CAN YOU HELP YOUR TEAM TONIGHT BY WATCHING ON TV? MORE EXPERIMENTAL METAPHYSICS FROM EINSTEIN, PODOLSKY, AND ROSEN

N. DAVID MERMIN

A few years ago I described (1981a)[1] a simple device that reveals in a very elementary way the extremely perplexing character the data from the Bohm-Einstein-Podolsky-Rosen experiment assumes in the light of the analysis of J. S. Bell. There is a second, closely related, form of that *Gedanken* demonstration,[2] which I would like to examine for several reasons.

1. It is simpler: there are only two (not three) settings for each switch.
2. The *Gedanken* data resemble more closely the data collected in actual realizations of the device.
3. None of the possible switch settings produce the perfect correlations found in the first version of the *Gedanken* demonstration, where the lights *always* flash the same color when the switches have the same setting. Since absolutely perfect correlations are never found in the imperfect experiments we contend with in the real world, an argument that eliminates this feature of the ideal *Gedanken* data can be

Some of the views expressed here were acquired in the course of occasional technical studies of Einstein-Podolsky-Rosen correlations supported by the National Science Foundation in a small corner of grant No. DMR 86–13368.

[1]Reprinted as an appendix below. An only slightly more technical but significantly more graceful version appeared a few years later (Mermin, 1985).

[2]What follows is my attempt to simplify some reformulations of EPR and Bell by Henry Stapp (for example, 1985a), but the interpretation I give differs from his, and any foolishness in what follows is entirely my own.

applied to real data from real experiments. (If you believe, however, along with virtually all physicists, that the quantum theory gives the correct ideal limiting description of all phenomena to which it can be applied, then this is not so important a consideration.)

4. Because the ideal perfect correlations are absent from this version of the *Gedanken* demonstration, one is no longer impelled to assert the existence of impossible instruction-sets. To establish that the new data nevertheless remain peculiar, it is necessary to take a different line of attack, which has again intriguing philosophical implications, but of a rather different character.[3]

1. The modified demonstration

In the modified *Gedanken* demonstration, there are only two switch settings (1 and 2) at each detector. Otherwise the setup is unchanged: there are two detectors (*A* and *B*) and a source (*C*), and the result of each run is the flashing of a red or green light. If one had actually built such a device according to the quantum-mechanical prescription, it could be transformed to run in this modified mode simply by readjusting the angle through which certain internal parts of each detector turned as the switch settings were changed.[4]

In its new mode of operation, the device produces the following data:

(*i*) When the experiment is run with both switches set to 2 (22 runs), the lights flash the same colors only 15% of the time; in 85% of the 22 runs different colors flash.

(*ii*) When the experiment is run with any of the other three possible switch settings (11, 12, or 21 runs) then the lights flash the same colors 85% of the time; in only 15% of these runs do different colors flash.

As in the earlier version of the *Gedanken* experiment, RR and GG are equally likely when the lights do flash the same colors, and RG and GR are equally likely when different colors flash. Also as earlier, the pattern of colors observed at any single detector is entirely random. There is no way to infer

[3]There are more orthodox ways of extracting the peculiar character of these data. The route I take here requires fewer formal probabilistic excursions, and leads to a rather different philosophical point, though I suspect a careful analysis of the use of probability distributions in the conventional arguments might uncover something quite similar.

[4]Physicists might note that if setting 1 at detector *A* corresponds to measuring the vertical spin component, then setting 2 at *A* measures the component at 90° to the vertical; setting 1 at *B*, 45° to the vertical, and setting 2 at *B*, −45° to the vertical, all four directions lying in the same plane. (In the earlier version the three switch settings at either detector corresponded to 0°, 120°, and −120°.) The fraction 85% is just $\cos^2(22.5°) = \frac{1}{4}(2 + \sqrt{2})$.

from the data at one detector how the switch was set at the other. Regardless of what is going on at detector *B*, the data for a great many runs at detector *A* is simply a random string of R's and G's, that might look like this:

Typical Data at Detector A

A: R G R R G G R G R R R G G G R G R R R G R G R G . . .

The choice of switch settings only affects the *relation* between the colors flashed at *both* detectors. If, for example, the above data had been obtained at detector *A* when its switch was set to 2, and in all those runs the switch at *B* had also been set to 2, then, as noted above, the color flashed at *B* would have agreed with that flashed at *A* in only 15% of the runs, and the lights flashed at both detectors together might thus have looked like this:[5]

Data from a Series of 22 Runs

A2: R G R R G G R G R R R G G G R G R R R G R G R G . . .
B2: G R G G G R G R G G R G R R G R G G R R R R G R . . .

Although the list of colors flashed at either detector remains quite random, the color flashed at *B* is highly (negatively) correlated with the color flashed at *A*. In the overwhelming majority (85%) of the runs the detectors flash different colors. Only in a few (15%) of the runs do the detectors flash the same colors.

On the other hand, for any of the other switch settings (take 21 as an example) the comparative data would have looked something like this:

Data from a Series of 21 Runs

A2: R G R R G G R G R R R G G G R G R R R G R G R G . . .
B1: R G R G G G R R R R R G G R G G R R R G G G R G . . .

Again we have two lists of colors, each entirely random, but they now agree with each other in 85% of the runs, disagreeing in only 15%.

There are various ways to run the modified *Gedanken* demonstration, but let me focus on the following procedure, which it seems to me makes a rather striking contribution to Abner Shimony's field of experimental metaphysics. Suppose we do a long series of runs in each of which both switches are set to 1:

---

[5]The numbers after *A* and *B* denote the fixed setting of their switches throughout the sequence of runs. In contrast to some earlier versions of the *Gedanken* demonstration, we now try out various fixed switch settings, rather than randomly resetting the switches after each run.

