Alan Turing has certainly contributed to a widespread belief that the quest for a perfect, unbreakable, cipher is a futile pursuit. The ancient art of concealing information has, in the past, been matched by the ingenuity of code-breakers, but no longer! With the advent of quantum cryptography, the hopes of would-be eavesdroppers have been dashed, perhaps for good. Moreover, recent research, building on schemes that were invented decades ago to perform quantum cryptography, shows that secure communication certified by a sufficient violation of a Bell inequality makes a seemingly insane scenario possible—devices of unknown or dubious provenance, even those that are manufactured by our enemies, can be safely used for secure communication, including key distribution. All that is needed to implement this bizarre and powerful form of cryptography is a loophole-free test of a Bell inequality, which is on the cusp of technological feasibility. We provide a brief overview of the intriguing connections between Bell inequalities and cryptography and describe how studies of quantum entanglement and the foundations of quantum theory influence the way we may protect information in the future.

Keywords: quantum cryptography; key distribution; non-locality

1. Introduction

Alan Turing’s first interest in mathematics came through physics. According to his biographer, Andrew Hodges, it was not the beauty of mathematical abstractions but rather the wonders of quantum physics that fired Turing’s imagination and creativity. Strongly influenced by the popular writings of an eminent Cambridge physicist, Arthur Eddington, a 16-year-old Turing became fascinated by quantum-mechanical indeterminism as a possible basis of free will [1]. The subject remained a recurring theme throughout his life. His early attempts to understand how a collection of ordinary atoms could become a thinking machine led him eventually to his seminal contributions to mathematics and computer science.

*Author for correspondence (j.pope1@physics.ox.ac.uk).

One contribution of 18 to a Theme Issue ‘The foundations of computation, physics and mentality: the Turing legacy’.
Judging from Turing’s occasional remarks and comments, quantum theory, with its inherent randomness, must have fascinated him. Like Albert Einstein, he found the standard Copenhagen interpretation, with its ridiculous reduction of the state vector, simply unacceptable. In 1953, shortly before his tragic death, Turing wrote to his friend and student Robin Gandy: ‘...I’m trying to invent a new quantum mechanics but it won’t really work. How about coming here next week and making it work for me?’

Today, sixty years later, there is no new quantum physics; we still rely on the same mathematical formalism, but our understanding of quantum phenomena and our ability to take advantage of them have greatly improved. One can only speculate whether Turing would find the current state of quantum physics up to his satisfaction, but we can be pretty certain that he would be pleased with its unexpected fusion with cryptography. This paper outlines the latest developments in this field and shows how they make a seemingly insane scenario possible—devices of unknown or dubious provenance, even those that are manufactured by our enemies, can be safely used for secure communication. This is a truly remarkable feat, also referred to as ‘device-independent cryptography’. The final twist to the story of secure communication would probably please Turing even more—it turns out that secure communication is possible even if the communicating parties are manipulated and their free will is somewhat limited.

2. Violating a Bell inequality

Our interests also originate from trying to describe the physical world. In order for this world to evolve in a fully definite, fully predictable manner, only certain types of correlations are permitted. The argument, originally proposed by Bell [2], and subsequently slightly modified by Clauser et al. [3], is deceptively simple and can be explained to anyone who knows something about probability but has never come across quantum physics. This is because the statement has nothing to do with quantum physics—it is all about the properties of classical correlations.

Alice and Bob are equipped with polarization analysers and sent to two distant locations. Somewhere in between them there is a source that emits pairs of photons that fly apart, one towards Alice and one towards Bob. We label the two photons as $A$ and $B$, respectively. If we do not know anything about quantum physics, we would assume that whatever measurements Alice and Bob might make, the results are already defined.$^1$ We ask Alice and Bob to each measure one of two pre-agreed polarizations. For every incoming photon, Alice and Bob choose randomly, and independently from each other, which particular polarization will be measured. Alice chooses between the bits 0 and 1 specifying each of the polarization measurements, and records the corresponding outcome $A_0$ or $A_1$, similarly for Bob recording either $B_0$ or $B_1$. Each polarization has value $+1$ or $-1$ (more specifically, there are only two possible values which we can map to $\pm 1$), thus we are allowed to think about them as random variables $A_j$ and $B_k$, $j, k = 0, 1$, which take values $\pm 1$. Let us define a new random variable $S$,

$$S = A_0(B_0 + B_1) + A_1(B_0 - B_1).$$

$^1$At the very least, the answers would be described by a probability distribution.

