

Article

Sensitivity to Context in Human Interactions

Oliver Waddup ¹, Pawel Blasiak ², James M. Yearsley ¹, Bartosz W. Wojciechowski ³
and Emmanuel M. Pothos ^{1,*}

¹ Department of Psychology, City, University of London, London EC1V 0HB, UK; Oliver.Waddup@city.ac.uk (O.W.); James.Yearsley@city.ac.uk (J.M.Y.)

² Institute of Nuclear Physics, Polish Academy of Sciences, 31-342 Kraków, Poland; pawel.blasiak@ifj.edu.pl

³ Institute of Applied Psychology, Jagiellonian University, 31-007 Kraków, Poland; b.wojciechowski@uj.edu.pl

* Correspondence: emmanuel.pothos.1@city.ac.uk

Abstract: Considering two agents responding to two (binary) questions each, we define sensitivity to context as a state of affairs such that responses to a question depend on the other agent's questions, with the implication that it is not possible to represent the corresponding probabilities with a four-way probability distribution. We report two experiments with a variant of a prisoner's dilemma task (but without a Nash equilibrium), which examine the sensitivity of participants to context. The empirical results indicate sensitivity to context and add to the body of evidence that prisoner's dilemma tasks can be constructed so that behavior appears inconsistent with baseline classical probability theory (and the assumption that decisions are described by random variables revealing pre-existing values). We fitted two closely matched models to the results, a classical one and a quantum one, and observed superior fits for the latter. Thus, in this case, sensitivity to context goes hand in hand with (epiphenomenal) entanglement, the key characteristic of the quantum model.

Keywords: bell inequalities; decision making; agent interaction; quantum theory



Citation: Waddup, O.; Blasiak, P.; Yearsley, J.M.; Wojciechowski, B.W.; Pothos, E.M. Sensitivity to Context in Human Interactions. *Mathematics* **2021**, *9*, 2784. <https://doi.org/10.3390/math9212784>

Academic Editor: Ronaldo Vigo

Received: 23 September 2021

Accepted: 29 October 2021

Published: 2 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction and Basic Definitions

Prisoner's dilemma (PD) games involve two players with a binary action each, typically denoted as cooperate (C) vs. defect (D). A usually symmetrical payoff matrix determines the reward of each player, depending on their combined action. Typically, payoffs are set so that it is most advantageous to D, if the other player Cs, but the mutual gain is highest if they both C (defection is then the Nash equilibrium). PD games have been extensively studied in psychology, partly because they can lead to apparent discrepancies with classical probability theory [1–4]. In the pioneering study by [4], participants were put in the shoes of one of the players in a PD game and were presented with three kinds of trials: first, trials for which participants were told the other player would defect; second, trials for which participants were told the other player would cooperate; third, trials for which participants were not given information about the other player. Results indicated that $Prob(D_{Participant} | unknown)$ was outside the bounds of $Prob(D_{Participant} | known C)$ and $Prob(D_{Participant} | known D)$, thus violating the law of total probability. Such results are not insurmountably inconsistent with classical probability theory, but they do challenge the ubiquitousness of classical probability theory in cognitive theory [5–8].

In standard PD paradigms, there is a Nash equilibrium for each participant to D, that is, neither participant can improve her position by unilaterally changing a D action. In this work, we do not consider such PD paradigms, but rather just the two-player interactions, based on a payoff matrix without a Nash equilibrium. We refer to such paradigms as PD variants. The surprising hypothesis we are interested in is whether there are PD variants for which choice statistics cannot be modelled with a four-way probability distribution (this statement will be qualified shortly). So, our paradigm reflects a minimal set up of interaction between two agents. While there is a vast literature on game theory, we avoid

engaging with this literature so as to focus on our specific objective: are there simple situations for the interaction between two agents, as just described, which might confound the straightforward expectation that behavior can be modelled with a four-way probability distribution?

Consider a PD variant, such that each of two players, Alice and Bob, has two binary questions; Alice's questions are $a1, a2$ and Bob's $b1, b2$, all having two possible outcomes ± 1 . A baseline classical expectation is that it is always possible to represent probabilities from such tasks as marginals from a four-way joint probability distribution. More conventionally, we expect that corresponding choice frequencies can be organized in a four-way table. So, our question is, are there PD variants for which participant behavior might be inconsistent with this expectation?

Noting that expectation values are computed as $\sum_{\text{all possible outcomes}} \text{Prob}(\text{outcome}_i) \cdot \text{Value}(\text{outcome}_i)$, for a pair of binary questions, x, y , with $\text{Prob}(\text{outcome}_i)$ being the probability of outcome_i and $\text{Value}(\text{outcome}_i)$ the value assigned to outcome_i , the expectation value is

$$\langle x \& y \rangle = \text{Prob}(++|x, y) \cdot 1 \cdot 1 + \text{Prob}(-+|x, y) \cdot (-1) \cdot 1 + \text{Prob}(- -|x, y) \cdot (-1) \cdot (-1) + \text{Prob}(+-|x, y) \cdot 1 \cdot (-1) \quad (1)$$

Define the quantity

$$S = |\langle a1 \& b1 \rangle + \langle a1 \& b2 \rangle + \langle a2 \& b1 \rangle - \langle a2 \& b2 \rangle| \quad (2)$$

Consider three conditions when computing these expectations. First, locality means that Alice answers her questions without any information about what Bob is doing, and vice versa. Locality means that Alice and Bob are separated in space and no communication between them is possible [9]. Second, free choice means that the question asked to Alice is determined independently from the one asked to Bob. Third, realism means that the outcomes to Alice and Bob's questions exist, whether Alice and Bob state them or not. One of the most significant results in theoretical physics is that, with locality, free choice, and realism, the maximum value of S is 2; this upper limit of S is called Bell's bound [10–12]. Let us take realism for granted, so henceforth we will focus on locality and free choice.

Note, locality and free choice are properties of the two systems producing the relevant statistics. So, in the example with Alice and Bob, locality means that the two agents are local relative to each other—there is no communication—so that Alice has no information about Bob when making her choices and vice versa. Likewise, free choice means that Alice's choices are not influenced by Bob's.

How can Bell's bound be broken? Consider Alice and Bob perfectly tuned to each other, so that $\langle a1 \& b1 \rangle = 1$, $\langle a1 \& b2 \rangle = 1$, and $\langle a2 \& b1 \rangle = 1$. Given this, if locality and free choice apply, then questions $a2, b2$ must correlate as well. This is because if $a1, b1$ perfectly correlate and $a1, b2$ perfectly correlate, then $b1, b2$ must perfectly correlate too. This, together with the fact that $a2, b1$ perfectly correlate with each other, leads to the conclusion that $a2, b2$ must perfectly correlate as well. But, if $\langle a2 \& b2 \rangle = 1$, then the $S = 2$, which is the maximum value that S can take, with realism, locality, and free choice. Therefore, the only way we can break Bell's bound is via some kind of *sensitivity to context*. For example, Bob's answers are sensitive to the context created by Alice's questions.

To explain sensitivity to context, suppose that the $b2$ question depends on whether Alice considers $a1$ or $a2$. If Alice considers $a1$, then Bob responds to $b2$ in a way that the two questions correlate with each other, $\langle a1 \& b2 \rangle = 1$. However, if Alice considers the $a2$ question, then Bob responds to $b2$ in a way that the outcomes of the two questions anticorrelate, $\langle a2 \& b2 \rangle = -1$. That is, there is no answer to the $b2$ question, independently of what Alice does. If we accept the possibility of sensitivity to context, then we can easily see that the Bell bound can be exceeded, in that $S = |\langle a1 \& b1 \rangle + \langle a1 \& b2 \rangle + \langle a2 \& b1 \rangle - \langle a2 \& b2 \rangle| = 1 + 1 + 1 - (-1) = 4$. In this simple situation, sensitivity to context means that the original set of questions $\{a1, a2, b1, b2\}$ is better understood as $\{a1, a2, b1, b2_{a1}, b2_{a2}\}$, where $b2$ has two different versions, depending on which question Alice responds to.

Cases when $S > 2$ reveal a case of correlation ‘stronger’ than classical correlation. For $S > 2$, it is not sufficient for pairs of questions to be responded perfectly in tune with each other (this would be a case of perfect, classical correlation). It is also required that responses are sensitive to the questions asked by the other agent. Thus, cases of $S > 2$ can be said to reflect *supercorrelation* (noting of course that correlation is a binary relation, whereas supercorrelation is a relation between answers amongst two sets of questions). As noted in the physics literature, the kind of correlation producing $S = 4$ is called a PR-box and refers to the strongest type of non-local correlation that is non-signaling, in the two-question, two-outcome scenario [13].

Especially in physics, this discussion is complicated by various inter-related notions, such as signaling, disturbance, and communication. Signaling is a statistical notion informing us of whether the choice of measurement on one side affects the statistics on the other side. The idea is that Alice and Bob have some device generating statistics $Prob(ab|xy)$, where a, b indicate outcomes and $x, y = 1, 2$ are the measurement settings for Alice and Bob respectively. Signaling is if Alice is able to send a meaningful signal to Bob concerning what her setting is, $x = 1, 2$. If signaling occurs, then Bob can infer Alice’s measurement setting by looking at the statistics on his side, i.e., depending on whether his statistics are different for different measurement settings for Alice: $Prob(b|1y) \neq Prob(b|2y)$. Let us note that, if Bob does not know the outcome of Alice’s measurement, then we have to marginalise across different possibilities for this outcome, writing, e.g., when we are interested in $x = 1$, $Prob(b|1y) = \sum_{a=+1,-1} Prob(ab|1y)$. So, the signaling condition is that $Prob(b|1y) = \sum_a Prob(ab|1y) \neq \sum_a Prob(ab|2y) = Prob(b|2y)$, that is, as noted, that Bob can tell whether Alice measures $x = 1$ or $x = 2$, by looking at the statistics on his side (later on, in the Signaling section we offer an equivalent way to compute signaling quantifiers).

When there is no signaling, another seminal result, Fine’s theorem [14], shows that one condition for the existence of a (four-way) joint probability distribution for four binary random variables is $S \leq 2$, which is called the Clauser, Horne, Shimony, and Holt (CHSH) inequality [12]. Note, there are four versions to the inequality, depending on which expectation is given a minus sign in Equation (2) and Fine’s result states that the bound 2 for all those four expressions is the sufficient condition for the existence of a joint probability distribution. When there is signaling, there is a corresponding generalized test of contextuality due to [15]; but see also [9]. Above we referred to sensitivity to context rather than contextuality. We will define sensitivity to context more precisely shortly and offer our rationale for why sensitivity to context is the more appropriate notion for the present work, as opposed to contextuality. Readers should note, however, that there is intense, ongoing debate on these issues.

Presently, what we are interested in is whether there is sensitivity to context, which can be defined as the non-existence of a joint probability distribution—informally, we can say that Alice changes her answer to her question, depending on the question that Bob has. When there is signaling, we can immediately conclude that there is sensitivity to context, regardless of whether $S > 2$ or $S \leq 2$. However, sometimes we may want to test for sensitivity to context without considering signaling. For example, this might be because signaling is low and hence our estimate of signaling is not necessarily reliable (for an example in physics, see [16]). In such cases, when $S > 2$, we can conclude that there is sensitivity to context (this follows from the usual proof of the Bell inequalities based only on the factorization property for conditional probabilities).

Here is the tricky point: Dzhamalov et al. [15] generalized test examines sensitivity to context when there is signaling (their expression can be seen as subtracting away the influence from signaling). But in the present case, the only interest is whether Alice employs the available information of what Bob does to demonstrate sensitivity to context (here and throughout as defined in the paragraph above), regardless of whether this is due to signaling or not. So Dzhamalov et al. [15] generalized test is not relevant here.

These distinctions are particularly relevant in psychology, since the only systems known to break Bell's bound are physical systems of microscopic particles, obeying the laws of quantum mechanics. By contrast, for macroscopic systems, it is generally (see shortly) accepted that violations of Bell's bound can be accounted for only by communication, disturbance or some other equivalent mechanism, between the two systems [9]. For example, demonstrably classical systems, such as containers with fluids at different levels, connected by tubes, allow the construction of variables which violate Bell's bound. But of course there is nothing peculiar going on and this is just a result of communication or influence between the systems (such examples have been known for a while, e.g., [17,18]). We can say that such systems demonstrate sensitivity to context. Note, there are subtleties to this discussion, for example see [17,19], who described possible systems for which a measurement (decision) itself can bring about the dependence to context needed for $S > 2$. An additional subtlety is whether communication is assumed to lead to signaling or not. In [18,19] there is no signaling, but in [17] there is signaling (as [18] note, in general, communication can be taken to be some influence of some sort, but it does not always have to lead to signaling). These ideas are interesting, though we think they do not apply to the present results (this issue is briefly considered in the General Discussion).

2. Psychological Implications and Outline

Bell's bound has an almost magical quality. Sensitivity to context means impossibility of describing the system in the usual way via a four-way probability distribution, with the marginal distributions representing the observed (conditional) statistics. But what exactly does this mean? Consider Table 1, wherein we assume that all marginal probabilities are 0.5. For the right-hand side, $S = 4$ and it can be shown that the corresponding probability information is not self-consistent (the same conjunction can be 'shown' to be both zero and non-zero, Appendix A). We think that, amongst experimental psychologists at least, it is a baseline expectation that probabilities can be organized in a table of this kind.

Table 1. Proportions of +, – responses for a PD variant.

	$b1 = +$	$b1 = -$	$b2 = +$	$b2 = -$		$b1 = +$	$b1 = -$	$b2 = +$	$b2 = -$
$a1 = +$	0.5		0.5		$a1 = +$	0.5		0.5	
$a1 = -$		0.5		0.5	$a1 = -$		0.5		0.5
$a2 = +$	0.5		0.5		$a2 = +$	0.5		0.5	
$a2 = -$		0.5		0.5	$a2 = -$		0.5		0.5

NB. Each table is four separate probability subtables, corresponding to different measurements for the two systems. For the left table, $S = 2$, and for the right, $S = 4 > 2$. It can be shown that the right table is inconsistent (Appendix A). The right table is a famous one, corresponding to the Popescu-Rohlich box (PR-box; [13]).

