True Contextuality Beats Direct Influences in Human Decision Making

Irina Basieva[1], Víctor H. Cervantes[2], Ehtibar N. Dzhafarov[3], Andrei Khrennikov[4]

[1]City University of London, UK (irina.basieva@gmail.com)

[2]Purdue University, USA (cervantv@purdue.edu)

[3]Purdue University, USA (ehtibar@purdue.edu)

[4]Linnaeus University, Sweden (andrei.khrennikov@lnu.se)

Abstract

In quantum physics there are well-known situations when measurements of the same property in different contexts (under different conditions) have the same probability distribution, but cannot be represented by one and the same random variable. Such systems of random variables are called contextual. More generally, true contextuality is observed when different contexts force measurements of the same property (in psychology, responses to the same question) to be more dissimilar random variables than warranted by the difference of their distributions. The difference in distributions is itself a form of context-dependence, but of another nature: it is attributable to direct causal influences exerted by contexts upon the random variables. The Contextuality-by-Default (CbD) theory allows one to separate true contextuality from direct influences in the overall context-dependence. The CbD analysis of numerous previous attempts to demonstrate contextuality in human judgments shows that all context-dependence in them can be accounted for by direct influences, with no true contextuality present. However, contextual systems in human behavior can be found. In this paper we present a series of crowdsourcing experiments that exhibit true contextuality in simple decision making. The design of these experiments is an elaboration of one introduced in the "Snow Queen" experiment (*Decision* 5, 193-204, 2018), where contextuality was for the first time demonstrated unequivocally.

KEYWORDS: concept combinations, context-dependence, contextuality, direct influences.

1   Introduction

A response to a stimulus (say, a question) is generally a *random variable* that can take on different values (say, Yes or No) with certain probabilities. The *identity* of a random variable, in nontechnical terms, is what uniquely distinguishes this random variable from other random variables.[1] The *distribution* of this random variable (probabilities with which it takes on different values) is part of this identity, but clearly not the entire identity: think of a handful of fair coins — a set of distinct random variables with the same distribution. Other stimuli (e.g., other questions posed together or prior to a given one) may *directly influence* the identity of the response to the given stimulus by changing its distribution. In fact, this change in the distribution, mathematically, is how the "directness" of the influence is defined. True *contextuality* is such dependence of the identity of a response to a stimulus on other stimuli that cannot be wholly explained by such direct influences. We will elaborate this definition below.

Contextuality is at the very heart of quantum mechanics (see, e.g., Liang, Spekkens, & Wiseman, 2011), where it can be observed by eliminating (or at least greatly reducing) all direct influences by experimental design. (In quantum physics "response to a stimulus" has to be replaced with "measurement of a property," but this is in essence the same input-output relation.) This paper addresses a question that ever since the 1990's interested researchers in physics, computer science, and psychology, the question of whether true contextuality can be observed outside quantum mechanics, with special interest (largely for philosophical reasons we will not be discussing) in whether it is present in human behavior. Many previous behavioral experiments designed to answer this question (e.g., Aerts, 2014; Aerts, Gabora, & Sozzo, 2013; Asano, Hashimoto, Khrennikov, Ohya, & Tanaka, 2014; Bruza, Kitto, Nelson, & McEvoy, 2009; Bruza, Kitto, Ramm, & Sitbon,

---

[1]In rigorous mathematical terms, a random variable is defined as a (measurable) function mapping a domain probability space into another (measurable) space. Its distribution is just one property of this function, the probability measure it induces on the codomain space.

2015; Bruza, Wang, & Busemeyer, 2015; Cervantes & Dzhafarov, 2017a, 2017b; Dzhafarov & Kujala, 2014b; Dzhafarov, Kujala, Cervantes, Zhang, & Jones, 2016; Dzhafarov, Zhang, & Kujala, 2015; Zhang & Dzhafarov, 2017) have been shown to result in systems of random variables that are noncontextual. This prompted Dzhafarov, Zhang, and Kujala (2015) to consider the possibility that human behavior may never exhibit true contextuality. It turns out, however, that contextual systems in human behavior can be found. In this paper we describe a series of experiments that, added to one previously conducted (Cervantes & Dzhafarov, 2018), demonstrate this unequivocally.

It should be emphasized at the outset that it would be incorrect to think of contextuality as being "surprising" and "strange" while noncontextuality is "trivial" and "expected." In the absence of constraints imposed by a general psychological theory, comparable to quantum mechanics, we have no justification for such judgements. One might argue in fact that it is most surprising that so many experiments in psychology are described by noncontextual systems of random variables. Nor would it be correct to assume that typical psychological models, even very simple ones, can only predict noncontextual systems: thus, in the concluding section of this paper we mention a simple model that, on the contrary, predicts only contextual systems (and has to be dismissed because of this). Contextuality analysis is not a predictive model of behavior, and both contextual and noncontextual systems are compatible with "ordinary" psychological models. In that, as we point out in Section 4, psychology is not different from quantum physics, where (non)contextuality of a system is established based on the laws of quantum physics but is not used to derive or revise them. What contextuality analysis elucidates is the nature and structure of random variables — arguably, the most basic and mandatory construct in the scientific analysis of empirical systems, whether in psychology or elsewhere. In a well-defined and mathematically rigorous sense, in a contextual system random variables form true "wholes" that cannot be reduced to sets of distinct random variables measuring or responding to specific elements of contexts while being also cross-influenced by other elements of contexts. This makes contextuality analysis inherently interesting, but we need much greater knowledge of which behavioral systems are contextual and which are not in order to determine what other properties of behavior these characteristics are related to. We will return to the role and meaning of contextuality after we introduce necessary definitions, theoretical results, and empirical evidence.

1.1 Direct influences and true contextuality

We introduce the basic notions related to contextuality analysis using a simple example — responses to three Yes/No questions asked two at a time. Most of the experiments reported below are of this kind. Let, e.g., the three questions be

q_1:   Do you like chocolate?

q_2:   Are you afraid of pain?

q_3:   Do you see your dentist regularly?

Let a very large group of people be divided into three subgroups: in the first subgroup each respondent is asked questions q_1 and q_2; in the second subgroup each respondent is asked questions q_2 and q_3; and in the third subgroup the questions are q_3 and q_1. We call these pairwise arrangements of questions *contexts*, and we denote them c_1, c_2, c_3, respectively. It does not matter for the example whether the questions are asked in a fixed order, randomized order, or (if in writing) simultaneously. A response to question q_i asked in context c_j is a random variable that we denote R_i^j: some of the people in the subgroup corresponding to context c_j will answer question q_i with Yes, others with No. Assuming the subgroups are so large that statistical issues can be ignored, by counting the numbers of responses we can get a good estimate of the probability distribution for our random variable:

$$R_i^j : \begin{array}{|c|c||c|} \hline Yes & No & \text{response} \\ \hline p_i^j & 1-p_i^j & \text{probability} \\ \hline \end{array}. \tag{1}$$

All in all we have six random variables in play, and they can be arranged in the form of the following *content-context* matrix (Dzhafarov & Kujala, 2016):

$$\begin{array}{|c|c|c||c|} \hline R_1^1 & R_2^1 & & c_1 \\ \hline & R_2^2 & R_3^2 & c_2 \\ \hline R_1^3 & & R_3^3 & c_3 \\ \hline \hline q_1 & q_2 & q_3 & \text{system } \mathcal{R}_3 \\ \hline \end{array}. \tag{2}$$

Now, the distributions of the responses to question q_i should be expected to differ depending on the context in which it is asked. For instance, when q_1 (Do you like chocolate?) is asked in combination with q_2 (Are you afraid of pain?), the probability of $R_1^1 =$"Yes, I like chocolate" may be relatively high, because chocolate is usually liked, and the mentioning

of pain in q_2 may make it sound especially comforting. However, when the same question q_1 is asked in context c_3, in combination with mentioning a dentist, the probability of $R_1^3 =$ "Yes, I like chocolate" may very well be lower. The same reasoning applies to the two other questions: the responses to each of them will generally be distributed differently depending on its context. This type of influence exerted by a context on the responses to questions within this context can be called *direct influence*. Indeed, the dependence of R_1^1 (responding to q_1) on q_2 (another question in the same context) is essentially of the same nature as the dependence of R_1^1 on q_1: a response to q_1 is based on the information contained in q_1 *and* (even if to a lesser extent) on the information contained in q_2. The other question in the same context can be viewed as part of the question to which a response is given.