Data from a Series of 11 Runs

A1: R G G G R G R G R R G R R G R G R R R R G R G G . . .
B1: R G G R R G G R R R G R R G R G R G R R G G G G . . .

About 85% of these 11 runs will produce the same colors, and 15%, different. Now because there are no connections of any kind between the detectors at *A* and *B*, it seems clear that whatever happens at *A* cannot in any way depend on how the switch was set at *B*, and vice versa. Let us elevate this commonsense remark into a principle, which I shall call the Baseball Principle. Before examining the implications of the *Gedanken* demonstration for the Baseball Principle, let us discuss it in the context from which its name derives, where it assumes (at least for me) an especially vivid character.

2. *The Baseball Principle*

I am a New York Mets fan, and when they play a crucial game I feel I should watch on television. Why? Not just to find out what is going on. Somewhere deep inside me, I feel that my watching the game makes a difference—that the Mets are more likely to win if I am following things than if I am not. How can I say such a thing? Do I think, for example, that by offering up little prayers at crucial moments I can induce a very gentle divine intervention that will produce the minute change in trajectory of bat or ball that makes the difference between a hit or an out? Of course not! My feeling is completely irrational. If you insisted that I calm down and think about it, I would have to admit that the outcome of the game does not depend in the least on whether I watch it or not. What I do or do not do in Ithaca, New York, can have no effect on what the Mets do or do not do in Flushing, New York. This is the Baseball Principle.

Now a pedant comes along and says, "What do you really mean by that Baseball Principle?" And then, being a pedant, he tells me what I really mean. What I really mean is this: If we examined a great many Mets games and divided them up into those I watched at least part of on TV and those I did not watch at all, and if my decision to watch or not was entirely independent of anything I knew about the game—made, for example, by tossing a coin—then we would find that the Mets were no more or less successful in those games I watched than in those games I did not.

Now I reply, "That's very nice, but I mean something much simpler. I mean that in each individual game, it doesn't make any difference whether I watch it or not. Tonight, for example, whatever the Mets do, will be exactly the same, whether or not I end up watching the game."

"C'mon," says the pedant, "that's silly. Either you watch the game or you don't. You can't say that what happens in the game in the case that didn't happen is the same as what happens in the case that did, because there's no way to check. What didn't happen *didn't happen*."

I say to the pedant: "Who's being silly here? Are you trying to tell me that it *does* make a difference in tonight's game whether I watch it or not?"

"No," says the pedant, "I'm saying that your statement that it doesn't make any difference whether or not you watch an individual game can only be viewed as a very convenient construct to summarize the more complex statistical statement about correlations between watching and winning over many games. All of its statistical implications are correct, but it has no meaning when applied to an individual game, because there is no way to verify it in the case of the individual game, which you cannot both watch and not watch."

But is it *wrong* to apply the Baseball Principle to an individual game?

3. *The Strong Baseball Principle*

Let us call the claim that the Baseball Principle applies to each individual game the Strong Baseball Principle. The Strong Baseball Principle insists that the outcome of any particular game does not depend on what I do with my television set—that whatever it is that happens *tonight* in Shea Stadium will happen in exactly the same way, whether or not I am watching on TV.

As a rational person, who is not superstitious, and does not believe in telepathy or the efficacy of prayer on the sporting scene, I am convinced of the Strong Baseball Principle. True, there is no way to verify it, since I cannot both watch and not watch tonight's game, and am therefore unable to compare how the game goes in both cases to make sure nothing changes. Nevertheless, deep in my heart, I do believe that because there is no mechanism connecting what I do with the TV at home to what happens in Shea Stadium, the outcome of tonight's game genuinely does not depend on whether I watch it or not: the Strong Baseball Principle. Try as you may to persuade me that the Strong Baseball Principle is meaningless, in my heart, I know it is right.

Remarkably, when run in the second mode, the *Gedanken* demonstration provides us with a case in which if it really does make no difference whether or not I watch the game, then it is not only meaningless, but demonstrably *wrong* to assert this principle in the individual case. If the Baseball Principle is right for the device, then the Strong Baseball Principle must be wrong, not merely because it naively compares possibilities only one of which can be realized, but because it is directly contradicted by certain observed facts. Such an experimental refutation of the Strong Baseball Principle would have been impossible before the discovery of the quantum theory; you cannot get into trouble using the Strong Baseball Principle in classical physics, and it can, in fact, be a powerful conceptual tool.[6] I believe that those who take the view that an experimental refutation is of no interest since reasoning from the Strong Baseball Principle was impermissible all along miss something of central importance for an understanding of the character of quantum phenomena.

4. *The device and the Baseball Principle*

We return from ball games to the device. There are no connections between the detectors or between the source and either detector. The Baseball Principle therefore applies, and asserts that what goes on at detector *A* does not depend on how the switch is set at detector *B*, and vice versa. This is readily verified in the statistical sense insisted on by the pedant. Keep the switch at *A* set to 1. Do a great many runs with the switch at *B* set to 1. Then, keeping the switch at *A* at 1, do a second series of runs with the switch at *B* set to 2. Compare the data at *A* in the two cases. It will have exactly the same character—namely a featureless sequence of R's and G's like the series of heads and tails you get by repeatedly flipping a coin. There is nothing in the outcome at *A* to distinguish between the runs in which *B* was set to 1 or to 2.

