The bottom line

- The Einstein–Podolsky–Rosen correlation can be simulated with local resources. A local simulation shows that it is possible for two systems to exhibit an Einstein–Podolsky–Rosen correlation if they both have their own ‘being–thus,’ in Einstein’s sense, and if the separate ‘being–thuses’ originate from a common source as the common cause of the correlation.

- In the case of Einstein–Podolsky–Rosen bananas, this will be the case if they originate on a banana tree as a particular pair, each with a shared value of some banana variable $\lambda$, the ‘being–thus’ of the bananas.

- The tastes for all possible pairs of peelings are determined by the shared value of $\lambda$, and the distribution of shared $\lambda$ values for many Einstein–Podolsky–Rosen pairs of bananas provides an explanation of the correlation. In a simulation of the correlation, the variable $\lambda$ corresponds to the simulation variable.

- The Einstein–Podolsky–Rosen argument is that
  
  (i) if you have a perfect correlation of this sort, and you can exclude a direct causal influence between the two systems (because the correlations persist if the systems are separated by any distance), then the only explanation for the correlation is the existence of a common cause or shared randomness

  (ii) since the description of the correlated systems in quantum mechanics lacks a representation of the common cause, quantum mechanics must be incomplete. What’s missing in a quantum state description like $|\phi^+\rangle$ is something corresponding to the banana variable $\lambda$, or the simulation variable in a local simulation.

3.2 Popescu–Rohrlich Bananas and Bell’s Theorem

The Einstein–Podolsky–Rosen argument lingered for about thirty years until John Bell showed that any common cause explanation of a probabilistic correlation between measurement outcomes on two separated systems would have to satisfy an inequality, now called Bell’s inequality. He also showed that the inequality is violated by measurements of certain two-valued observables of a pair of quantum systems in an entangled state. So Einstein’s intuition about the correlations of entangled quantum states, and the Einstein–Podolsky–Rosen argument that depends on this intuition, turn out to be wrong. There are correlations between quantum systems, where a causal influence from one system to the other can be excluded, that have no common cause explanation. The result is known as Bell’s theorem.

It took a while before the implications of Bell’s theorem penetrated mainstream physics, probably around the time that Alain Aspect and colleagues confirmed the violation of Bell’s inequality in a series of experiments on entangled photons in the early 1980’s. It took several more years for
3.2 Popescu–Rohrlich Bananas and Bell’s Theorem

the result to trigger a revolution in quantum information in the 1990’s, when it became respectable again to discuss foundational questions in quantum mechanics. Some of the initial discussion took place in a sort of samizdat publication that was hand-typed, mimeographed, and mailed to a select list of readers by a Swiss foundation, the Association F. Gonseth, as a ‘Written Symposium’ on ‘Hidden Variables and Quantum Uncertainty’ that went through 36 issues, from November, 1973 to October, 1984. The back of each issue carried the statement (in English, French, and German):

‘Epistemological Letters’ are [sic] not a scientific journal in the ordinary sense. They want to create a basis for an open and informal discussion allowing confrontation and ripening of ideas before publishing in some adequate journal.

Einstein, Podolsky, and Rosen argued for the incompleteness of quantum mechanics from the perfect correlation between the outcomes of the same observable measured on two separated systems in an entangled state. Bell drew a different conclusion from the probabilistic correlation between the outcomes of measurements of different observables on the two systems. Instead of following Bell’s argument, I’ll derive a version of Bell’s theorem as a limitation on the ability of Alice and Bob to simulate a correlation proposed by Sandu Popescu and David Rohrlich if they are restricted to local resources.

Popescu and Rohrlich imagine a hypothetical device, now known as a PR box, with an input and output port for Alice and an input and output port for Bob. There are two possible inputs to these ports, labeled 0 and 1, and two possible outputs, also labeled 0 and 1, that are randomly related to the inputs: the outputs 0 and 1 for Alice and Bob separately occur with equal probability for any input. In particular, Alice’s output occurs randomly and independently of Bob’s input, and similarly Bob’s output occurs randomly and independently of Alice’s input. A PR box can be used only once: after an input by Alice, no further Alice-input is possible, and similarly for Bob.

Now, Popescu and Rohrlich suppose that the Alice part of the box and the Bob part of the box can be separated, without any physical connection between them, and that even when separated by an arbitrary distance, the two parts of the box produce the correlation (see Figure 3.3):

- the outputs for Alice and Bob are the same, except when both inputs are 1, in which case the outputs are different
- the marginal probabilities of the outputs 0 or 1 for any input by Alice or Bob separately are 1/2 (so the no-signaling condition is satisfied)

A PR box functions in such a way that if Alice inputs 0 or 1, her output is 0 or 1 with probability 1/2, irrespective of Bob’s input, and irrespective of whether Bob inputs anything at all. Similarly for Bob. The requirement is simply that whenever there is in fact an input by Alice and an input by Bob, in whatever temporal order, the inputs and outputs are correlated in the required way. So for Alice and Bob separately, the output appears to be random for any input, but if Alice and Bob get together and compare inputs and outputs, they find that there are three cases when their outputs are the same, both 0 or both 1 (when they both input 0, or when Alice inputs 0 and Bob inputs 1, or when Alice inputs 1 and Bob inputs 0), and one case where their outputs are different, one output is 0 and the other output is 1 (when Alice and Bob both input 1).
In Bananaworld, Popescu–Rohrlich pairs of bananas grow on Popescu–Rohrlich trees which, like Einstein–Podolsky–Rosen trees, have two-banana bunches. Tastes and peelings are correlated like the correlation of a PR box and persist if the bananas are separated by any distance:

- if the peelings are $SS, ST, TS$, the tastes are the same, 00 or 11
- if the peelings are $TT$, the tastes are different, 01 or 10
- the marginal probabilities for the tastes 0 or 1 if a banana is peeled $S$ or $T$ are 1/2, irrespective of whether or not the paired banana is peeled (so the no-signaling principle is satisfied)

What could Alice and Bob do to simulate the correlation of Popescu–Rohrlich pairs of bananas? Here’s a strategy that would work in three out of four rounds of the simulation game, on average, if the prompts are random. Alice and Bob generate a sufficiently long list of random 0’s and 1’s before the simulation game starts (in this case, just one random list), and they each take a copy of the list with them to consult during the game. So they share a list of random bits. If they each respond to a prompt with the bit on the shared list, in order for each successive round, they will satisfy the condition on the marginal probabilities (because they each respond 0 or 1 with probability 1/2, since the lists are random) and they will win all the rounds of the simulation game for which the prompts are $SS, ST, \text{ or } TS$ (because they both ignore the prompt and respond with the same bit, either 0 or 1), and they will lose all the rounds for which the the prompts are $TT$.
(because they always respond the same, and the requirement for this combination of prompts is to respond differently). If the prompts are random, the probability of winning the game with this strategy is $3/4$.

It turns out that this is the optimal strategy for winning the PR simulation game with local resources, as I’ll show below. This is a version of Bell’s theorem. You might be puzzled about how a result about hypothetical bananas could be relevant to a theorem about quantum mechanics. Firstly, this is not so much a result about hypothetical bananas as a result about a game with a moderator and two players. The optimal probability of winning the game with local resources is $3/4$. This limitation on winning the game—on simulating the PR correlation with local resources—is a fact about the world, a limitation on correlations with a common cause. In a simulation with local resources, the correlation arises from a shared random variable, where each shared value of the random variable labels a local instruction set or local response strategy for Alice and Bob separately. Common causes are nature’s local resources, so if there’s no way in principle for Alice and Bob to simulate a correlation with local resources, then nature can’t do it either. Secondly, Bell’s theorem is about common cause correlations, not directly about quantum mechanics. Bell proved that a correlation between measurement outcomes on two separated systems satisfies a certain inequality if the correlation has a common cause. The relevance for quantum mechanics is that the inequality can be violated by the correlations of entangled quantum states, as Bell showed (and as I’ll show in next section 3.3, Simulating Popescu–Rohrlich Bananas). In terms of the PR
simulation game, it’s also a fact about the world that Alice and Bob can do a better job of winning
the game if they are allowed to use entangled quantum states instead of two copies of the same
random list. It follows that there are nonlocal quantum correlations between separated systems
that can’t arise from a common cause (and can’t arise from a direct causal influence between the
systems either, if the systems are sufficiently far apart and the measurements are sufficiently close
together in time, because any signal carrying information between the systems would have to move
faster than the speed of light).

The Popescu–Rohrlich correlation

- A PR box (after Popescu and Rohrlich) has an input and output port for Alice and an
input and output port for Bob. There are two possible inputs to these ports, labeled 0
and 1, and two possible outputs, also labeled 0 and 1. The Alice part of the box and the
Bob part of the box can be separated, without any physical connection between them,
even when separated the two parts of the box produce the correlation

(i) if the inputs are 00, 01, 10, the outputs are the same, but if the inputs are 11 the
outputs are different

(ii) the marginal probabilities of the outputs 0 or 1 for any input by Alice or Bob
separately are 1/2 (so the no-signaling condition is satisfied)

- In Bananaworld, Popescu–Rohrlich pairs of bananas occur on trees with bunches of
just two bananas. Tastes and peelings are correlated like the correlation of a PR box
and persist if the bananas are separated by any distance:

(i) if the peelings are SS, ST, TS, the tastes are the same, but if the peelings are
TT, the tastes are different

(ii) the marginal probabilities for the tastes 0 or 1 if a banana is peeled S or T are
1/2 (so the no-signaling condition is satisfied)

There are lots of proofs of Bell’s theorem in the literature on quantum mechanics, many of them
informative only to the authors. The simplest proof I know is from a short but illuminating book
for non-specialists, Quantum Chance by Nicolas Gisin, which, as the subtitle indicates, is about
‘nonlocality, teleportation, and other quantum marvels.’ Gisin’s proof is the sort of immediately
obvious, back–of–the–envelope proof you will be able to reproduce for your friends over a beer
once you get the basic idea. Gisin was awarded the first Bell prize in 2009 ‘for his theoretical and
experimental work on foundations and applications of quantum physics, in particular: quantum
non-locality, quantum cryptography and quantum teleportation.’ With the group he leads, Gisin
has performed some spectacular demonstrations of long-distance entanglement of photons using
fiber optic cables under Lake Geneva.