Is all context-dependence of this direct influence variety? As it turns out, the answer is negative. Imagine, e.g., that all direct influences are eliminated by some procedural trick, and each question in each context is answered Yes with probability 1/2. This means, in particular, that R_1^1 and R_1^3 have one and the same distribution,

$$R_1^1 : \begin{array}{|c|c|} \hline Yes & No \\ \hline 1/2 & 1/2 \\ \hline \end{array}, \quad R_1^3 : \begin{array}{|c|c|} \hline Yes & No \\ \hline 1/2 & 1/2 \\ \hline \end{array}, \tag{3}$$

and if one does not take into account their relations to R_2^1 (in context c_1) and to R_3^3 (in context c_3), one could consider R_1^1 and R_1^3 as if they were always equal to each other — essentially one and the same random variable.[2] And similarly for R_2^1 and R_2^2, and for R_3^2 and R_3^3. If one looks at each column of matrix (2) separately, ignoring the row-wise joint distributions, then one can write

$$\begin{array}{l} R_1^1 = R_1^3 \\ R_2^1 = R_2^2 \\ R_3^2 = R_3^3 \end{array}. \tag{4}$$

Consider, however, the possibility that no respondent ever gives the same answer to both questions posed to her. Thus, if she answers Yes to q_1 in context c_1 (which can happen with probability 1/2), she always answers No to q_2, and vice versa. Denoting Yes and No by $+1$ and -1, respectively, we have a chain of equalities

$$\begin{array}{l} R_1^1 = -R_2^1 \\ R_2^2 = -R_3^2 \\ R_1^3 = -R_3^3 \end{array}, \tag{5}$$

and it is clear that (4) and (5) cannot be satisfied together: combining them would lead to a numerical contradiction. We should conclude therefore that when the joint distributions within contexts are taken into account, R_1^1 and R_1^3, or R_2^1 and R_2^2, or R_3^2 and R_3^3 cannot be considered always equal to each other. In at least one of these pairs, the two random variables should be more different than it is warranted by their individual distributions (which are, in this example, identical). This is a situation in which we can say that the system exhibits *true contextuality*, the kind of context-dependence that is not reducible to direct influences (in this example, absent).

Empirical data, especially outside quantum physics, almost always involve some direct influences, but the logic of finding out whether they also involve true contextuality remains the same. Continuing to use matrix (2) as a demonstration tool, we first look at the columns of the matrix one by one, ignoring the contexts. For each pair of random variables in a column (responses to the same question), we find out how close to each other they could be made if they were jointly distributed. In other words, we find the maximal probabilities with which each of the equalities in (4) can be satisfied. Then we investigate whether all the variables in our system can be made jointly distributed while preserving these maximal probabilities. If the answer is negative, we conclude that the contexts force the random variables sharing a column to be more dissimilar than warranted by direct influences (differences in their individual distributions). We then call such a system *contextual*. Otherwise it is *noncontextual*. This is the gist of the approach to contextuality called Contextuality-by-Default (CbD), and we illustrate it in the next section by a detailed numerical example.

CbD forms the theoretical basis for the design and analysis of our experiments. For completeness, however, another approach to the notion of contextuality should be mentioned, one treating context-dependent probabilities as a generalization of conditional probabilities defined through Bayes's formula (Khrennikov, 2009). With some additional assumptions these contextual probabilities can be represented by quantum-theoretical formalisms — state vectors in complex Hilbert space and Hermitian operators or their generalizations. Applications of such approach to cognitive psychology can be found in Khrennikov (2010) and Busemeyer and Bruza (2012), among other monographs and papers. CbD, by contrast, is squarely within classical probability theory. Although contextuality in CbD can be called "quantum-like" due to the origins of the concept in quantum physics, CbD uses no quantum formalisms.

---

[2] The "as if" here serves to circumvent the technicalities associated with the fact that, strictly speaking, we are dealing here not with R_1^1 and R_1^3 themselves but with their probabilistic copies (*couplings*) that are jointly distributed. See Dzhafarov and Kujala (2014a, 2017b) for details.

1.2 A numerical example and interpretation

The following numerical example illustrates how CbD works. Let there be just two dichotomous questions, q_1 and q_2, answered in two contexts, c_1 and c_2 (e.g., in two different orders, as in Wang & Busemeyer, 2013). The content-context matrix here is

$$
\begin{array}{|c|c||c|}
\hline
R_1^1 & R_2^1 & c_1 \\
\hline
R_1^2 & R_2^2 & c_2 \\
\hline\hline
q_1 & q_2 & \text{system } \mathcal{R}_2 \\
\hline
\end{array}. \tag{6}
$$

Assume that the joint distributions along the rows of the matrix are as shown:

c_1	$R_2^1 = Yes$	$R_2^1 = No$	
$R_1^1 = Yes$	$1/2$	0	$1/2$
$R_1^1 = No$	0	$1/2$	$1/2$
	$1/2$	$1/2$	

c_2	$R_1^2 = Yes$	$R_1^2 = No$	
$R_2^2 = Yes$	a	$1/2 - a$	$1/2$
$R_2^2 = No$	$3/4 - a$	$a - 1/4$	$1/2$
	$3/4$	$1/4$	

$$\tag{7}$$

where a is some value between $1/4$ and $1/2$. Knowing these distributions means that, for any filling of the matrix (6) with values of the random variables $R_1^1, R_2^1, R_1^2, R_2^2$ (Yes or No, for a total of 16 combinations), we know the row-wise probabilities: e.g.,

$$
\begin{array}{|c|c||c|}
\hline
R_1^1 = Yes & R_2^1 = Yes & p_1(Yes, Yes) = 1/2 \\
\hline
R_1^2 = Yes & R_2^2 = No & p_2(Yes, No) = 3/4 - a \\
\hline
\end{array}. \tag{8}
$$

We see from (7) that R_2^1 and R_2^2 (the responses to question q_2) are distributed identically. Because of this, if they were jointly distributed (see footnote 1), the maximal probability with which they could be equal to each other would be 1:

q_2	$R_2^2 = Yes$	$R_2^2 = No$	
$R_2^1 = Yes$	$1/2$	0	$1/2$
$R_2^1 = No$	0	$1/2$	$1/2$
	$1/2$	$1/2$	

$$\tag{9}$$

The responses to question q_1, however, are distributed differently, and in the imaginary matrix of their joint distribution,

q_1	$R_1^2 = Yes$	$R_1^2 = No$	
$R_1^1 = Yes$	$1/2$	0	$1/2$
$R_1^1 = No$	$1/4$	$1/4$	$1/2$
	$3/4$	$1/4$	

$$\tag{10}$$

the maximal possible probability of $R_1^1 = R_1^2 = Yes$ is $1/2$, and the maximal possible value of $R_1^1 = R_1^2 = No$ is $1/4$. Therefore, if they were jointly distributed, the maximal probability with which $R_1^1 = R_1^2$ would be $3/4$. Now, with these imaginary distributions, for any filling of the matrix (6) with Yes-No values of the random variables $R_1^1, R_2^1, R_1^2, R_2^2$, we also have the column-wise probabilities: e.g.,

$$
\begin{array}{|c|c|}
\hline
R_1^1 = Yes & R_2^1 = Yes \\
\hline
R_1^2 = Yes & R_2^2 = No \\
\hline
p_1'(Yes, Yes) = 1/2 & p_2'(Yes, No) = 0 \\
\hline
\end{array}. \tag{11}
$$

The problem we have to solve now is: are these column-wise probabilities compatible with the row-wise probabilities in (8)? The compatibility means that, to any of the 16 filling of the matrix (6) with values of the random variables $R_1^1, R_2^1, R_1^2, R_2^2$, we can assign a probability, e.g., $p\left(R_1^1 = Yes, R_2^1 = Yes, R_1^2 = Yes, R_2^2 = No\right)$, such that the row-wise sums of these probabilities agree with (8) and the column-wise sums agree with (11). This is a classical linear programming problem: for any given value of a it is guaranteed that either such an assignment of probabilities will be found (so that the system is noncontextual) or the determination will be made that such an assignment does not exist (the system is contextual). In our case, however, one need not resort to linear programing to see that no such assignment of probabilities is possible for any value of a other than $1/2$. Indeed, we see from the c_1-distribution in (7) and from (9) that, with probability 1,

$$
R_1^1 = R_2^1 = R_2^2. \tag{12}
$$

So, R_1^1 and R_2^2 are essentially the same random variable, say X. But, from (10), this X equals R_1^2 with probability $3/4$, whereas from the c_2-distribution in (7), this X equals R_1^2 with probability $2a - 1/4$, which is not $3/4$ if $a \neq 1/2$. The conclusion is that the joint distributions along the two rows of the content-context matrix (6) prevent the responses to

the same questions in the two columns of the matrix to be as close to each other as they can be if the two columns are viewed separately. The system therefore is contextual for any $a \neq 1/2$.