But what about the Strong Baseball Principle? Given the lack of any connection between the detectors, can we not also assert that what goes on at one detector in any *individual* run of the experiment does not depend on how the switch is set at the other detector? Granted, there is no way to test this stronger assertion, but surely, for the same reason, there is also no way to refute it. But here, remarkably in my opinion, we have a case in which the Strong Baseball Principle is directly contradicted by the data. Consider what happened when the device was run with both switches set to 1:

Actual Data from a Series of 11 Runs

A1: R G G G R G R G R R G R R G R G R R R R G R G G . . .
B1: R G G R R G G R R R G R R G R G R G R R G G G G . . .

If there are really no connections between *A* and *B*, and no spooky actions at a distance, then what happens at detector *A* cannot depend on how the switch is set at detector *B* (and vice versa). The Strong Baseball Principle takes this to

[6]In a deterministic world in which the future can be calculated from present conditions, the Strong Baseball Principle can be given an unambiguous meaning.

mean that in the first run of this sequence (in which both lights flashed R) the light at detector *A* would have flashed R even if the switch on detector *B* had been set to 2 instead of 1, and similarly, for every other run in the series, if *B* had been set to 2 nothing would have changed at *A*. In no individual run can the outcome at *A* depend on how the switch was set at *B*. (Compare this with "In no individual baseball game can the outcome at Shea Stadium depend on how the switch was set on my TV.")

Well, if that is so, we can say something about what would have happened if the run had been 12 (*A*1 and *B*2) rather than 11 (*A*1 and *B*1)—namely the outcomes at *A* would have been exactly the same as before:[7]

The 11 Runs and What the Strong Baseball Principle Can Say About What Would Have Happened Had They Been 12 Runs

B2: ? . . .
A1: R G G G R G R G R R G R R G R G R R R R G R G G . . .

A1: R G G G R G R G R R G R R G R G R R R R G R G G . . .
B1: R G G R R G G R R R G R R G R G R G R R G G G G . . .

Note that in this application of the Strong Baseball Principle we make no commitment at all to what colors flashed at *B* in the case that did not take place (with the switch at *B* set to 2) since, after all, that did not happen. We merely assert that whatever might have taken place at *B* in that unrealized experiment, nothing would have turned out any differently at *A*.

We can also say the same thing about what would have happened at *B*, if we had set the switch differently at *A*. This gives us one more pair of rows:

The 11 Runs and What the Strong Baseball Principle Can Say About What Would Have Happened Had They Been 12 Runs Or What Would Have Happened Had They Been 21 Runs

B2: ? . . .
A1: R G G G R G R G R R G R R G R G R R R R G R G G . . .

A1: R G G G R G R G R R G R R G R G R R R R G R G G . . .
B1: R G G R R G G R R R G R R G R G R G R R G G G G . . .

B1: R G G R R G G R R R G R R G R G R G R R G G G G . . .
A2: ? . . .

---

[7]This does not imply determinism—indeed, I am not convinced that what happens in a baseball game *is* deterministic; it simply says, in the baseball case, that whatever it is that does happen is not going to depend on what a television set 300 miles away is doing.

Consider now what we have laid out here. The middle two (third and fourth) rows show what actually happened: both switches were set to 1, and the first run gave RR, the second, GG, the third GG, the fourth GR, etc. The top two rows (first and second) express the Strong Baseball Principle in the form that asserts that the outcome of *each individual* run at *A* does not depend on how the switch is set at *B*. The bottom two (fifth and sixth) express it as an assertion that the outcome of each run at *B* does not depend on the switch setting at *A*.

Now what about the question marks? They appear in the top (first) and bottom (sixth) rows because those rows represent what would have happened at *B* and *A* had the switches there been other than what they actually were. Evidently *some* sequence of R's and G's would have been produced in either case,[8] but we have no way of telling which. Experience with the device, however, tells us some of the features these sequences would have had, if the runs had been 12 or 21 runs rather than the 11 run that actually took place. An acceptable sequence of R's and G's for the first (*B*2) row, must agree with the sequence of R's and G's in the second (*A*1) row in about 85% of the positions, since that is the way 12 runs always work. Similarly a sequence of R's and G's replacing the question marks in the sixth row must agree in about 85% of the positions with the sequence in the fifth row, since that is what always happens in 21 runs. These considerations cut down on the number of ways of replacing question marks with R's and G's, but many different possibilities are still allowed.

A final application of the Strong Baseball Principle can be made to restrict these possibilities further. Suppose both switches had been set to 2 rather than 1. We can regard this 22 series of runs either as a modification of a 21 series (modified by changing the switch setting at *B* without changing anything at *A*) or as a modified 12 series (in which the switch was changed at *A* without anything having been done at *B*). We do not know, of course, what would have happened at *B* in the hypothetical 12 series (top row of question marks) or at *A* in the hypothetical 21 series (bottom row of question marks). The Strong Baseball Principle asserts that whatever series of R's and G's at *A* the question marks in the bottom row might stand for in the 21 run, that same series of R's and G's would also have happened at *A* in that series of runs had

[8]At this moment in my talk at the conference, there were cries of protest from the philosophers in the house. I was told that "If I were hungry, I would eat a candy bar" does not imply the proposition "There exists a candy bar which is the one I would eat were I hungry" (the Candy Bar Principle). I affirmed my commitment to the Candy Bar Principle. I said I wanted to make a rather different point, but I think they all stopped listening then and there. I hope you will not stop reading here and now. If you insist on talking candy, I would suggest that a more accurately analogous proposition is "Either there exists a candy bar which is the one I would eat were I hungry or there does not."

the switch at B been set to 2 instead of 1—i.e., had the runs been 22 runs instead. By the same token, whatever sequence of R's and G's the question marks in the top row represented for the results at B in a series of 12 runs, that same sequence would also have described the results at B had the runs been 22 runs.