We are interested in how these ideas translate to two individuals playing a game, corresponding to a Bell scenario (i.e., each individual has two binary questions). Of course, an interaction between two individuals is an extremely common decision situation. With the locality and free choice assumptions, in general it is impossible to break Bell's bound [19–21]. For two agents, the only way Bell's bound can be exceeded is if at least one of the free choice or locality assumptions is violated. For example, suppose we retain free choice and allow violations of locality. Then, Bob needs to adjust his answers depending on knowledge of which question Alice receives. So, the decision to stay local or not is 'outsourced' to Bob—in the experimental paradigm we employ, it is up to the participants (on a trial by trial basis) to decide whether to stay local or not. This is the essence of the paradigm we will shortly present.

So far, while there have been several studies concerning Bell's bound in psychology, these studies have focused on the thought processes of individual participants. Specifically, there have been several examinations of sensitivity to context, for the same participant answering all four questions, $a1$, $a2$, $b1$, $b2$ (for an early example see [22]). The issue of compositionality in conceptual combination concerns whether the constituent concepts

combine in a way that their meaning independently determines the meaning of the composite concept. For example, in considering the novel conceptual combination ‘spring plant’, under a compositionality assumption we would look for some meaning from ‘spring’ and some from ‘plant’, independently combined together. A contrasting hypothesis is that a constituent in a conceptual combination acquires meaning contextually, depending on the other constituent. For example, in the case of boxer-bat, whether we consider a sporting or animal sense for ‘bat’, will impact on the how we interpret ‘boxer’ [23]. A number of theorists have employed the CHSH inequality or variants to conclude in favor of non-compositionality in conceptual combination [23,24], an issue of considerable significance concerning conceptual representation [25–27]. Similar ideas have been pursued in memory associations [28,29] and in decision making [24,30].

There has been no research exploring Bell’s ideas for interacting agents. Our purpose is to develop a paradigm based on a PD variant involving the interaction of a participant with a hypothetical counterpart. The payoff matrices can be set up in a way that optimal performance (relative to overall payoff) requires sensitivity to the counterpart’s choices in some cases, but not others. Allowing participants to choose whether to communicate or not with their counterpart on every trial, we can examine participant’s sensitivity to context and the capacity of different modeling approaches to capture behavior.

We propose two models for modeling choice behavior, based on the models widely employed in physics for Bell paradigms. The classical model (specifically, a local hidden variables one) is based on an assumption of perfect coordination between the interacting agents, but without communication of the questions each agent receives on any trial. It allows for no sensitivity to context. The quantum model is also based on an assumption of perfect coordination between the agents, but, additionally, it allows sensitivity to context up to a certain degree (quantified by Tsirelson’s bound [31]). In physics, such quantum models are interesting, because they allow sensitivity to context, even though there is no obvious physical mechanism violating locality and free choice (and there is no signaling). In psychology, such a quantum model offers a particular hypothesis of the extent to which any communication between participants can translate to sensitivity to context.

Note that we could construct more elaborate classical models, in which the causal role of communication on the observed statistics is included, and such models could (in principle) be reconciled with the sort of paradigm we have outlined above. However, we think it is more interesting to explore a baseline classical model (perfect coordination, but no sensitivity to context) vs. the standard quantum model (perfect coordination and some sensitivity to context), to inform our understanding of the extent to which participants could employ their information resource. We think it is surprising and interesting that, when $S > 2$, as we shall see, a superficially reasonable classical model cannot offer a good description of behavior. Examining violations of Bell’s bound while allowing for interacting participants to break locality mimics attempts in physics to describe experimental statistics in Bell paradigms, by allowing violations of free choice and locality [32].

More generally, the use of quantum probability theory in cognitive modeling follows an assumption that, in some cases, quantum principles offer better descriptions to human behavior [33–35]. Quantum cognitive models have been explored for many kinds of cognitive processes, including decision making, categorization, similarity, perception, and memory. What is common amongst such diverse applications is a handful of characteristics which researchers have taken to be indicative of quantum-like processes. For example, sometimes behavior appears to be subject to interference effects, so that the law of total probability is violated—the PD games and analogous situations in [4] are good examples. In other cases, when participants are asked to make a decision, it appears that the underlying mental state changes. Social psychologists have been aware of such processes for a long time [36]. The added value from quantum models is that in quantum theory there is a specific requirement for how the state ought to change as a result of measurements (in behavior, decisions) and various researchers have taken advantage of these processes to build cognitive models (e.g., [37,38]). Of course, as outlined above, there have also

been behavioral results indicative of sensitivity to context, for which the Bell framework and corresponding quantum models have been invoked to construct relevant theory (e.g., [23,24]). Quantum cognitive models have had good generative value, for example, in terms of anticipating biases from prior decisions [38] or a surprising constraint for question order effects [39].

As per our comments for Bell inequality violations above, in quantum cognitive models any quantum processes are epiphenomenal and are underwritten by an assumption of classical neurophysiology [40]. Moreover, there have been some compelling proposals of heuristic models mimicking quantum models [41]. So, why invoke the (unfamiliar) concepts of quantum theory at all? There are two reasons. First, it appears that in some behavioural cases quantum models can offer particularly simple explanations. Such cases tend to be ones for which behaviour is sensitive to context (as in the present case) or there are conflicting biases for behaviour, which appear to interfere with each other. Second, different quantum models generally employ the same set of principles and so have been used to identify commonalities between findings which, up to that point, had been considered separate [42]. So, even assuming that there is no ‘real’ quantum structure in the brain, and even if there are compelling mimicries between a specific quantum model and models based on other principles (as in [41]), we think there is explanatory value in considering such models.

3. Experiment 1

3.1. Participants

Participants were recruited using Prolific Academic and we restricted sampling to UK nationals. They were paid £2.25 for their involvement. Sample size was set a priori to 100 participants (50 males, 49 females and 1 participant who self-identified as ‘other’). Participants were between 18 and 62 years old ($M_{\text{Age}} = 31.08$ years old, $SD = 11.70$). Participants also reported their English fluency on a scale from 1 (extremely uncomfortable) to 5 (extremely comfortable), with the majority of participants reporting 5 ($n = 97$) and only a few others ($n = 3$) reporting 4 or lower.

3.2. Materials and Procedure

We employed a one-shot, PD variant, such that there were two possible questions for each player. Participants were told to imagine they were arrested with an associate and were both under suspicion for a minor crime in the Old Wild West. The sheriff of the town would interrogate them and their associate separately, asking one question to each. The sheriff would ask either: (1) “Did you know the victim?”, or (2) “Were you at the scene of the crime?” The first question corresponds to $a1$ or $b1$ and the second to $a2$ or $b2$ (the participant and his/her associate questions are denoted by ‘ a ’ and ‘ b ’ respectively). Participants had two possible actions: to confess (equivalent to ‘D’ in the standard PD paradigm; coded with a minus sign) or deny (equivalent to ‘C’; coded with a plus sign). Depending on the combination of questions, a different sentencing policy would apply. Participants were told that their sentencing policy would depend on their question and their response, as well as their associate’s question and response. Participants were expected to favor decisions leading to lower sentences (fewer days spent in prison) for just themselves or for both themselves and their associate [2,43]. We created ‘Good’ and ‘Bad’ payoff matrices, such that the sentencing would bias participants to deny or confess respectively. Note, participants were shown the payoff matrix just for themselves, but were told that their associate would receive the same payoff matrix.

There were eight unique trials which can be denoted as $a1b1$ good, $a1b1$ bad, etc. Each participant received all eight trials and was told to respond independently (e.g., each trial contained a different payoff matrix and each associate had a different name across the trials). Table 2 shows an example of a good and bad matrix in $a2b1$. For $a2b1$ bad, the payoff bias has been created with a bias towards confessing the crime. For $a2b1$ good matrix, this bias is towards denying the crime. Note, the assumption that participants are

responding independently may appear unrealistic. However, it is only marginally relevant to the present purpose, which was to collect data on choice behavior ostensibly inconsistent with a simple classical model.

Table 2. The good and bad matrices for *a2b1* and *a2b2* trials in Experiment 1. The payoff matrices for *a1b1* and *a1b2* trials had a format closely analogous to that for *a2b1*, and variations were employed to create the requisite number of trials.

a2b1 (Bad)			a2b1 (Good)																				
Participant Does Not Check			Participant Does Not Check																				
You did not check on Isabel.			You did not check on Rick.																				
Isabel will be asked whether she knew the victim of the crime or whether she was at the scene of the crime. Since you don't know what Isabel's question will be, the following sentencing policy will apply. Please note, the numbers in the sentencing policies refer to the number of days you will serve in prison.			Rick will be asked whether he knew the victim of the crime or whether he was at the scene of the crime. Since you don't know what Rick's question will be, the following sentencing policy will apply. Please note, the numbers in the sentencing policies refer to the number of days you will serve in prison																				
<table><tr><td>You Deny</td><td>25</td></tr><tr><td>You Confess</td><td>4</td></tr></table>			You Deny	25	You Confess	4	<table><tr><td>You Deny</td><td>3</td></tr><tr><td>You Confess</td><td>20</td></tr></table>			You Deny	3	You Confess	20										
You Deny	25																						
You Confess	4																						
You Deny	3																						
You Confess	20																						
Were you at the scene of the crime?			Were you at the scene of the crime?																				
Participant Checks			Participant Checks																				
You checked on Isabel and found that she will be asked about whether she was at the scene of the crime. So, you know that the following policy for sentencing will apply. Please note, the numbers in the sentencing policies refer to the number of days you will serve in prison.			You checked on Rick and found that he will be asked about whether he knew the victim. So, you know that the following policy for sentencing will apply. Please note, the numbers in the sentencing policies refer to the number of days you will serve in prison																				
<table><tr><td></td><td>Isabel Denies</td><td>Isabel Confesses</td></tr><tr><td>You Deny</td><td>29</td><td>21</td></tr><tr><td>You Confess</td><td>5</td><td>2</td></tr></table>				Isabel Denies	Isabel Confesses	You Deny	29	21	You Confess	5	2	<table><tr><td></td><td>Rick Denies</td><td>Rick Confesses</td></tr><tr><td>You Deny</td><td>3</td><td>2</td></tr><tr><td>You Confess</td><td>20</td><td>19</td></tr></table>				Rick Denies	Rick Confesses	You Deny	3	2	You Confess	20	19
	Isabel Denies	Isabel Confesses																					
You Deny	29	21																					
You Confess	5	2																					
	Rick Denies	Rick Confesses																					
You Deny	3	2																					
You Confess	20	19																					
Were you at the scene of the crime?			Were you at the scene of the crime?																				
a2b2 (Bad)			a2b2 (Good)																				
Participant Does Not Check			Participant Does Not Check																				
Were you at the scene of the crime?			Were you at the scene of the crime?																				
<table><tr><td>You Deny</td><td>52</td></tr><tr><td>You Confess</td><td>52</td></tr></table>			You Deny	52	You Confess	52	<table><tr><td>You Deny</td><td>52</td></tr><tr><td>You Confess</td><td>52</td></tr></table>			You Deny	52	You Confess	52										
You Deny	52																						
You Confess	52																						
You Deny	52																						
You Confess	52																						
Participant Checks			Participant Checks																				
<table><tr><td></td><td>Jack Denies</td><td>Jack Confesses</td></tr><tr><td>You Deny</td><td>101</td><td>2</td></tr><tr><td>You Confess</td><td>2</td><td>101</td></tr></table>				Jack Denies	Jack Confesses	You Deny	101	2	You Confess	2	101	<table><tr><td></td><td>Lucy Denies</td><td>Lucy Confesses</td></tr><tr><td>You Deny</td><td>101</td><td>2</td></tr><tr><td>You Confess</td><td>2</td><td>101</td></tr></table>				Lucy Denies	Lucy Confesses	You Deny	101	2	You Confess	2	101
	Jack Denies	Jack Confesses																					
You Deny	101	2																					
You Confess	2	101																					
	Lucy Denies	Lucy Confesses																					
You Deny	101	2																					
You Confess	2	101																					

The participant's associate was hypothetical and he/she was always assumed to behave as expected, e.g., in the case of trial *a1b1* good, the associate would deny the crime in the *b1* question. How would a participant know what the associate is likely to be doing? In most cases, there would be a choice associated with a lower sentence and so the participant would/ should guess that her associate would be selecting this option. This would be applicable for *a1b1*, *a1b2*, and *a2b1* trials. For *a2b2* trials, the payoffs would not uniquely identify an action as optimal. For these trials, the participant would receive a hint of what the associate is likely to be doing: participants were told that the sheriff

does not know much about the crime, but he does know that exactly one between the participant and his/her associate, was at the scene of the crime. Participants therefore were cautioned that if the participant and his/her associate were to both confess or both deny for these trials, the sheriff would punish them with a high penalty. For example, for the *a2b2* good trial, the sentencing matrix would be biasing towards anticorrelation between the participant and his/her associate, and the participant would receive an additional hint that the associate is ‘likely’ to deny the crime. So, sensitivity to context is built into the structure of the problem, in the simple sense that the participant’s action needs to be informed by the associate’s action when his/her question is *a2*, but not when it is *a1*.

To clarify, given each payoff matrix, there is an ‘obvious’ response for what the hypothetical participant should be doing: we just assume that the hypothetical participant follows this action. The exception is the *a2b2* case, where we offered an additional hint of what the hypothetical participant is doing.