Phil. Trans. R. Soc. A (2012)
It is easy to see that one of the terms $B_0 \pm B_1$ must be equal to zero and the other to $\pm 2$, hence $S = \pm 2$. The average value of $S$ must lie somewhere in between, i.e.

$$-2 \leq \langle S \rangle \leq 2.$$  \hspace{1cm} (2.2)

That’s it! This result, named after its founders as the Clauser–Horne–Shimony–Holt (CHSH) inequality [3], is a simple mathematical statement about correlations and yet is so profound. No quantum theory is involved, and no specific physical process is invoked. This is simply a statement of the properties of correlations that can appear between the binary values of $A_0$, $A_1$, $B_0$ and $B_1$ if those results are pre-determined.

In fact, instead of photons and polarization analysers Alice and Bob may each be given sealed, impregnable boxes. The inner workings of the boxes are unknown but the exterior design is simple—two buttons to press, and two light bulbs labelled $+1$ and $-1$. Alice and Bob can then take the boxes into their respective locations, select one of the two buttons to press and watch the boxes responding with flashes of light. For example, for the first reading, Alice and Bob may record outcomes $A_0$ and $B_1$, respectively. They do this by pressing the corresponding buttons and the boxes generate outcomes, say, Alice’s box flashes $+1$ and Bob’s $-1$. They record the settings and the results, and repeat the process until they accumulate a sufficiently large amount of data to evaluate relative frequencies of different outcomes, and hence $\langle S \rangle$. In fact, a Bell test is not limited to the CHSH inequality. There is an entire class of Bell inequalities where the number of experimenters, inputs and outputs exceed two of each, and what they all have in common is a real bound that must not be exceeded if we treat the measurement outcomes as classical random variables. The boxes may respond in a correlated manner, the correlations can be estimated, and a Bell inequality can be checked. Technical details of the hardware are irrelevant. The focus is on correlations alone. Why, then, the interest? Surprisingly enough, there are correlations in Nature that violate Bell inequalities.

So, what does it take to violate the CHSH inequality (2.2)? We know that a violation is inconsistent with assigning numerical values to $A$ and $B$ prior to these values being actually registered. This said, we still may be able to assign numerical values to correlations $\langle AB \rangle$. If we take this unorthodox approach, then the expression

$$\langle S \rangle = \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle$$

(2.3)

admits $\pm 4$ as its two extreme values. Predicated on the tenet of relativity that forbids the possibility of superluminal signalling, which is to say that making a measurement on one particle cannot affect the properties of any particles outside of the particle’s future light-cone, we must let the results registered by Alice and Bob be completely random and unpredictable to achieve maximal violation, though they may be correlated. For example, in a purely abstract, hypothetical, device introduced by Tsirelson [4] and commonly referred to as a $P–R$ box due to further discussion by Popescu & Rohrlich [5], the values of $A_1$ and $B_1$ when measured together are random but always different from each other, $A_1 B_1 = -1$, and all the remaining variables are random but always identical, $A_0 B_0 = A_0 B_1 = A_1 B_0 = 1$, yielding $\langle S \rangle = 4$. This inherent randomness is mind-boggling. If things just happen and numerical values appear out of the blue then they did not ‘exist’ prior to measurement.

*Phil. Trans. R. Soc. A* (2012)
Nature, it seems, embraces randomness and permits correlations that violate the CHSH inequality but the violation is a modest one, far short of $\langle S \rangle = 4$. This has been observed in a number of painstaking experiments [6–9]. Today, however, such experiments are routine. For example, in a process called ‘parametric down conversion’ a photon from a laser beam enters a beta-barium–borate crystal and gets absorbed while it excites an atom in the crystal. The atom subsequently decays, emitting two ‘polarization-entangled’ photons, so that if the polarization analysers $A$ and $B$ are set $\theta$ degrees apart, then the results agree ($AB = 1$) with probability $\sin^2 \theta$ and hence differ ($AB = -1$) with probability $\cos^2 \theta$. This experimental fact is consistent with quantum mechanical predictions, assuming that the photons are prepared in the so-called singlet state. This gives the correlation coefficient $\langle AB \rangle = \sin^2 \theta - \cos^2 \theta = -\cos 2\theta$. Correlations of this kind cannot be used to send instantaneous messages but they do violate the CHSH inequality. Choose angles $0$, $\pi/4$, $\pi/8$ and $3\pi/8$ for $A_0$, $A_1$, $B_0$ and $B_1$, respectively, and you obtain $|\langle S \rangle| = 2\sqrt{2}$.