This last application of the Strong Baseball Principle, by comparing hypothetical cases, has a different character than the first two, which compare a hypothetical case with the real one, and here it might more accurately be termed the Very Strong Baseball Principle. Returning to the sporting analogy, the Very Strong Baseball Principle applies when the game is, in fact, canceled because of rain. I nevertheless maintain that had the game been played, it would have taken place in exactly the same way, whether or not I watched it. This last assertion may elicit an even more violent objection from the pedant. Is it really reasonable to insist that something should happen in exactly the same way when conditions change very far away from it, when in actual fact it never happened at all?

But is it really any more reasonable, I hasten to add, to insist that such an assertion is impermissible? I maintain that if last night's game had not been rained out, it would have happened the same way whether or not I had watched it on television. Can you prove me wrong when I say this? Wouldn't most unsuperstitious people regard the proposition as true? Indeed, as uninterestingly true? To be sure, the pedant will translate it into a series of harmless statistical assertions, but is it really *wrong* to apply it to the individual case as well? The hallmark of the Strong Baseball Principle at work is this nagging conviction, to which only a pedant could object. For how can one possibly get into any trouble asserting relations between two things neither of which actually happened?

One can. It is worse than bad form; it is bad physics. Let us try it out. We have to replace the first row with some sequence of R's and G's and the sixth row with some other such sequence in such a way that the first and second rows give the right statistics for 12 runs, the fifth and sixth, for 21 runs, *and* the first and sixth for 22 runs. We do not insist that any particular way of doing this is preferable to or any more deserving of some hypothetical reality than any other, but for the Strong Baseball Principle to survive, *some* among the various possibilities must be consistent with these statistics.

Now in 22 runs the colors disagree 85% of the time, so whatever goes into the first row has to disagree with whatever goes into the sixth in about 85% of the positions.

On the other hand the set of R's and G's in the top row can differ from that in the second row in only about 15% of the positions (since they must have the correlations appropriate to a series of 12 runs). The second row is the same as the third row (by the Strong Baseball Principle). The third row differs from the fourth row in only about 15% of the positions, since they give the data in a 11 run. The fourth row is the same as the fifth row (by the Strong Baseball Principle). And the fifth row can differ from the set of R's and G's appearing in the bottom row in only 15% of the positions (since those rows must have the correlations appropriate to a series of 21 runs).

A moment's reflection on the last paragraph is enough to reveal that whatever sequence of R's and G's is in the top row, it can differ from whatever sequence is in the bottom row, in at most about 15% + 15% + 15% = 45% of the positions. But according to the next to the last paragraph whatever is in the top row must differ from whatever is in the bottom row in about 85% of the positions. You cannot have it both ways. Thus the (Very) Strong Baseball Principle is so restrictive as to rule out *every* possibility for the unrealized switch settings. Far from merely being meaningless nonsense, an application of the Strong Baseball Principle to the *Gedanken* demonstration contradicts the observed facts.

5. *Conclusion*

In this demolition of the Strong Baseball Principle, we did not interpret it as demanding the existence in some cosmic bookkeeping office of a list of data for the unperformed runs. We only took it to require that if the *actual* experiment consists of a long series of 11 runs, then among all the *possible* sets of data that *might* have been collected had the experiment instead consisted of 12, 21, or 22 runs, there should be *some* satisfying the condition that, run by run, what happens at one detector does not depend on how the switch is set at the other. If the Strong Baseball Principle is valid, it should be possible to *imagine* sets of *B*2 and *A*2 data such that the *B*2 data produce the right statistics (85% same and 15% different) when combined with the actual *A*1 data, the *A*2 data produce the right statistics (85% same and 15% different) when combined with the actual *B*1 data, and the two sets of imagined data produce the right statistics (15% same and 85% different) for a 22 experiment.[9]

Since it is impossible to imagine *any* such sets of data, then the Strong Baseball Principle has to be abandoned not because it is bad form, unjustifiable, or frivolous to argue from what might have happened but did not, but

[9]In Candy Bar terms, the Strong Baseball Principle does not say that there exists a particular sequence of R's and G's which are the colors that would have flashed had a detector been set differently. It only says that among all the mutually exclusive and exhaustive possibilities for such sequences should be *some* that are consistent with the frequencies of flashings characteristic of the four different pairs of switch settings.

because there are no conceivable sets of data for the cases that might have happened but did not, which are consistent with the numerical constraints imposed by the known behavior of the device, when those constraints are further restricted by the Strong Baseball Principle.

This attack is inherently nonclassical. If, in the best *Gedanken* demonstration I could devise, the 85% and 15% had been replaced by 75% and 25%, then the argument would have collapsed. For instead of the top row being able to differ from the bottom by no more than 15% + 15% + 15% = 45%, which is manifestly less than the required 85%, it would only have been possible to bound the difference by 25% + 25% + 25%, which is just enough to provide the required 75%. Only by exploiting *quantum* correlations can one construct an 85%–15% *Gedanken* demonstration. Any model of the device one might devise based on classical physics would necessarily result in 75%–25% or less extreme statistics, and the Strong Baseball Principle would be immune from this kind of refutation by physicists, no matter how dim a view of it philosophers took. I assert this with confidence because classical physics is local and deterministic and in a deterministic world the Strong Baseball Principle makes perfect sense as a manifestation of locality.