Why is this interesting? In psychological terms, the interpretation of the question order effect seems straightforward: the first question reminds something or draws one's attention to something that is relevant to the second question. What is shown by the contextuality analysis of our hypothetical question-order system is that this interpretation is only sufficient for $a = 1/2$, being incomplete in all other cases. The responses to two questions posed in a particular order form a "whole" that cannot be reduced to an action of the first question upon the second response: the identity of the two random variables changes beyond the effect of this action on their distributions. We will return to this issue in the concluding section of the paper.

The reader should not forget that we are discussing a numerical example rather than experimental data. The large body of experimental data on the question-order effect collected by Wang and Busemeyer (2013) has been subjected to contextual analysis in Dzhafarov, Zhang, and Kujala (2015), the result being that the responses to any of the many pairs of questions studied exhibit no contextuality. In fact, almost all question pairs are in a good agreement with the "QQ law" discovered by Wang and Busemeyer (2013),

$$\Pr\left[R_1^1 = R_2^1\right] = \Pr\left[R_1^2 = R_2^2\right], \tag{13}$$

and, as shown in Dzhafarov, Zhang, and Kujala (2015), this law implies no contextuality: this system of random variables is entirely describable in terms of each response being dependent on "its own" question, plus the second respond being also influenced by the first question. The idea of a "whole" being irreducible to interacting parts is not therefore an automatically applicable formula. To see if it is applicable at all, in psychology, one should look for empirical evidence elsewhere. Such evidence is presented below.

1.3 Contextuality-by-Default

CbD was developed (Dzhafarov, Cervantes, Kujala, 2017; Dzhafarov & Kujala 2014a, 2016, 2017a, 2017b; Kujala, Dzhafarov, & Larsson, 2015) as a generalization of the quantum-mechanical notion of contextuality (Abramsky & Brandenburger, 2011; Fine, 1982; Kochen & Specker, 1967; Kurzynski, Ramanathan, & Kaszlikowski, 2012). The latter only applies to *consistently connected* systems, those in which direct influences are absent, i.e., responses to the same stimulus (or measurements of the same property) in different contexts are distributed identically. In physics this requirement is known by such names as "no-signaling," "no-disturbance," etc.; in psychology it is known as *marginal selectivity* (Dzhafarov, 2003; Townsend & Schweickert, 1989). This requirement is never satisfied in behavioral experiments (Dzhafarov & Kujala, 2014b; Dzhafarov, Kujala, Cervantes, Zhang, & Jones, 2016; Dzhafarov, Zhang, & Kujala, 2015), and it is often violated in quantum physical experiments too (Adenier & Khrennikov, 2017; Kujala, Dzhafarov, & Larsson, 2015). The main difficulty faced by many previous attempts to reveal contextuality in human behavior was that they could not apply mathematical tests predicated on the assumption of consistent connectedness to systems in which this requirement does not hold. As mentioned in the introduction, a CbD-based analysis of these experiments (Dzhafarov & Kujala, 2014b; Dzhafarov, Kujala, Cervantes, Zhang, & Jones, 2016; Dzhafarov, Zhang, & Kujala, 2015) showed that all context-dependence in them was attributable to direct influences. The first unequivocal evidence of the existence of contextual systems in human behavior was provided by Cervantes and Dzhafarov's (2018) "Snow Queen" experiment.

The idea underlying the design of the "Snow Queen" experiment (and all the experiments reported below) is suggested by the criterion (necessary and sufficient condition) of contextuality when CbD is applied to cyclic systems with dichotomous random variables (Dzhafarov, Kujala, & Larsson, 2015; Kujala & Dzhafarov, 2016; Kujala, Dzhafarov, & Larsson, 2015). In such a system n questions and n contexts can be arranged as

$$q_1 \overset{c_1}{\smile} q_2 \overset{c_2}{\smile} \cdots \overset{c_{n-2}}{\smile} q_{n-1} \overset{c_{n-1}}{\smile} q_n \tag{14}$$
$$\underset{c_n}{}$$

The number n is referred to as the rank of the system. The question-order system (6) considered in Section 1.2 is the smallest possible cyclic system, of rank 2,

$$q_1 \overset{c_1}{\underset{c_2}{\smile}} q_2. \tag{15}$$

The system (2) in Section 1.1 is a cyclic system of rank 3,

$$q_1 \overset{c_1}{\smile} q_2 \overset{c_2}{\smile} q_3, \tag{16}$$
$$\underset{c_3}{}$$

5

and it is used in four of the six experiments reported below. The remaining two are analyzed as cyclic systems of rank 4,

$$q_1 \overset{c_1}{=\!=\!=} q_2 \overset{c_2}{=\!=\!=} q_3 \overset{c_{n-1}}{=\!=\!=} q_4 \,, \tag{17}$$

with the content-context matrix

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	q_3	q_4		system $\mathcal{R}_4$

$$\tag{18}$$

To formulate the criterion of contextuality in cyclic systems, we encode the values of our random variables by $+1$ and -1. Then the products of the random variables in the same context, such as $R_1^1 R_2^1$, are well-defined, and so are the expected values $\mathsf{E}\left[R_1^1 R_2^1\right]$, $\mathsf{E}\left[R_2^2 R_3^2\right]$, etc. For instance, if the joint distribution of R_1^1 and R_2^1 (responses to questions q_1 and q_2 in context c_1) is

c_1	$R_2^1 = +1$	$R_2^1 = -1$	
$R_1^1 = +1$	a	b	$a+b$
$R_1^1 = -1$	c	d	$c+d$
	$a+c$	$b+d$	

$$\tag{19}$$

then $R_1^1 R_2^1$ has the distribution

$R_1^1 R_2^1 = +1$	$R_1^1 R_2^1 = -1$
$a+d$	$b+c$

$$\tag{20}$$

and the distribution of R_1^1 and R_2^1 is described by the expected values

$$
\begin{aligned}
\mathsf{E}\left[R_1^1\right] &= (a+b) - (c+d), \\
\mathsf{E}\left[R_2^1\right] &= (a+c) - (b+d), \\
\mathsf{E}\left[R_1^1 R_2^1\right] &= (a+d) - (b+c).
\end{aligned}
\tag{21}
$$

We will also need a special function, s_{odd}: given some real numbers $x_1, \ldots, x_n$,

$$s_{odd}\left(x_1, \ldots, x_n\right) = \max\left(\pm x_1 \pm \ldots \pm x_n\right), \tag{22}$$

where each $\pm$ is to be replaced with $+$ or $-$, and the maximum is taken over all choices that contain an odd number of minus signs. Thus,

$$
\begin{aligned}
s_{odd}\left(x, y\right) &= \max\left(-x + y, x - y\right), \\
s_{odd}\left(x, y, z\right) &= \max\left(-x + y + z, x - y + z, x + y - z, -x - y - z\right), \\
&\quad\text{etc.}
\end{aligned}
\tag{23}
$$

The theorem proved by Kujala and Dzhafarov (2016) says that a cyclic system of rank n is contextual (exhibits true contextuality) if and only if

$$D = s_{odd}\left(\mathsf{E}\left[R_1^1 R_2^1\right], \mathsf{E}\left[R_2^2 R_3^2\right], \ldots, \mathsf{E}\left[R_n^n R_1^n\right]\right) - (n-2) - \Delta > 0, \tag{24}$$

where

$$\Delta = \left|\mathsf{E}\left[R_1^1\right] - \mathsf{E}\left[R_1^n\right]\right| + \left|\mathsf{E}\left[R_2^1\right] - \mathsf{E}\left[R_2^2\right]\right| + \ldots + \left|\mathsf{E}\left[R_n^{n-1}\right] - \mathsf{E}\left[R_n^n\right]\right|. \tag{25}$$

The value of Δ is a measure of direct influences, or of inconsistent connectedness. It shows how much, overall, the distributions of responses to one and the same question differ in different contexts. If $\Delta = 0$, the system is consistently connected: the response to a given question is not influenced by the other questions with which it co-occurs in the same context.[3] One can loosely interpret s_{odd} as a measure of the "potential true contextuality": it shows how much, overall, the identities of the random variables responding to the same question differ in different contexts. The contextuality test for a cyclic system therefore can be viewed as a test of whether these differences exceed those due to direct influences alone. The failure of the previous attempts to find contextuality in behavioral data may be described by saying that the

---

[3]The special case of (24) for $\Delta = 0$ was proved, by very different mathematical means, in Araújo, Quintino, Budroni, Cunha, & Cabello (2013).

empirical situations chosen for investigation had too strong direct influences for the amount of potential true contextuality they contained.