On each trial, participants were allowed to choose whether to communicate (i.e., violate locality) or not. They had the option to try to check, so as to discover the question that their counterpart was going to be asked. We discouraged participants from checking frequently by telling them that a check involved a risk of being caught and automatically receiving a high sentence. The first four trials in the experiment always attracted a penalty if a participant checked on his/her counterpart (these trials were fixed and different from the main experimental trials). Following these first four trials, without a noticeable break in the procedure, participants went through the eight trials corresponding to each of the four combinations of questions in each of their Good/Bad instantiation. The recorded data concerned only these eight trials and participants never experienced the penalty for checking during these trials.

On each trial, a participant was shown a 2×1 matrix for just their payoffs. If he/she decided to check, then he/she would be told which question was assigned to the associate (*b1* or *b2*) and the matrix would expand to show the payoffs for all combination of answers for the participant and associate (Table 2). If the participant did not decide to check, then he/she would just be shown again the initial 2×1 matrix for just their payoffs. Either way, on each trial they had to decide whether to deny or to confess.

In this experiment, for a particular trial (e.g., *a1b2*) the payoffs in the 2×1 matrices were the approximate average of the payoffs in the 2×2 one. For example, looking at the top left of Table 2, $(29 + 21) \div 2 = 25$. Note, averaged decimal payoffs were rounded up to the nearest whole number. But we did not create true averages across different trials. That is, for trial *a1b2*, there would be a 2×1 matrix which would be the average of payoffs in a corresponding 2×2 one. However, the *a1* payoff would not be an average from the *a1b1* and *a1b2* payoff matrices. These considerations are somewhat unimportant (in any case, they are addressed in Experiment 2, what matters is the bias for action, which was to Deny in all good matrices and Confess in the bad ones).

Initial instructions explained the format of the PD game. Participants then responded to a few practice trials, but with detailed additional instructions for each step of a trial. After these trials, participants were told that the main experiment would start. They first received the four consequence-checking rounds, and then the eight PD trials, after which the experiment concluded.

3.3. Results

We observed a significant difference in the overall proportion of trials when participants checked vs. not checked, $\chi^2(1, n = 800) = 121.69, p = < 0.001$ (Tables 3 and 4). Additionally, participants were more likely to check with *a2b2* trials than for other ones. Note, we carried out these comparisons so that Good question combinations were compared only with other Good question combinations and analogously for the Bad ones. Minimally, these results show that participants were sensitive to the context of their associate’s questions, necessary to achieve higher performance.

We further show the choice probabilities to deny for all question combinations and separately for checking vs. non-checking trials (Table 5). Consider the Deny/Good/Checking column. In this case, because the matrices are good, by design the participant's associate is meant to be denying; the participant should also recognize that it is better to deny. As expected, choice proportions reveal high probability for the participant to deny in pairs $a1b1$, $a1b2$, $a2b1$. For the last pair, $a2b2$, however, the participant and his/her associate are biased to anticorrelate and, given the associate will be denying, we observe a low proportion for deny choices, again as expected (0.07). We observe the reverse pattern in the Deny/Bad/Checking column.

Table 3. Frequencies of checking for each question combination.

	Deny (Good)				Confess (Bad)				
	$a1b1$	$a1b2$	$a2b1$	$a2b2$	$a1b1$	$a1b2$	$a2b1$	$a2b2$	
Check	26	13	23	69	25	30	33	63	Exp. 1
No Check	74	87	77	31	75	70	67	37	
Check	27	27	67	64	20	23	63	73	Exp. 2
No Check	74	74	34	37	81	78	38	28	

Table 4. Chi square tests for comparisons of rates of checking for all Good and Bad question combinations.

Good Matrices	Exp. 1	Exp. 2
$a1b1, a1b2$	5.38 *	0
$a1b1, a2b1$	0.24	31.84 ***
$a1b1, a2b2$	37.07 ***	27.38 ***
$a1b2, a2b1$	3.39	31.84 ***
$a1b2, a2b2$	64.82 ***	27.38 ***
$a2b1, a2b2$	42.59 ***	0.2
Bad Matrices		
$a1b1, a1b2$	0.63	0.27
$a1b1, a2b1$	1.55	37.81 ***
$a1b1, a2b2$	29.3 ***	55.97 ***
$a1b2, a2b1$	0.21	32.40 ***
$a1b2, a2b2$	21.89 ***	49.63 ***
$a2b1, a2b2$	18.03 ***	2.25

Note. * $p < 0.05$, *** $p < 0.001$. The n for the tests are 100 for Experiment 1 and 101 for Experiment 2.

Table 5. Observed probabilities for all question combinations in Experiment 1 ($n = 100$ for each cell), split by decision to deny or confess and whether participants checked or not.

		Checking		No Checking	
		Deny	Confess	Deny	Confess
Good	$a1b1$	0.92	0.08	0.95	0.05
	$a1b2$	0.92	0.08	0.98	0.02
	$a2b1$	0.83	0.17	0.97	0.03
	$a2b2$	0.07	0.93	0.74	0.26
Bad	$a1b1$	0.04	0.96	0.15	0.85
	$a1b2$	0.07	0.93	0.11	0.89
	$a2b1$	0.12	0.88	0.1	0.9
	$a2b2$	0.83	0.17	0.65	0.35

Note, when the participant is not checking, he/she ostensibly does not know which question the associate will be asked, and therefore there is no basis for the participant to distinguish between cases when he/she should correlate with the associate ($a1b1$, $a1b2$, $a2b1$) vs. anticorrelate ($a2b2$). If this assumption were entirely correct, we should be observing

identical deny proportions across all four question combinations, when not checking, but this is not the case (e.g., for the bad matrices, 0.65 is higher than the choice proportions for the other question combinations). As noted, the issue is that the reduced payoff matrix when not checking should be identical for $a1b1$ and $a1b2$ (and likewise for $a2b1$ and $a2b2$), but this was not the case (because reduced payoff matrices were constructed separately for each question combination). We address this issue in Experiment 2.

Despite this point, the results of Experiment 1 are still useful for modeling and for exploring the question of whether the particular classical vs. quantum models we will propose are adequate. Relatedly, the empirical result is perhaps unsurprising: participants seek more information when existing information is inadequate for a decision. On one level, this is certainly true, since the task was designed to incorporate sensitivity to context in a particular way. On another level, our objective is less so to offer a surprising empirical finding, but to show that choice probabilities from this seemingly innocuous situation cannot be modeled by a classical model incorporating the (assumed) perfect coordination between the participant and her associate.

4. Experiment 2

Experiment 1 showed that participants recognized that there would be different biases for action depending on whether their associate's question was $b1$ or $b2$. In this experiment, we constructed payoff matrices so that the reduced matrix for e.g., $a1$ would be the collapsed matrix across the $a1b1$ and $a1b2$ possibilities (Table 6).

Whereas previously there were only eight main trials (four question combinations in good and bad versions), for which participants were free to decide whether to check or not check, in this experiment we added eight trials when participants were forced to check and another eight trials in which participants were forced to not check (e.g., on some trials they were told that they had to check on their associate). Recall that to use Equation (1), we need probabilities for e.g., $Prob(++|a1b1)$, which is computed by considering the number of times the participant denies when given question $a1$ together with his/her counterpart denying when given question $b1$. Trivially, $Prob(++ \text{ good } | a1b1 \text{ checking}) = \frac{\text{counts of deny } a1b1 \text{ good checking}}{\text{all counts of } a1b1 \text{ good checking}}$. With this approach, we can compute S values for the entire sample, but it is difficult to do so for individual participants, because e.g., a participant may have not checked in the case of the $a1b1$ good trial. With the additional trials in this experiment, all relevant probabilities can be computed within participants, e.g., $Prob(++ \text{ good } | a1b1 \text{ checking}) = \frac{\text{counts of deny } a1b1 \text{ good checking for the participant}}{\text{all counts of } a1b1 \text{ good checking for the participant}}$, and so S values can be computed within participants (which enables us to conduct some statistical tests). For the example of this probability, $Prob(++ \text{ good } | a1b1 \text{ checking})$, for a particular participant there would be a max of two relevant trials and a min of one trial, depending on whether the participant decided to check when he/she had the option to do so.

We also included three questionnaires. First, we included the Toronto Empathy Questionnaire (TEQ, [44]), since the present task is one of guessing what a (hypothetical) associate is planning to do. The questionnaire asks participants to rate 16 questions on a five-point scale, ranging from Never (1), Rarely (2), Sometimes (3), Often (4), to Always (5). Items include, "Other people's misfortunes do not disturb me a great deal" and "It upsets me to see someone being treated disrespectfully". Second, we include the 17-item Cognitive Uncertainty (CU) subscale from the Uncertainty Response Scale [45]. The CU asks participants to state how well a series of statements describe them, including, "I like to plan ahead in detail rather than leaving things to chance" and "I like to know exactly what I'm going to do next" on a four-point scale of Never (1), Sometimes (2), Often (3) and Always (4). This questionnaire assesses the possibility that checking behavior is driven by uncertainty aversion. Finally, we employed the Cognitive Reflection Task (CRT) to test for engagement and reflection with our PD tasks. However, the original CRT has been massively overused [46,47]. To reduce the likelihood that participants had encountered the original CRT in the past, we used three of the word problems presented in the appendices

of [47]. Participants read each of the questions and were asked to provide an answer in the text box.

Table 6. The good and bad matrices for the *a2b1* and *a2b2* trials in Experiment 2.

<i>a2b1</i> (Bad)	<i>a2b1</i> (Good)																		
Participant Does Not Check	Participant Does Not Check																		
You did not check on Isabel.	You did not check on Rick.																		
Isabel will be asked whether she knew the victim of the crime or whether she was at the scene of the crime. Since you don't know what Isabel's question will be, the following sentencing policy will apply. Please note, the numbers in the sentencing policies refer to the number of days you will serve in prison.	Rick will be asked whether he knew the victim of the crime or whether he was at the scene of the crime. Since you don't know what Rick's question will be, the following sentencing policy will apply. Please note, the numbers in the sentencing policies refer to the number of days you will serve in prison.																		
<table><tr><td>You Deny</td><td>52</td></tr><tr><td>You Confess</td><td>52</td></tr></table>	You Deny	52	You Confess	52	<table><tr><td>You Deny</td><td>51</td></tr><tr><td>You Confess</td><td>51</td></tr></table>	You Deny	51	You Confess	51										
You Deny	52																		
You Confess	52																		
You Deny	51																		
You Confess	51																		
Were you at the scene of the crime?	Were you at the scene of the crime?																		
Participant Checks	Participant Checks																		
You checked on Isabel and found that she will be asked about whether she was at the scene of the crime. So, you know that the following policy for sentencing will apply. Please note, the numbers in the sentencing policies refer to the number of days you will serve in prison.	You checked on Rick and found that he will be asked about whether he knew the victim. So, you know that the following policy for sentencing will apply. Please note, the numbers in the sentencing policies refer to the number of days you will serve in prison.																		
<table><tr><td></td><td>Isabel Denies</td><td>Isabel Confesses</td></tr><tr><td>You Deny</td><td>52</td><td>52</td></tr><tr><td>You Confess</td><td>2</td><td>102</td></tr></table>		Isabel Denies	Isabel Confesses	You Deny	52	52	You Confess	2	102	<table><tr><td></td><td>Rick Denies</td><td>Rick Confesses</td></tr><tr><td>You Deny</td><td>2</td><td>100</td></tr><tr><td>You Confess</td><td>51</td><td>51</td></tr></table>		Rick Denies	Rick Confesses	You Deny	2	100	You Confess	51	51
	Isabel Denies	Isabel Confesses																	
You Deny	52	52																	
You Confess	2	102																	
	Rick Denies	Rick Confesses																	
You Deny	2	100																	
You Confess	51	51																	
Were you at the scene of the crime?	Were you at the scene of the crime?																		
<i>a2b2</i> (Bad)	<i>a2b2</i> (Good)																		
Participant Does Not Check	Participant Does Not Check																		
<table><tr><td>You Deny</td><td>52</td></tr><tr><td>You Confess</td><td>52</td></tr></table>	You Deny	52	You Confess	52	<table><tr><td>You Deny</td><td>51</td></tr><tr><td>You Confess</td><td>51</td></tr></table>	You Deny	51	You Confess	51										
You Deny	52																		
You Confess	52																		
You Deny	51																		
You Confess	51																		
Participant Checks	Participant Checks																		
<table><tr><td></td><td>Jack Denies</td><td>Jack Confesses</td></tr><tr><td>You Deny</td><td>101</td><td>3</td></tr><tr><td>You Confess</td><td>3</td><td>101</td></tr></table>		Jack Denies	Jack Confesses	You Deny	101	3	You Confess	3	101	<table><tr><td></td><td>Lucy Denies</td><td>Lucy Confesses</td></tr><tr><td>You Deny</td><td>100</td><td>2</td></tr><tr><td>You Confess</td><td>2</td><td>100</td></tr></table>		Lucy Denies	Lucy Confesses	You Deny	100	2	You Confess	2	100
	Jack Denies	Jack Confesses																	
You Deny	101	3																	
You Confess	3	101																	
	Lucy Denies	Lucy Confesses																	
You Deny	100	2																	
You Confess	2	100																	

4.1. Participants

Participants were recruited using Prolific Academic and we restricted sampling to UK nationals only. They were paid £4.50 for their involvement. Sample size was set a priori to 100 participants, and we recruited 101 participants, 50 males, 50 females and 1 participant who self-identified as 'other'. Participants were between 18 and 78 years old ($M_{\text{Age}} = 32.13$ years old, $SD = 12.54$). Participants also reported their English fluency on a scale from 1 (extremely uncomfortable) to 5 (extremely comfortable), with the majority of participants reporting 5 ($n = 95$) and only a few others ($n = 6$) reporting 4 or lower. None of the participants for this experiment had taken part in Experiment 1.