The idea of the proof, applied to the simulation game for Popescu–Rohrlich bananas, is that
there are really only four possible local strategies available to Alice at each round of the game,
since she has no access to Bob’s prompt or his response to the prompt, and similarly for Bob. Alice
3.2 Popescu–Rohrlich Bananas and Bell’s Theorem

Table 3.1: All possible response strategies for simulating the Popescu–Rohrlich correlation. The correct responses are indicated in bold.

<table>
<thead>
<tr>
<th>Alice’s strategy</th>
<th>Bob’s strategy</th>
<th>response for input SS</th>
<th>response for input ST</th>
<th>response for input TS</th>
<th>response for input TT</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td><strong>00</strong></td>
<td><strong>00</strong></td>
<td><strong>00</strong></td>
<td>00</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>01</td>
<td>01</td>
<td>01</td>
<td><strong>01</strong></td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>same</td>
<td><strong>00</strong></td>
<td>01</td>
<td><strong>00</strong></td>
<td><strong>01</strong></td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>different</td>
<td>01</td>
<td><strong>00</strong></td>
<td>01</td>
<td>00</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td><strong>10</strong></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>same</td>
<td>10</td>
<td><strong>11</strong></td>
<td>10</td>
<td><strong>11</strong></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>different</td>
<td><strong>11</strong></td>
<td>10</td>
<td><strong>11</strong></td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>same</td>
<td>0</td>
<td><strong>00</strong></td>
<td><strong>00</strong></td>
<td>10</td>
<td><strong>10</strong></td>
<td>3</td>
</tr>
<tr>
<td>same</td>
<td>1</td>
<td>01</td>
<td>01</td>
<td><strong>11</strong></td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>same</td>
<td>same</td>
<td><strong>00</strong></td>
<td>01</td>
<td>10</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>same</td>
<td>different</td>
<td>01</td>
<td><strong>00</strong></td>
<td><strong>11</strong></td>
<td><strong>10</strong></td>
<td>3</td>
</tr>
<tr>
<td>different</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td><strong>00</strong></td>
<td>00</td>
<td>1</td>
</tr>
<tr>
<td>different</td>
<td>1</td>
<td><strong>11</strong></td>
<td><strong>11</strong></td>
<td>01</td>
<td><strong>01</strong></td>
<td>3</td>
</tr>
<tr>
<td>different</td>
<td>same</td>
<td>10</td>
<td><strong>11</strong></td>
<td><strong>00</strong></td>
<td><strong>01</strong></td>
<td>3</td>
</tr>
<tr>
<td>different</td>
<td>different</td>
<td><strong>11</strong></td>
<td>10</td>
<td>01</td>
<td>00</td>
<td>1</td>
</tr>
</tbody>
</table>

can respond 0 or 1 without reference to her prompt, or she can respond 0 or 1 depending on her prompt. Re-labeling the prompt $S$ as 0 and the prompt $T$ as 1 (as in the original Popescu–Rohrlich labeling of the inputs to a PR box), this amounts to saying that there are only two ways to respond depending on the prompt: Alice’s response can either be the same as the prompt (0 for $S$ and 1 for $T$), or different from the prompt (1 for $S$ and 0 for $T$). Since there are two possible prompts, there are four possible local strategies for Alice (two for each prompt). Similarly, there are four possible local strategies for Bob, which yields sixteen possible combined strategies. All you need to do now is count how many times each strategy is successful in simulating the Popescu–Rohrlich correlation for the four possible pairs of prompts.

Call the two possible strategies that are independent of the prompt 0 and 1, according to the response, and the two possible strategies that depend on the prompt ‘same’ and ‘different,’ according to whether the strategy is to match the numerical value of the prompt (to respond 0 for $S$ and 1 for $T$), or differ from it (to respond 1 for $S$ and 0 for $T$). Table 3.1 shows that no combination of Alice’s strategy and Bob’s strategy is successful for more than three out of four possible pairs of prompts.

If the prompts are random, each of the four possible pairs of prompts $SS, ST, TS, TT$ occurs
with probability $1/4$, so

$$p(\text{successful simulation}) = \frac{1}{4}(p(\text{responses same}|SS) + p(\text{responses same}|ST) + p(\text{responses same}|TS) + p(\text{responses different}|TT))$$

The notation $p(\ldots)$ stands for the probability of the expression in parenthesis, and the vertical slash sign $|$ indicates a conditional probability, which can be read as ‘given that.’ So $p(\text{responses same}|SS)$ is the probability of ‘responses same,’ given that Alice peels $S$ and Bob peels $S$, and so on. Since the most any strategy is successful is for three out of four pairs of prompts (and the least any strategy is successful is for one out of four pairs of prompts), Table 3.1 can be summed up in the statement:

$$1/4 \leq p_L(\text{successful simulation}) \leq 3/4$$

This is a version of Bell’s inequality. It says that the probability of successfully simulating the correlation of Popescu–Rohrlich bananas with local resources available to Alice alone or to Bob alone (‘$L$’ for local), lies between $1/4$ and $3/4$. So the optimal probability is $3/4$. If Alice and Bob each separately respond randomly to the prompts, the success probability is $1/2$. In this case they will give the same response in half the rounds, and different responses in half the rounds (see below). In a simulation restricted to local resources, the correlation arises from a shared random variable, where the shared values of the random variable range over different common causes of the correlation (in this case, the sixteen possible combined strategies). So the impossibility of perfectly simulating the Popescu–Rohrlich correlation if Alice and Bob are restricted to local resources means that the correlation can’t arise from a common cause. If Alice and Bob are allowed to use entangled quantum states as a shared resource, they can break the $3/4$ barrier (see the next section). So it must be impossible to simulate the correlations of entangled states with local resources, which means that, incredibly, we live in a nonlocal quantum world in which there are correlations without a common cause.

Where does a common cause come into the argument? Each row of Table 3.1 represents a deterministic Alice strategy and a deterministic Bob strategy for responding to the possible prompts. The combined strategy for Alice and Bob is equivalent to a deterministic common cause of the correlation in the responses, a local instruction set for Alice and a local instruction set for Bob on the basis of which they produce the correlated responses. The table lists the sixteen possible combined deterministic strategies or deterministic common causes and the corresponding score, the number of correct responses, for the four possible input pairs. Alice and Bob will have to use the values of a shared random variable with sixteen possible values, corresponding to the sixteen combined strategies, to choose a combined strategy randomly for each round of the game. Computer scientists prefer the term ‘shared randomness’ to ‘common cause,’ and refer to Bell’s result as showing that the correlations of entangled quantum states can’t arise from shared randomness in the correlated quantum systems.

If the values of the random variable occur with a probabilities between 0 and 1 that sum to 1, then $1/4 \leq p_L(\text{successful simulation}) \leq 3/4$ for any probabilities, provided the values of the random variable are independent of the choice of prompts. For example, Alice and Bob could toss
3.2 Popescu–Rohrlich Bananas and Bell’s Theorem

four fair coins many times before the start of the simulation game. The outcome of each quadruple toss is a quadruple of heads and tails (or a quadruple of bits, taking heads as 0 and tails as 1) and there are $2 \times 2 \times 2 \times 2 = 16$ possible quadruples, which occur with equal probability of $1/16$ if the coins are fair. Alice and Bob each record the list of random quadruples and consult the shared list in order during successive rounds of the simulation game. In this case, the shared values of the random variable are the sixteen quadruples, which occur with equal probability. Eight out of sixteen strategies correspond to a score of 1 out of 4, and these are equally weighted with a probability of $1/16$, and eight equally weighted strategies correspond to a score of 3 out of 4. So

$$p_L(\text{successful simulation}) = \frac{8}{16} \cdot \frac{1}{4} + \frac{8}{16} \cdot \frac{3}{4} = \frac{1}{2}$$

If Alice and Bob use biased coins, the probabilities will be different, but even if the strategy represented by the top row has unit probability and all the other fifteen strategies have zero probability, $p_L(\text{successful simulation}) = 3/4$. So Alice and Bob can achieve a successful simulation with probability $3/4$, but there’s no way the probability can exceed $3/4$ for any probability distribution of strategies.

Bell’s theorem

- The optimal probability of successfully simulating the correlation of Popescu–Rohrlich bananas with local resources is $3/4$.
- In a simulation with local resources, the correlation arises from a shared random variable, where the shared values of the random variable range over different common causes of the correlations.
- The impossibility of perfectly simulating the Popescu–Rohrlich correlation if Alice and Bob are restricted to local resources means that the correlation can’t have a common cause explanation.
- If Alice and Bob are allowed to use entangled quantum states as a shared resource, they can break the $3/4$ barrier and achieve an optimal probability of about .85.
- So, incredibly, we live in a nonlocal quantum word in which there are correlations without a common cause.

There are various Bell inequalities for different sets of observables besides the specific inequality in Bell’s original 1964 paper. John Clauser, Michael Horne, Abner Shimony, and Richard Holt derived an inequality in 1969 that plays an important role in some of the subsequent chapters. It’s also easier to test the Clauser–Horne–Shimony–Holt inequality experimentally than Bell’s original inequality. Here’s a proof of the Clauser–Horne–Shimony–Holt version of Bell’s theorem.

Consider again the problem of simulating the correlation of a pair of Popescu–Rohrlich bananas. It’s convenient to re-label the responses $\pm 1$ instead of 0, 1. ‘Responses are the same’ and ‘responses are different’ mean the same thing whatever the units. So the probabilities—
3 Bananaworld

\( p(\text{responses same}|A, B) \), the conditional probability that Alice and Bob produce the same response, given the prompts \( A \) and \( B \), and \( p(\text{responses different}|A, B) \), the conditional probability that Alice and Bob produce different responses, given the prompts \( A \) and \( B \)—take the same values whatever the units. Here \( A \) represents Alice’s prompt, \( S \) or \( T \), and \( B \) represents Bob’s prompt, and I’ve separated the prompts by commas to avoid confusion with products in the following argument.