We show in the next section the quantum-mechanical justification that this is the maximal violation of the CHSH inequality that quantum correlations can offer, based on a proof by Tsirelson [4]. Moreover, any physical realization of such a maximal violation is effectively equivalent to performing measurements on pairs of maximally entangled qubits, e.g. two maximally entangled photons or spin one-half particles. These experimental data allow us to conclude that as long as the measurements are selected randomly, the outcomes observed are also random. We may find some consolation in harnessing this randomness and putting it to good use. This, finally, brings us to cryptography.

The first practical application of Bell inequalities was in the art of secret communication [10]. With the benefit of hindsight it is not very surprising. Information is always represented by measurable physical properties and, if such properties exist before they are measured, then measurement of these properties cannot change their values. Such a property was termed by Einstein as an ‘element of reality’ [11] and its value can be predicted with certainty ‘...without in any way disturbing a system...’. This is just a description of perfect eavesdropping. Conversely, if such properties do not exist prior to measurements (as proved by violation of a Bell inequality), then there is nothing to eavesdrop on. This was the basic idea that led to the development of a new tool for detecting eavesdropping.

### 3. A key distribution protocol

While the security of all modern-day ciphers is only conditional on the limitations of an eavesdropper with finite resources, Claude Shannon had already proved in 1949 that use of a one-time pad, wherein a perfectly random binary string as long as the message to be encrypted is added in base 2 to the message, gives unconditional security provided the string is used only once [12]. Despite this incredible power, there is a practical issue that arises when using the one-time pad; both parties in the encryption protocol must know the key in order to both encrypt and decrypt the message. Therefore, the problem to be concerned with is *key distribution*, where extremely long binary keys must not only be generated efficiently between Alice and Bob but also kept secret from any adversaries. Either Alice and Bob must meet in advance and share a sufficiently long random
key, or they must generate this key as and when it is needed. Whichever they choose, the paranoid can always ask ‘how can we know that an adversary has not compromised the key?’, to which, for classical information, there is no answer. Below we describe a key distribution protocol which provides a proof (subject to certain limited assumptions) that a distributed key really is secure, by way of violating a Bell inequality [10].

Alice and Bob randomly select measurements to perform which generate outcomes from a choice of \( A_0, A_1, B_0 \) and \( B_1 \). Alice may also choose to perform a third measurement with outcome \( A_2 \). The two repeatedly sample the devices and make a record of their measurement inputs and outputs for each run—see figure 1a. After a sufficient number of runs, they publicly announce their choice of measurement inputs, and collect statistical data as follows.

**Determining security.** Alice and Bob reveal all of their outcome pairs \((A_0, B_0), (A_0, B_1), (A_1, B_0)\) and \((A_1, B_1)\). They calculate the relative frequencies of each \( A_i B_j \) and determine the value of \( \langle S \rangle \) described by (2.3). If a violation of the CHSH inequality is observed, then the remaining results, which were recorded but not communicated in public, remain secret. They were never read by anyone, because before they were registered they did not exist. These remaining results can now be used to generate the key. However, if a violation of the CHSH inequality is not observed, Alice and Bob abort the protocol, discard their results and try again later.

**Key generation.** Whenever Alice records the outcome \( A_2 \) and Bob \( B_0 \), the outcomes are used to generate the key. We define the bit error rate to be \( Q = P(A_2 = B_0) \) and this is determined by calculating the relative frequency of correlated bits over a small publicly announced portion of the collected data.

*Phil. Trans. R. Soc. A* (2012)
The case $Q = 0$ corresponds to Alice and Bob generating perfectly anticorrelated data. This is only possible with classical devices if Alice and Bob have a predetermined strategy for generating these outcomes, and the elements of reality in the $B_0$ outcomes will prevent the CHSH inequality from being violated. However, anticorrelated yet locally random outcomes can be obtained by devices performing quantum measurements on shared states, for instance, if the outcomes $A_2$ and $B_0$ are owing to identical measurements performed on a shared singlet state $\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. In this case, Bob simply flips all bits in his keystring to obtain an identical key to Alice’s. If the bit error rate is not 0, an identical keystring can still be obtained, with $Q$ acting as an indicator for how much error correction is needed.

No matter how an eavesdropper, Eve, attempts to subvert the protocol, she could not pre-determine the results registered by Alice and Bob. We know this, because if the results had existed in the program in any form, be it directly or as a computational procedure, then the Bell inequality would have been satisfied, we would not have obtained a violation, and would have aborted the protocol. Furthermore, despite not being used for testing the inequality, she could not have pre-programmed values for the outcomes $A_2$. This is because Eve cannot pre-program outcomes for $B_0$ lest they be subject to a CHSH test, thus they are random and the outcomes for $A_2$ must also be random to achieve sufficient anticorrelation.