The idea of the "Snow Queen" experiment was to make the value of s_{odd} as large as possible, increasing its chances of "beating" Δ, a quantity that cannot be controlled by experimental design.[4] The formal structure of the experiment was a cyclic system of rank 4, with q_1 and q_3 being two choices of characters from a story (Snow Queen, by H.C. Andersen), and q_2 and q_4 being two choices of attributes of these characters.

$$
\begin{array}{|c|c|c|c||c|}
\hline
R_1^1 & R_2^1 & & & c_1 \\
\hline
& R_2^2 & R_3^2 & & c_2 \\
\hline
& & R_3^3 & R_4^3 & c_3 \\
\hline
R_1^4 & & & R_4^4 & c_4 \\
\hline
\hline
q_1: \begin{array}{c} \text{Gerda} \\ \text{Troll} \end{array} & q_2: \begin{array}{c} \text{beautiful} \\ \text{unattractive} \end{array} & q_3: \begin{array}{c} \text{Snow Queen} \\ \text{old Finn woman} \end{array} & q_4: \begin{array}{c} \text{kind} \\ \text{evil} \end{array} & \text{system } \mathcal{SQ}_4 \\
\hline
\end{array} \tag{26}
$$

For instance, in context c_3, a respondent could choose either Snow Queen or old Finn woman, and also choose either "kind" or "evil." The instruction said the choices had to match the story line. The respondents knew, e.g., that Snow Queen is beautiful and evil, and that the old Finn woman is unattractive and kind.[5] It is easy to show that if all respondents followed the instruction correctly, s_{odd} in this experiment had to have the maximal possible value of 4. The amount of direct influences measured by Δ was considerable, but the left-hand side expression in (24) was well above zero, with very high statistical reliability (evaluated by 99.99% bootstrap confidence intervals).

One possible criticism of the "Snow-Queen" experiment can be that the paired choices were too "asymmetric": choice of a character, such as Gerda, and choice of a characteristic, such as "beautiful," seem too different in nature. In the experiments reported below the paired choices were "on a par." Otherwise, the experiments followed the same logic, ensuring the highest possible value for s_{odd}. This value equals n, the rank of the cyclic system. In quantum physics, the systems with this property (if, additionally, they are consistently connected, i.e., $\Delta = 0$), are called PR-boxes, after Popescu and Rohrlich (1994). In our experiments n was 3 or 4.

2 Method

Participants

We recruited 6192 participants on CrowdFlower (2018) between February 7 and 12, 2018. They agreed to participate in this study by accepting a standard consent from. The consent form and the interactive experimental procedure were provided via a Qualtrics survey hosted by City University London. The study was approved by City University London Research Ethics Committee, PSYETH (S/L) 17/18 09. (The number of participants was chosen so that we could construct reliable 99.99% bootstrap confidence intervals for each context in each experiment, as described below.)

Materials and procedure

Each respondent participated in all six experiments, in a random order. For each of the experiments, each participant was randomly and independently assigned to one of the conditions (contexts). In each context, a participant was introduced to a pair of choices to be made by a fictional Alice; each choice was between two alternatives. There were three contexts in Experiments 1-4, and four contexts in Experiments 5 and 6. Figure 1 shows the way the instruction and choices were presented to respondents in one context of Experiment 1.

Experiments 1 to 4

In experiments 1-4, in each context, the character Alice was faced with two choices out of a set of three dichotomous choices. The participant was asked to select a pair of responses that respected Alice's preferences as stated in the instructions (see Fig. 1). The system would not allow the respondent to make only one choice or two choices contradicting the instructions. The following depicts the situations presented, while table 1 summarizes the sets of dichotomous choices.

---

[4]In physics the situation is different: one can eliminate or greatly reduce direct influences by, e.g., separating two entangled particles by a space-time interval that prevents transmission of a signal between them.

[5]This instruction is an analogue of the quantum-mechanical *preparation*, an empirical procedure preceding an experiment with the aim of creating a specific pattern of high correlations between measurements.

Figure 1: The appearance of the computer screen to the participant if assigned to context c_1 in experiment 1. The participant was required to choose an option for each question, in this case each menu section; the next experiment or the end of the survey would be reached by clicking the 'Next' arrow. If the participant had made both choices in accordance with the instructions, in this case having chosen Soup (H) with Beans (L) or Salad (L) with Burger (H), clicking the 'Next' arrow allowed the survey to continue; otherwise the participants were prompted to revise or complete their responses.

Table 1: Dichotomous choices in experiments 1 to 4. Each respondent was asked to make two choices (q_1&q_2 or q_2&q_3 or q_3&q_1), randomly and independently assigned to this respondent in each experiment.

	q_1	q_2	q_3
1. Meals	Starters: Soup (H)* or Salad (L)	Main course: Burger (H)* or Beans (L)	Dessert: Cake (H)* or Coffee (L)
2. Clothes	Skirt: Plain* or Fancy	Blouse: Plain* or Fancy	Jacket: Plain* or Fancy
3. Presents	Book: Big expensive book (E)* or Smaller book(C)	Soft toy (bear): (E)* or (C)	Construction set: (E)* or (C)
4. Exercises	Arms: Hard* or Easy	Back: Hard* or Easy	Legs: Hard* or Easy

* Denotes the response encoded with +1

Experiment "Meals." Alice wishes to order a two-course meal. For each course she can choose a high-calorie option (indicated by H) or a low-calorie option (indicated by L). Alice does NOT want both courses to be high-calorie nor does she want both of them to be low-calorie.

Experiment "Clothes." Alice is dressing for work, and chooses two pieces of clothing. She does NOT want both of them to be plain, nor does she want both of them to be fancy.

Experiment "Presents." Alice wishes to buy two presents for her nephew's birthday. She can choose either a more expensive option (indicated by E) or a cheaper option (indicated by C). Alice does NOT want both presents to be expensive or both presents to be cheap.

Experiment "Exercises." Alice is doing two physical exercises. Alice does NOT want both exercises to be hard or both to be easy.

Experiments 5 and 6

In experiments 5 and 6, in each context, the character Alice was faced with two choices out of a set of four. In all other respects the procedure was similar to that in Experiments 1-4. The participant was asked to select a pair of responses

Table 2: Dichotomous choices in experiments 5 and 6. Each respondent was asked to make two choices ($q_1\&q_2$ or $q_2\&q_3$ or $q_3\&q_4$ or $q_4\&q_1$), randomly and independently assigned to this respondent in each experiment.

	q_1	q_2	q_3	q_4
5. Directions	West—East fork $\leftarrow$ or $\rightarrow$	NorthWest—SouthEast fork $\nwarrow$ or $\searrow$	North—South fork $\uparrow$ or $\downarrow$	NorthEast—SouthWest fork $\nearrow$ or $\swarrow$
6. Colored figures	one of	one of	one of	one of

For each choice q_i, the response encoded by $+1$ is the one on the left: e.g., for q_1 in Experiment 5, the response $\leftarrow$ was encoded by $+1$.

that respected the character's preferences as stated in the instructions. The following depicts the situations presented, while table 2 summarizes the sets of dichotomous choices.

Experiment "Directions." Alice goes for a walk, and has to choose path directions at forks. Alice wants the two directions to be as similar as possible (i.e., the angle between them to be as small as possible).

Experiment "Colored figures." Alice is taking a drawing lesson, and is presented with two pairs consisting of a square and a circle (the pairs being labeled as "Section 1" and "Section 2"). Alice needs to choose one figure from each section, and she wants the two figures chosen to be of similar color.