The Clauser–Horne–Shimony–Holt inequality involves the expectation value or average value of the product of the responses for a particular pair of prompts \( A \) and \( B \), denoted by the symbol \( \langle AB \rangle \). The expectation value \( \langle AB \rangle \) is the weighted sum of the products of the four possible pairs of responses, with the probabilities for the paired responses, \( p(-1, -1|A, B) \), \( p(-1, 1|A, B) \), \( p(1, -1|A, B) \), \( p(1, 1|A, B) \) as the weights. (So the expectation value, unlike the probabilities, does depend on the units for the responses, because the products of the responses are different for different units.) This sounds complicated, but there’s a simple way of expressing the expectation value. The products of the responses are 1 if the responses are the same (\(-1, -1\) or \(1, 1\)), and \(-1\) if the responses are different (\(-1, 1\) or \(1, -1\)), so the expectation value is just a sum of probabilities multiplied by \(\pm 1\) appropriately:

\[
\langle AB \rangle = p(-1, -1|A, B) - p(-1, 1|A, B) - p(1, -1|A, B) + p(1, 1|A, B)
\]

Since \( p(\text{responses same}|A, B) = 1 - p(\text{responses different}|A, B) \) and \( p(\text{responses different}|A, B) = 1 - p(\text{responses same}|A, B) \), the expectation value \( \langle AB \rangle \) can equally well be expressed as

\[
\langle AB \rangle = 2p(\text{responses same}|A, B) - 1 = 1 - 2p(\text{responses different}|A, B)
\]

so

\[
p(\text{responses same}|A, B) = \frac{1 + \langle AB \rangle}{2}
\]

\[
p(\text{responses different}|A, B) = \frac{1 - \langle AB \rangle}{2}
\]

To keep track of whether \( S \) or \( T \) refers to Alice or Bob in the following proof, I’ll need to add subscripts: \( S_A \) or \( S_B \), \( T_A \) or \( T_B \). With the common cause or shared random variable \( \lambda \) to consider as well, the notation becomes clumsy. So I’ll simply write \( S_A \) as \( A \) and \( T_A \) as \( A' \), and \( S_B \) as \( B \) and \( T_B \) as \( B' \), indicating the prompts corresponding to Alice’s two ways of peeling her banana as \( A \) and \( A' \) for peeling by the stem end (\( S \)) or the top end (\( T \)), and similarly I’ll indicate the prompts corresponding to Bob’s two ways of peeling his banana as \( B \) and \( B' \) for \( S \) and \( T \).

The relationship between the expectation value \( \langle AB \rangle \) and the probabilities of the responses also
3.2 Popescu–Rohrlich Bananas and Bell’s Theorem

holds for the expectation values $\langle AB \rangle$, $\langle A'B \rangle$, and $\langle A'B' \rangle$. So

$$p(\text{successful simulation}) = \frac{1}{4}(p(\text{responses same}|A, B) + p(\text{responses same}|A, B'))$$
$$+ p(\text{responses same}|A', B) + p(\text{responses different}|A', B'))$$
$$= \frac{1}{4} \left( \frac{1 + \langle AB \rangle}{2} + \frac{1 + \langle AB' \rangle}{2} + \frac{1 + \langle A'B \rangle}{2} + \frac{1 - \langle A'B' \rangle}{2} \right)$$
$$= \frac{1}{2} \left( 1 + \frac{K}{4} \right)$$

where $K = \langle AB \rangle + \langle AB' \rangle + \langle A'B \rangle - \langle A'B' \rangle$.

Suppose Alice and Bob have a strategy for simulating the correlation of a pair of Popescu–Rohrlich bananas using a shared random variable $\lambda$. (The point of the theorem, of course, is to prove that there can be no such strategy.) The possible values of $\lambda$ needn’t be the sixteen values corresponding to the sixteen combined strategies in Gisin’s proof—they could be any values, with any probabilities, as long as the $\lambda$ probabilities are independent of the probabilities of the prompts, which are assumed to be freely chosen by a moderator or randomly chosen by a randomizing device. The $\lambda$ values could even be a continuous set with a probability distribution $\rho(\lambda)$ (a probability function defined on the continuous set of $\lambda$ values that integrates to 1 over the set, just as the set of probabilities for a finite set of $\lambda$ values sums to 1). The responses could be deterministic functions of $\lambda$, or it could be that, given prompts $A$ and $B$, Alice responds $a$ with a probability $p_\lambda(a|A)$ that depends on $\lambda$ and Bob responds $b$ with a probability $p_\lambda(b|B)$ that depends on $\lambda$. (I represent $\lambda$ as a subscript here because it’s easier to read an expression like $p_\lambda(a|A)$ rather than $p(a|A, \lambda)$ in the proof below.)

The tastes and peelings of Popescu–Rohrlich bananas are correlated (statistically dependent) and not randomly related. So if $a, b$ represent tastes and $A, B$ peelings, the probability of a pair of tastes given a pair of peelings won’t be equal to the product of the probability of the taste of Alice’s banana given Alice’s peeling and the probability of the taste of Bob’s banana given Bob’s peeling:

$$p(a, b|A, B) \neq p(a|A)p(b|B)$$

But in a simulation of these probabilities, the probability of a joint response to a pair of prompts, given a particular value of the shared random variable $\lambda$, is equal to the product of the probability of Alice’s response to her prompt given $\lambda$ and the probability of Bob’s response to his prompt given $\lambda$:

$$p_\lambda(a, b|A, B) = p_\lambda(a|A)p_\lambda(b|B)$$

That’s because Alice’s response to a prompt for a given value of $\lambda$ can’t depend on Bob’s prompt or on Bob’s response to his prompt, and similarly for Bob’s response with respect to Alice. So the responses, conditional on the shared random variable $\lambda$, are statistically independent, and the joint probability factorizes to a product of marginal probabilities for Alice and Bob separately.

This is where the classical resource of shared randomness plays a role in the argument. Putting it differently, this is where the assumption that the correlations are due to a common cause, or
a local hidden variable, plays a role in limiting the extent to which Alice and Bob can simulate the correlation. The first equation says that the tastes and peelings are correlated or statistically dependent. The second equation says that the prompts and responses in a simulation with local resources are conditionally statistically independent, given the common cause represented by the shared random variable $\lambda$. (For more on conditional statistical independence, see the subsection Correlations in the More section at the end of the chapter.)

The expectation value or average value of the responses for a pair of prompts $A$ and $B$, and a particular value of $\lambda$, is $\langle AB \rangle_\lambda$. Assuming conditional statistical independence, the joint probabilities in the expression for $\langle AB \rangle_\lambda$ can each be expressed as a product of probabilities for Alice and Bob separately: $p_\lambda(-1, -1|A, B) = p_\lambda(-1|A) \cdot p_\lambda(-1|B)$, and so on, so:

$$
\langle AB \rangle_\lambda = p_\lambda(-1, -1|A, B) - p_\lambda(-1, 1|A, B) - p_\lambda(1, -1|A, B) + p_\lambda(1, 1|A, B)
= p_\lambda(-1|A) \cdot p_\lambda(-1|B) - p_\lambda(-1|A) \cdot p_\lambda(1|B) - p_\lambda(1|A) \cdot p_\lambda(-1|B) + p_\lambda(1|A) \cdot p_\lambda(1|B)
= (p_\lambda(1|A) - p_\lambda(-1|A))(p_\lambda(1|B) - p_\lambda(-1|B))
= \langle A \rangle_\lambda \langle B \rangle_\lambda
$$

where $\langle A \rangle_\lambda$ and $\langle B \rangle_\lambda$ represent averages over the two possible values $\pm 1$ of $A$ and $B$, with weights equal to the $\lambda$-probabilities for 1 and $-1$. 

3 Bananaworld
The Clauser–Horne–Shimony–Holt version of Bell’s theorem

- The tastes and peelings of Popescu–Rohrlich bananas are correlated (statistically dependent) and not randomly related. So if \( a, b \) represent tastes and \( A, B \) peelings, the probability of a pair of tastes given a pair of peelings won’t factorize to the product of the marginal probabilities for Alice and Bob separately: \( p(a, b|A, B) \neq p(a|A)p(b|B) \).

- In a local simulation, the joint probability of a pair of responses to a pair of prompts, conditional on a shared random variable \( \lambda \) (representing a common cause of the correlation), does factorize to a product of the marginal probabilities, conditional on \( \lambda \), for Alice and Bob separately: \( p_\lambda(a, b|A, B) = p_\lambda(a|A)p_\lambda(b|B) \).

- The condition that the joint probability factorizes is Bell’s locality condition. It says that the prompts and responses are uncorrelated (or conditionally statistically independent), given the common cause represented by the shared random variable \( \lambda \). That’s because Alice’s response to a prompt for a given value of \( \lambda \) can’t depend on Bob’s prompt or on Bob’s response to his prompt, and similarly for Bob’s response with respect to Alice.

- Clauser, Horne, Shimony, and Holt consider the quantity \( K = \langle AB \rangle + \langle AB' \rangle + \langle A'B \rangle - \langle A'B' \rangle \), where \( \langle AB \rangle \) denotes the expectation value of the product of the responses for a pair of prompts \( A, B \).

- The expectation value of the product of the responses for a pair of prompts factorizes for a particular value of \( \lambda \). So \( K_\lambda = \langle A \rangle_\lambda \langle B \rangle_\lambda + \langle A \rangle_\lambda \langle B' \rangle_\lambda + \langle A' \rangle_\lambda \langle B \rangle_\lambda - \langle A' \rangle_\lambda \langle B' \rangle_\lambda \).

- This can be expressed as \( K_\lambda = \langle A \rangle_\lambda \left[ \langle B \rangle_\lambda + \langle B' \rangle_\lambda \right] + \langle A' \rangle_\lambda \left[ \langle B \rangle_\lambda - \langle B' \rangle_\lambda \right] \). Since each of the conditional expectation values is between \(-1\) and \(+1\), and one of the bracketed expressions with the \( B \) and \( B' \) expectation values is \( 2 \) or \(-2\) (in which case the other bracketed expression is \( 0 \)), it follows that \(-2 \leq K_\lambda \leq 2 \).

- In a simulation of the Popescu–Rohrlich correlation with local resources, \( K \) is \( K_\lambda \) averaged over \( \lambda \), with the probabilities of the \( \lambda \) values as the weights. Since averaging over \( \lambda \) won’t change the inequality, \(-2 \leq K_L \leq 2 \) (‘L’ for local).