Going in the other direction, it is easy to invent fictitious *Gedanken* demonstrations that produce data that refute the Strong Baseball Principle even more resoundingly than does the device. Consider, for example, a hypothetical device in which 85% and 15% were replaced by 100% and 0%, so that the lights always (not just most of the time) flashed the same colors in 11, 12, and 21 runs, and never (not just infrequently) flashed the same colors in 22 runs. Then the argument refuting the Strong Baseball Principle would be even simpler. An 11 run would necessarily result in the same color (say R) at *A* and *B*. Suppose instead the switch at *A* had been set to 2. The Strong Baseball Principle would then assert that R would still have flashed at *B* and since the same colors always flash in 12 runs, *A* would still have flashed R. By the same token *B* would still have flashed R had its switch been set to 2. Therefore, since the setting of the switch at one detector cannot affect what happens at the other, both would have flashed R if both had been set to 2. But when both are set to 2, both have to flash different colors.

No experiment is known that can provide this more compact refutation. Even quantum miracles can go only so far. The 85%–15% statistics are the most extreme I know how to extract from the quantum theory, and although they are strong enough to demolish the Strong Baseball Principle, the argument we went through is somewhat less direct than that available for the 100%–0% statistics.

It is a characteristic feature of all quantum conundrums that something has to have a nonvanishing probability of happening in two or more mutually exclusive ways for startling behavior to emerge. The viewpoints of quantum and classical physics are distinguished, more than anything else, by the impropriety in quantum physics of reasoning from an exhaustive enumeration of two or more such possibilities in cases that might have happened but did not. We are startled when such reasoning fails because as an analytical tool in classical physics and everyday life, it is not only harmless but often quite fruitful. The most celebrated of all quantum conundrums—how can there be a diffraction pattern when the electron had to go through one slit or the other?—is based on precisely this impropriety. It is just where there is room for some interplay between various unrealized possibilities, that one can look for the quantum world to perform for us the most magical of its tricks.

Therefore it is wrong to apply the principle that what happens at *A* does not depend on how the switch is set at *B* to individual runs of the experiment. Many people want to conclude from this that what happens at *A does* depend on how the switch is set at *B*, which is disquieting in view of the absence of any connections between the detectors. The conclusion can be avoided, if one renounces the Strong Baseball Principle, maintaining that indeed what happens at *A* does not depend on how the switch is set at *B*, but that this is only to be understood in its statistical sense, and most emphatically cannot be applied to individual runs of the experiment. To me this alternative conclusion is every bit as wonderful as the assertion of mysterious actions at a distance. I find it quite exquisite that, setting quantum metaphysics entirely aside, one can demonstrate directly from the data and the assumption that there are no mysterious actions at a distance, that there is no conceivable way consistently to apply the Baseball Principle to individual events.

APPENDIX: QUANTUM MYSTERIES FOR ANYONE

> We often discussed his notions on objective reality. I recall that during one walk Einstein suddenly stopped, turned to me and asked whether I really believed that the moon exists only when I look at it.
>
> —A. Pais[10]

> As O. Stern said recently, one should no more rack one's brain about the problem of whether something one cannot know anything about exists all the same, than about the ancient question of how many angels are able to sit on the point of a needle. But it seems to me that Einstein's questions are ultimately always of this kind.
>
> —W. Pauli[11]

---

[10]Pais (1979, 907).

[11]From a 1954 letter to Max Born (Born 1971, 223).

Pauli and Einstein were both wrong. The questions with which Einstein attacked the quantum theory do have answers; but they are not the answers Einstein expected them to have. We now know that the moon is demonstrably not there when nobody looks.

The impact of this discovery on philosophy may have been blunted by the way in which it is conventionally stated, which leaves it fully accessible only to those with a working knowledge of quantum mechanics. I hope to remove that barrier by describing this remarkable aspect of nature in a way that presupposes no background whatever in the quantum theory or, for that matter, in classical physics either. I shall describe a piece of machinery that presents without any distortion one of the most strikingly peculiar features of the atomic world. No formal training in physics or mathematics is needed to grasp and ponder the extraordinary behavior of the device; it is only necessary to follow a simple counting argument on the level of a newspaper brain-twister.

Being a physicist, and not a philosopher, I aim only to bring home some strange and simple facts which might raise issues philosophers would be interested in addressing. I shall try, perhaps without notable success, to avoid raising and addressing such issues myself. What I describe should be regarded as something between a parable and a lecture demonstration. It is less than a lecture demonstration for technical reasons: even if this were a lecture, I lack the time, money, and particular expertise to build the machinery I shall describe. It is more than a parable because the device could in fact be built with an effort almost certainly less than, say, the Manhattan project, and because the conundrum posed by the behavior of the device is no mere analogy, but the atomic world itself, acting at its most perverse.

There are some black boxes within the device whose contents can be described only in highly technical terms. This is of no importance. The wonder of the device lies in what it does, not in how it is put together to do it. One need not understand silicon chips to learn from playing with a pocket calculator that a machine can do arithmetic with superhuman speed and precision; one need not understand electronics or electrodynamics to grasp that a small box can imitate human speech or an orchestra. At the end of the essay I shall give a brief technical description of what is in the black boxes. That description can be skipped. It is there to serve as an existence-proof only because you cannot buy the device at the drugstore. It is no more essential to appreciating the conundrum of the device than a circuit diagram is to using a calculator or a radio.

The device has three unconnected parts. The question of connectedness lies near the heart of the conundrum, but I shall set it aside in favor of a few simple practical assertions. There are neither mechanical connections (pipes, rods, strings, wires) nor electromagnetic connections (radio, radar, telephone, or light signals) nor any other relevant connections. Irrelevant connections may be hard to avoid. All three parts might, for example, sit atop a single table. There is nothing in the design of the parts, however, that takes advantage of such connections to signal from one to another, for example, by inducing and detecting vibrations in the table top.