3 Results

In Experiments 1-4, irrespective of the specific content of the questions, there were three dichotomous choices, q_1, q_2, q_3, offered to the respondents two at a time. Denoting, for each of the choices, one of the response options $+1$ and the other -1, the results have the following form:

$$
\begin{array}{c|c|c|c}
c_1 & R_2^1 = 1 & R_2^1 = -1 & \\
\hline
R_1^1 = 1 & 0 & p_1 & p_1 \\
R_1^1 = -1 & 1-p_1 & 0 & 1-p_1 \\
\hline
 & 1-p_1 & p_1 &
\end{array}
\qquad
\begin{array}{c|c|c|c}
c_2 & R_3^2 = 1 & R_3^2 = -1 & \\
\hline
R_2^2 = 1 & 0 & p_2 & p_2 \\
R_2^2 = -1 & 1-p_2 & 0 & 1-p_2 \\
\hline
 & 1-p_2 & p_2 &
\end{array}
$$

$$
\begin{array}{c|c|c|c}
c_3 & R_1^3 = 1 & R_1^3 = -1 & \\
\hline
R_3^3 = 1 & 0 & p_3 & p_3 \\
R_3^3 = -1 & 1-p_3 & 0 & 1-p_3 \\
\hline
 & 1-p_3 & p_3 &
\end{array}
\tag{27}
$$

In reference to the CbD criterion (24)-(25), it follows that in these experiments

$$
s_{odd}\left(\mathsf{E}\left[R_1^1 R_2^1\right], \mathsf{E}\left[R_2^2 R_3^2\right], \mathsf{E}\left[R_3^3 R_1^3\right]\right) = s_{odd}\left(-1, -1, -1\right) = 3,
\tag{28}
$$

so that D in (24) is

$$
D = 2 - \Delta,
\tag{29}
$$

where

$$
\begin{aligned}
\Delta &= \left|\mathsf{E}\left[R_1^1\right] - \mathsf{E}\left[R_1^3\right]\right| + \left|\mathsf{E}\left[R_2^2\right] - \mathsf{E}\left[R_2^1\right]\right| + \left|\mathsf{E}\left[R_3^3\right] - \mathsf{E}\left[R_3^2\right]\right| \\
&= 2\left|p_1 + p_3 - 1\right| + 2\left|p_2 + p_1 - 1\right| + 2\left|p_3 + p_2 - 1\right|.
\end{aligned}
\tag{30}
$$

Table 3 presents the observed values of $\hat{p}_1, \hat{p}_2$ and $\hat{p}_3$ for each context of each of Experiments 1-4, and the corresponding numbers of participants from which these probabilities were estimated.

In Experiments 5 and 6 there were four dichotomous choices, q_1, q_2, q_3, q_4, and each respondent was offered two of them, forming one of four possible contexts. Denoting, again, for each of the choices, one of the response options $+1$ and

Table 3: Probability estimates $\hat{p}_1, \hat{p}_2, \hat{p}_3$ that determine the outcomes of Experiments 1-4 in accordance with (27), and the sizes N_1, N_2, N_3 of the samples from which these estimates were computed.

Experiment	c_1		c_2		c_3	
	$\hat{p}_1$	N_1	$\hat{p}_2$	N_2	$\hat{p}_3$	N_3
1. Meals	0.349	2090	0.658	2052	0.653	2050
2. Clothes	0.639	1996	0.566	2086	0.435	2110
3. Presents	0.547	2081	0.387	2052	0.515	2059
4. Exercises	0.590	2058	0.306	2024	0.580	2110

Table 4: Probability estimates $\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4$ that determine the outcomes of Experiments 5 and 6 in accordance with (31), and the sizes N_1, N_2, N_3, N_4 of the samples from which these estimates were computed.

Experiment	c_1		c_2		c_3		c_4	
	$\hat{p}_1$	N_1	$\hat{p}_2$	N_2	$\hat{p}_3$	N_3	$\hat{p}_4$	N_4
5. Directions	0.471	1549	0.706	1504	0.645	1537	0.750	1602
6. Colored figures	0.419	1603	0.819	1589	0.360	1482	0.154	1517*

* One participant assigned to context c_4 was excluded from Experiment 6 because she or he did not complete the responses in accordance with the instructions.

another -1, the results have the following form:

$$
\begin{array}{c|c|c|c}
c_1 & R_2^1 = 1 & R_2^1 = -1 & \\ \hline
R_1^1 = 1 & p_1 & 0 & p_1 \\ \hline
R_1^1 = -1 & 0 & 1-p_1 & 1-p_1 \\ \hline
& p_1 & 1-p_1 &
\end{array}
\qquad
\begin{array}{c|c|c|c}
c_2 & R_3^2 = 1 & R_3^2 = -1 & \\ \hline
R_2^2 = 1 & p_2 & 0 & p_2 \\ \hline
R_2^2 = -1 & 0 & 1-p_2 & 1-p_2 \\ \hline
& p_2 & 1-p_2 &
\end{array}
$$

$$
\begin{array}{c|c|c|c}
c_3 & R_4^3 = 1 & R_4^3 = -1 & \\ \hline
R_3^3 = 1 & p_3 & 0 & p_3 \\ \hline
R_3^3 = -1 & 0 & 1-p_3 & 1-p_3 \\ \hline
& p_3 & 1-p_3 &
\end{array}
\qquad
\begin{array}{c|c|c|c}
c_4 & R_1^4 = 1 & R_1^4 = -1 & \\ \hline
R_4^4 = 1 & 0 & p_4 & p_4 \\ \hline
R_4^4 = -1 & 1-p_4 & 0 & 1-p_4 \\ \hline
& 1-p_4 & p_4 &
\end{array}
\tag{31}
$$

In reference to the CbD criterion (24)-(25), it follows that in these experiments

$$
s_{odd}\left(\mathsf{E}\left[R_1^1 R_2^1\right], \mathsf{E}\left[R_2^2 R_3^2\right], \mathsf{E}\left[R_3^3 R_4^3\right], \mathsf{E}\left[R_4^4 R_1^4\right]\right) = s_{odd}\left(1,1,1,-1\right) = 4, \tag{32}
$$

whence, once again,

$$
D = 2 - \Delta, \tag{33}
$$

where

$$
\begin{aligned}
\Delta &= \left|\mathsf{E}\left[R_1^1\right] - \mathsf{E}\left[R_1^4\right]\right| + \left|\mathsf{E}\left[R_2^1\right] - \mathsf{E}\left[R_2^2\right]\right| + \left|\mathsf{E}\left[R_3^2\right] - \mathsf{E}\left[R_3^3\right]\right| + \left|\mathsf{E}\left[R_4^3\right] - \mathsf{E}\left[R_4^4\right]\right| \\
&= 2\left|p_1 + p_4 - 1\right| + 2\left|p_2 - p_1\right| + 2\left|p_3 - p_2\right| + 2\left|p_4 - p_3\right|.
\end{aligned}
\tag{34}
$$

Table 4 presents the observed values of $\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4$ in Experiment 5 and 6, and the corresponding numbers of participants from which these probabilities were estimated.

Table 5 shows the estimated values of $D = 2 - \Delta$ in all our experiments. We see that contextuality is observed in Experiments 1-4 and 5. Experiment 6, however, shows no contextuality: the negative value in the bottom row indicates that direct influences here are all one needs to account for the results.

We evaluate statistical reliability of these results in two ways. The first way is to compute an upper bound for the standard deviation of $\hat{D}$ and use it to conservatively test the null-hypothesis $D = 0$ (the maximal noncontextual value) against $D > 0$ (contextuality). In Experiment 6 the alternative hypotheses changes to $D < 0$ (noncontextuality), with $D = 0$ in the null hypothesis interpreted as the infimum of contextual values. We begin by observing that each $\hat{p}_i$ has variance $\frac{p(1-p)}{N_i} \leq \frac{1}{4N_i} \leq \frac{1}{4N_{min}}$, where N_{min} is the smallest among N_i for a given experiment, as shown in Tables 3 and

Table 5: Estimated values of $D = 2 - \Delta$ in Experiments 1-4 ($n = 3$) and 5-6 ($n = 4$). Positive (negative) values of D indicate contextuality (resp., noncontextuality).

	1. Meals	2. Clothes	3. Presents	4. Exercises	5. Directions	6. Colored figures
$\hat{D} = 2 - \hat{\Delta}$	1.361	1.440	1.548	1.223	0.758	-0.984

Table 6: Statistical significance of contextuality in Experiment 1-5 and of noncontextuality in Experiment 6.

Experiment:	1. Meals	2. Clothes	3. Presents	4. Exercises	5. Directions	6. Colored
$\hat{D} = 2 - \hat{\Delta}$	1.361	1.440	1.548	1.223	0.758	-0.984
N_{min}	2050	1996	2052	2024	1504	1482
Upper bound for st. dev. of $\hat{D}$	0.094	0.095	0.094	0.095	0.146	0.147
Number of st. dev. from zero	> 14.5	> 15.1	> 16.5	> 12.9	> 5.1	> 6.6
t-distribution p-value	$< 10^{-45}$	$< 10^{-48}$	$< 10^{-57}$	$< 10^{-36}$	$< 10^{-6}$	$< 10^{-10}$
Chebyshev p-value	< 0.005	< 0.005	< 0.004	< 0.006	< 0.038	< 0.023

4. Using the independent coupling of stochastically unrelated $\hat{p}_i$'s, commonly adopted in statistics, each summand in (30) and (34) has a variance bounded by $\frac{2}{N_{min}}$. The different summands are not independent, but the standard deviation of the sum cannot exceed the sum of their standard deviations. This means that $3\sqrt{\frac{2}{N_{min}}}$ for Experiments 1-4 and $4\sqrt{\frac{2}{N_{min}}}$ for Experiments 5-6 are upper bounds for the standard deviation of $\hat{D}$. These values are reported in Table 6. If we assume applicability of the central limit theorem, given the very large sample sizes, the t-distribution-based p-values are essentially zero. If we make no assumptions, the maximally conservative p-values based on Chebyshev's inequality are still below the conventional significance levels.