The quantity \( K = \langle AB \rangle + \langle AB' \rangle + \langle A'B \rangle - \langle A'B' \rangle \) for a particular value of \( \lambda \) can therefore be expressed as

\[
K_\lambda = \langle A \rangle_\lambda \langle B \rangle_\lambda + \langle A \rangle_\lambda \langle B' \rangle_\lambda + \langle A' \rangle_\lambda \langle B \rangle_\lambda - \langle A' \rangle_\lambda \langle B' \rangle_\lambda = \langle A \rangle_\lambda \left[ \langle B \rangle_\lambda + \langle B' \rangle_\lambda \right] + \langle A' \rangle_\lambda \left[ \langle B \rangle_\lambda - \langle B' \rangle_\lambda \right]
\]

This says that \( K_\lambda \) is equal to the product of \( \langle A \rangle_\lambda \) and the sum \( \left[ \langle B \rangle_\lambda + \langle B' \rangle_\lambda \right] \), plus the product of \( \langle A' \rangle_\lambda \) and the difference \( \left[ \langle B \rangle_\lambda - \langle B' \rangle_\lambda \right] \). Since \( \langle B \rangle_\lambda, \langle B' \rangle_\lambda \) are both between \(-1\) and \(1\), the sum in brackets \( \left[ \langle B \rangle_\lambda + \langle B' \rangle_\lambda \right] \) can take a maximum value of \(2\) (when the terms in the sum are...
both 1) or a minimum value of $-2$ (when the terms in the sum are both $-1$), and in both cases $\langle B \rangle_\lambda - \langle B' \rangle_\lambda = 0$. Similarly, $\langle B \rangle_\lambda - \langle B' \rangle_\lambda$ can take a maximum value of 2 (when $\langle B \rangle_\lambda = 1$ and $\langle B' \rangle_\lambda = -1$) or a minimum value of $-2$ (when $\langle B \rangle_\lambda = -1$ and $\langle B' \rangle_\lambda = 1$), and in both these cases $\langle B \rangle_\lambda + \langle B' \rangle_\lambda = 0$. So the maximum and minimum values of one of the bracketed expressions with the $B$ and $B'$ expectation values is 2 or $-2$, in which case the other bracketed expression is 0. Since the smallest value of $\langle A \rangle_\lambda$ or $\langle A' \rangle_\lambda$ is $-1$ and the largest value is 1, it follows that $K_\lambda$ lies between $-2$ and 2.

The quantity $K$ is just $K_\lambda$ averaged over $\lambda$. Averaging over $\lambda$ won’t change the inequality, because an average is just a weighted sum of the quantities $K_\lambda$ with the probabilities of the $\lambda$ values as the weights, and these are all positive numbers between 0 and 1. So in a simulation of the Popescu–Rohrlich correlation with local resources, $K_L$ ('L' for local) is similarly bounded:

$$-2 \leq K_L \leq 2$$

This is the Clauser–Horne–Shimony–Holt version of Bell’s theorem.

Since $p_L$(successful simulation) = $\frac{1}{2}(1 + \frac{K_L}{4})$ as I showed above, and the maximum value of $K_L$ is 2, it follows, as before, that the optimal probability of simulating the Popescu–Rohrlich correlation with local resources is

$$p_L$(successful simulation) $\leq \frac{1}{2}(1 + \frac{2}{4}) = 3/4$$

I’ll show in the next section that if Alice and Bob are allowed to base their strategy on shared entangled quantum states prepared before they separate, then they can achieve a value of $K_Q = 2\sqrt{2}$. So they can do better than classical players or players restricted to local resources and win the simulation game with probability $\frac{1}{2}(1 + \frac{2\sqrt{2}}{4}) \approx .85$. The value $K_Q = 2\sqrt{2}$ is called the Tsirelson bound, after Boris Tsirelson (sometimes spelled Cirel’son) who first proved that this is in fact the optimal quantum value. The demonstration in the next section doesn’t show this. Rather, I show only the possibility of simulating the correlations of Popescu–Rohrlich bananas with probability $\frac{1}{2}(1 + \frac{2\sqrt{2}}{4})$ using entangled quantum states, and so of achieving a value $K = 2\sqrt{2}$ with quantum resources—not that this is the optimal quantum value, which is harder to prove. The optimal quantum value can be achieved with appropriate measurements on quantum states that are said to be ‘maximally entangled.’ The state $|\phi^+\rangle$ is maximally entangled, and appropriate measurements on $|\phi^+\rangle$ produce the maximal violation of the Clauser–Horne–Shimony–Holt inequality $-2 \leq K \leq 2$.

It’s convenient to write $E = K/4$. Then the optimal value of $E$ for classical common cause correlations, or correlations that can be simulated with local resources, is $E_L = 1/2$, and the optimal quantum value is $E_Q = 1/\sqrt{2}$. The difference is the square root in the quantum case! After a similar analysis showing why a classical computer can’t simulate arbitrary quantum correlations, Feynman commented:

I’ve entertained myself always by squeezing the difficulty of quantum mechanics into a smaller and smaller place, so as to get more and more worried about this particular item. It seems to be almost ridiculous that you can squeeze it to a numerical question that one thing is bigger than another. But there you are—it is bigger than any logical argument can produce, if you have this kind of [classical] logic.
3.3 Simulating Popescu–Rohrlich Bananas

The Popescu–Rohrlich correlation is designed to achieve a value of $E_{PR} = 1$, which is the maximum value of $E$ for no-signaling correlations (corresponding to $K = 4$, when the first three terms take the value 1 and the term with the $-$ sign takes the value $-1$).

In Bananaworld there are bananas that exhibit classical correlations, with a value of $E$ that is less than or equal to $1/2$, as well as quantum correlations, where this correlation can take values up to $1/\sqrt{2}$, and superquantum no-signaling correlations, which can take values between $1/\sqrt{2}$ and 1. The question raised by Popescu and Rohrlich was why we live in a world in which correlations are limited by the Tsirelson bound $1/\sqrt{2}$, rather than the no-signaling bound 1. Is there some principle about our world that limits the Clauser–Horne–Shimony–Holt correlation to values that are less than or equal to $1/\sqrt{2}$? I’ll take up this question in Chapter 9, Why the Quantum?

The bottom line

- The probability of successfully simulating the Popescu–Rohrlich correlation is $\frac{1}{2}(1 + \frac{K}{4})$. In a simulation with local resources, $K_L = 2$, so the optimal probability of a successful simulation with local resources is $3/4$.

- If Alice and Bob are allowed to base their simulation strategy on shared entangled quantum states, they can achieve a value of $K_Q = 2\sqrt{2}$ with maximally entangled states like $|\phi^+\rangle$. So with quantum resources, the optimal probability of a successful simulation increases to $\frac{1}{2}(1 + \frac{2\sqrt{2}}{4}) \approx 0.85$.

- The quantum value $K_Q = 2\sqrt{2}$ is called the Tsirelson bound. It’s convenient to write $E = K/4$. Then the optimal value of $E$ for classical common cause correlations, or equivalently correlations that can be simulated with local resources, is $E_L = 1/2$, the optimal quantum value is $E_Q = 1/\sqrt{2}$, and the optimal value for a PR box or Popescu–Rohrlich bananas is $E_{PR} = 1$, which is also the maximum value for no-signaling correlations.

3.3 Simulating Popescu–Rohrlich Bananas

In the simulation game, a moderator gives Alice and Bob separate prompts, $S$ or $T$, at each round of the game, and Alice and Bob are each supposed to respond with a 0 or a 1. They win the round if their responses agree with the correlations for Popescu–Rohrlich bananas:

- if the peelings are $SS, ST, TS$, the tastes are the same, 00 or 11

- if the peelings are $TT$, the tastes are different 01 or 10

- the marginal probabilities for the tastes 0 or 1 if a banana is peeled $S$ or $T$ are $1/2$ (so the no-signaling constraint is satisfied)

If Alice and Bob are allowed access to local resources, like shared random lists of 0’s and 1’s prepared before the start of the simulation, the probability of winning a round of the game is at most
With entangled quantum states as a resource, Alice and Bob can achieve a success probability of approximately .85.

Figure 3.5: The polarization directions $A, A'$ and $B, B'$ for the optimal quantum simulation of Popescu-Rohrlich bananas.

Here’s the way they do it. Before the simulation starts, Alice and Bob prepare many pairs of photons in the maximally entangled state $|\phi^+\rangle = |0\rangle|0\rangle + |1\rangle|1\rangle$. They each store one photon from each pair for later measurements during the simulation, and they keep track of the different pairs of entangled photons, so that the photons they measure at each round belong to the same entangled state. The strategy is for Alice to measure the polarization of her photons in directions $0$ and $\pi/4$, represented by the polarization observables $A$ and $A'$, when she gets the prompt $S$ or $T$, respectively, and for Bob to measure the polarization of his photons in directions $\pi/8$ and $-\pi/8$, represented by the polarization observables $B$ and $B'$, when he gets the prompt $S$ or $T$, respectively. (Up to now, I’ve indicated the angles of polarization directions in degrees. It’s more convenient here and in subsequent sections to use radians, where $90^\circ = \pi/2, 45^\circ = \pi/4,$ and $22.5^\circ = \pi/8$.)

For measurements of a pair of polarization observables $A, B$ or $A', B'$ or $A, B'$, $A', B$ on two entangled photons, the angle between the polarization directions is $\pi/8$, and so the probability that they get the same outcome, both horizontally polarized or both vertically polarized in the directions in which the polarizations are measured, is $\cos^2(\frac{\pi}{8}) \approx .85$, and the probability that they get different outcomes is $\sin^2(\frac{\pi}{8}) \approx .15$. For a measurement of the polarization observables $A', B'$, though, the angle between the polarization directions is $3\pi/8$, and so the probability that they get the same outcome is $\cos^2(\frac{3\pi}{8}) = \sin^2(\frac{\pi}{8})$ and the probability that they get different outcomes is $\sin^2(\frac{3\pi}{8}) = \cos^2(\frac{\pi}{8})$. (If your trigonometry is a little rusty, see the summary of useful trigonometric relations in the subsection Some Useful Trigonometry in the More section at the end of the chapter.)