By insisting so on the absence of connections, I am inevitably suggesting that the wonders to be revealed can be fully appreciated only by experts on connections or their lack. This is not the right attitude to take. Were we together and had I the device at hand, you could pick up the parts, open them up, and poke around as much as you liked. You would find no connections. Neither would an expert on hidden bugs, the Amazing Randi, or any physicists you called in as consultants. The real worry is unknown connections. Who is to say that the parts are not connected by the transmission of unknown Q-rays and their detection by unrecognizable Q-detectors? One can only offer affidavits from the manufacturer testifying to an ignorance of Q-technology and, in any event, no such intent.

Evidently it is impossible to rule out conclusively the possibility of connections. The proper point of view to take, however, is that it is precisely the wonder and glory of the device that it impels one to doubt these assurances from one's own eyes and hands, professional magicians, and technical experts of all kinds. Suffice it to say that there are no connections that suspicious lay people or experts of broad erudition and unimpeachable integrity can discern. If you find yourself questioning this, then you have grasped the mystery of the atomic world.

Two of the three parts of the device (*A* and *B*) function as detectors. Each detector has a switch that can be set in one of three positions (1, 2, and 3) and a red and a green light bulb (figure 1). When a detector is set off it flashes either its red light or its green. It does this no matter how its switch is

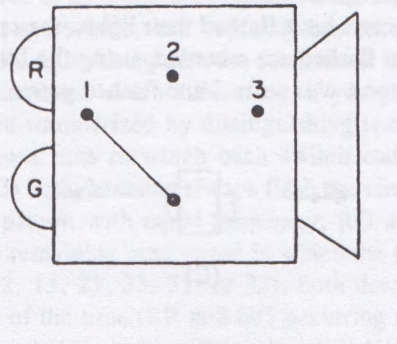


Figure 1. A detector. Particles enter on the right. The red (R) and green (G) lights are on the left. The switch is set to 1.

set, though whether it flashes red or green may well depend on the setting. The only purpose of the lights is to communicate information to us; marks on a ribbon of tape would serve as well. I mention this only to emphasize that the unconnectedness of the parts prohibits a mechanism in either detector that might modify its behavior according to the color that may have flashed at the other.

The third and last part of the device is a box (*C*) placed between the detectors. Whenever a button on the box is pushed, shortly thereafter two particles emerge, moving off in opposite directions toward the two detectors (figure 2). Each detector flashes either red or green whenever a particle reaches it. Thus within a second or two of every push of the button, each detector flashes one or the other of its two colored lights.

Because there are no connections between parts of the device, the link between pressing the button on the box and the subsequent flashing of the detectors can be provided only by the passage of the particles from the box to the detectors. This passage could be confirmed by subsidiary detectors between the box and the main detectors *A* and *B*, which can be designed so as not to alter the functioning of the device. Additional instruments or shields could also be used to confirm the lack of other communication between the box and the two detectors or between the detectors themselves (figure 3).

The device is operated repeatedly in the following way. The switch on each detector is set at random to one of its three possible positions, giving nine equally likely settings for the pair of detectors: 11, 12, 13, 21, 22, 23, 31, 32, and 33. The button on the box is then pushed, and somewhat later each detector flashes one of its lights. The flashing of the detectors need not be simultaneous. By changing the distance between the box and the detectors we can arrange that either flashes first. We can also let the switches be given their random settings either before or after the particles leave the box. One could even arrange for the switch on *B* not to be set until after *A* had flashed (but, of course, before *B* flashed).

After both detectors have flashed their lights, the settings of the switches and the colors that flashed are recorded, using the following notation: 31 GR means that detector *A* was set to 3 and flashed green, while *B* was set to 1

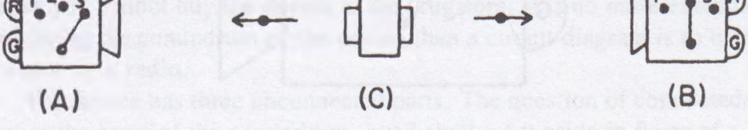


Figure 2. The complete device. *A* and *B* are the two detectors. *C* is the box from which the two particles emerge.

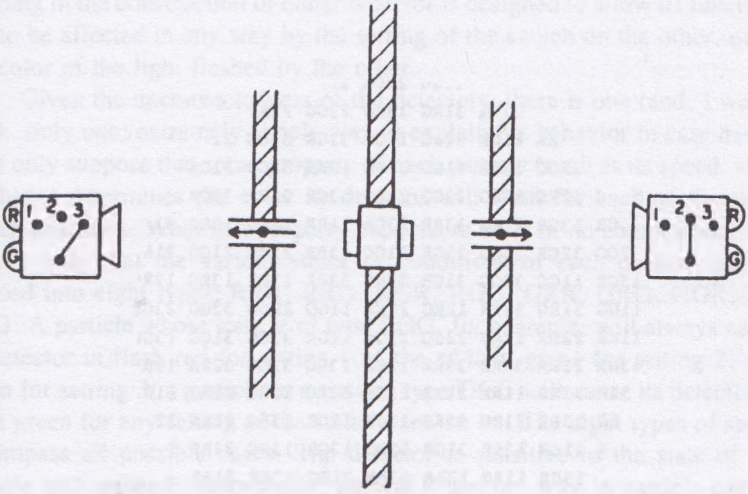


Figure 3. Possible refinement of the device. The box is embedded in a wall that cuts off one detector from the other. Subsidiary detectors confirm the passage of the particles to the main detectors.