4.2. Materials and Procedure

In Experiment 2, the payoff matrices were set up so that if the participant did not check, the reduced 2×1 payoff matrix would be identical across the two possible question combinations, e.g., $a1b1$, $a1b2$ (Table 6). Additionally, there were 24 trials in total: eight choice trials, where the participant can choose or not to check on their counterpart (as in Experiments 1), eight trials for which the participant is forced to check, and eight trials for which the participant is forced to not check.

4.3. Results

As expected, when participants did not check, choice proportions were nearly identical across matched pairs of question combinations (e.g., $a1b1$ good and $a1b2$ good, Table 7). Once again, we were interested in the extent to which participants check on their counterpart when they were meant to, notably in the case of $a2b1$ and $a2b2$ trials. For this experiment, this analysis will only examine the trials when participants could decide whether to check or not. We first confirmed that there was a difference in the overall proportion of trials when participants checked vs. did not check, $\chi^2(1, n = 808) = 148.31$, $p < 0.001$ (Table 3). Moreover, participants were more likely to check with $a2b1$ and $a2b2$ trials than for other ones (Table 4).

Table 7. Observed probabilities for all forced question combinations in Experiment 2 ($n = 101$ for each cell), split by decision to deny or confess.

		Checking		No Checking	
		Deny	Confess	Deny	Confess
Good	$a1b1$	0.94	0.06	0.93	0.07
	$a1b2$	0.95	0.05	0.94	0.06
	$a2b1$	0.66	0.34	0.69	0.31
	$a2b2$	0.10	0.90	0.67	0.33
Bad	$a1b1$	0.05	0.95	0.04	0.96
	$a1b2$	0.05	0.95	0.06	0.94
	$a2b1$	0.39	0.61	0.67	0.33
	$a2b2$	0.78	0.22	0.68	0.32

We next consider the individual differences measures. We computed d' , empathy (TEQ), aversion (CU), engagement/reflection (CRT) scores and S for each participant, using Equation (1) (focused on the trials when participants could choose whether to check or not). The d' coefficient was calculated as $d' = \Phi^{-1}(H) - \Phi^{-1}(F)$, where H and F are hits and false alarms, respectively, and the Φ^{-1} function converts raw scores to z scores by fitting a normal distribution (0, 1 mean and standard deviation) to scores from each participant and then inverting [48]. Hits are considered instances of checking when the participants are meant to be checking (on $a2b1$ and $a2b2$ trials) and false alarms instances of checking when there would be no need for participants to check (on $a1b1$ and $a1b2$ trials). Note, due to the small number of trials per participants, we had a large number of probabilities of 0 or 1, which we corrected by adding 1 to the number of trials and 0.5 to the counts of hits and false alarms ([48], p. 144). Indeed, participants checked more so on $a2b1$ and $a2b2$ trials (hits) than they did in the other trials (false alarms). This is evident from the mean d' ($M = 0.995$, $SD = 1.25$) being above zero. All measures were then correlated with each other, without a multiple comparisons correction, as the intention is exploratory. There are two notable results. First, there was no relationship between individual participant S scores and d' , $r = -0.135$, $p = 0.18$. Second, there was a negative relationship between S and empathy, $r = -0.23$, $p < 0.05$. Higher values of S imply higher sensitivity to context, which in this case means that a participant is better at recognizing when he/she should reverse decisions, based on what his/her counterpart is doing. One possible explanation for this result is that participants higher in empathy try to over-guess their counterpart's

action, at the expense of considering the statistical properties of the game. There were no other significant results.

5. Modeling

It appears that participants are sensitive to the context of their associate's decisions in the PD variants we employed, but does this sensitivity to context push choice statistics beyond the descriptive adequacy of classical models (of a certain kind) and, if yes, in what way? This is the key research question in the present work. The aim of the two models we will shortly present is to describe as closely as possible average choice statistics across trials.

The data produced by the experiments has the form of eight probabilities, corresponding to the decision of the participants to deny (plus) or confess (minus), when encountering the different PD payoff matrices (sentencing policies). The recorded probabilities always correspond to the participant deciding to plus. Therefore, for the $a1b1$ good matrix, the observed probability is recorded as $Prob(++)$ and in the case of the $a1b1$ bad matrix, the observed probability is recorded as $Prob(+-)$. Of course, we further inferred $Prob(-+)$, $Prob(--)$, etc.

We will present two models for the observed data, referred to as the classical hidden variables model (or just classical model) and the quantum model. It is more standard to formulate these models assuming a stochastic, rather than deterministic, associate. Accordingly, we combined choice statistics from the good and bad trials using e.g., $Prob(++|a1b1) = Prob(++|a1b1 \text{ Good}) \cdot Prob(\text{Good}|a1b1) + Prob(++|a1b1 \text{ Bad}) \cdot Prob(\text{Bad}|a1b1) = Prob(++|a1b1 \text{ Good}) \cdot Prob(\text{Good}|a1b1)$, where $Prob(\text{Good}|a1b1)$, $Prob(\text{Bad}|a1b1)$ refer to the probability of a good, bad game for a given choice of questions, respectively, and $Prob(++|a1b1 \text{ Bad}) = 0$. Since in all cases we employed equal proportions of good, bad trials, for each choice of questions, then $Prob(\text{Good}|a1b1) = Prob(\text{Bad}|a1b1) = 0.5$.

6. Hidden Variables Classical Model

According to this model, for each of the two agents, there is a hidden variable λ describing each sub-system, such that $\lambda_A = -\lambda_B$, with λ_A uniformly distributed over a 3D sphere. Note, this is an expression of perfect anti-correlation of the hidden variables corresponding to the agents, as opposed to perfect correlation, but this difference is immaterial (this is illustrated for the quantum model in Appendix B, but the case is analogous for the classical model). So, the main assumptions of the model are as follows. First, if the same questions are asked, the participant will always perfectly coordinate in the same way with the counterpart, that is, either always correlate or always anticorrelate; assuming always-correlation, if the participant denies, it is assumed the counterpart will deny as well, etc. Second, there is a specific value for all question outcomes at all times. The implication of this more subtle assumption is that the participant should produce an outcome to her question, independently of which question is asked to her counterpart. In physics, this is the key realism assumption. Third, this model assumes locality and free choice. In the present experiments, we endow participants with a means of violating locality, so if they do this in a certain way, we expect the model to perform poorly. A final, minor assumption is that the participant will generally recognize the optimal action in each trial (corresponding to a lower sentence), and that she will always assume that her associate will also take the optimal action. This assumption is minor because of the way the payoff matrices were constructed, but if it is wrong, the model will just fail (both models will fail). In what follows, instead of a participant and her counterpart, we sometimes talk about two interacting agents, Alice and Bob.

The first agent is measured in two directions, $a1$, $a2$ and the second agent is measured in two different directions, $b1$, $b2$. In the present psychological context, 'directions' just correspond to the steer for action from each question, which is a function of the information in the payoff matrix and the agent's interpretation of this information (which will depend

on his/her personality etc.). Non-trivial algebra shows that (e.g., [10]; note, the assumption concerning the existence of the hidden variable λ impacts on how these probabilities are derived):

$$\text{Prob}(++|a,b) = \frac{\theta}{2\pi} = \text{Prob}(--|a,b), \text{Prob}(+-|a,b) = \frac{1}{2} - \frac{\theta}{2\pi} = \text{Prob}(-+|a,b) \quad (3)$$

The key parameter in Equation (3) is the angle θ , in radians, corresponding to the correlation between a measurement direction a for Alice and b for Bob. So, the joint probability for Alice and Bob to deny for question combination ab depends on the relation between how Alice perceives question a and Bob question b . Note that when $\theta = 0$, there is an equal chance for Alice and Bob to anticorrelate in one way (plus, minus) vs. the opposite way (minus, plus), which is just an expression of the assumption $\lambda_A = -\lambda_B$, in the considered hidden variable model.

Since we have four pairs of measurement directions, $a1b1$, $a1b2$, $a2b1$, $a2b2$, then there are four angles as the parameters of this model. But these parameters are not independent. In the original physics set up they are actual measurement directions—psychologically, there is a corresponding assumption regarding the extent to which the two agents align or not in their consideration of questions. Suppose we have co-planar measurement directions, without much loss of generality. Then, the Figure 1 arrangement is a plausible representation of the four directions. Without loss of generality, we set $\theta_a = 0$ and θ_{b1} , θ_{b2} and θ_{a2} as shown in Figure 1. Then, the four angles needed for the classical model are given as $a1b1 = \theta_{b1} \bmod \pi$, $a1b2 = \theta_{b2} \bmod \pi$, $a2b1 = (\theta_{a2} - \theta_{b1}) \bmod \pi$, and $a2b2 = (\theta_{a2} - \theta_{b2}) \bmod \pi$. The mod π function simply ensures that the angles for the four question pairs stay within the $0 < \text{angle} < \pi$ limit. It is defined as:

$$\text{mod } \pi(x) = \begin{cases} \text{if } x > 0, \begin{cases} \text{if } x - \pi < 0, x \\ \text{if } x - \pi > 0, 2\pi - x \end{cases} \\ \text{if } x < 0, \begin{cases} \text{if } x + \pi > 0, -x \\ \text{if } x + \pi < 0, 2\pi + x \end{cases} \end{cases} \quad (4)$$

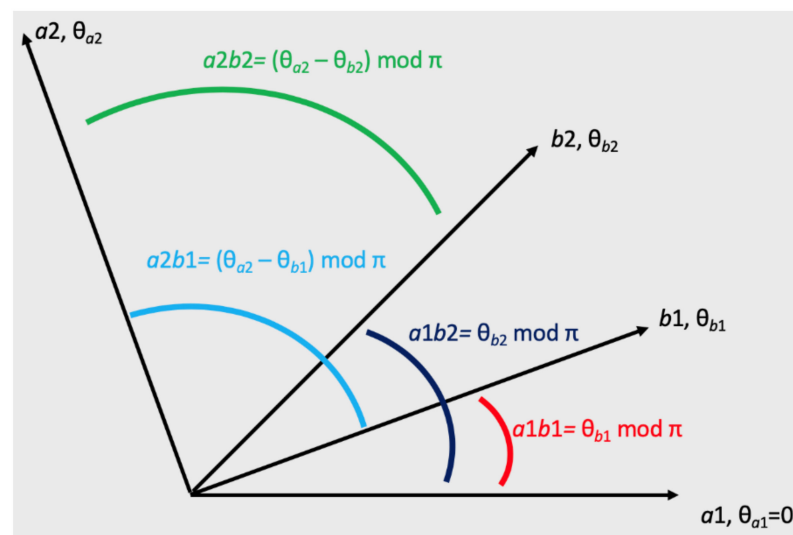


Figure 1. The arrangement of the four measurement directions.

We next consider the S value given this classical model. $\text{Prob}(++|a1, b1)$ is the probability for both agents to +, when the questions are $a1$, $b1$, $\text{Prob}(+-|a1, b1)$ the probability for Alice to + and Bob to—etc. Each expectation value is given by $\langle a \& b \rangle = \frac{2\theta}{\pi} - 1$, where θ is the angle between the measurement directions a , b . The overall result for the classical model is then:

$$S = \left| -2 + \frac{2}{\pi} [\theta_{a1b1} + \theta_{a1b2} + \theta_{a2b1} - \theta_{a2b2}] \right| \quad (5)$$

Note, we have mentioned that for this classical model S is bounded by 2. It can be shown that for $\theta_{a1b1} + \theta_{a1b2}$ the max is $2\pi - \theta$ and the min is θ , where θ is the angle between $b1, b2$, and for $\theta_{a2b1} - \theta_{a2b2}$ the max, min are θ and $-\theta$. Together these results deliver the classical limits for S .

7. Quantum Model

One of the most significant discoveries in the history of quantum theory has been the capacity of the theory to break the classical $S \leq 2$ bound, seemingly without violating either locality or free choice. In the present paradigm, the situation is less philosophically challenging, since we endow the two agents with a communication capacity to break locality. Since the statistics produced by the quantum model are equivalent to classical ones, but with a degree of violation of locality (or free choice; [32]), the quantum model is a reasonable option for the present paradigm. The assumptions of the quantum model are equivalent to those of the classical one, but for two differences. First, instead of the Bayesian probability rules, we employ the probability rules from quantum theory. Second, instead of a hidden variable capturing perfect coordination between the two agents, we have the quantum property of *entanglement* (see just below). However, this is not true (physical) quantum entanglement, but rather one of a more epiphenomenal flavor [40].

A column vector is denoted as $|x\rangle$, its conjugate transpose as $\langle x|$, and an inner product between two vectors as $\langle x|y\rangle$. Since we are concerned with two systems (agents), we need to employ tensor products to construct the joint state from the individual states, for example, $|x\rangle \otimes |y\rangle$ which can be written for brevity as $|xy\rangle$. We employ a qubit representation such that 0 means an intention for a ‘−’ (minus) action (Confess) and 1 a ‘+’ action (Deny). States are represented as $|\psi\rangle = a|x\rangle + b|y\rangle$. Measurements can change the state, so if on measuring ψ we obtain x the new state becomes $|\psi\rangle = |x\rangle$.

We start with state, $|\psi_+\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}$, where the tensor structure is so that the first index corresponds to Alice and the second to Bob (the subscript ‘+’ in $|\psi_+\rangle$ simply indicates a ‘correlation’ state). So, $|00\rangle$ means that Alice is intending to minus and Bob to minus etc. Note, in physics, the state used is typically the singlet state, which is an anticorrelation state, $|\psi_-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$. However, the predictions from $|\psi_+\rangle$ are essentially identical but for a fixed rotation of the measurement directions; so, for the purposes of model fitting, this issue is irrelevant (in a way analogous to that for the classical model). The state $|\psi_+\rangle$ is called entangled and is one of perfect coordination between the two agents, but now using the rules of quantum theory. The predictions from the quantum model are then (Appendix C).

$$Prob(++|a, b; \psi_-) = \frac{1}{2} \sin^2\left(\frac{\theta}{2}\right) = Prob(--|a, b; \psi_-), \quad Prob(+-|a, b; \psi_-) = \frac{1}{2} \cos^2\left(\frac{\theta}{2}\right) = Prob(-+|a, b; \psi_-) \quad (6)$$

As before, the crucial parameter is the angle θ for each measurement direction. The four angles are constrained as for the classical model (Figure 1), so that the quantum model also has three parameters.