If the possible values of the polarization observables $A$ and $A'$ for Alice’s photons are represented as 1 for horizontal polarization in the directions of the polarization measurements, and $-1$ for vertical polarization in the directions orthogonal to the polarization measurements, and similarly for the polarization observables $B$ and $B'$ for Bob’s photons, then the expectation value of $AB$ in
3.4 Loopholes

the state $|\phi^+\rangle$ is:

$$
\langle AB \rangle_{|\phi^+\rangle} = p(\text{outcomes same}|A, B) - p(\text{outcomes different}|A, B)
= \cos^2\left(\frac{\pi}{8}\right) - \sin^2\left(\frac{\pi}{8}\right)
= \cos\left(2 \cdot \frac{\pi}{8}\right)
= \cos \frac{\pi}{4} = \frac{1}{\sqrt{2}}
$$

and similarly for $\langle AB' \rangle_{|\phi^+\rangle}$ and $\langle A'B \rangle_{|\phi^+\rangle}$. But $\langle A'B' \rangle_{|\phi^+\rangle} = \sin^2\left(\frac{\pi}{8}\right) - \cos^2\left(\frac{\pi}{8}\right) = -\cos \frac{\pi}{4} = -\frac{1}{\sqrt{2}}$.

The Clauser–Horne–Shimony–Holt quantity $K$ for these polarization measurements is therefore:

$$
K = \langle AB \rangle_{|\phi^+\rangle} + \langle AB' \rangle_{|\phi^+\rangle} + \langle A'B \rangle_{|\phi^+\rangle} - \langle A'B' \rangle_{|\phi^+\rangle}
= 4 \cdot \frac{1}{\sqrt{2}} = 2\sqrt{2}
$$

So with this strategy, exploiting polarization measurements on many pairs of photons in the maximally entangled state $|\phi^+\rangle$, Alice and Bob can successfully simulate the Popescu–Rohrlich correlation with probability $\frac{1}{2}(1 + K) = \frac{1}{2}(1 + 2\sqrt{2}) \approx .85$. This turns out to be the optimal quantum probability, and $2\sqrt{2}$ is the maximum quantum value of $K$, the Tsirelson bound.

3.4 Loopholes

There are basically two possible loopholes to worry about in an experiment designed to test Bell’s inequality: a locality loophole and a detection loophole. If the time interval between measurements on the two entangled photons isn’t short enough to prevent a signal from transmitting information between the photons, there’s a locality loophole. The first attempt to close the locality loophole was an experiment by Aspect, Dalibard, and Roger in 1982. They used an ingenious setup where the directions in which the polarizations of the two photons are measured changes rapidly while the photons are moving towards the analyzers, so information about the polarization observable measured on one photon, and the outcome of the measurement, would have to be transmitted faster than light to reach the remote system by the time that system’s polarization is measured.

To get a feel for the detection loophole, suppose that in simulating the correlation of Popescu–Rohrlich bananas Alice is allowed to reject half her prompts and ask for a new prompt. Then Alice and Bob could successfully simulate the correlation with local resources. They begin by sharing two random variables, $\lambda$ and $\mu$, which can each be 0 or 1. For half the rounds, on average, $\mu = 0$. The strategy is for Alice to accept the prompt $T$ for these rounds, but to reject the prompt $S$, in which case the round is aborted. Then for all the accepted rounds of the simulation when $\mu = 0$, the prompts are either $TS$ or $TT$. Alice responds with the value $\lambda$. Bob responds with $\lambda$ if his prompt is $S$, but if his prompt is $T$, he flips the value of $\lambda$ and responds with 1 if $\lambda = 0$ and 0 if
$\lambda = 1$. So Alice and Bob agree for the prompts $TS$ and disagree for the prompts $TT$. When $\mu = 1$, the strategy is for Alice to reject the prompt $T$. Then for all accepted rounds of the simulation with $\mu = 1$, the prompts are $SS$ or $ST$. For these rounds, Alice and Bob both respond with the value of $\lambda$. So they agree for the prompts $SS$ and $ST$. Taking the two cases together, Alice and Bob agree for the prompts $SS, ST, TS$, and disagree for the prompt $TT$, as required.

The point here is that if a photon counter or other detection device simply fails to register an outcome in half the cases in an experimental test of Bell’s inequality—corresponding to Alice rejecting her prompt and aborting the round—the correlation of entangled photons could be derived from a theory in which the correlated photons share hidden variables. In other words, a common cause wouldn’t be ruled out. Of course, a 50% efficiency is pretty bad, but it raises the question about how efficient measuring instruments need to be to detect a violation of Bell’s inequality. Nicolas Gisin and Bernard Gisin have shown that if Alice and Bob are each allowed to reject a prompt in $1/3$ of the rounds, corresponding to detectors with an efficiency of a much as $2/3$, they could simulate the Popescu–Rohrlich correlation perfectly. Good photon detectors are better than this and can achieve an efficiency of more than 90%.

There’s a fairly large literature on loophole-free tests of Bell’s inequality, involving the question of detector efficiency as well as the problem of ensuring that the entangled quantum systems are spacelike separated at the times of their respective measurements, so that any signal carrying information between them would have to violate the relativistic constraint on traveling faster than the speed of light. There are experiments that close the locality loophole and experiments that close the detection loophole. An ingenious loophole-free 2015 experiment performed by Hensen et al.

3.5 More

3.5.1 Some Useful Trigonometry

For those who have forgotten their trigonometry, here are some trigonometric relations that are used at various places in the book. I’ll represent angles as radians rather than degrees: $\pi = 180^\circ$, $\pi/2 = 90^\circ$, $\pi/4 = 45^\circ$, $\pi/8 = 22.5^\circ$.

To begin:

$$\cos(x \pm y) = \cos x \cos y \mp \sin x \sin y$$

So

$$\cos 2\theta = \cos^2 \theta - \sin^2 \theta = 1 - 2 \sin^2 \theta = 2 \cos^2 \theta - 1$$

since $\cos^2 \theta + \sin^2 \theta = 1$. 
Here’s another way of putting these expressions, in terms of the relation between the angle and the half angle, rather than the relation between twice the angle and the angle:

\[
1 - \cos \theta = 2 \sin^2 \frac{\theta}{2} \\
1 + \cos \theta = 2 \cos^2 \frac{\theta}{2}
\]

The expression for \( \cos(x + y) \) as \( \cos x \cos y - \sin x \sin y \) is useful for deriving relations like \( \cos(\theta + \frac{\pi}{2}) = \cos \theta \cos \frac{\pi}{2} - \sin \theta \sin \frac{\pi}{2} = -\sin \theta \), because \( \cos \frac{\pi}{2} = 0 \) and \( \sin \frac{\pi}{2} = 1 \). From the expression for \( \cos(x - y) \) as \( \cos x \cos y + \sin x \sin y \) you can derive

\[
\cos \frac{3\pi}{8} = \cos(\frac{\pi}{2} - \frac{\pi}{8}) = \cos \frac{\pi}{2} \cos \frac{\pi}{8} + \sin \frac{\pi}{2} \sin \frac{\pi}{8} = \sin \frac{\pi}{8}
\]

Also

\[
\sin(x \pm y) = \sin x \cos y \pm \cos x \sin y
\]

so

\[
\sin 2\theta = 2 \sin \theta \cos \theta
\]

and

\[
\sin(\frac{3\pi}{8}) = \sin(\frac{\pi}{2} - \frac{\pi}{8}) = \cos \frac{\pi}{8}
\]

Finally,

\[
\cos \frac{\pi}{4} = \sin \frac{\pi}{4} = \frac{1}{\sqrt{2}}
\]

and

\[
\cos^2 \frac{\pi}{8} = \frac{2 + \sqrt{2}}{4} \\
\sin^2 \frac{\pi}{8} = \frac{2 - \sqrt{2}}{4}
\]

3.5.2 Correlations

A remarkable feature of quantum mechanics is that there are nonlocal quantum correlations associated with entangled quantum states that are inconsistent with any explanation by a direct causal connection between the events or by a common cause.

Correlations that have a common cause are \textit{conditionally statistically independent} with respect to the common cause. I’ll show below that conditional statistical independence is equivalent to two conditions, sometimes called ‘outcome independence’ and ‘parameter independence.’

First a brief review of some elementary background on conditional probability to clarify terminology and notation. Consider again the playing card example. A card is randomly selected from a deck of cards and cut in half. The two half-cards are placed in envelopes and mailed to different addresses. Opening the envelopes after their arrival will reveal perfectly correlated colors, but
nothing travels between the two addresses as the cause of the correlation. The common cause in this case is a variable associated with the card from which the two half cards were cut, with values, black or red, that fix the color of the half-cards. In the general case, a common cause could be probabilistic and fix only the probabilities of the states of the correlated systems, with probabilities between 0 and 1.

Call the color observables of the half-cards mailed to different addresses $A$ and $B$. I’ll represent particular colors, values of $A$ and $B$, by $a$ and $b$. The probability that a half card at the first address has a particular color, $p(a)$, where $a$ is ‘red’ or ‘black,’ is 1/2. But the conditional probability, $p(a|b)$, that a half-card at the first address has a particular color $a$, given that the half-card at the second address has a particular color $b$, is 1 if $a = b$ and 0 if $a \neq b$, because the two half-cards always have the same color. The vertical slash sign ‘$|$’ indicates a conditional probability: in this case, the probability of the event $a$ to the left of the slash, given the condition indicated by the event $b$ to the right of the slash.

The joint probability $p(a, b)$ that $A$ has the value $a$ and $B$ has the value $b$, and the conditional probability, are related as in the following way:

$$p(a, b) = p(a|b)p(b) = p(b|a)p(a)$$

where $p(a|b)$ is the conditional probability that $A$ has the value $a$, given that $B$ has the value $b$, and $p(b|a)$ is the conditional probability that $B$ has the value $b$, given that $A$ has the value $a$.

The colors are uncorrelated or statistically independent if the color of a half card is irrelevant to the color of the half card at the remote address—if $p(a|b) = p(a)$ or $p(b|a) = p(b)$. These conditions are equivalent: one follows from the other because of their relation to the joint probability. Equivalently, the colors are uncorrelated if $p(a, b) = p(a)p(b)$. This follows if you replace $p(a|b)$ by $p(a)$ or $p(b|a)$ by $p(b)$ in the expression for the joint probability in terms of the conditional probability.

The colors are correlated or statistically dependent if $p(a, b) \neq p(a)p(b)$—if the joint probability of a pair of colors, $a$ and $b$, is not equal to the product of the probabilities of the colors at the two locations.