In our second statistical analysis, we computed bootstrap distributions and constructed the 99.99% bootstrap confidence intervals for D from 500000 independent resamples for each context of each experiment (Davison & Hinkley, 1997). These are presented in Figure 2. As we see, the left endpoints of the confidence intervals for experiments 1-5 are well above zero. For experiment 6, the 99.99% bootstrap confidence interval (Fig. 2) has the right endpoint well below zero, indicating reliable lack of contextuality.

4   Discussion

Our results confirm beyond doubt the presence of true contextuality, separated from direct influences, in simple decision making. Compared to the "Snow Queen" experiment (Cervantes & Dzhafarov, 2018), where the paired choices belonged to different categories (choice of characters, such as "Gerda or Troll," was paired with the choice of characteristics, such as "kind or evil"), in our experiments the paired choices belonged to the same category (e.g., two levels of arm exercises were paired with two levels of leg exercises). The fact that our results are similar to those of the "Snow Queen" experiment shows that this difference is immaterial. What is material is the design that ensures a very large value of s_{odd} in the contextuality criterion (24). In our experiments it was in fact the largest possible value, one equal to the rank of the cyclic system, n. This value in all but one of our experiments was sufficient to "beat" direct influences, measured by Δ (in the sense that their difference exceeded $n - 2$). The one exception we got, with "Colored figures," is also valuable, as it shows that the presence of true contextuality in our experiment is an empirical finding rather than mathematical consequence of the design: even with s_{odd} maximal in value, direct influences may very well exceed the value of $s_{odd} - (n - 2)$, making the the value of D in (24) negative.

As explained in Cervantes and Dzhafarov (2018), in much greater detail than in the present brief recap, it is important that the design we used was between-subjects, i.e. each respondent in each experiment was assigned to a single context only. The reason for this is that if a single respondent were asked to make pairs of choices in all three contexts (in Experiments 1-4) or in all four contexts (in Experiments 5 and 6), it would have created an empirical joint distribution of all the random variables in the respective systems. This would contravene the logic of CbD, in which different contexts are mutually exclusive, and the random variables in different rows of content-context matrices are stochastically unrelated (have no joint distribution).

One might question another aspect of our experimental design: the fact that the respondents were not allowed to contravene their instructions and make incorrect choices (e.g., choose two "high" options or two "low" options in Experiments
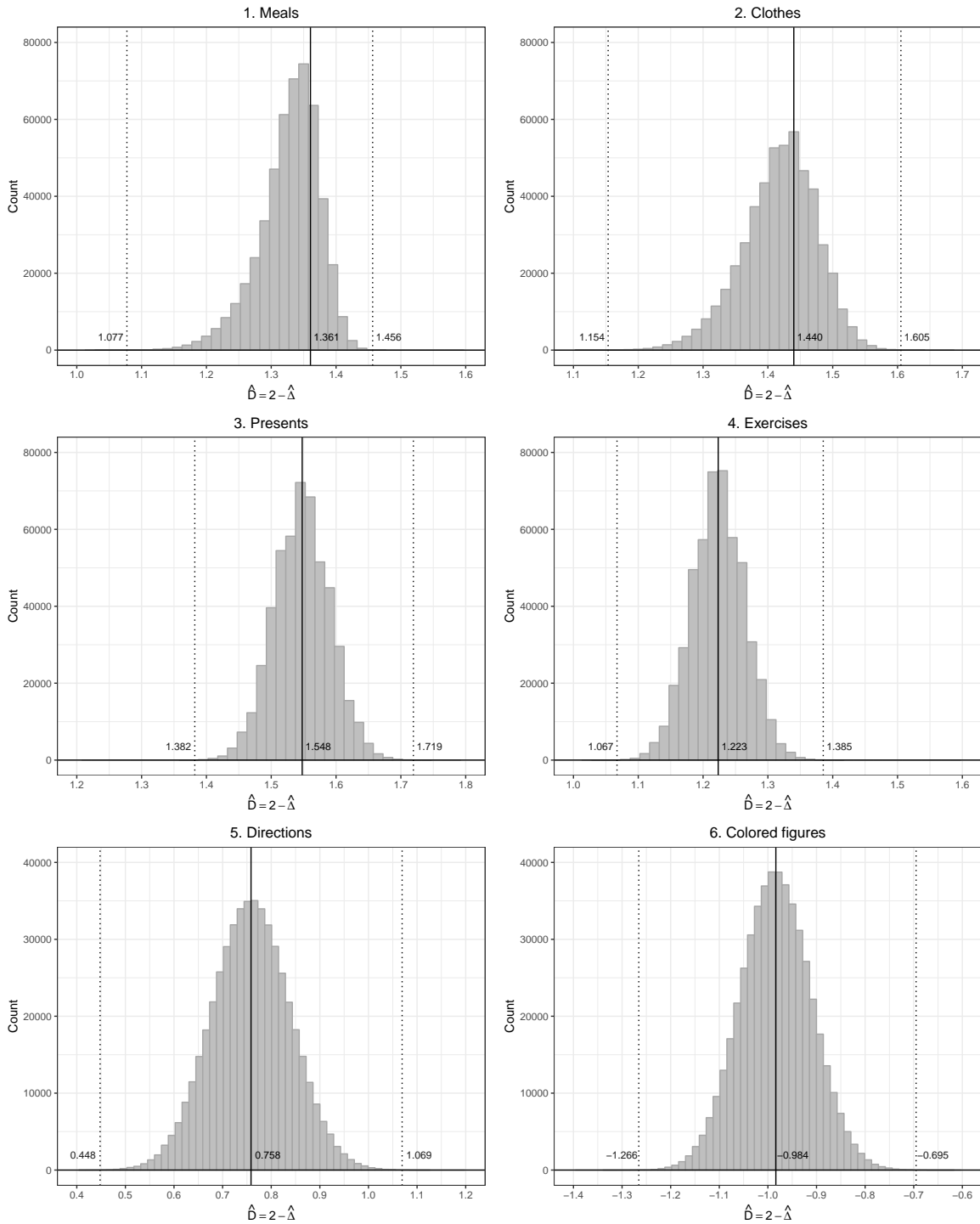
Figure 2: Histograms of the bootstrap values of $\hat{D} = 2 - \hat{\Delta}$ for Experiments 1-6. The solid vertical line indicates the location of the observed sample value. The vertical dotted lines indicate the locations of the 99.99% bootstrap confidence intervals.

1-4). The main reason for this is that in a crowdsourcing experiment, with no additional information about the respondents, it is difficult to understand what could lead a person not to follow the simple instructions. Ideally, one would want to separate data due to deliberate non-compliance or disregard from "honest mistakes," and this is impossible. In fact, it is hard to fathom what an "honest mistake" in a situation as simple as ours might be. In the "Snow Queen" experiment (Cervantes and Dzhafarov, 2018), where the choices were, arguably, less simple than in the present experiments, incorrect responses were allowed, and their percentage was just over 8%. Their inclusion or exclusion did not make any difference for analysis and conclusions.

In the opening of the paper and at the end of Section 1.2 we alluded to the interpretation of true contextuality in terms of the "wholes" irreducible to interacting "parts." One must not mistake this interpretation for the old adage that "the whole is something besides the parts" (Aristotle) or, as reformulated by Kurt Koffka (1935), "the whole is something else than the sum of its parts" (p.176). These and similar statements are not only vague, they have also been rendered essentially meaningless by their indiscriminate application to all kinds of situations. In most of cases one has a justifiable suspicion that what is meant is that parts interact, or that someone can discern a pattern in them. This is probably always true when the parts are deterministic entities. In the case of random variables, however, there is a rigorous analytic meaning of saying that the whole is different from, and indeed greater than a system of parts with all their interactions. Random variables measuring or responding to one and the same "part" (property or stimulus) have different identities in different "wholes" (contexts), with the difference being greater than warranted by the mere distributional differences caused by their interactions with other elements of the "wholes." If this sounds too philosophical to be of importance in scientific practice, we have an example of quantum mechanics to counter this view.