We can consider the computation for the Bell bound from the quantum model. We have that the expectation values are given by $\langle a \& b \rangle = -\cos \theta$, where θ is the angle between the two measurement directions. Then,

$$S = [-\cos \theta_{a1b1} - \cos \theta_{a1b2} - \cos \theta_{a2b1} - (-\cos \theta_{a2b2})] \quad (7)$$

It can be immediately seen that if we set the angle for $a1b1, a2b1, a1b2$ to $\frac{\pi}{4}$, with the arrangement as in Figure 1, $a2b2$ is $\frac{3\pi}{4}$. Then $S = |-\cos \frac{\pi}{4} - \cos \frac{\pi}{4} - \cos \frac{\pi}{4} - (-\cos \frac{3\pi}{4})| = 2\sqrt{2} > 2$. In fact, though not obvious from the present discussion, a quantum model cannot produce S values greater than $2\sqrt{2}$ and $2\sqrt{2}$ is called Tsirelson’s bound [31].

8. Overview of the Two Models

The question we are interested in is whether a model satisfying realism, locality, and free choice can model this data—this is the hidden variables classical model. The answer is not automatically no because, even though locality is violated, it is an empirical question whether participants recognize the need to employ non-local resources and use the available information efficiently. If participants do not employ the non-local information, then the results could still be described by a local model and $S < 2$. That is, in this situation, the possibility of communication (checking) is clearly a necessary condition for participant data to violate Bell's bound, but it is not a sufficient one.

A related question is whether any use of local information can be modelled by a quantum model (which is constrained by Tsirelson's bound) or not. If not, then participants' checking behavior and use of the corresponding information would be greater than what is allowed by quantum theory.

Because there is communication in this case, it is likely that there is signaling as well. If there is signaling, the bound of $S = 2$ is clearly not a fundamental limitation on how a system behaves. However, there is still an empirical question on how people behave, and we can ask the question (as above) of whether human behavior can be characterized by a local model ($S < 2$), a nonlocal model constrained by Tsirelson's bound (the quantum model), or something else.

Table 8 shows the predictions from both models, where probabilities correspond to averaged data across multiple trials. This is easier to show by retaining the reference to the good, bad matrices, bearing in mind that in the fitted data we average probabilities across these two experimental situations to better match the actual models.

Table 8. Correspondence between observed probabilities and predictions from the classical and quantum models.

Term	Observed Probability	Classical Prediction	Quantum Prediction
<i>a1b1</i> good	$Prob(++)$	$Prob(++) = \frac{\theta}{2\pi}$	$Prob(++) = \frac{1}{2} \sin^2\left(\frac{\theta}{2}\right)$
<i>a1b1</i> bad	$Prob(+-)$	$Prob(+-) = \frac{1}{2} - \frac{\theta}{2\pi}$	$Prob(+-) = \frac{1}{2} \cos^2\left(\frac{\theta}{2}\right)$
<i>a1b2</i> good	$Prob(++)$	$Prob(++) = \frac{\theta}{2\pi}$	$Prob(++) = \frac{1}{2} \sin^2\left(\frac{\theta}{2}\right)$
<i>a1b1</i> bad	$Prob(+-)$	$Prob(+-) = \frac{1}{2} - \frac{\theta}{2\pi}$	$Prob(+-) = \frac{1}{2} \cos^2\left(\frac{\theta}{2}\right)$
<i>a2b1</i> good	$Prob(++)$	$Prob(++) = \frac{\theta}{2\pi}$	$Prob(++) = \frac{1}{2} \sin^2\left(\frac{\theta}{2}\right)$
<i>a2b1</i> bad	$Prob(+-)$	$Prob(+-) = \frac{1}{2} - \frac{\theta}{2\pi}$	$Prob(+-) = \frac{1}{2} \cos^2\left(\frac{\theta}{2}\right)$
<i>a2b2</i> good (Bob '+'s)	$Prob(++)$	$Prob(++) = \frac{\theta}{2\pi}$	$Prob(++) = \frac{1}{2} \sin^2\left(\frac{\theta}{2}\right)$
<i>a2b2</i> bad (Bob '-'s)	$Prob(+-)$	$Prob(+-) = \frac{1}{2} - \frac{\theta}{2\pi}$	$Prob(+-) = \frac{1}{2} \cos^2\left(\frac{\theta}{2}\right)$

9. Model Fitting

Fits were assessed with Maximum Likelihood Estimation (MLE), using the G^2 expression for summary statistics in an experiment, $G^2 = 2N \sum_{\text{trial types}} \left(o_i \ln \frac{o_i}{e_i} + (1 - o_i) \ln \frac{1 - o_i}{1 - e_i} \right) = 2N \sum_{i,j=\{+,-\}} \left(Prob(ij, \text{observed}) \ln \frac{Prob(ij, \text{observed})}{Prob(ij, \text{model})} \right)$, where N is the number of observations and o_i, e_i observed and expected probabilities for each trial type. Best fit for the models was identified through directed grid search with a step size for angle differences of 0.1; all parameters were taken to be uniformly distributed in a $[0, 2\pi]$ range. For simplicity, since N was nearly identical for the two experiments, we ignored it in computing G^2 .

10. Fit Results

Table 9 shows observed, classical predicted, and quantum predicted probabilities. Observe that for the $a1b1$, $a1b2$, and $a2b1$ pairs we recorded higher probabilities along the diagonals of the corresponding cells, but for the $a2b2$ pair, the opposite is true. This is the essential impression of supercorrelation and sensitivity to context: participants respond differently to question $a2$ depending on whether his/her counterpart received question $b1$ (correlation) vs. $b2$ (anticorrelation).

Table 9. The observed and fitted results for Experiments 1 (left) and 2 (right).

	b1+	b1−	b2+	b2−		b1+	b1−	b2+	b2−
a1+	0.47	0.06	0.485	0.05	a1+	0.47	0.035	0.46	0.025
a1−	0.03	0.44	0.015	0.45	a1−	0.03	0.465	0.04	0.475
a2+	0.47	0.055	0.14	0.38	a2+	0.34	0.25	0.155	0.425
a2−	0.03	0.445	0.36	0.12	a2−	0.16	0.25	0.345	0.075
Empirical probabilities					Empirical probabilities				
	b1+	b1−	b2+	b2−		b1+	b1−	b2+	b2−
a1+	0.398	0.102	0.427	0.073	a1+	0.446	0.054	0.459	0.041
a1−	0.102	0.398	0.073	0.427	a1−	0.054	0.446	0.041	0.459
a2+	0.398	0.102	0.223	0.277	a2+	0.255	0.245	0.159	0.341
a2−	0.102	0.398	0.277	0.223	a2−	0.245	0.255	0.341	0.159
Probabilities predicted by the classical model					Probabilities predicted by the classical model				
	b1+	b1−	b2+	b2−		b1+	b1−	b2+	b2−
a1+	0.448	0.052	0.45	0.05	a1+	0.474	0.026	0.476	0.024
a1−	0.052	0.448	0.05	0.45	a1−	0.026	0.474	0.024	0.476
a2+	0.45	0.05	0.159	0.341	a2+	0.307	0.193	0.095	0.405
a2−	0.05	0.45	0.341	0.159	a2−	0.193	0.307	0.405	0.095
Probabilities predicted by the quantum model					Probabilities predicted by the quantum model				

We computed three S values, one for the observed choice probabilities, one for the predicted probabilities based on the classical model, and one for the predicted probabilities based on the quantum model. Note that, for Experiment 2, empirical S was computed on the basis of the trials for which participants could freely choose whether to check on their associate or not. For Experiment 1, the empirical S , best fit classical S , and best fit quantum S were, respectively, 3, 2 ($G^2 = 0.46$), and 2.76 ($G^2 = 0.08$). For Experiment 2, the corresponding values were 2.46, 2 ($G^2 = 0.17$), and 2.65 ($G^2 = 0.09$). Bootstrapped 95% confidence intervals for the empirical S values were [2.73, 3.23] for Experiment 1 and [2.23, 2.71] for Experiment 2. The confidence intervals were computed by first calculating individual S values for each participant (only choice trials were used in this computation). Means were then calculated from each of the 1000 bootstrap samples created (each bootstrapped sample was a random choice of N values from the original sample, with replacement, where N = number of values in the sample, i.e., the number of participants). Finally, the bootstrapped means were sorted and quantiles of 0.025 and 0.975 were utilized to indicate the 95% confidence intervals for each participant. In all cases, the empirical data show $S > 2$, which demonstrates sensitivity to context and the impossibility of a four-way classical probability distribution to explain the data. The classical model resulted in worse fits than the quantum one, with the latter producing S values closer to the observed ones. Note that while the quantum model is able to capture a certain kind of sensitivity to context, of course it cannot describe any behavior [31].

Using the forced checking and non-checking trials in Experiment 2, we computed S values for checking and non-checking trials for each participant. Note, in this case, it is only checking trials that should allow a violation of the $S \leq 2$ bound—therefore, for non-checking trials, it must be the case that $S \leq 2$. When participants were not checking on their associate, S for the good and bad trials respectively were 1.78 and 1.82; when

checking, we observed 2.91 and 2.59, respectively. The difference in S between checking (averaged across good, bad matrices 2.75) and non-checking trials (averaged across good, bad matrices 1.80) was reliable, $Z = -6.44$, $p < 0.001$ (using the Wilcoxon Signed Rank Test, as the normality assumption would be suspect here).

11. Signaling

We finally, briefly consider the issue of signaling, for completeness. We can define a signaling quantity as:

$$I_S = \sum_{i=1,2} |\langle a_i \rangle_{b1} - \langle a_i \rangle_{b2}| + \sum_{j=1,2} |\langle b_j \rangle_{a1} - \langle b_j \rangle_{a2}| \quad (8)$$

$$= |\langle a_1 \rangle_{b1} - \langle a_1 \rangle_{b2}| + |\langle a_2 \rangle_{b1} - \langle a_2 \rangle_{b2}| + |\langle b_1 \rangle_{a1} - \langle b_1 \rangle_{a2}| + |\langle b_2 \rangle_{a1} - \langle b_2 \rangle_{a2}|$$

where the expectation values are defined as expected, for example, $\langle a_1 \rangle_{b1} = (+1) \cdot (\text{Prob}(++|a1b1) + \text{Prob}(+-|a1b1)) + (-1) \cdot (\text{Prob}(-+|a1b1) + \text{Prob}(|a1b1))$. Note, the max value for I_S is 8, when communication in both directions is considered (this is relevant in evaluating the size of the observed I_S values). We review a point which may lead to confusion: the probabilities in Tables 5 and 7 are not exactly the ones appearing in these expectation values. This is because, in Tables 5 and 7, we counted probabilities separately for the Good and Bad matrices, i.e., the probabilities in Tables 5 and 7 are e.g., $\text{Prob}(++|a1b1, \text{Good})$. Therefore, as seen above too, we need to compute $\text{Prob}(++|a1b1) = \text{Prob}(++ \text{ Good}|a1b1) + \text{Prob}(++ \text{ Bad}|a1b1)$, but recall $\text{Prob}(++ \text{ Bad}|a1b1) = 0$. So, $\text{Prob}(++|a1b1) = \text{Prob}(++ \text{ Good}|a1b1) = \text{Prob}(++|a1b1, \text{Good}) \cdot \text{Prob}(\text{Good}|a1b1) = \text{Prob}(++|a1b1, \text{Good}) \cdot \frac{1}{2}$, because in the present design $\text{Prob}(\text{Good}|a1b1) = \text{Prob}(\text{Bad}|a1b1) = 1/2$ (meaning the probability of having a ‘good’ payoff matrix etc.; the same applies for all question combinations). The probabilities $\text{Prob}(++|a1b1, \text{Good})$ etc. are the ones in Tables 5 and 7 and so in computing the expectation values for I_S , all probabilities from Tables 5 and 7 need to be multiplied by a factor of $\frac{1}{2}$ (the same applies to the calculations for the S values presented in Table 10).

Table 10. Contextuality tests for Experiments 1 and 2.

	S	I_S	$ S - I_S$
Exp1 checking	3.2	0.08	3.12
Exp1 no-checking	2.45	0.33	2.12
Exp2 checking	2.74	0.18	2.56
Exp2 no-checking	1.8	0.04	1.76

We computed I_S separately for each experiment and for the checking vs. no checking trials. For Experiment 1, for the checking and no checking trials we observed, respectively, that $I_S = 0.08$ and $I_S = 0.33$. The corresponding values in Experiment 2 were $I_S = 0.08$ and $I_S = 0.04$. In Experiment 2, the results are as expected, since there is more signaling in the checking trials (ostensibly as a result of communication). In Experiment 1, even though for the no checking trials there was no communication, we still observed sizeable signaling. Signaling in Experiment 1 would be the result of the lack of balancing between the payoff matrices (as discussed in detail above). A consideration of signaling is clearly useful as a way to establish whether there might be unintended causal influences in the experimental statistics (as in Experiment 1). However, the non-zero I_S in Experiment 2 in the no checking trials indicates that signaling may be apparent even when there is no plausible corresponding mechanism, perhaps as a result of noise [16]. This does recommend caution when employing signaling in such experiments, especially when the N is small (as would be the case in behavioral experiments).

The calculation of the signaling quantifiers I_S allows us to test for contextuality in the sense of [15], which we do here for completeness. According to this work, contextuality is present whenever $|S| - I_S > 2$ (the S here refers to the maximum one between the four pos-

sible ways to compute it; here, we focused only on $S = |\langle a1 \& b1 \rangle + \langle a1 \& b2 \rangle + \langle a2 \& b1 \rangle - \langle a2 \& b2 \rangle|$, which is most relevant to our experimental design). In Table 10, we offer a complete record of relevant S values for the checking/ no checking quantifiers separately, for both experiments, as well as the quantities $|S| - I_S$, which are, as it happens, indicative of contextuality.