In the scenario considered, where a card is cut in half and the two half-cards are mailed to different addresses, the colors of the half-cards are correlated according to this definition. Labeling the two addresses as 1 and 2, $p(\text{red, red}) = 1/2$ and $p_1(\text{red})p_2(\text{red}) = 1/4$, because $p_1(\text{red}) = p_2(\text{red}) = 1/2$, so $p(\text{red, red}) \neq p_1(\text{red})p_2(\text{red})$, and similarly, $p(\text{black, black}) \neq p_1(\text{black})p_2(\text{black})$. Also, $p(\text{red, black}) \neq p_1(\text{red})p_2(\text{black})$ because $p(\text{red, black}) = 0$ and $p_1(\text{red})p_2(\text{black}) = 1/4$. So $p(a, b) \neq p(a)p(b)$, since $p(a, b)$ is either 1/2 or 0, and $p(a)p(b) = 1/4$.

Now for conditional statistical independence. To leave open the possibility that observables might be indefinite before they are measured, $p(a, b)$ could written explicitly as $p(a, b|A, B)$, the conditional probability that the values of the two observables are $a$ and $b$, given that the observables $A$ and $B$ are measured at the separate locations. This notation is redundant for the playing card example, since cards have a definite color even before opening the envelope at either address. In the case of Bananaworld correlations the notation isn’t redundant, because a banana only has a definite
3.5 More

taste after it’s peeled a certain way, which corresponds to measuring an observable and recording a
definite outcome. Conditional statistical independence is expressed formally as

\[ p(a, b|A, B, \lambda) = p(a|A, \lambda)p(b|B, \lambda) \]

Given the common cause \( \lambda \), peelings and tastes are uncorrelated: the joint probability of a pair
of tastes for the pair of bananas, conditional on the common cause \( \lambda \) and a pair of peelings,
\( p(a, b|A, B, \lambda) \), is equal to the product of the conditional probabilities for the separate bananas.
Here \( A \) and \( B \) denote Alice’s peeling and Bob’s peeling, respectively, each \( S \) or \( T \), and \( a \) and \( b \)
denote the respective tastes of Alice’s banana and Bob’s banana, ordinary (0) or intense (1). The
joint conditional probability is said to be factorizable. This is Bell’s locality condition.

Averaging over \( \lambda \)—adding the terms weighted by the \( \lambda \)-probabilities—produces the correlations
for the averaged probabilities. So the joint probability—the observed or operational or ‘surface’
joint probability (as opposed to the ‘hidden’ conditional \( \lambda \) probability)—is not equal to the product
of the marginal probabilities for Alice and Bob separately:

\[ p(a, b|A, B) \neq p(a|A)p(b|B) \]

Conditional statistical independence, the mark of a common cause, is equivalent to two condi-
tions on the probabilities:\[21\]

- **outcome independence**: the probability that a banana tastes ordinary or intense, given \( \lambda \) and
  a particular peeling (\( S \) or \( T \)), is independent of the taste of the paired banana after it is peeled

- **parameter independence**: the probability that a banana tastes ordinary or intense, given \( \lambda \)
  and a particular peeling (\( S \) or \( T \)) is independent of how you peel the paired banana

The point of the distinction is that the taste of a peeled banana (the outcome of a particular
peeling) is not under Alice’s or Bob’s control, but we suppose that how Alice or Bob peel their
bananas is something they can choose freely (the ‘parameter’ choice can be \( S \) or \( T \)). See the next
chapter, Really Random, for more on this ‘free choice’ assumption.

What’s the difference between parameter independence and the no-signaling principle? Parameter
independence is the condition ‘no signaling, given \( \lambda \).’ What’s excluded is the possibility of
signaling by exploiting access to \( \lambda \). Parameter independence refers explicitly to \( \lambda \), the common
cause. The no-signaling principle is an observational or operational condition characterizing the
surface phenomenon—it doesn’t refer to \( \lambda \). If there is a common cause \( \lambda \), the no-signaling principle
is ‘no signaling, conditional on \( \lambda \), averaged over \( \lambda \).’

I’ll write \( p^{AB}_\lambda(a, b) \) for the joint probability \( p(a, b|A, B, \lambda) \), and \( p^{AB}_\lambda(a|b) \) for the conditional
probability \( p(a|b, A, B, \lambda) \). Since \( A, B, \) and \( \lambda \) appear throughout, it’s easier to see what’s going on
if these symbols are represented as subscripts and superscripts in this way, rather than putting them
on the right hand side of the conditional symbol ‘|’. Then outcome independence is the condition

\[
\begin{align*}
p^{AB}_\lambda(a|b) &= p^{AB}_\lambda(a) \\
p^{AB}_\lambda(b|a) &= p^{AB}_\lambda(b)
\end{align*}
\]
and parameter independence is the condition

\[ p_{\lambda}^{AB}(a) = p_{\lambda}^{A}(a) \]
\[ p_{\lambda}^{AB}(b) = p_{\lambda}^{B}(b) \]

The ‘parameter’ here is the remote observable.

To derive conditional statistical independence from parameter independence and outcome independence, begin by expressing the joint probability \( p_{\lambda}^{AB}(a, b) \) in terms of the conditional probability \( p_{\lambda}^{AB}(a|b) \) as

\[ p_{\lambda}^{AB}(a, b) = p_{\lambda}^{AB}(a|b)p_{\lambda}^{AB}(b) \]

(Don’t be thrown off by the superscripts and subscripts here. Think of the superscripts and subscripts in \( p_{\lambda}^{AB}(a|b) \) and \( p_{\lambda}^{AB}(b) \) as defining new probabilities \( p'(a|b) \) and \( p'(b) \). Then the equation reads: \( p'(a, b) = p'(a|b)p'(b) \), which is just the definition of conditional probability.) By outcome independence, \( p_{\lambda}^{AB}(a, b) = p_{\lambda}^{AB}(a)p_{\lambda}^{AB}(b) \), and by parameter independence, \( p_{\lambda}^{AB}(a, b) = p_{\lambda}^{A}(a)p_{\lambda}^{B}(b) \), which is conditional statistical independence (expressed above as \( p(a, b|A, B, \lambda) = p(a|A, \lambda)p(b|B, \lambda) \)).

To go the other way and derive parameter independence from conditional statistical independence, begin with conditional statistical independence:

\[ p_{\lambda}^{AB}(a, b) = p_{\lambda}^{A}(a)p_{\lambda}^{B}(b) \]

In Bananaworld, the two values of \( b \) are 0 and 1, so \( p_{\lambda}^{AB}(a, 0) = p_{\lambda}^{A}(a)p_{\lambda}^{B}(0) \) and \( p_{\lambda}^{AB}(a, 1) = p_{\lambda}^{A}(a)p_{\lambda}^{B}(1) \). Adding \( p_{\lambda}^{AB}(a, 0) \) and \( p_{\lambda}^{AB}(a, 1) \) gives the marginal probability \( p_{\lambda}^{AB}(a) \), the sum of the probabilities of \( a \) for all possible values of \( b \). So \( p_{\lambda}^{AB}(a) = p_{\lambda}^{A}(a)(p_{\lambda}^{B}(0) + p_{\lambda}^{B}(1)) \). Since \( p_{\lambda}^{B}(0) + p_{\lambda}^{B}(1) = 1 \) (because 0 and 1 are the only two possibilities), it follows that

\[ p_{\lambda}^{AB}(a) = p_{\lambda}^{A}(a) \]

which is parameter independence for the outcome \( a \). Similarly, \( p_{\lambda}^{AB}(b) = p_{\lambda}^{B}(b) \) follows by adding the two expressions \( p_{\lambda}^{AB}(0, b) = p_{\lambda}^{A}(0)p_{\lambda}^{B}(b) \) and \( p_{\lambda}^{AB}(1, b) = p_{\lambda}^{A}(1)p_{\lambda}^{B}(b) \) to get the marginal probability \( p_{\lambda}^{AB}(b) \).

To derive outcome independence from conditional statistical independence, begin again with conditional statistical independence, \( p_{\lambda}^{AB}(a, b) = p_{\lambda}^{A}(a)p_{\lambda}^{B}(b) \), and use the equalities just derived for parameter independence, \( p_{\lambda}^{AB}(a) = p_{\lambda}^{A}(a) \) and \( p_{\lambda}^{AB}(b) = p_{\lambda}^{B}(b) \), to write conditional statistical independence as

\[ p_{\lambda}^{AB}(a, b) = p_{\lambda}^{A}(a)p_{\lambda}^{B}(b) \]

From the relation between joint probability and conditional probability, \( p_{\lambda}^{AB}(a, b) \) can also be expressed as \( p_{\lambda}^{AB}(a, b) = p_{\lambda}^{A}(a)p_{\lambda}^{AB}(b|a) \). Taking these two expressions together gives

\[ p_{\lambda}^{AB}(b|a) = p_{\lambda}^{AB}(b) \]

which is outcome independence for the outcome \( b \) with respect to the outcome \( a \). Similarly, \( p_{\lambda}^{AB}(a, b) \) can be expressed as \( p_{\lambda}^{AB}(a, b) = p_{\lambda}^{A}(a|b)p_{\lambda}^{AB}(b) \) from the relation between joint probability and conditional probability, which gives \( p_{\lambda}^{AB}(a|b) = p_{\lambda}^{AB}(a) \), outcome independence for the outcome \( a \) with respect to the outcome \( b \).
3.5 More

The bottom line

- For correlations with a common cause \( \lambda \), the joint probabilities are conditionally statistically independent with respect to \( \lambda \), which is to say that the joint probabilities conditional on \( \lambda \) factorize to a product of the marginal probabilities conditional on \( \lambda \):
  \[
  p(a, b|A, B, \lambda) = p(a|A, \lambda)p(b|B, \lambda).
  \]

- Conditional statistical independence is equivalent to two conditions on the probabilities. In Bananaworld, the two conditions are
  
  **Outcome independence**: the probability that a banana tastes ordinary or intense, given \( \lambda \) and a particular peeling (\( S \) or \( T \)), is independent of the taste of the paired banana after it is peeled.

  **Parameter independence**: the probability that a banana tastes ordinary or intense, given \( \lambda \) and a particular peeling (\( S \) or \( T \)), is independent of how you peel the paired banana.

- The point of the distinction is that the taste of a peeled banana (the outcome of a particular peeling) is not under Alice’s or Bob’s control, but we suppose that how Alice and Bob peel their bananas is something they can choose freely (the ‘parameter’ choice can be \( S \) or \( T \)).