Contextuality in quantum mechanics is not a predictive theory, and it is never used to derive any parts of quantum-mechanical theory. Rather the other way around, quantum-mechanical theory is used to determine a system's behavior, from which it is possible to establish if the system is contextual. Thus, in the most famous example of quantum contextuality, involving spins of entangled particles (Bell, 1964, 1966), the correlations between spins are computed by standard quantum-theoretic formulas, and the results are used to establish that, for certain choices of axes along which the spins are measured, the system is contextual. The computations themselves make no use of contextuality, nor are they being amended in any way as a result of establishing contextuality or lack thereof. Nevertheless, the contextuality analysis of spins of entangled particles (Bell, 1964, 1966; Clauser, Horne, Shimony, & Holt, 1969; Fine 1982), mathematically related to a special case of our contextuality criterion (24), with $n = 4$ and $\Delta = 0$, is considered highly significant. A prominent experimental physicist, Alain Aspect, called it "one of the profound discoveries of the [20th] century" (Aspect, 1999), and teams of experimentalists have put much effort into verifying that the quantum-mechanical predictions used to derive it are correct (Handsteiner et al., 2017). The reason for this is, of course, that contextuality reveals something about one of the most fundamental aspects of quantum theory: the nature of random variables used to describe quantum phenomena. Thus, it is significant that typical systems of random variables describing classical mechanics happen to be noncontextual, while some quantum-mechanical systems are contextual. In time it has also become clear that, in addition to its foundational significance, quantum contextuality correlates with physical properties that can be used for practical purposes. Physicists and computer scientists at present are beginning to pose the question of "contextuality advantage" or "contextuality as a resource," which is the question of whether contextuality or noncontextuality of a system can be utilized for practical purposes. It is argued, e.g., that the degree of contextuality (a notion we have not discussed in this paper, see Dzhafarov, Cervantes & Kujala, 2017; Kujala & Dzhafarov, 2016) is directly related to computational advantage of quantum computing over conventional one (Abramsky, Barbosa, & Mansfield, 2017; Frembs, Roberts, & Bartlett, 2018).

Psychology shares the mandatory use of random variables with quantum physics: stochasticity of responses in most areas of psychology is inherent, it cannot be reduced by progressively greater control of stimuli and conditions. The status and role of contextuality therefore can be expected to be similar. The same as in quantum physics, contextuality analysis is not a predictive model competing with other models. Thus, in constructing a model to fit our data, contextuality analysis can help only in the trivial sense: as with any other property of the data, if contextuality or noncontextuality of them is established, a model is to be rejected if it fails to predict this property. As an example, one could attempt to fit our data by a model with responses being chosen from some "covertly" evoked initial responses actualized with the aid of some conflict resolution scheme. Assume that each question q has a probability h of being "covertly" answered $+1$ (standing here for one of the two options), and that in a context $c = (q, q')$ these covert responses occur independently, so that $(+1, +1)$ occurs with probability hh', $(+1, -1)$ with probability $h(1 - h')$ etc. If the combination of covert responses is allowed by the instructions (e.g., West and North-West in Experiment 5, or Red and Orange in Experiment 6), they turn into observed responses; if the combination is prohibited (say, West and South-East, or Red and Blue), the respondent randomly flips one of the two responses, say, with probability $1/2$. Then the observed probability of choosing an allowed combination $(+1, -1)$ is computed as $h(1 - h') + hh'/2 + (1 - h)(1 - h')/2$. This model can be shown to predict that a system in our experiments is contextual, but it is incompatible with the noncontextuality in Experiment 6. This was only one example, however. Simple models that can predict both contextual and noncontextual outcomes

in our experiments can be readily constructed, because all one has to predict are three probabilities (p_1, p_2, p_3) in (27) for Experiments 1-4, and four probabilities (p_1, p_2, p_3, p_4) in (31) for Experiments 5-6. Consider, e.g., a model with eight triples $(+1, +1, +1), (+1, +1, -1), \ldots, (-1, -1, -1)$, mental states evoked with certain probabilities, with the following decision rule: if the context is (q_i, q_j), $i, j = 1, 2, 3$ and the mental state contains r_i (+1 or -1) and r_j (+1 or -1) in the ith and jth positions, respectively, then respond (r_i, r_j) if this response combination is allowable; if the combination is forbidden, choose one of the allowable combinations with probability $1/2$. The model has 7 free parameters, and it can fit (p_1, p_2, p_3) in Experiments 1-4 precisely. For Experiments 5 and 6, the eight triples have to be replaced with 16 quadruples. We need not get into discussing such models here: it was not a purpose of our experiments to achieve a deeper understanding of how someone chooses to eat soup and beans over burger and salad. Rather our aim was to capitalize on the psychological transparency and modeling simplicity of such choices to firmly establish that "quantum-like" contextuality can be observed outside quantum physics, in human behavior. Recall that many previous attempts to demonstrate behavioral contextuality have failed, so our paper is only one of the first two steps (the other one being the "Snow Queen" experiment in Cervantes & Dzhafarov, 2018) on the path of identifying contextual systems in human behavior.

Thinking by analogy with the "contextuality advantage" mentioned above, can we, at this early stage of exploration, point out any properties of human behavior as correlating with or being indicated by contextuality? One obvious fact is that in our experiments contextuality is negatively related to the value of Δ, the amount of direct influences. Lack of direct influences means that the probability of choosing a particular option, say, burger, is the same irrespective of what context this option is included in (e.g., whether the plain skirt is chosen in the skirt-blouse combination or in the jacket-skirt one). The lack of direct influences would result in the maximal possible value of $D = 2$. This simplicity, however, is specific to our design, in which s_{odd} function does not vary. For a more general class of systems of random variables, one cannot simply replace contextuality with a measure inversely related to the amount of direct influences (we even have examples when the two are synergistic rather than antagonistic). Another dimension of human behavior that can be related to contextuality can be called the degree of "similarity" or "unanimity" of decisions across pools of respondents, or across repeated responses by the same person when a within-subject design is possible (as in Cervantes & Dzhafarov, 2017a, b, and Zhang & Dzhafarov, 2017). Consider, e.g., one of our Experiments 1-4, and assume that the respondents agreed among themselves on what option to choose in each context. The system then would become deterministic and noncontextual, with $D = -4$ or $D = 0$, depending on the pattern of choices agreed upon. Small deviations from an agreed-on pattern would result in small deviations from the corresponding values of D. On the other extreme we have maximal diversity, when in each context the opposite options are chosen with equal probabilities. In this case the system would reach the maximal possible degree of contextuality. Again, it is not possible to simply replace contextuality with some measure of unanimity, such as variance: the maximal value of contextuality can also be achieved without maximal diversity of responses, and "deep noncontextuality," with D between -4 and 0, can be achieved with non-deterministic systems. With due caution, one can conjecture that the degree of (non)contextuality, for a given format of the content-context matrix, may reflect a combination of the two dimensions mentioned: (in)consistency of choices across contexts (reflecting the amount of direct influences) and unanimity/diversity of choices made in each context across a pool of respondents or repeated in a within-subject design (reflecting the amount of determinism/stochasticity). We will not know if this or other relations of contextuality to various aspects of behavior can be established until we broaden our knowledge of the degree of (non)contextuality to a much larger class of behavioral systems.

REFERENCES

1. Abramsky, S., & Brandenburger, A. (2011). The sheaf-theoretic structure of non-locality and contextuality. New Journal of Physics, 13(11), 113036. `https://doi.org/10.1088/1367-2630/13/11/113036`

2. Abramsky, S., Barbosa, R., Mansfield, S. (2017). Contextual fraction as a measure of contextuality. Physical Review Letters 119, 050504. `https://doi.org/10.1103/PhysRevLett.119.050504`

3. Adenier, G., & Khrennikov, A. Y. (2017). Test of the no-signaling principle in the Hensen "loophole-free CHSH experiment." Fortschritte der Physik, 65, 1600096. `https://doi.org/10.1002/prop.201600096`

4. Aerts, D. (2014). Quantum theory and human perception of the macro-world. Frontiers in Psychology, 5, 1–19. `https://doi.org/10.3389/fpsyg.2014.00554`

5. Aerts, D., Gabora, L., & Sozzo, S. (2013). Concepts and their dynamics: A quantum-theoretic modeling of human thought. Topics in Cognitive Science, 5(4), 737–772. `https://doi.org/10.1111/tops.12042`

6. Araújo, M., Quintino, M.T. , Budroni, C. , Cunha, M. T., Cabello, A. (2013). All noncontextuality inequalities for the n-cycle scenario, Physical Review A 88, 022118. `https://doi.org/10.1103/PhysRevA.88.022118`