So how come Eve, who has control upon just about everything, cannot learn the key? Surely there must be something upon which she has no control, something that is essential for the security of the whole scheme? Indeed, there is, and this something is called ‘free will’. Each binary outcome registered by Alice is obtained for a randomly chosen setting of the knob; her choice is free and independent from that of Bob. The same holds for Bob. If Alice’s and Bob’s choices are known in advance, then Eve can easily pre-program the results of the pre-determined measurements, so that the CHSH inequality is violated and Alice and Bob foolishly believe that they generated a secret key. As with the bit error rate, the key is to try and identify how much control Eve has over the free will of Alice and Bob, and hence compensate for it.

The fact that the conclusions drawn from a Bell inequality are independent from how exactly the recorded results were generated takes cryptography to an entirely different level, even when compared with the quantum ‘prepare and measure cryptography’ [13]. Although the key distribution protocol is basically the same as the one proposed some time ago [10], more recent work first proposed by Mayers & Yao [14] and further developed by Acín et al. [15] gives it an entirely new twist. It shows that the original protocol is in fact much more powerful than originally anticipated. It makes a seemingly insane scenario possible—devices of unknown or dubious provenance, even those that are manufactured by our enemies, can safely be used for secure key distribution. This is a truly remarkable feat, also referred to as the device-independent key distribution (see figure 1b). Can we implement it? Yes! But there are several questions of practical nature that must be addressed before we can push cryptography to the device-independent limit.

*Recall that measurements by Alice and Bob on a singlet state give opposite results if $\theta = 0$. 

Phil. Trans. R. Soc. A (2012)
Does any violation of the CHSH inequality guarantee secrecy? If quantum theory is all that there is, then the maximal violation allowed, \(|\langle S \rangle| = 2\sqrt{2}\), implies both perfect randomness and perfect security. Anything less than that may contain corrupted correlations, in which some bits may be known to an adversary. On the other hand, anything more than that (beyond plausible statistical fluctuations) should immediately send alarm bells ringing, as this suggests Alice’s and Bob’s measurement choices are not perfectly random, and could be correlated with Eve’s pre-programmed outputs. Still, as long as the violation lies between 2 and \(2\sqrt{2}\), assuming perfect free will, we can estimate how many bits are compromised and use a number of cryptographic techniques, such as error correction combined with hashing, to distill a secret key. As one may expect, the amount of information available to the adversary is related directly to the degree of violation of a Bell inequality \([15]\), as shown below. We assume here that Eve’s strategy is restricted to collective attacks where she acts independently on each round of testing.

Consider the CHSH inequality in operational terms—instead of treating the outcomes \(A_0, A_1, B_0\) and \(B_1\) that Alice and Bob may obtain as random variables, we consider them to be measurement operators, determining the measurement outcomes on a quantum state. Their eigenvalues \(\pm 1\) correspond to the \(\pm 1\) outcomes of the random variables that we had before. Each such measurement can be represented by a three-dimensional unit vector on the Bloch sphere, \(\hat{a}_0, \hat{a}_1, \hat{b}_0\) and \(\hat{b}_1\), respectively, so that \(A_0 = \hat{a}_0 \cdot \sigma\) where \(\sigma\) is a vector of Pauli matrices. In the device-independent scenario, Eve pre-selects these measurements as well as the quantum state to be distributed between Alice and Bob on each round (for the CHSH scenario, we may restrict the shared state to being two entangled qubits), our aim being to calculate the largest possible violation \(S_{\text{max}}\), defined to be the maximum of \(\langle \psi | S | \psi \rangle\) over all two-qubit states \(|\psi\rangle\) and all measurements \(\hat{a}_0, \hat{a}_1, \hat{b}_0, \hat{b}_1\), where \(S\) is the CHSH operator

\[
S = (\hat{a}_0 \cdot \sigma) \otimes (\hat{b}_0 + \hat{b}_1) \cdot \sigma + (\hat{a}_1 \cdot \sigma) \otimes (\hat{b}_0 - \hat{b}_1) \cdot \sigma.
\]  

By unitary freedom, Eve can select \(|\psi\rangle = \sin \theta |00\rangle + \cos \theta |11\rangle\) where \(0 \leq \theta < \pi/4\). Rewriting \(|\psi\rangle\) in the form of eqn (1) in \([16]\), we see that

\[
S_{\text{max}} = \max_{\hat{a}_0, \hat{a}_1, \hat{b}_0, \hat{b}_1} (\hat{a}_0 \cdot t \cdot (\hat{b}_0 + \hat{b}_1) + \hat{a}_1 \cdot t \cdot (\hat{b}_0 - \hat{b}_1)),
\]  

where \(t\) is a \(3 \times 3\) diagonal matrix with diagonal entries \((\sin 2\theta, -\sin 2\theta, 1)\). Now define the unit vectors \(\hat{c}, \hat{c}'\) and scalars \(l, l'\) by

\[
\hat{b}_0 + \hat{b}_1 = l\hat{c},
\]  

and

\[
\hat{b}_0 - \hat{b}_1 = l'\hat{c}'.
\]