12. General Discussion

Sensitivity to context is an important insight concerning the representation of information, whether in physics, data science, or psychology. Outside the physics of microscopic particles, it is assumed that there are no true quantum processes, and the study of sensitivity to context leads one to question the mechanism that supports it. In psychology, some pioneering work has been carried out so that both sets of questions, $\{a1, a2\}$ and $\{b1, b2\}$, would be answered by the same participant, or in any case concern mental processes focused on the individual (e.g., [24,28]). Such approaches cannot be adapted to the interaction between separate agents because, in general, without communication there is no possibility of breaking Bell's bound (or without rigging the choice of the questions asked to each agent).

For the first time, in this study we developed an approach enabling the application of the Bell framework in the interaction of two cognitive (and so macroscopic) agents. We considered putative locality violations as an information resource, that two interacting agents can employ at will (cf. [32]). We developed a simple empirical paradigm which embodied sensitivity to context in its structure, as a variant of a PD task [2] Empirical results showed that participants were sensitive to this context and the empirical S values exceeded Bell's bound. As noted, this is not surprising, given the structure of the payoff matrices we employed. The more surprising implication is that this sensitivity prevented fits by a simple classical model and therefore shows another way in which PD tasks and variants can produce results problematic for baseline expectation from classical probability theory. 'Baseline' is a key qualification here since, as noted above, a classical model incorporating communication could be developed to account for the present results. Therefore, the present situation is not unlike most so-called paradoxes in probabilistic inference, for which a baseline classical probability approach appears erroneous, but it is always possible to offer accommodating elaborations (e.g., faced with a result such as $Prob(X \& Y) > Prob(X)$, one could write $Prob(X \& Y | A) > Prob(X | B)$).

Theoretically, we fitted two closely matched models, a classical and a quantum one. The latter produced superior fits. This conclusion adds to the body of evidence that quantum theory sometimes offers a good descriptive framework for behavior [33,34]. Elsewhere, we have suggested that this is because quantum theory looks like Bayesian inference, but in a local way [49]. That is, a set of questions for which it is impossible to have a complete joint probability distribution (e.g., because of resource limitations) is divided into subsets, such that within each subset—locally—we have Bayesian inference, but across subsets apparent classical errors arise. The idea that behavior is 'locally rational' has a precedent in psychology [50,51].

Note that the immediate availability of locality violations to the participants makes it unlikely that any results showing $S > 2$ would be due to 'correlations of the second kind', as discussed by S. Aerts and D. Aerts [19,21,52]. In Experiment 2, when participants would check on the hypothetical counterpart we observed $S > 2$ and when they would not, $S < 2$, showing that any apparent sensitivity to context was not brought about just by the measurements (decisions) themselves.

From the point of view of a physicist, the present results are interpreted as sensitivity to context, due to communication, regardless of whether this sensitivity to context is due to signaling or not. As noted, rather than considering signaling a nuisance influence, in this case we are interested in it, as a possible way in which Alice makes use of the information she has about Bob's questions.

There have been several challenges in realizing this project. First, the notion of applying the Bell framework to the interaction of cognitive agents superficially goes against the grain of Bell's work in physics. To address this problem, we had to formalize a notion of violations of locality or free choice, as information resources, which can be adopted vs. not at will (our formal work on this topic is reported in [32], as well as consider the distinction between context sensitivity and contextuality (for the latter see [15]). Second, adapting the classical and quantum models developed for systems of microscopic particles in physics to behavioral data required careful consideration of the underlying assumptions of the models and how they could be matched to behavioral situations. Third, the difference between contextuality and sensitivity to context and the restrictive (or not) role of signaling in Bell-type paradigms are highly contentious issues. We think the approach we chose is justified, but equally we have offered additional analyses which we hope will allow researchers of differing opinions to still appreciate the results. Finally, reporting the research was challenging: the primary audience for this work is cognitive scientists, but we also hope to interest physicists and mathematicians familiar with Bell who might be intrigued by applications outside physics. But the mathematics is likely to be unfamiliar and challenging to cognitive scientists, while the details of the behavioral paradigm unfamiliar to physicists and mathematicians. Overall, interdisciplinary work of this kind, while conceptually exciting and potentially rewarding, is fraught with challenges—we can only hope that we have been at least partly successful in overcoming them.

The present analysis has practical potential. Consider two agents, Alice and Bob, for whom it is in their interest to supercorrelate, but such that they are not meant to break locality and free choice, e.g., they are not meant to communicate. Alice and Bob might be an employee in a tech firm and a stockbroker considering investment opportunities in that firm, respectively. The present framework could be employed to determine whether Alice and Bob benefit from supercorrelation, either on the basis of violations of locality (which may reveal illegal insider trading) or free choice (which could correspond to Alice and Bob independently being sensitive to market conditions which determine the 'questions' each one of them has to respond to, at a given time). Clearly, the applicability of such an analysis depends largely on how the questions for each agent are specified and whether there is advantage in supercorrelation, which may not be very often.

In closing, we hope that the present work will further encourage researchers to employ the notion of contextuality and the corresponding technical tools in the study of the interaction between multiple agents.

Author Contributions: Conceptualization, E.M.P., P.B., and J.M.Y.; methodology, O.W. and E.M.P.; software, E.M.P. and P.B.; validation, all authors; formal analysis, P.B. and J.M.Y.; investigation, O.W. and B.W.W.; resources, E.M.P., P.B., B.W.W.; data curation, O.W.; writing—original draft preparation, E.M.P.; writing—review and editing, all authors; visualization, O.W. and E.M.P.; supervision, O.W.; project administration, O.W. and E.M.P.; funding acquisition, E.M.P., P.B., and B.W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Office of Naval Research Global grant N62909-19-1-2000.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of the Psychology Department at City, University of London (ETH1920-0169; approved on 4 September 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are openly available in the OSF at DOI: 10.17605/OSF.IO/R743P.

Acknowledgments: E.M.P., P.B., and B.W.W. were funded by Office of Naval Research Global grant N62909-19-1-2000.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

An example of how $S > 2$ means that the counts or proportions for four binary questions are inconsistent, that is, they do not obey the expected sum rules (if a 4×4 table is employed for representation).

Below we show a table of probabilities, which could be computed from frequencies in a behavioral experiment (Table A1). We try to fill it in assuming maximum correlation for the three first pairs and then maximum anticorrelation for the last pair. For a pair of questions, e.g., a and b , there are four cells, and the sum of the corresponding probabilities must be one. That is, $Prob(a+ \wedge b+) + Prob(a+ \wedge b-) + Prob(a- \wedge b+) + Prob(a- \wedge b-) = 1$ (the notation $a+$ means a plus outcome for question a). This constraint is a simple implication of the fact that these probabilities span all the space of possibilities for the outcomes of any pair of two questions. Let us first consider the highlighted cells, corresponding to the a, b pair of questions.

Table A1. A sequence of tables to illustrate the implications of $S > 2$: a and b .

	$b = +$	$b = -$	$b' = +$	$b' = -$
$a = +$				
$a = -$				
$a' = +$				
$a' = -$				

Because we are assuming maximum correlation, we can set $Prob(a+ \wedge b+) = p$ and $Prob(a- \wedge b-) = 1 - p = q$, with the other probabilities equal to 0 (we leave blank the cells corresponding to 0 probabilities). Table A1 then becomes:

Table A2. A sequence of tables to illustrate the implications of $S > 2$: computing a and b .

	$b = +$	$b = -$	$b' = +$	$b' = -$
$a = +$	p			
$a = -$		q		
$a' = +$				
$a' = -$				

We then consider the cells corresponding to the a, b' pair of questions, and the cells corresponding to the a', b pair of questions.

Table A3. A sequence of tables to illustrate the implications of $S > 2$: a, b' and a', b .

	$b = +$	$b = -$	$b' = +$	$b' = -$
$a = +$	p			
$a = -$		q		
$a' = +$				
$a' = -$				

Note the highlighted parts below are constrained from the white part. For example, consider the highlighted $a+$ row. If we set one of the cells, e.g., $Prob(a+ \wedge b'+)$, then the other probability follows, from the law of total probability:

$$Prob(a+) = Prob(a+ \wedge b'+) + Prob(a+ \wedge b'-) = Prob(a+ \wedge b+) + Prob(a+ \wedge b-)$$

The second part of the equation is known (from the white part of the table), $Prob(a+ \wedge b+) + Prob(a+ \wedge b-) = p$. In physics this condition is called non-signaling, i.e., the marginal probability $Prob(a+)$ does not depend on the other question b or b' . So, in filling in the highlighted parts of the table we only need to worry about one probability in each row. This one probability can be set on the basis of the logic above, p and $q = 1 - p$ for the terms which are meant to be correlating, so ending up with:

Table A4. A sequence of tables to illustrate the implications of $S > 2$: computing a, b' and a', b .

	$b = +$	$b = -$	$b' = +$	$b' = -$
$a = +$	p		p	
$a = -$		q		q
$a' = +$	p			
$a' = -$		q		

We note that the locality and free choice assumptions are equivalent to an assumption of lack of contextuality for all questions. That is, lack of contextuality must mean that e.g., $Prob(a')$ is the same regardless of context, that is, regardless of whether it is measured with b or b' . Therefore, we would have that

$$Prob(a'+) = Prob(a' + \wedge b+) + Prob(a' + \wedge b-) = Prob(a' + \wedge b'+) + Prob(a' + \wedge b'-) = p$$

$$Prob(a'-) = 1 - Prob(a'+) = 1 - p = q$$

$$Prob(b'+) = Prob(a + \wedge b'+) + Prob(a - \wedge b'+) = Prob(a' + \wedge b'+) + Prob(a' - \wedge b'+) = p$$

$$Prob(b'-) = 1 - Prob(b'+) = 1 - p = q$$

We highlight the cells relevant to the first constraint, for $Prob(a'+)$, in Table A5 below.

Table A5. A sequence of tables to illustrate the implications of $S > 2$: illustrating constraints.

	$b = +$	$b = -$	$b' = +$	$b' = -$
$a = +$	p		p	
$a = -$		q		q
$a' = +$	p			
$a' = -$		q		

Note that the above marginal conditions lead to:

$$Prob(a'+) = Prob(b'+) = p$$

And likewise

$$Prob(a'-) = Prob(b'-) = q$$

What remains is to fill the bottom right part of the table in a way that reflects the anticorrelation pattern for $a'b'$, i.e., so that the diagonal elements vanish

$$Prob(a' + \wedge b'+) = Prob(a' + \wedge b'-) = 0$$

Now, by the same logic as above, we arrive at

$$Prob(a'+) = Prob(a' + \wedge b'-) = Prob(b'-),$$

$$Prob(a'-) = Prob(a' + \wedge b'+) = Prob(b'+),$$

which entails that $p = q = 0.5$. Thus, we obtain Table A6b. The reader may compare this table with the case when all questions maximally correlate. The derivation follows the same pattern except the last step, which in this case does not fix probabilities p and q . See Table A6a for comparison.

Note that in Table A6a, with $p = q = 0.5$, individually each answer is a coin toss, but looking at Alice and Bob together they perfectly coordinate for questions ab , ab' and $a'b$ and perfectly anticorrelate for questions $a'b'$. This is an instance of the famous PR-box (Popescu-Rohrlich box) considered in the physics literature [13]. In the present work, we aim to provide an analogue for the interaction of two agents.

However, contextuality, or its lack of it, does not help us see how violation of the Bell bound is inconsistent with the existence of a four-way probability distribution. This can be readily seen by comparing the two tables below, Table A6a,b.

Table A6. A sequence of tables to illustrate the implications of $S > 2$. Note, (a) is ‘classical’, while (b) is ‘contextual’ (the response to a' depends on whether Bob answers b or b').

	$b = +$	$b = -$	$b' = +$	$b' = -$		$b = +$	$b = -$	$b' = +$	$b' = -$
$a = +$	p		p		$a = +$	0.5		0.5	
$a = -$		q		q	$a = -$		0.5		0.5
$a' = +$	p		p		$a' = +$	0.5			0.5
$a' = -$		q		q	$a' = -$		0.5	0.5	
(a)					(b)				

In order to show that the right table is contextual and therefore inconsistent with classical probability theory, we need to make use of the overarching assumption that there exists a complete (four-way in this case) classical probability distribution.

$$\begin{aligned}
 & Prob(a' + \wedge b' -) \\
 &= Prob(a' + \wedge b' - \wedge a + \wedge b +) + Prob(a' + \wedge b' - \wedge a + \wedge b -) \\
 &+ Prob(a' + \wedge b' - \wedge a - \wedge b +) + Prob(a' + \wedge b' - \wedge a - \wedge b -)
 \end{aligned}$$

If we know a part of a conjunction to be 0 then a more restrictive conjunction will be 0 too:

$$Prob(a' + \wedge b' - \wedge a + \wedge b +) \text{ is 0 because } Prob(b' - \wedge a +) = 0$$

$$Prob(a' + \wedge b' - \wedge a + \wedge b -) \text{ is 0 because } Prob(b' - \wedge a +) = 0$$

$$Prob(a' + \wedge b' - \wedge a - \wedge b +) \text{ is 0 because } Prob(a - \wedge b +) = 0$$

$$Prob(a' + \wedge b' - \wedge a - \wedge b -) \text{ is 0 because } Prob(a' + \wedge b -) = 0$$

$$\text{So, } Prob(a' + \wedge b' -) = 0 \text{ and analogously } Prob(a' - \wedge b' +) = 0$$

But then:

$$Prob(a' + \wedge b' +) + Prob(a' + \wedge b' -) = Prob(a' + \wedge b +) + Prob(a' + \wedge b -) \Leftrightarrow Prob(a' + \wedge b' +) = 0.5$$

$$Prob(a' - \wedge b' -) + Prob(a' - \wedge b' +) = Prob(a' - \wedge b +) + Prob(a' - \wedge b -) \Leftrightarrow Prob(a' - \wedge b' -) = 0.5$$

This is in contradiction with Table A6b where $Prob(a' + \wedge b' +) = Prob(a' + \wedge b' -) = 0$. The above shows that (baseline classically) the probabilities are constrained to be set in a certain way. The contextuality needed to break the Bell bound is not allowed.