- Parameter independence is the condition ‘no signaling, given \( \lambda \).’ The no-signaling principle is an observational or operational condition characterizing the surface phenomenon—it doesn’t refer to \( \lambda \). If there is a common cause \( \lambda \), the no-signaling principle is ‘no signaling, conditional on \( \lambda \), averaged over \( \lambda \).’

3.5.3 Boolean Algebras

In everyday language and classical theories, we talk about objects as having properties that fit together in a certain way when we use words like ‘and,’ ‘or,’ and ‘not,’ and we assume that events fit together in a corresponding way. There’s an implicit structure here that we take for granted without thinking about it.

Consider the playing card example again, and suppose the state of a card is defined by whether it’s a club, spade, diamond, or heart, so the state space of a card is a set of four states: \( \{ \text{club, spade, diamond, heart} \} \). The symbol \( \{ \ldots \} \) denotes the set of elements represented by the entries between the braces. The properties ‘black’ or ‘red’ divide the state space into two subsets: black = \( \{ \text{club, spade} \} \) and red = \( \{ \text{diamond, heart} \} \).

Suppose Alice and Bob play a game of cards where they each play a card at each round of the game. Since a round is associated with a pair of cards, the state space of the game is the Cartesian product of the four-element state space of Alice and the four-element state space of Bob—the set of ordered pairs, \( S \), where the first member of each pair is a state in Alice’s state space and the second member is a state in Bob’s state space. There are \( 4 \times 4 = 16 \) states in the state space.
3 Bananaworld

$S$: \{(club, club), (club, spade), (club, diamond), (club, heart), (spade, club), (spade, spade), \ldots, (heart, heart)\}. Every event—that Alice plays a heart, or Bob plays a black card—is represented by a subset of the state space. For example, the event ‘Alice plays a heart’ is represented by the subset \{(heart, club), (heart, spade), (heart, diamond), (heart, heart)\}.

A classical disjunction (‘or,’ in the inclusive sense: ‘this or that, or both’), say ‘Alice played a heart or a red card,’ is represented by the set-theoretical union of the subsets associated with ‘Alice played a heart’ and ‘Alice played a red card.’ The union of two subsets $S_1$ and $S_2$ is the subset, denoted by $S_1 \cup S_2$, consisting of all the elements (ordered pairs of states) that are in either $S_1$ or $S_2$. In this case the union is just the eight-element subset associated with ‘Alice played a red card,’ since the ‘Alice played a heart’ subset is included in the ‘Alice played a red card’ subset: \{(diamond, club), (diamond, spade), (diamond, diamond), (diamond, heart), (heart, club), (heart, spade), (heart, diamond), (heart, heart)\}. This eight-element subset of the state space is the set-theoretical complement of the subset associated with ‘Alice played a black card.’ The complement of a subset $S$ is the subset $S'$ of elements not in $S$, and every card that is not a red card is a black card.

A classical conjunction (‘and’), say ‘Alice played a club and Bob played a red card,’ is represented by the set-theoretical intersection of the subsets associated with ‘Alice played a club’ and ‘Bob played a red card.’ The intersection of two subsets $S_1$ and $S_2$ is the subset, denoted by $S_1 \cap S_2$, consisting of all the elements that are in both $S_1$ and $S_2$. In this case the intersection is the two-element subset: \{(club, diamond), (club, heart)\}.

If the proposition ‘Alice played a club’ is true, then it follows that the proposition ‘Alice played a club or a red card’ is true. This logical relation is represented by the inclusion of the ‘Alice played a club’ subset $T$ in the ‘Alice played a club or a red card’ subset $U$, denoted by $T \subseteq U$.

These set-theoretical structural relations were formalized by George Boole in the mid-19th century as an algebraic calculus, now called a Boolean algebra. Here’s the idea. The union of two subsets, $S_1$ and $S_2$, is the smallest subset containing both $S_1$ and $S_2$. So in the collection of subsets of the state space $S$, it’s the least upper bound of $S_1$ and $S_2$ with respect to the ordering defined by set inclusion $\subseteq$. It’s an upper bound of $S_1$ and $S_2$ because $S_1 \subseteq (S_1 \cup S_2)$ and $S_2 \subseteq (S_1 \cup S_2)$. It’s the least upper bound because $S_1 \cup S_2$ is the smallest subset that includes both $S_1$ and $S_2$. Similarly, since the intersection of $S_1$ and $S_2$ is the largest subset contained in both $S_1$ and $S_2$, it’s the greatest lower bound of $S_1$ and $S_2$ with respect to the ordering defined by set inclusion. It’s a lower bound of $S_1$ and $S_2$ because $(S_1 \cap S_2) \subseteq S_1$ and $(S_1 \cap S_2) \subseteq S_2$. Since $S_1 \cap S_2$ is the largest subset that is included in both $S_1$ and $S_2$, it’s the greatest lower bound.

A Boolean algebra $B$, defined abstractly, is a set of elements $a, b, \ldots$ with a partial ordering relation $\leq$ that holds for some elements of the algebra, $x \leq y$, just as the relation $\subseteq$ holds for some subsets of the set $S$. A least upper bound, denoted by $x \lor y$, and a greatest lower bound, denoted by $x \land y$, can be defined for any two elements in the algebra with respect to this partial ordering, just as a least upper bound and a greatest lower bound was defined for subsets of the set $S$ with respect to the partial ordering defined by set inclusion $\subseteq$. So now you have a structure with two binary operations $\lor$ and $\land$. To complete the algebra you need a complement, $x'$, for every element $x$ in the algebra (corresponding to the complement of a subset in the state space $S$), a maximum element with respect to the ordering, represented by $1$ (corresponding to the entire state space $S$),
and a minimum element with respect to the ordering, represented by 0 (corresponding to the empty set): $x \leq 1$ and $0 \leq x$ for every element in the algebra.

The Boolean algebra corresponding to the four-element state space {club, spade, diamond, heart} of a playing card is a sixteen-element Boolean algebra. The minimal non-0 elements in a Boolean algebra are referred to as Boolean atoms. These are the one-element subsets \{club\}, \{spade\}, \{diamond\}, \{heart\}. They are minimal elements in the algebra, because there are no elements between the one-element subsets and the empty set or 0 element. Apart from the four atoms, there are the two maximum and minimum elements 0 and 1, the four complements of the atoms (called co-atoms) corresponding to the subsets representing not-club (\{spade, diamond, heart\}), not-spade (\{club, diamond, heart\}), not-diamond (\{club, spade, heart\}), not-heart (\{club, spade, diamond\}), and the six elements corresponding to the subsets: \{club, spade\}, \{club, diamond\}, \{club, heart\}, \{spade, diamond\}, \{spade, heart\}, \{diamond, heart\}.

The operations $\land$ and $\lor$ are associative: $(a \land b) \land c = a \land (b \land c)$, and similarly for $\lor$, so you can drop the parentheses without ambiguity. If the atoms corresponding to club, spade, diamond, heart, are denoted by $c, s, d, h$, the co-atoms are represented by $c' = s \lor d \lor h$, $s' = c \lor d \lor h$, $d' = c \lor s \lor h$, $h' = c \lor s \lor d$. If there are $n$ atoms in a Boolean algebra, the number of elements in the algebra is $2^n$, so the Boolean algebra corresponding to the four-atom state space of a card has $2^4 = 16$ elements, and the sixteen-atom state space of the card game has $2^{16} = 65,536$ elements.

A Boolean algebra defined in this way as a partially ordered structure is referred to as a Boolean lattice. There’s a standard way of representing a Boolean lattice called a ‘Hasse diagram,’ where points represent the Boolean elements and lines represent the partial ordering. If two points are connected by a line, the Boolean element represented by the top point is above the element represented by the lower point in the ordering. To illustrate, the Boolean algebra with four atoms is represented as the Hasse diagram in Figure 3.6.

Classical logic has the structure of a Boolean algebra, and every Boolean algebra is isomorphic to an algebra of subsets of some set. Formally, this means that the semantics of classical logic is set-theoretic: we interpret classical propositions as subsets of a set. A classical proposition that says that a system has a certain property is true if and only if the state of the system is in the set of states associated with the property. The dynamics of a theory describes how the state evolves in time and so describes how a system’s properties change in time. Classical probability theory assumes that the underlying property structure over which probabilities are defined is a Boolean algebra.

This might all seem rather abstract, but here’s the connection with physics: if the observables of a theory all commute, the properties, represented by ‘yes–no’ observables, form a Boolean algebra. The observables in classical Newtonian mechanics all commute, so the properties of classical systems form a Boolean algebra. The state of a classical particle is specified by the position and momentum of the particle, represented by a point in the six-dimensional state space or ‘phase space’ of the particle, with three coordinates for position in the $x, y,$ and $z$ directions, and three corresponding momentum coordinates. The dynamics of classical mechanics describes the trajectory of the particle in phase space, and so how the position and momentum changes over time. Other quantities, like energy, angular momentum, and so on, are functions of position and momentum. So specifying that the energy lies in a certain range amounts to specifying the subset
Figure 3.6: Hasse diagram of a Boolean lattice with four atoms, ♠, ♣, ♦, ♥. The two elements on the top line without labels are (in order, from left to right): not-♦ = {♠, ♣, ♥} and not-♣ = {♠, ♦, ♥}. The four elements without labels on the line below that are (in order, from left to right): {♠, ♦}, {♠, ♥}, {♣, ♦}, {♣, ♥}.

of the state space containing all the states in which the energy, as a function of the state, lies in this range, just as specifying the color of a card as red amounts to specifying the subset of card states, diamonds and hearts, that have the property red. Position and momentum take a continuous range of values, but you could consider just two possible positions for a particle, whether it is in a certain spatial region or not, and two possible momenta, whether the momentum is in a certain range or not. Then the state space is a four-element space, and the structure of properties is the same as the Boolean algebra represented in Figure 3.6. Taking finer subdivisions for position and momentum, or many particles, increases the number of elements in the state space and the number of elements in the associated Boolean algebra, but it doesn’t change any structural feature. Going from a finite or denumerable set of values to a continuum of values introduces technical issues involving infinities, but again this doesn’t change the Boolean character of the underlying structure.