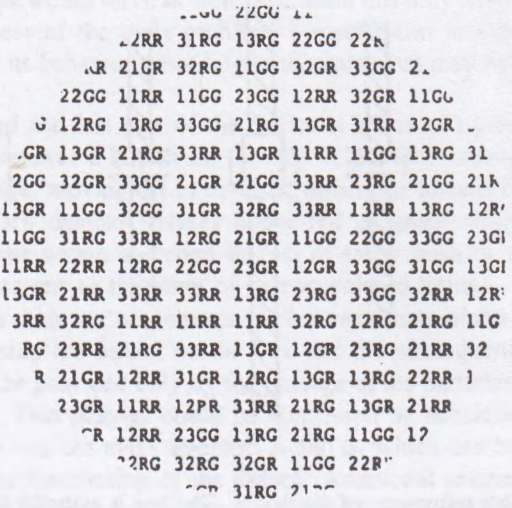
and flashed red; 12 RR describes a run in which *A* was at 1, *B* at 2, and both flashed red; 22 RG describes a run in which both detectors were set to 2, *A* flashed red and *B* flashed green; and so on. A typical fragment from a record of many runs is shown in figure 4.

The accumulated data have a random character, but, like data collected in many tossings of a coin, they reveal certain unmistakable features when enormously many runs are examined. The statistical character of the data should not be a source of concern or suspicion. Blaming the behavior of the device on repeated, systematic, and reproducible accidents is to offer an explanation even more astonishing than the conundrum it is invoked to dispel.

The data accumulated over millions (or, if you prefer, billions or trillions) of runs can be summarized by distinguishing two cases:

Case a. In those runs in which each switch ends up with the same setting (11, 22, or 33) both detectors always flash the same color. RR and GG occur in a random pattern with equal frequency; RG and GR never occur.

Case b. In the remaining runs, those in which the switches end up with different settings (12, 13, 21, 23, 31, or 32), both detectors flash the same color only a quarter of the time (RR and GG occurring randomly with equal frequency); the other three quarters of the time the detectors flash different colors (RG and GR occurring randomly with equal frequency).



..GG 22GG 1.
.2RR 31RG 13RG 22GG 22R
.R 21GR 32RG 11GG 32GR 33GG 2.
22GG 11RR 11GG 23GG 12RR 32GR 11GG
G 12RG 13RG 33GG 21RG 13GR 31RR 32GR .
.GR 13GR 21RG 33RR 13GR 11RR 11GG 13RG 31
2GG 32GR 33GG 21GR 21GG 33RR 23RG 21GG 21R
13GR 11GG 32GG 31GR 32RG 33RR 13RR 13RG 12R
11GG 31RG 33RR 12RG 21GR 11GG 22GG 33GG 23G
11RR 22RR 12RG 22GG 23GR 12GR 33GG 31GG 13G
13GR 21RR 33RR 33RR 13RG 23RG 33GG 32RR 12R
3RR 32RG 11RR 11RR 11RR 32RG 12RG 21RG 11G
RG 23RR 21RG 33RR 13GR 12GR 23RG 21RR 32
R 21GR 12RR 31GR 12RG 13GR 13RG 22RR 1
23GR 11RR 12RR 33RR 21RG 13GR 21RR
.R 12RR 23GG 13RG 21RG 11GG 12
.2RG 32RG 32GR 11GG 22R
..R 31RG 21..

Figure 4. Fragment of a page of a volume from the set of notebooks recording a long series of runs.

These results are subject to the fluctuations accompanying any statistical predictions, but, as in the case of a coin-tossing experiment, the observed ratios will differ less and less from those predicted, as the number of runs becomes larger and larger.

This is all it is necessary to know about how the device operates. The particular fractions $\frac{1}{4}$ and $\frac{3}{4}$ arising in case *b* are of critical importance. If the smaller of the two were $\frac{1}{3}$ or more (and the larger $\frac{2}{3}$ or less) there would be nothing wonderful about the device. To produce the conundrum it is necessary to run the experiment sufficiently many times to establish with overwhelming probability that the observed frequencies (which will be close to 25% and 75%) are not chance fluctuations away from expected frequencies of $33\frac{1}{3}\%$ and $66\frac{2}{3}\%$. (A million runs is more than enough for this purpose.)

These statistics may seem harmless enough, but some scrutiny reveals them to be as surprising as anything seen in a magic show, and leads to similar suspicions of hidden wires, mirrors, or confederates under the floor. We begin by seeking to explain why the detectors invariably flash the same colors when the switches are in the same positions (case *a*). There would be any number of ways to arrange this were the detectors connected, but they are not.

Nothing in the construction of either detector is designed to allow its functioning to be affected in any way by the setting of the switch on the other, or by the color of the light flashed by the other.

Given the unconnectedness of the detectors, there is one (and, I would think, only one) extremely simple way to explain the behavior in case *a*. We need only suppose that some property of each particle (such as its speed, size, or shape) determines the color its detector will flash for each of the three switch positions. What that property happens to be is of no consequence; we require only that the various states or conditions of each particle can be divided into eight types; RRR, RRG, RGR, RGG, GRR, GRG, GGR, and GGG. A particle whose state is of type RGG, for example, will always cause its detector to flash red for setting 1 of the switch, green for setting 2, and green for setting 3; a particle in a state of type GGG will cause its detector to flash green for any setting of the switch; and so on. The eight types of states encompass all possible cases. The detector is sensitive to the state of the particle and responds accordingly; putting it another way, a particle can be regarded as carrying a specific set of flashing instructions to its detector, depending on which of the eight states the particle is in.