7. Asano, M., Hashimoto, T., Khrennikov, A. Y., Ohya, M., & Tanaka, Y. (2014). Violation of contextual generalization of the Leggett-Garg inequality for recognition of ambiguous figures. Physica Scripta, T163, 14006. `https://doi.org/10.1088/0031-8949/2014/T163/014006`

8. Aspect, A. (1999). Bell's inequality tests: More ideal than ever. Nature 398, 189-190. `https://doi.org/10.1038/18296`

9. Bell, J. S. (1964). On the Einstein-Podolsky-Rosen paradox. Physics, 1(3), 195–200.

10. Bell, J. S. (1966). On the problem of hidden variables in quantum mechanics. Review of Modern Physic 38, 447-453. `https://doi.org/10.1103/RevModPhys.38.447`

11. Bruza, P. D., Kitto, K., Nelson, D., & McEvoy, C. (2009). Is there something quantum-like about the human mental lexicon? Journal of Mathematical Psychology, 53(5), 362–377. `https://doi.org/10.1016/j.jmp.2009.04.004`

12. Bruza, P. D., Kitto, K., Ramm, B. J., & Sitbon, L. (2015). A probabilistic framework for analysing the compositionality of conceptual combinations. Journal of Mathematical Psychology, 67, 26–38. `https://doi.org/10.1016/j.jmp.2015.06.002`

13. Bruza, P. D., Wang, Z., & Busemeyer, J. R. (2015). Quantum cognition: a new theoretical approach to psychology. Trends in Cognitive Sciences, 19(7), 383–393. `https://doi.org/10.1016/j.tics.2015.05.001`

14. Busemeyer, J.R., & Bruza, P.D. (2012). Quantum Models of Cognition and Decision. Cambridge: Cambridge University Press. `https://doi.org/10.1017/CBO9780511997716`

15. Cervantes, V. H., & Dzhafarov, E. N. (2017a). Exploration of contextuality in a psychophysical double-detection experiment. In J. A. de Barros, B. Coecke, E. Pothos (Eds.), Quantum Interaction. LNCS (Vol. 10106, pp. 182-193). Dordrecht: Springer. . `https://doi.org/10.1007/978-3-319-52289-0_15`

16. Cervantes, V. H., & Dzhafarov, E. N. (2017b). Advanced analysis of quantum contextuality in a psychophysical double-detection experiment. Journal of Mathematical Psychology 79, 77-84. `https://doi.org/10.1016/j.jmp.2017.03.003`

17. 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. `https://doi.org/10.1037/dec0000095`

18. 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. `https://doi.org/10.1103/PhysRevLett.23.880`

19. CrowdFlower (2018). `http://www.crowdflower.com/survey`

20. Davison, A. C., & Hinkley, D. V. (1997). Bootstrap Methods and their Application (1st ed.). New York, NY, USA: Cambridge University Press. `https://doi.org/10.1017/CBO9780511802843`

21. Dzhafarov, E.N. (2003). Selective influence through conditional independence. Psychometrika, 68, 7-26. `https://doi.org/10.1007/BF02296650`

22. 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. `https://doi.org/10.1098/rsta.2016.0389`

23. Dzhafarov, E. N., & Kujala, J. V. (2014a). Contextuality is about identity of random variables. Physica Scripta, 2014(T163), 14009. `https://doi.org/10.1088/0031-8949/2014/T163/014009`

24. Dzhafarov, E. N., & Kujala, J. V. (2014b). On selective influences, marginal selectivity, and Bell/CHSH inequalities. Topics in Cognitive Science, 6(1), 121–128. `https://doi.org/10.1111/tops.12060`

25. 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. `https://doi.org/10.1016/j.jmp.2016.04.010`

26. Dzhafarov, E. N., & Kujala, J. V. (2017a). Contextuality-by-Default 2.0: Systems with Binary Random Variables. In J. A. de Barros, B. Coecke, & E. Pothos (Eds.), Quantum Interaction. LNCS (Vol. 10106, pp. 16–32). Dordrecht: Springer. `https://doi.org/10.1007/978-3-319-52289-0_2`

27. Dzhafarov, E. N., & Kujala, J. V. (2017b). Probabilistic foundations of contextuality. Fortschritte Der Physik, 65, 1600040 `https://doi.org/10.1002/prop.201600040`

28. Dzhafarov, E. N., Kujala, J. V., Cervantes, V. H., Zhang, R., & Jones, M. (2016). On contextuality in behavioural data. Philosophical Transactions of the Royal Society A, 374, 20150234. `https://doi.org/10.1098/rsta.2015.0234`

29. Dzhafarov, E. N., Kujala, J. V., & Larsson, J.-Å. (2015). Contextuality in three types of quantum-mechanical systems. Foundations of Physics, 45(7), 762–782. `https://doi.org/10.1007/s10701-015-9882-9`

30. Dzhafarov, E. N., Zhang, R., & Kujala, J. V. (2015). Is there contextuality in behavioural and social systems? Philosophical Transactions of the Royal Society A, 374, 20150099. `https://doi.org/10.1098/rsta.2015.0099`

31. Fine, A. (1982). Hidden variables, joint probability, and the Bell inequalities. Physical Review Letters, 48(5), 291–295. `https://doi.org/10.1103/PhysRevLett.48.291`

32. Frembs, M., Roberts, S., & Bartlett, S. D. (2018). Contextuality as a resource for measurement-based quantum computation beyond qubits. `https://arxiv.org/abs/1804.07364`

33. Handsteiner, J., Friedman, A. S., Rauch, D., Gallicchio, J., Liu, B., Hosp, H., Kofler, J., Bricher, D., Fink, M., Leung, C., Mark, A., Nguyen, H. T., Sanders, I., Steinlechner, F., Ursin, R., Wengerowsky, S., Guth, A. H., Kaiser, D. I., Scheidl, T., & Zeilinger, A. (2017). Cosmic Bell Test: Measurement Settings from Milky Way Stars. Physical Review Letters, 118, 060401. `https://doi.org/10.1103/PhysRevLett.118.060401`

34. Khrennikov, A. (2009). Contextual approach to quantum formalism. Springer, Berlin-Heidelberg-New York. `https://doi.org/10.1007/978-1-4020-9593-1`

35. Khrennikov, A. (2010). Ubiquitous quantum structure: from psychology to finances. Springer, Berlin-Heidelberg-New York. `https:doi.org/10.1007/978-3-642-05101-2`

36. Kochen, S., & Specker, E. P. (1967). The problem of hidden variables in quantum mechanics. Journal of Mathematics and Mechanics, 17(1), 59–87. `http://doi.org/10.1512/iumj.1968.17.17004`

37. Koffka, K. (1935). Principles of Gestalt psychology. Routledge, London, UK.

38. 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. `https://doi.org/10.1007/s10701-015-9964-8`

39. Kujala, J. V., Dzhafarov, E. N., & Larsson, J-Å. (2015). Necessary and sufficient conditions for an extended noncontextuality in a broad class of quantum mechanical systems. Physical Review Letters, 115(15), 150401. `https://doi.org/10.1103/PhysRevLett.115.150401`

40. Kurzynski, P., Ramanathan, R., & Kaszlikowski, D. (2012). Entropic test of quantum contextuality. Physical Review Letters, 109(2), 020404. `https://doi.org/10.1103/PhysRevLett.109.020404`

41. Liang, Y.-C., Spekkens, R. W., Wiseman, H. M. (2011). Specker's parable of the overprotective seer: A road to contextuality, nonlocality and complementarity. Physics Reports 506, 1-39. `https://doi.org/10.1016/j.physrep.2011.05.001`

42. Popescu, S., & Rohrlich, D. (1994). Quantum nonlocality as an axiom. Foundations of Physics, 24(3), 379–385. `https://doi.org/10.1007/BF02058098`

43. Townsend, J.T., & Schweickert, R. (1989). Toward the trichotomy method of reaction times: Laying the foundation of stochastic mental networks. Journal of Mathematical Psychology, 33, 309-327. `https://doi.org/10.1016/0022-2496(89)90012-6`

44. 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(4), 689–710. `https://doi.org/10.1111/tops.12040`

45. Zhang, R., & Dzhafarov, E. N. (2017). Testing contextuality in cyclic psychophysical systems of high ranks. In J. A. de Barros, B. Coecke, E. Pothos (Eds.) Quantum Interaction. LNCS (Vol. 10106, pp. 151–162). Dordrecht: Springer. `https://doi.org/10.1007/978-3-319-52289-0_12`