Appendix B

We show that quantum predictions are equivalent for all maximally entangled states (note that for the classical model it is straightforward to see that the model can easily adjust itself depending on whether perfect correlation or anti-correlation between the agents is assumed). This equivalence is up to a simple transformation of the angles employed. Therefore, it has little impact in model fits whether we employ a perfect correlation state (which fits our empirical situation well) or a perfect anticorrelation one (which is the standard state for this kind of analysis).

We know that for the Bell singlet state

$$|\psi_{-}\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$$

the statistics of measurement outcomes in directions a for Alice and b for Bob is given by the formula

$$\langle ab \rangle_{\psi_{-}} = \langle \psi_{-} | a \cdot \sigma_A \otimes b \cdot \sigma_B | \psi_{-} \rangle = -a \cdot b = -\cos \theta_{ab}$$

Suppose we want to get statistics for the same measurement on a different maximally entangled state, such as

$$|\psi\rangle = \frac{|\phi_0\rangle|\xi_0\rangle + |\phi_1\rangle|\xi_1\rangle}{\sqrt{2}}$$

with some orthogonal basis states $|\phi_0\rangle, |\phi_1\rangle$ for Alice and $|\xi_0\rangle, |\xi_1\rangle$ for Bob. The trick is to observe that there exist two unitaries U and V such that

$$\begin{aligned} |0\rangle_A &\xrightarrow{U} |\phi_0\rangle_A, |1\rangle_A \xrightarrow{U} |\phi_1\rangle_A \\ |1\rangle_B &\xrightarrow{V} |\xi_0\rangle_B, |0\rangle_B \xrightarrow{V} |\xi_1\rangle_B \end{aligned}$$

Note that these unitaries are realized as rotations of the Bloch sphere for the respective Alice and Bob's qubit, i.e., we have (cf. [53], Chapter 4.2)

$$U = R_{\hat{n}_A}(\theta_A), V = R_{\hat{n}_B}(\theta_B)$$

Clearly, we have

$$|\psi\rangle = U \otimes V |\psi_-\rangle$$

Now, we compute

$$\begin{aligned} \langle ab_\psi \rangle &= \langle \psi | a \otimes \sigma_A \otimes b \otimes \sigma_B | \psi \rangle = \langle \psi_- | (U^\dagger \otimes V^\dagger) a \cdot \sigma_A \otimes b \cdot \sigma_B (U \otimes V) | \psi_- \rangle \\ &= \langle \psi_- | (U^\dagger a \cdot \sigma_A U) \otimes (V^\dagger b \cdot \sigma_B V) | \psi_- \rangle = \langle \psi_- | a' \cdot \sigma_A \otimes b' \cdot \sigma_B | \psi_- \rangle = \langle a' b' \rangle_{\psi_-} = -a' \cdot b' \\ &= -\cos \theta_{a'b'} \end{aligned}$$

where $a' = R_{\hat{n}_A}(-\theta_A)a$ and $b' = R_{\hat{n}_B}(-\theta_B)b$. The bottom line is that we treat $\langle ab \rangle_\psi$ as if it was $\langle a' b' \rangle_{\psi_-}$ (i.e., just measuring in different measurement basis).

That is, we are able to conclude that $\langle ab \rangle_\psi = \langle a' b' \rangle_{\psi_-}$. We either look for angles which produce best fit to our data, assuming ψ or different angles which produce best fit to our data, assuming ψ_- . The two pictures are equivalent and the two sets of angles are linearly related to each other.

Appendix C

We review the derivation of choice probabilities for the quantum model (which is fairly standard in physics).

In text, we employed $|\psi_+\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}$, but recall that for a maximally entangled state, such as $|\psi_+\rangle$, joint probabilities are unaffected, but for a fixed rotation on the angles, regardless of which maximally entangled state we employ, whether it is an anti-correlation one, such as $|\psi_-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$ or a correlation one, such as $|\psi_+\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}$ (this can be shown fairly easily). Specifically, below we proceed with $|\psi_-\rangle$, as is standard in physics discussions, but if a reader wishes to know the exact predictions for $|\psi_+\rangle$ all that is needed is to transform the angles as $\theta' = \theta + \pi$.

First, we need to show that a state $|\psi_-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$ can be written in any alternative, equivalent basis.

Starting with the Bell state

$$|\psi_-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$$

We seek to express it in the alternative basis

$$|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$$

where $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |\uparrow\rangle$, $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |\downarrow\rangle$. We have

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

$$|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

Which means

$$|+\rangle + |-\rangle = \frac{2}{\sqrt{2}}|0\rangle \Leftrightarrow |0\rangle = \frac{\sqrt{2}}{2}(|+\rangle + |-\rangle)$$

$$|+\rangle - |-\rangle = \frac{2}{\sqrt{2}}|1\rangle \Leftrightarrow |1\rangle = \frac{\sqrt{2}}{2}(|+\rangle - |-\rangle)$$

So,

$$\begin{aligned} |\psi_-\rangle &= \frac{|01\rangle - |10\rangle}{\sqrt{2}} = \frac{1}{2\sqrt{2}}((|+\rangle + |-\rangle) \otimes (|+\rangle - |-\rangle) - (|+\rangle - |-\rangle) \otimes (|+\rangle + |-\rangle)) \\ &= \frac{1}{2\sqrt{2}}(|++\rangle - |+-\rangle + |-+\rangle - |--\rangle - |++\rangle - |+-\rangle + |-+\rangle + |--\rangle) \\ &= \frac{1}{2\sqrt{2}}(-|+-\rangle + |-+\rangle - |++\rangle + |--\rangle) = \frac{1}{\sqrt{2}}(|-+\rangle - |++\rangle) \end{aligned}$$

Therefore, as long we can write

$$|\pm n\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$$

then we can have

$$|\psi_-\rangle = \frac{1}{\sqrt{2}}(|n_-n_+\rangle - |n_+n_-\rangle)$$

Second, we derive the expression for the joint probabilities for two measurement directions, e.g., $a \cdot \sigma_A$ on particle A, $a \cdot \sigma_A = a_x\sigma_x + a_y\sigma_y + a_z\sigma_z$. To find $Prob(+|a; \psi_-)$, to mean + outcome for the $a \cdot \sigma_A$ observable, we rewrite ψ_- in the $|\pm a\rangle$ basis,

$$|\psi_-\rangle = \frac{1}{\sqrt{2}}(|a_-a_+\rangle - |a_+a_-\rangle)$$

Therefore,

$$Prob(+|a; \psi_-) = |P_{a+} \otimes I \frac{1}{\sqrt{2}}(|a_-a_+\rangle - |a_+a_-\rangle)|^2 = \frac{1}{2}|a_+a_-\rangle|^2 = \frac{1}{2}$$

And clearly $Prob(-|a; \psi_-) = \frac{1}{2}$ too.

By analogy, we have $Prob(+|b; \psi_-) = Prob(-|b; \psi_-) = \frac{1}{2}$. In order to compute probabilities for simultaneous measurements on both particles for $a \cdot \sigma_A$ on the first particle and $b \cdot \sigma_B$ on the second particle, we must express $|\psi_-\rangle$ in the basis for $a \cdot \sigma_A$ and then compute the dot product between P_{b+} and $|a_+\rangle, |a_-\rangle$.

$$Prob(++|a, b; \psi_-) = |P_{a+} \otimes P_{b+} \frac{1}{\sqrt{2}}(|a_-a_+\rangle - |a_+a_-\rangle)|^2 = \frac{1}{2}||a_+\rangle \otimes P_{b+}|a_-\rangle|^2 = \frac{1}{2}|P_{b+}|a_-\rangle|^2 = \frac{1}{2}|\langle b_+|a_-\rangle|^2$$

$$\begin{aligned} Prob(+-|a, b; \psi_-) &= |P_{a+} \otimes P_{b-} \frac{1}{\sqrt{2}}(|a_-a_+\rangle - |a_+a_-\rangle)|^2 = \frac{1}{2}||a_+\rangle \otimes P_{b-}|a_-\rangle|^2 = \frac{1}{2}|P_{b-}|a_-\rangle|^2 = \frac{1}{2}|\langle b_-|a_-\rangle|^2 \\ &= \frac{1}{2}|\langle b_+|a_+\rangle|^2 \end{aligned}$$

In order to compute $|\langle b_+ | a_+ \rangle|^2$ we need to identify the eigenvectors of the operator $a \cdot \sigma_A = \begin{pmatrix} a_z & a_x - ia_y \\ a_x + ia_y & -a_z \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta e^{-i\varphi} \\ \sin \theta e^{i\varphi} & -\cos \theta \end{pmatrix}$, where θ is the polar and φ is the azimuthal angle. Given this, the eigenvectors of $a \cdot \sigma_A$ are $|+a\rangle = \begin{pmatrix} \cos \frac{\theta}{2} \\ \sin \frac{\theta}{2} e^{i\varphi} \end{pmatrix}$, $|-a\rangle = \begin{pmatrix} \sin \frac{\theta}{2} \\ -\cos \frac{\theta}{2} e^{i\varphi} \end{pmatrix}$.

Then,

$$|\langle b_+ | a_+ \rangle|^2 = \left| \cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} + \sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} e^{i(\varphi_a - \varphi_b)} \right|^2 = \left| \cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} + \sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \cos(\varphi_a - \varphi_b) + i \sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \sin(\varphi_a - \varphi_b) \right|^2$$

where we use $e^{ix} = \cos x + i \sin x$. Then,

$$\begin{aligned} |\langle b_+ | a_+ \rangle|^2 &= \left(\cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} + \sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \cos(\varphi_a - \varphi_b) \right)^2 + \left(\sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \sin(\varphi_a - \varphi_b) \right)^2 \\ &= \left(\cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \right)^2 + \left(\sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \cos(\varphi_a - \varphi_b) \right)^2 \\ &\quad + 2 \cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \cos(\varphi_a - \varphi_b) + \left(\sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \sin(\varphi_a - \varphi_b) \right)^2 \\ &= \left(\cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \right)^2 + \left(\sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \right)^2 (\cos^2(\varphi_a - \varphi_b) + \sin^2(\varphi_a - \varphi_b)) \\ &\quad + 2 \cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \cos(\varphi_a - \varphi_b) \\ &= \left(\cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \right)^2 + \left(\sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \right)^2 + 2 \cos \frac{\theta_a}{2} \cos \frac{\theta_b}{2} \sin \frac{\theta_a}{2} \sin \frac{\theta_b}{2} \cos(\varphi_a - \varphi_b) \end{aligned}$$

We now need to employ some trigonometric identities: $\left(\cos \frac{\theta}{2} \right)^2 = \frac{1+\cos \theta}{2}$, $\left(\sin \frac{\theta}{2} \right)^2 = \frac{1-\cos \theta}{2}$, $\sin \theta = 2 \sin \frac{\theta}{2} \cos \frac{\theta}{2}$, $\cos(\varphi_a - \varphi_b) = \cos(\varphi_a) \cos(\varphi_b) + \sin(\varphi_a) \sin(\varphi_b)$, $\sin \theta = 2 \sin \frac{\theta}{2} \cos \frac{\theta}{2}$

$$\begin{aligned} |\langle b_+ | a_+ \rangle|^2 &= \frac{1+\cos \theta_a}{2} \frac{1+\cos \theta_b}{2} + \frac{1-\cos \theta_a}{2} \frac{1-\cos \theta_b}{2} \\ &\quad + \frac{1}{2} \sin \theta_a \sin \theta_b (\cos(\varphi_a) \cos(\varphi_b) + \sin(\varphi_a) \sin(\varphi_b)) \\ &= \frac{1}{4} (1 + \cos \theta_a + \cos \theta_b + \cos \theta_a \cos \theta_b) + \frac{1}{4} (1 - \cos \theta_a - \cos \theta_b + \cos \theta_a \cos \theta_b) \\ &\quad + \frac{1}{2} \sin \theta_a \sin \theta_b (\cos(\varphi_a) \cos(\varphi_b) + \sin(\varphi_a) \sin(\varphi_b)) \\ &= \frac{1}{2} (1 + \cos \theta_a \cos \theta_b + \sin \theta_a \sin \theta_b \cos(\varphi_a) \cos(\varphi_b) + \sin \theta_a \sin \theta_b \sin(\varphi_a) \sin(\varphi_b)) \end{aligned}$$

We next have to use next the identities $a_x = \sin \theta \cos \varphi$, $a_y = \sin \theta \sin \varphi$, $a_z = \cos \theta$

$$|\langle b_+ | a_+ \rangle|^2 = \frac{1}{2} (1 + a_z b_z + a_x b_x + a_y b_y) = \frac{1}{2} (1 + \cos \theta_{ab}) = \cos^2 \frac{\theta_{ab}}{2}$$

where θ_{ab} is the angle between directions a, b .