Taking the dot product of (3.3) and (3.4) reveals that \(\hat{c}\) and \(\hat{c}'\) are orthogonal. Squaring them yields \(l^2 + l'^2 = 4\), suggesting we write \(l = 2\cos \phi\) and \(l' = 2\sin \phi\).
for some \( \phi \). We then ascertain the maximum achievable CHSH value to be

\[
S_{\text{max}}(\theta) = \max_{a_0, a_1, c, c'} 2 (\hat{a}_0 \cdot t \cdot \hat{c} \cos \phi + \hat{a}_1 \cdot t \cdot c' \sin \phi)
\]

\[
= \max_{\hat{c}, c'} 2 (\| t \cdot \hat{c} \| \cos \phi + \| t \cdot c' \| \sin \phi)
\]

\[
= \max_{\hat{c}, c'} 2 \sqrt{\| t \cdot \hat{c} \|^2 + \| t \cdot c' \|^2}
\]

\[
= 2 \sqrt{1 + \sin^2 2\theta},
\]

where the steps for maximization are choosing \( \hat{a}_0 \) and \( \hat{a}_1 \) to be parallel to \( t \cdot \hat{c} \) and \( t \cdot c' \), respectively, in (3.5), defining \( \phi \) in (3.6) by \( \tan \phi = \| t \cdot c' \| / \| t \cdot \hat{c} \| \), and selecting \( \hat{c} \) and \( c' \) in (3.7) as the eigenvectors of \( t \) with largest eigenvalues (remembering \( \hat{c} \) and \( c' \) must be orthogonal). Note that this maximization has selected Alice’s measurements to be the Pauli operations \( A_0 = Z \) and \( A_1 = X \). Equation (3.8) is further maximized by selecting \( \theta = \pi/4 \), yielding the familiar

\[
S_{\text{max}} = 2 \sqrt{2},
\]

but we shall see that varying \( \theta \) allows us to find a trade-off between \( S_{\text{max}}(\theta) \) and the probability with which an eavesdropper can guess the measurement outcome.

The guessing probability \( G \) is the maximum probability that Eve guesses the outcome of the key bit generated by the pair \((A_0, B_2)\) (in a key distribution protocol that differs from the above one only in that Bob, not Alice, has the third measurement setting). Since the measurements have only \( \pm 1 \) outcomes, we have that

\[
G = \frac{1}{2} \left( 1 + \text{Tr}(A_0 \rho_A) \right),
\]

where \( \rho_A \) is the reduced density matrix on Alice, \( \rho_A = \cos^2 \theta |0 \rangle \langle 0| + \sin^2 \theta |1 \rangle \langle 1| \). Choosing \( A_0 = Z \) gives the highest expectation \( \text{Tr}(\rho_A A_0) = \cos^2 \theta - \sin^2 \theta \), yielding \( G = \cos^2 \theta \). Since the optimal strategies for calculating \( G \) and maximizing \( S \) are compatible by selecting \( A_0 = Z \), the maximum guessing probability has therefore been bounded above by

\[
G \leq \frac{1}{2} \left( 1 + \sqrt{2 - \frac{S^2}{4}} \right).
\]

Alice and Bob can assume this worst-case scenario and still guarantee a private key from the protocol. They use the bound they have obtained for Eve’s accessible information on the raw key in a classical error correction process.

Privacy amplification. Private key distillation is possible with the usual price to pay, namely, the length of the key. Typically, this would function by taking an initial random seed (which would be unknown to any adversary), and using it to process the generated bit values in such a way that, given Eve’s partial information about the bits, she cannot guess the output bit with probability greater than \( \frac{1}{2} (1 + \epsilon) \), where \( \epsilon \) is a small parameter that we choose, defining the trade-off between how confident we are that she has no information about our bits, and the rate at which these bits are generated. However, we are going to be interested in the scenario where this initial random seed is not available, so we shall give an explicit protocol for privacy amplification that functions under the same assumption that we make throughout this paper; Eve only attacks each round of the experiment independently, but does not use any seed randomness.
Suppose that Alice and Bob have succeeded in generating a string of bits \( x_k \). As a result of their evaluation of the Bell inequality, they also have a bound on Eve’s maximum probability for guessing any one of these bits, \( G \). If Alice and Bob take \( N \) such bits and add them modulo 2, the resulting output bit can be guessed by Eve only if she has incorrectly guessed an even number of the outcomes of the individual measurements. This occurs with probability \( (1 + (2G - 1)^N)/2 \). Evidently, as \( N \) increases, this tends to \( \frac{1}{2} \), which is the case providing ideal security but can only be approximated with a finite number of runs. By setting an allowable threshold \( \epsilon \) and picking \( N = \log(\epsilon)/\log(2G - 1) \), Alice and Bob can pick their desired bound on security of the generated bit as a compromise on the number of raw key bits required to calculate it.