For a playing card as an entity in a game, the most precise specification of the state of a card is the suit and the value of the card: a two of hearts, or an ace of clubs. Any other features are irrelevant. You could consider an even more precise specification, given by the positions and momenta of all the particles in a card, but this would be massively redundant if the only distinctions are those
relevant to the game. The notion of a ‘system’ and associated state space depends in part on the set of properties you want to include in a theory. What the playing card example has in common with classical mechanics is that a state corresponds to a complete list of all the properties of a system. Since each property is associated with a subset of the state space of the system, whether the system is a playing card or a classical particle, specifying all the properties of a system at a particular time amounts to specifying all the subsets containing the state, which is equivalent to specifying the state.

A measurement in a classical theory is then simply a procedure for finding out whether or not a system has a certain property, and since the state is associated with a list of all the properties of the system, nothing prevents you from finding out as many properties of a system as you like, in principle, and so fixing the classical state to an arbitrary degree of accuracy. Even if a measurement procedure involves a dynamical interaction that correlates a property of a system with the pointer value of a measuring instrument and the interaction disturbs the state of the system, you could reverse the measurement interaction and restore the original state, since the dynamics of classical mechanics is reversible, and so build up the list by further measurements.

The point here is that quantum observables don’t all commute, and quantum properties don’t have the structure of a Boolean algebra. The way in which quantum probabilities and quantum correlations differ from classical probabilities and classical correlations arises because of the specific non-Boolean character of the underlying quantum property structure. For example, a classical disjunction is true if and only if one or both disjuncts is true (one or both of the propositions on either side of the ‘or’), but this is not the case for a quantum disjunction. A quantum disjunction can be true without either of the disjuncts having a well-defined truth value, and this difference between the classical and quantum case shows up in the enhanced computational power of quantum computers relative to classical computers. (See section 8.1, Quantum Computation, in Chapter 8, Quantum Feats.) The structure of quantum properties is not even embeddable into a Boolean algebra—there is no one-to-one structure-preserving map from the quantum property structure into a Boolean algebra, which is another way of saying that the probabilities of quantum mechanics do not arise because something has been left out of a Boolean or classical story. Instead of a Boolean algebra, what you have is a collection of Boolean algebras, each corresponding to a mutually commuting set of ‘yes–no’ observables. Observables in different Boolean algebras generally don’t commute, but the Boolean algebras are ‘entwined’ in a particular way, so that an observable can belong to more than one Boolean algebra. What’s meant by a ‘measurement’ is then not the same thing as a measurement in a classical theory, because there is no state in the sense corresponding to a classical list of properties in a non-Boolean theory.

The non-embeddability result is known as the Kochen–Specker theorem. Simon Kochen and Ernst Specker, and Bell in an independent argument, proved that no ‘noncontextual’ hidden variable theory, in which the observables all have definite values prior to measurement fixed by the hidden variables, can reproduce the quantum probabilities of a qutrit or any quantum system more complex than a qubit. (A context is defined by a Boolean algebra of mutually commuting ‘yes–no’ observables. In a contextual theory, the value of an observable can be different for each of the Boolean contexts to which it belongs.) This is equivalent to the non-embeddability of the quantum property structure into a Boolean algebra. See sections 6.2 and 6.3, The Aravind–Mermin Magic
The bottom line

- If the observables of a theory all commute, as in classical Newtonian physics, the properties, represented by ‘yes–no’ observables, form a Boolean algebra. Classical logic has the structure of a Boolean algebra, and classical probability theory assumes that the underlying property structure over which probabilities are defined is a Boolean algebra.

- A measurement in a classical theory is a procedure for finding out whether or not a system has a certain property, and since the state is associated with a list of all the properties of the system, nothing in principle prevents you from finding out as many properties of a system as you like, and so fixing the classical state to an arbitrary degree of accuracy.

- Quantum properties don’t have the structure of a Boolean algebra, and the way in which quantum probabilities and quantum correlations differ from classical probabilities and classical correlations arises because of the specific non-Boolean character of the underlying quantum property structure.

- Instead of a Boolean algebra, what you have is a collection of Boolean algebras, each corresponding to a mutually commuting set of ‘yes–no’ observables. Observables in different Boolean algebras generally don’t commute, but the Boolean algebras are ‘entwined’ in a particular way, so that an observable can belong to more than one Boolean algebra.

- Simon Kochen and Ernst Specker, and Bell in an independent argument, proved that no ‘noncontextual’ hidden variable theory, in which the observables all have definite values prior to measurement fixed by the hidden variables, can reproduce the quantum probabilities of a qutrit or any quantum system more complex than a qubit. (A context is defined by a Boolean algebra of mutually commuting ‘yes–no’ observables. In a contextual theory, the value of an observable can be different for each of the Boolean contexts to which it belongs.) This result, known as the Kochen–Specker theorem, is equivalent to the impossibility of embedding the quantum property structure into a Boolean algebra.

Notes


3. Richard Feynman’s quote ‘It has not yet become obvious to me that there’s no real problem . . .’ is from his paper ‘Simulating physics with computers,’ International Journal of Theoretical Physics 21, 467 (1982). The quote is on p. 471.


   One elegant family of nonlocal hidden variable theories has been investigated in some detail by Bohm and Bub (1966) (their interest in this family having probably arisen more from its amenability to mathematical analysis than from physical heuristics). This family has the property of agreeing statistically with quantum mechanics after the hidden variables become ‘randomized.’ However, immediately after a measurement is performed the hidden variables are in a kind of nonequilibrium distribution, and hence new measurements performed before a certain relaxation time has elapsed can be expected to disagree statistically with quantum mechanics. Bohm and Bub conjecture that this relaxation time is of the order of $\hbar/kT$, where $T$ is the temperature of the apparatus. At room temperature the relaxation time would be about $10^{-13}$ seconds. Papaliolios [‘Experimental test of a hidden-variable quantum theory,’ Physical Review Letters 18, 622-625 (1967)] ingeniously tested their conjecture by measuring the intensity of light passing through a stack of three extremely thin sheets of polaroid with varying orientations of their axes of polarization. The results were entirely in accordance with quantum mechanics, even though the transit time of light through each sheet of polaroid was about $7.5 \times 10^{-14}$ seconds. He remarks that ‘it is also possible to perform a more definitive test of Bohm and Bub’s choice of $\hbar/kT$ as the relaxation time, by repeating the experiment at lower temperatures. The lack of a theoretical understanding of this choice of $\tau$, however, does not at this time justify cooling the apparatus to liquid air (or lower) temperatures.’

6. The original version of the Ghirardi–Rimini–Weber theory was published in GianCarlo C.


8. Bell’s critical review of ‘no go’ results about the impossibility of hidden variable extensions of quantum mechanics satisfying various constraints: ‘On the problem of hidden variables in quantum mechanics,’ *Reviews of Modern Physics* 38, 447–452 (1966). Bell’s comment, referring to Bohm’s hidden variable theory, that the theory evades these no go results because the equations of motion are nonlocal, so that ‘an explicit causal mechanism exists whereby the disposition of one piece of apparatus affects the results obtained with a distant piece,’ is on p. 452. His further comment that ‘the Einstein–Podolsky–Rosen paradox is resolved in the way in which Einstein would have liked least,’ and his query whether ‘any hidden variable account of the quantum mechanics must have this extraordinary character’ are both on p. 452. Bell’s analysis concludes with the proposal:

   It would therefore be interesting, perhaps, to pursue some further ‘impossibility proofs,’ replacing the arbitrary axioms objected to above by some condition of locality, or of separability of distant systems.

   His 1964 proof—Bell’s theorem—was his response to this proposal. It was published two years prior to the long-delayed publication of the paper that raised the issue of locality as a constraint on any hidden variable extension of quantum mechanics.


11. Einstein’s comment in a letter to Max Born, ‘That which really exists in B should therefore not depend on what kind of measurement is carried out in part of space A . . .’ is on p. 164 of *The Born-Einstein Letters*. Born reproduces the comment from Einstein’s marginal ‘caustic comments’ (including ‘Ugh!’ and ‘Blush, Born, Blush’) to the last chapter, ‘Metaphysical

12. Nicolas Gisin’s characterization of ‘no signaling’ as ‘no communication without transmission’ is from his book *Quantum Chance: Nonlocality, Teleportation, and Other Quantum Marvels* (Copernicus, Göttingen, 2014).

13. Sandu Popescu and David Rohrlich introduced the nonlocal box now referred to as a PR box in their paper ‘Quantum nonlocality as an axiom,’ *Foundations of Physics* 24, 379 (1994). They pointed out that relativistic causality by itself does not rule out simulating a PR-box with a probability greater than 3/4. This was actually shown earlier by L. A. Khalfi and B. S. Tsirelson, ‘Quantum and quasi-classical analogs of Bell inequalities,’ in P. Lahti and P. Mittelstaedt (eds.), *Symposium on the Foundations of Modern Physics* (World Scientific, Singapore, 1985), pp. 441–460, and independently by Peter Rastall, ‘Locality, Bell’s theorem, and quantum mechanics,’ *Foundations of Physics* 15, 963–972 (1985). Other nonlocal boxes with interesting correlations include the millionaire box introduced by Andrew C. Yao in ‘Protocols for secure computations,’ *SFCS ’82, Proceedings of the 23rd Annual Symposium on Foundations of Computer Science* pp. 160–164 (1982). The millionaire box is a generalization of a PR box, where the inputs can take any value in the interval between 0 and 1, as well as 0 and 1 (which are the only two possible inputs in the case of a PR box). If Alice’s input is less than or equal to Bob’s input, the outputs of the box are the same. If Alice’s input is greater than Bob’s input, the outputs are different. So two millionaires, Alice and Bob, who are not equally rich, could use such a box to decide who is richer without revealing how much money they each have. They both input the amount of money they own (converted to decimals between 0 and 1). If the outputs are the same, Bob is richer. If the outputs are different, Alice is richer. Needless to say, although the correlation is logically possible, the box is an imaginary device.

14. Nicolas Gisin’s proof of Bell’s theorem is from *Quantum Chance: Nonlocality, Teleportation, and Other Quantum Marvels* (Copernicus, Göttingen, 2014).


17. Feynman’s comment about ‘squeezing the difficulty of quantum mechanics into a smaller and smaller place’ is from Feynman’s article ‘Simulating physics with computers,’ *International Journal of Theoretical Physics* 21, 467 (1982).


21. The terms ‘parameter independence’ and ‘outcome independence’ are due to Abner Shimony, from ‘Contextual hidden variables and Bell’s inequalities,’ *British Journal for the Philosophy of Science* 3, 24–45 (1984).