So, as required, we have

$$\text{Prob}(+-|a, b; \psi_-) = \frac{1}{2} \cos^2 \left(\frac{\theta}{2} \right) = \text{Prob}(-+|a, b; \psi_-)$$

In order to compute $|\langle b_+ | a_- \rangle|^2$ we note again $|+a\rangle = \begin{pmatrix} \cos \frac{\theta}{2} \\ \sin \frac{\theta}{2} e^{i\varphi} \end{pmatrix}$, $|-a\rangle = \begin{pmatrix} \sin \frac{\theta}{2} \\ -\cos \frac{\theta}{2} e^{i\varphi} \end{pmatrix}$, so that

$$|\langle b_+ | a_- \rangle|^2 = \left| \cos \frac{\theta_a}{2} \sin \frac{\theta_b}{2} - \sin \frac{\theta_a}{2} \cos \frac{\theta_b}{2} e^{i(\varphi_a - \varphi_b)} \right|^2$$

Which then gives us

$$\text{Prob}(- - | a, b; \psi_-) = \frac{1}{2} \sin^2\left(\frac{\theta}{2}\right) = \text{Prob}(+ + | a, b; \psi_-)$$

With a quick ‘sanity’ check on the above calculations, namely if $\theta = 0$ then we have the same direction of measurement for both particles. However, such an angle gives us $\text{Prob}(+ + | a, b; \psi_-) = 0$, consistent with the assumption, since the two sub-systems are meant to be anti-correlated.

Finally, we can use the above to compute the correlator for the S quantity for the quantum model. Note, first of all, the definition of the expected value of an observable in quantum theory, if a physical quantity A and a state of the system are represented respectively by the self-adjoint operator A and the normalized vector $\psi \in H$, then the expected value $\langle A_\psi \rangle$ of A is $\langle A \rangle_\psi = \langle \psi, A\psi \rangle$.

The observable $a \cdot \sigma_A \otimes b \cdot \sigma_B$ has a spectral decomposition

$$a \cdot \sigma_A \otimes b \cdot \sigma_B = \sum_i \lambda_i P_i = P_+ \otimes P_+ - P_+ \otimes P_- - P_- \otimes P_+ + P_- \otimes P_-$$

The expectation value of observable $a \cdot \sigma_A \otimes b \cdot \sigma_B$ is then given by,

$$\begin{aligned} \psi_- | a \cdot \sigma_A \otimes b \cdot \sigma_B | \psi_- &= \psi_- | P_+ \otimes P_+ - P_+ \otimes P_- - P_- \otimes P_+ + P_- \otimes P_- | \psi_- \\ &= \text{Prob}(+ + | a, b; \psi_-) - \text{Prob}(+ - | a, b; \psi_-) - \text{Prob}(- + | a, b; \psi_-) + \text{Prob}(- - | a, b; \psi_-) \\ &= \frac{1}{2} \sin^2\left(\frac{\theta}{2}\right) - \frac{1}{2} \cos^2\left(\frac{\theta}{2}\right) - \frac{1}{2} \cos^2\left(\frac{\theta}{2}\right) + \frac{1}{2} \sin^2\left(\frac{\theta}{2}\right) = \sin^2\left(\frac{\theta}{2}\right) - \cos^2\left(\frac{\theta}{2}\right) \\ &= \frac{1 - \cos \theta}{2} - \frac{1 + \cos \theta}{2} = -\cos \theta = -a \cdot b \end{aligned}$$

The interpretation of this expectation value is that it is the average value of the product of outcome value for the first sub-system times outcome value for the second sub-system, where outcome for the first sub-system is measured along a and outcome for the second sub-system is measured along b.

References

1. Broekaert, J.; Busemeyer, J.; Pothos, E. The disjunction effect in two-stage simulated gambles. An experimental study and comparison of a heuristic logistic, Markov and quantum-like model. *Cogn. Psychol.* **2020**, *117*, 101262. [\[CrossRef\]](#)
2. Chater, N.; Vlaev, I.; Grinberg, M. A new consequence of Simpson’s paradox: Stable cooperation in one-shot prisoner’s dilemma from populations of individualistic learners. *J. Exp. Psychol. Gen.* **2008**, *137*, 403–421. [\[CrossRef\]](#)
3. Pothos, E.M.; Busemeyer, J. A quantum probability explanation for violations of ‘rational’ decision theory. *Proc. R. Soc. B Biol. Sci.* **2009**, *276*, 2171–2178. [\[CrossRef\]](#)
4. Shafir, E.; Tversky, A. Thinking through uncertainty: Nonconsequential reasoning and choice. *Cogn. Psychol.* **1992**, *24*, 449–474. [\[CrossRef\]](#)
5. Griffiths, T.L.; Chater, N.; Kemp, C.; Perfors, A.; Tenenbaum, J.B. Probabilistic models of cognition: Exploring representations and inductive biases. *Trends Cogn. Sci.* **2010**, *14*, 357–364. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Khrennikov, A. On quantum-like probabilistic structure of mental information. *Open Syst. Inf. Dyn.* **2004**, *11*, 267–275. [\[CrossRef\]](#)
7. Oaksford, M.; Chater, N. Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behav. Brain. Sci.* **2009**, *32*, 69–84. [\[CrossRef\]](#)
8. Tenenbaum, J.B.; Kemp, C.; Griffiths, T.L.; Goodman, N.D. How to grow a mind: Statistics, structure, and abstraction. *Science* **2011**, *331*, 1279–1285. [\[CrossRef\]](#)
9. Atmanspacher, H.; Filk, T. Contextuality revisited: Signaling may differ from communicating. In *Synthese Library*; Springer: Cham, Switzerland, 2019; pp. 117–127.
10. Bell, J.S. On the einstein-podolsky rosen paradox. *Phys. Phys. Fiz.* **1964**, *1*, 195–200. [\[CrossRef\]](#)
11. Bell, J.S.; Horne, M.A.; Zeilinger, A. Speakable and unspeakable in quantum mechanics. *Am. J. Phys.* **1989**, *57*, 567. [\[CrossRef\]](#)
12. Clauser, J.F.; Horne, M.A.; Shimony, A.; Holt, R.A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **1969**, *23*, 880–884. [\[CrossRef\]](#)
13. Popescu, S.; Rohrlich, D. Quantum nonlocality as an axiom. *Found. Phys.* **1994**, *24*, 379–385. [\[CrossRef\]](#)
14. Fine, A. Joint distributions, quantum correlations, and commuting observables. *J. Math. Phys.* **1982**, *23*, 1306–1310. [\[CrossRef\]](#)
15. Dzhafarov, E.N.; Zhang, R.; Kujala, J.V. Is there contextuality in behavioural and social systems? *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150099. [\[CrossRef\]](#)

16. Adenier, G.; Khrennikov, A.Y. Test of the no-signaling principle in the Hensen loophole-free CHSH experiment. *Fortschr. Phys.* **2017**, *65*, 1600096. [CrossRef]
17. Aerts, D. Example of a macroscopical classical situation that violates Bell inequalities. *Lett. Nuovo Cimento* **1982**, *34*, 107–111. [CrossRef]
18. Toner, B.F.; Bacon, D. Communication cost of simulating bell correlations. *Phys. Rev. Lett.* **2003**, *91*, 187904. [CrossRef]
19. Aerts, S. A Realistic Device that Simulates the Non-Local PR Box without Communication. 2005. Available online: <http://arxiv.org/abs/quant-ph/0504171> (accessed on 20 September 2021).
20. Aerts, D.; Arguëlles, J.A.; Beltran, L.; Geriente, S.; De Bianchi, M.S.; Sozzo, S.; Veloz, T. Spin and wind directions I: Identifying entanglement in nature and cognition. *Found. Sci.* **2018**, *23*, 323–335. [CrossRef] [PubMed]
21. Aerts, D. Quantum and concept combination, entangled measurements, and prototype theory. *Top. Cogn. Sci.* **2014**, *6*, 129–137. [CrossRef] [PubMed]
22. Conte, E.; Khrennikov, A.; Todarello, O.; Federici, A.; Zbilut, J.P. A preliminary experimental verification on the possibility of bell inequality violation in mental states. *NeuroQuantology* **2008**, *6*, 214–221. [CrossRef]
23. Bruza, P.D.; Kitto, K.; Ramm, B.J.; Sitbon, L. A probabilistic framework for analysing the compositionality of conceptual combinations. *J. Math. Psychol.* **2015**, *67*, 26–38. [CrossRef]
24. Aerts, D.; Sozzo, S.; Veloz, T. New fundamental evidence of non-classical structure in the combination of natural concepts. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150095. [CrossRef]
25. Hampton, J.A. Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *J. Exp. Psychol. Learn. Mem. Cogn.* **1988**, *14*, 12–32. [CrossRef]
26. Osherson, D.N.; Smith, E.E. On the adequacy of prototype theory as a theory of concepts. *Cognition* **1981**, *9*, 35–58. [CrossRef]
27. Storms, G.; De Boeck, P.; Hampton, J.A.; Van Mechelen, I. Predicting conjunction typicalities by component typicalities. *Psychon. Bull. Rev.* **1999**, *6*, 677–684. [CrossRef]
28. Bruza, P.; Kitto, K.; Nelson, D.; McEvoy, C. Is there something quantum-like about the human mental lexicon? *J. Math. Psychol.* **2009**, *53*, 362–377. [CrossRef] [PubMed]
29. Nelson, D.L.; McEvoy, C. Entangled associative structures and context. In *AAAI Spring Symposium: Quantum Interaction*; The AAAI Press: Menlo Park, CA, USA, 2007; pp. 98–105.
30. Basieva, I.; Cervantes, V.H.; Dzhamalov, E.N.; Khrennikov, A. True contextuality beats direct influences in human decision making. *J. Exp. Psychol. Gen.* **2019**, *148*, 1925–1937. [CrossRef]
31. Cirel'son, B.S. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.* **1980**, *4*, 93–100. [CrossRef]
32. Blasiak, P.; Pothos, E.M.; Yearsley, J.M.; Gallus, C.; Borsuk, E. Violations of locality and free choice are equivalent resources in Bell experiments. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, 118. [CrossRef] [PubMed]
33. Busemeyer, J.R.; Bruza, P. *Quantum Models of Cognition and Decision Making*; Cambridge University Press: Cambridge, UK, 2011.
34. Pothos, E.M.; Busemeyer, J.R.; Trueblood, J.S. A quantum geometric model of similarity. *Psychol. Rev.* **2013**, *120*, 679–696. [CrossRef]
35. Haven, E.; Khrennikov, A. *Quantum Social Science*; Cambridge University Press: Cambridge, UK, 2013.
36. Schwarz, N. Attitude construction: Evaluation in context. *Soc. Cogn.* **2007**, *25*, 638–656. [CrossRef]
37. Kvam, P.D.; Pleskac, T.J.; Yu, S.; Busemeyer, J. Interference effects of choice on confidence: Quantum characteristics of evidence accumulation. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10645–10650. [CrossRef]
38. White, L.C.; Pothos, E.M.; Jarrett, M. The cost of asking: How evaluations bias subsequent judgments. *Decision* **2020**, *7*, 259–286. [CrossRef]
39. Wang, Z.; Solloway, T.; Shiffrin, R.M.; Busemeyer, J. Context effects produced by question orders reveal quantum nature of human judgments. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 9431–9436. [CrossRef]
40. Yearsley, J.; Pothos, E.M. Challenging the classical notion of time in cognition: A quantum perspective. *Proc. R. Soc. B Biol. Sci.* **2014**, *281*, 20133056. [CrossRef]
41. Kellen, D.; Singmann, H.; Batchelder, W.H. Classic-probability accounts of mirrored (quantum-like) order effects in human judgments. *Decision* **2018**, *5*, 323–338. [CrossRef]
42. Yearsley, J.M.; Trueblood, J.S. A quantum theory account of order effects and conjunction fallacies in political judgments. *Psychon. Bull. Rev.* **2017**, *25*, 1517–1525. [CrossRef] [PubMed]
43. Vlaev, I.; Chater, N. Game relativity: How context influences strategic decision making. *J. Exp. Psychol. Learn. Mem. Cogn.* **2006**, *32*, 131–149. [CrossRef] [PubMed]
44. Spreng, R.N.; McKinnon, M.C.; Mar, R.A.; Levine, B. The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *J. Pers. Assess.* **2009**, *91*, 62–71. [CrossRef] [PubMed]
45. Greco, V.; Roger, D. Coping with uncertainty: The construction and validation of a new measure. *Pers. Individ. Differ.* **2001**, *31*, 519–534. [CrossRef]
46. Frederick, S. Cognitive reflection and decision making. *J. Econ. Perspect.* **2005**, *19*, 25–42. [CrossRef]
47. Primi, C.; Morsanyi, K.; Chiesi, F.; Donati, M.A.; Hamilton, J. The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *J. Behav. Decis. Mak.* **2016**, *29*, 453–469. [CrossRef]
48. Stanislaw, H.; Todorov, N. Calculation of signal detection theory measures. *Behav. Res. Meth. Instrum. Comput.* **1999**, *31*, 137–149. [CrossRef] [PubMed]

-
49. Pothos, E.M.; Lewandowsky, S.; Basieva, I.; Barque-Duran, A.; Tapper, K.; Khrennikov, A. Information overload for (bounded) rational agents. *Proc. R. Soc. B Biol. Sci.* **2021**, *288*, 20202957. [[CrossRef](#)] [[PubMed](#)]
 50. Fernbach, P.M.; Sloman, S.A. Causal learning with local computations. *J. Exp. Psychol. Learn. Mem. Cogn.* **2009**, *35*, 678–693. [[CrossRef](#)] [[PubMed](#)]
 51. Lewandowsky, S.; Kalish, M.; Ngang, S.K. Simplified learning in complex situations: Knowledge partitioning in function learning. *J. Exp. Psychol. Gen.* **2002**, *131*, 163–193. [[CrossRef](#)] [[PubMed](#)]
 52. Aerts, D. An attempt to imagine parts of the reality of the micro-world. In *Problems in Quantum Physics II*; Mizerski, J., Posiewnik, A., Pykacz, J., Zukowski, M., Eds.; World Scientific Publishing Company: Singapore, 1990; pp. 3–25.
 53. Nielsen, M.A.; Chuang, I.L. *Quantum Computation and Quantum Information*; Cambridge University Press: New York, NY, USA, 2000.