### 4. Loopholes

We have already mentioned that violation of Bell inequalities is an experimental fact. What is it then, that prevents us from running the experiments that violated Bell inequalities again, but this time under the label of ‘device-independent key distribution’? As convincing as they were, these experiments still left some loopholes. For example, it is in principle possible that the photons detected in the experiments did not represent a fair sample of all photons emitted by the source (the so-called detection loophole) or that the various parts and components of the experiment were causally connected (the locality loophole). Some of these concerns were addressed in more recent experiments \([17,18]\); however, truth be told, the ultimate violation of a Bell inequality, that single experiment closing all the loopholes at once, is still missing. Nevertheless, Nature would have to be very malicious if it were to cheat us selectively; on locality in some experiments and exploiting detection loopholes in some other. In contrast, an eavesdropper has all the rights to be malicious, so we wish to either develop technology sophisticated enough to close these loopholes or ask exactly how much of an advantage an eavesdropper can have by exploiting them. We focus here on the latter question, by looking closer at the detection and locality loopholes and also how Eve can manipulate the ‘free will’ of Alice’s and Bob’s measurement selections.

Before we do this, it is worth rephrasing Bell inequalities in the context of game theory, which will give us slightly more intuition for what Alice and Bob hope to achieve when using their devices for Bell tests. Imagine a referee choosing at random a question pair \((j, k) \in \{0, 1\}^2\) for each round of the game. The referee queries Alice with \(j\) and Bob with \(k\) such that Alice and Bob are only aware of their own questions. They each send an answer, \(a\) for Alice and \(b\) for Bob, back to the referee, where \(a, b \in \{-1, 1\}\). The round is won by Alice and Bob if their outputs satisfy the condition \(ab = (-1)^{jk}\), which corresponds to Alice and Bob delivering correlated bits for \((j, k) \neq (1, 1)\) and anticorrelated bits when asked \((j, k) = (1, 1)\). Note that these are precisely the outcome correlations that maximally violate the CHSH inequality to the stronger-than-quantum limit. The game is repeated for arbitrarily many rounds and Alice and Bob want to win as many rounds as they can. Their winning probability is defined to be

\[
P^w = \frac{1}{4} \sum_{(j,k)\in\{0,1\}^2} P(ab = (-1)^{jk}|jk). \tag{4.1}
\]
We see that this game is equivalent to a test of the CHSH inequality where \( a \) and \( b \) correspond to the outcomes \( A_j \) and \( B_k \), respectively, which Alice and Bob obtain, if we view the ‘referee’ to be a representation of the free will the users have to select their inputs in each round. It is then easy to see that \( \langle S \rangle = 4(P^w \cdot (+1) + (1 - P^w) \cdot (-1)) \), since the expectation of the correlation of outcomes for each pair of measurements is either +1 or −1. Hence

\[
P^w = \frac{1}{2} + \frac{\langle S \rangle}{8} \tag{4.2}
\]

and the winning probabilities for players with classical, quantum and non-signalling capabilities are \( \frac{3}{4}, (\frac{1}{2} + \frac{1}{\sqrt{2}}) \) and 1, respectively.

(a) Locality loophole

To discuss the first loophole, that of locality, we consider the instance in which the measurement pairs of the Bell test are not distantly separated. In the game formulation, this corresponds to Alice and Bob being causally connected, which is to say that they can communicate with one another between receiving their questions and having to return the answers. Equivalently, in the device-independent scenario, we can visualize this as a single black box, where Eve can immediately know both inputs for a run of the experiment and subsequently choose both outputs within the required timeframe. Upon an input pair \((j, k)\), Eve can randomly select Alice’s output as \( a \in \{\pm 1\} \) and then calculate the value \( b = a \cdot (-1)^j \), which clearly maximizes (4.1). Since using \( b \) for Bob’s output on every run will provide the maximal CHSH value of 4 (and similarly \(-b \) yields \(-4\)), Eve can dilute this value and avoid suspicion from the users of the device by displaying \( b \) with probability \( p \) and otherwise \(-b \) (for instance, over many runs a CHSH value of \( 2\sqrt{2} \) will be achieved if \( p = \frac{1}{2} + \frac{1}{\sqrt{2}} \)).

