributions to Logic. Springer Nature. (available as arXiv:1906.02718.)

[58] Moore, D.W. (2002). Measuring new types of question-order effects. Public Opinion Quarterly 66:80-91.

[59] Wang, Z., Busemeyer, J.R. (2013). A quantum question order model supported by empirical tests of an a priori and precise prediction. Topics in Cognitive Science 5:689–710.

[60] Bohm, D, Aharonov, Y. (1957). Discussion of experimental proof for the paradox of Einstein, Rosen and Podolski. Physical Review 108:1070-1076.

[61] Dzhafarov, E.N., Kon, M. (2018). On universality of classical probability with contextually labeled random variables. Journal of Mathematical Psychology 85:17-24