(b) Detection loophole

In practical implementations of a Bell test, not all photons will be detected. This is represented by a detector efficiency \( \eta \), which is the probability that a photon will be detected on any given run. We should therefore be concerned that an eavesdropper might be faking this inefficiency in the detectors to mask some tampering. Imagine that we have a source emitting photons within fixed time intervals, so that Alice and Bob know when their devices should receive a photon. If detection fails, they decide to register a +1 outcome in place of the missed measurement result. With probability \( \eta^2 \) the CHSH test runs as normal and (assuming the optimal state and measurements) \( 2\sqrt{2} \) is observed. However, with probability \( 2\eta(1 - \eta) \), one of the two detectors fails and the resulting outcomes are entirely uncorrelated; for this subset of runs, the CHSH test yields 0. Finally, a proportion \( (1 - \eta)^2 \) of runs will have failed detections in both devices, for which both users register +1 and a CHSH value of 2 is observed for these runs.

It follows that a CHSH violation (i.e. any value above 2) can only be achieved if \( \eta > \frac{2}{\sqrt{2} + 1} \approx 83 \) per cent, and a maximal violation can only be achieved when the detectors are perfect. Assuming the detector efficiency is above this threshold, we proceed as normal and, when we find the CHSH violation, assume Eve followed her optimal eavesdropping strategy as specified in the previous section.
Table 1. Deterministic model specifying, for each subroutine $\lambda$, the probability $p(jk|\lambda)$ that the measurement pair $(j, k)$ is performed and the outcomes $A_j$ and $B_k$ to be given for each input.

| $\lambda$ | $A_0$ | $A_1$ | $B_0$ | $B_1$ | $p(00|\lambda)$ | $p(01|\lambda)$ | $p(10|\lambda)$ | $p(11|\lambda)$ |
|-----------|-------|-------|-------|-------|----------------|----------------|----------------|----------------|
| $\lambda_1$ | +1    | +1    | +1    | +1    | $P$         | $P$            | $P$            | $1 - 3P$       |
| $\lambda_2$ | +1    | -1    | +1    | +1    | $P$         | $P$            | $1 - 3P$       | $P$            |
| $\lambda_3$ | +1    | +1    | +1    | -1    | $P$         | $1 - 3P$       | $P$            | $P$            |
| $\lambda_4$ | +1    | -1    | -1    | +1    | $1 - 3P$   | $P$            | $P$            | $P$            |

this, we determine how much classical post-processing is necessary in order to completely exclude Eve. There are several studies of note in the literature which extend this consideration to non-locality tests using non-maximally entangled states [19] and asymmetric detection efficiencies [20].

(c) Experimental free will

As previously mentioned, the violation of a Bell inequality is guaranteed by Alice and Bob possessing complete ‘free will’ in selecting which measurements they perform. There is a wealth of literature, both philosophical and technical, about the meaning of free will but in the present context it suffices to consider the process by which the users select their measurements. While you might consider them making random choices for each run of the experiment just from their own internal thought processes, in practice they will individually use random number generators (RNGs) to inform their selections as this will allow many runs of the experiment to be performed, ideal for high-speed key distribution. Alice and Bob must therefore trust that these devices are capable of generating perfect randomness, and this reliance on further technology provides Eve with another avenue for sneakiness.

Consider the case where Eve is in control of both the testing devices and the RNGs used to select the measurements performed. By modifying the distribution of measurement choices to a non-uniform one, she can use a completely deterministic model in the devices to fake a CHSH violation. Since Alice and Bob announce their choices after the key distribution protocol, they would be highly suspicious if their selections did not estimate a uniform distribution so Eve must guarantee this is still the case. However, by specifying a program defined by a variable $\lambda$, Eve is able to implement subroutines in the experiment within which the measurement selections can deviate from uniform. On each run, the program picks $\lambda$ according to a probability distribution $\rho(\lambda)$. For example, consider the model described by table 1, adapted from [21,22], in which $\rho(\lambda)$ is taken to be uniform over four possible values.

The parameter $P \in [\frac{1}{4}, \frac{1}{3}]$ represents the deviation of the measurement selection distribution from uniform—at $P = \frac{1}{4}$ we are back in the scenario where the measurement selection is uniform regardless of $\lambda$, corresponding to perfect randomness in the RNGs, while $P = \frac{1}{3}$ corresponds to being able to ensure that one specific pair of measurement settings is not chosen. Recalling that the observed CHSH value is largest when the outcomes $(A_1, B_1)$ are anticorrelated and all other outcome pairs are correlated, notice that for a given $\lambda$ the only

Phil. Trans. R. Soc. A (2012)
outcome pair specified by the table for which this is not the case is precisely for the input pair with probability \( 1 - 3P \). Devices operating according to this model will yield \( S_{\text{max}} = 24P - 4 \), which increases from 2 to 4 as \( P \) increases from \( \frac{1}{4} \) to \( \frac{1}{3} \).

Since the column sums for each input pair are equal, the impression of complete experimental free will is given to Alice and Bob, who have no knowledge of the underlying \( \lambda \). A value of \( P \approx 0.285 \) yields \( S_{\text{max}} = 2\sqrt{2} \), and yet Eve has complete knowledge of the generated outcomes.

Before discussing the implications of this, a couple of technical points. Firstly, one may be concerned that this model always specifies \( A_0 = +1 \), which is not true in the non-adversarial CHSH experiment. However, since we are concerned only with the correlations of the outcomes rather than the actual values, the same result is achieved, for instance, with a model which with probability \( \frac{1}{2} \) corresponds to the given table and with probability \( \frac{1}{2} \) uses the same table but takes the negative of each output value. Secondly, we could specify a more general input distribution for each \( \lambda \) with four parameters (or three using normalization), but it has been proved that the given one-parameter distribution in fact yields the largest CHSH value [22,23].

The model above simulates a violation of the CHSH inequality by using pre-programming so Alice and Bob need to be wary of these tactics when relying on a CHSH violation to guarantee security in cryptographic tasks. If nothing is known about the parameter \( P \), then no matter that value of the CHSH violation, the generated key cannot be trusted. However, if \( a \text{ priori} \) we know that \( P \) can only take on a limited range, the derived bounds inform the maximum amount of information that Eve could have obtained for a given CHSH violation, and hence privacy amplification can be used to exclude her, provided she does not have perfect information. How that parameter \( P \) might be assessed in an experiment remains an open problem.

Evidently, Alice and Bob also need a trusted source of random numbers for such tasks. Again, the device-independent scenario, in which we do not trust the provenance of the devices, is desirable and, again, it turns out that the CHSH test can be used to authenticate the randomness and non-pre-determined nature of measurement results, permitting the expansion of initial random seeds (necessary to be able to select the measurement bases for the test) into longer random strings [24–26].

(d) Collective attacks

In analysing the possible performance of an eavesdropper, Eve, we have assumed that she acts independently on each round of the experiment. Recently, progress has been made in lifting this restriction [27,28]. Conceptually, the starting point of these approaches has been that by implementing each run of the experiment on a different, causally disconnected device, it essentially becomes impossible for an eavesdropper to do anything other than a collective attack. While such restrictions severely impair the practicality of the scheme, relaxing them risks introducing further loopholes, such as those detailed in [29]. Equally, these papers have assumed complete free will, and hence the availability of randomness for the purposes of privacy amplification. In the previous scenario of limited free will, we avoided this by giving a simple protocol for privacy
amplification which relied on the assumption of collective attacks but without using any randomness. The study of [30] has provided a suitable starting point for overcoming this restriction in the presence of limited free will.

5. Conclusions

The field of cryptography has developed massively over many centuries, with ever more sophisticated methods of encryption emerging, but until recent decades there were some misconceptions about the pre-determined nature of information that prevented researchers from taking the ultimate leap into achieving unconditionally secure cryptographic protocols. In this paper, we have shown how arguments refuting this lack of chance in Nature not only prevailed but provided an exciting new direction for cryptography, and how even more recent arguments have removed the reliance on a quantum mechanical framework to ensure that these protocols will stand the test of time even if quantum theory does not. Even if one day quantum physics is superseded by a new theory, even then, as long as the new theory does not admit any instant communication, we can use Bell inequality violations as an indicator of security, transcending the borders of classical and quantum theory alike, such as in the first security proof by Barrett et al. [31] of key distribution in an arbitrary no-signalling theory. What is more, such security guarantees are independent from the trustworthiness of an external party manufacturing and distributing these devices (even when the users have limited experimental free will), providing cryptographic protocols that even Alan Turing himself could not have hoped to crack.

This work was supported by the National Research Foundation and the Ministry of Education, Singapore. The work of J.P. was also supported by an EPSRC postgraduate studentship.

References