The absence of RG or GR when the two switches have the same settings can then be simply explained by assuming that the two particles produced in a given run are both produced in the same state; i.e., they carry identical instruction-sets. Thus if both particles in a run are produced in states of type RRG, then both detectors will flash red if both switches are set to 3. The detectors flash the same colors when the switches have the same settings because the particles carry the same instructions.

This hypothesis is the obvious way to account for what happens in case *a*. I cannot prove that it is the only way, but I challenge the reader, given the lack of connections between the detectors, to suggest any other. The apparent inevitability of this explanation for the perfect correlations in case *a* forms the basis for the conundrum posed by the device. For the explanation is quite incompatible with what happens in case *b*.

If the hypothesis of instruction-sets were correct, then both particles in any given run would have to carry identical instruction-sets whether or not the switches on the detectors were set the same. At the moment the particles are produced there is no way to know how the switches are going to be set. For one thing, there is no communication between the detectors and the particle-emitting box, but in any event the switches need not be set to their random positions until after the particles have gone off in opposite directions from the box. To ensure that the detectors invariably flash the same color every time the switches end up with the same settings, the particles leaving the box in each run must carry the same instructions even in those runs (case *b*) in which the switches end up with different settings.

Let us now consider the totality of all case *b* runs. In none of them do we ever learn what the full instruction sets were, since the data reveal only the colors assigned to two of the three settings. (The case *a* runs are even less informative.) Nevertheless, we can draw some nontrivial conclusions by examining the implications of each of the eight possible instruction sets for those runs in which the switches end up with different settings. Suppose, for example, that both particles carry the instruction-set RRG. Then out of the six possible case *b* settings, 12 and 21 will result in both detectors flashing the same color (red), and the remaining four settings, 13, 31, 23, and 32, will result in one red flash and one green. Thus both detectors will flash the same color for two of the six possible case *b* settings. Since the switch settings are completely random, the various case *b* settings occur with equal frequency. Both detectors will therefore flash the same color in a third of those case *b* runs in which the particles carry the instruction-sets RRG.

The same is true for case *b* runs where the instruction-set is RGR, GRR, GGR, GRG, or RGG, since the conclusion rests only on the fact that one color appears in the instruction-set once and the other color, twice. In a third of the case *b* runs in which the particles carry any of these instruction-sets, the detectors will flash the same color. The only remaining instruction-sets are RRR and GGG; for these sets both detectors will evidently flash the same color in every case *b* run.

Thus, regardless of how the instruction-sets are distributed among the different runs, in the case *b* runs *both detectors must flash the same color at least a third of the time*. (This is a bare minimum; the same color will flash more than a third of the time, unless the instruction sets RRR and GGG never occur.) As emphasized earlier, however, when the device actually operates, the same color is flashed only a quarter of the time in the case *b* runs.

Thus the observed facts in case *b* are incompatible with the only apparent explanation of the observed facts in case *a*, leaving us with the profound problem of how else to account for the behavior in both cases. This is the conundrum posed by the device, for there is no other obvious explanation of why the same colors always flash when the switches are set the same. It would appear that there must, after all, be connections between the detectors—connections of no known description which serve no purpose other than relieving us of the task of accounting for the behavior of the device in their absence.

I shall not pursue this line of thought, since my aim is only to state the conundrum of the device, not to resolve it. The lecture demonstration is over. I shall only add a few remarks on the device as a parable.

One of the historic exchanges between Einstein (Einstein, Podolsky, Rosen 1935) and Bohr (1935), which found its surprising denouement in the work of J. S. Bell (1964) nearly three decades later, can be stated quite clearly in terms of the device. I stress that the transcription into the context of the device is only to simplify the particular physical arrangement used to raise the issues. The device is a direct descendant of the rather more intricate but conceptually similar *Gedanken* experiment proposed in 1935 by Einstein, Podolsky, and Rosen. We are still talking physics, not descending to the level of analogy.

The Einstein, Podolsky, Rosen experiment amounts to running the device under restricted conditions in which both switches are required to have the same setting (case *a*). Einstein would argue (as was argued above) that the perfect correlations in each run (RR or GG but never RG or GR) can be explained only if instruction-sets exist, each particle in a run carrying the same instructions. In the Einstein, Podolsky, Rosen version of the argument, the analogue of case *b* was not evident, and its fatal implications for the hypothesis of instruction-sets went unnoticed until Bell's paper.

The *Gedanken* experiment was designed to challenge the prevailing interpretation of the quantum theory, which emphatically denied the existence of instruction-sets, insisting that certain physical properties (said to be complementary) had no meaning independent of the experimental procedure by which they were measured. Such measurements, far from revealing the value of a preexisting property, had to be regarded as an inseparable part of the very attribute they were designed to measure. Properties of this kind have no independent reality outside the context of a specific experiment arranged to observe them: the moon is *not* there when nobody looks.

In the case of my device, three such properties are involved for each particle. We can call them the 1-color, 2-color, and 3-color of the particle. The *n*-color of a particle is red if a detector with its switch set to *n* flashes red when the particle arrives. The three *n*-colors of a particle are complementary properties. The switch on a detector can be set to only one of the three positions, and the experimental arrangements for measuring the 1-, 2-, or 3-color of a particle are mutually exclusive. (We may assume, to make this point quite firm, that the particle is destroyed by the act of triggering the detector, which is, in fact, the case in many recent experiments probing the principles that underly the device.)

To assume that instruction-sets exist at all is to assume that a particle has a definite 1-, 2-, and 3-color. Whether or not all three colors are known or knowable is not the point; the mere assumption that all three have values violates a fundamental quantum-theoretic dogma.

No basis for challenging this dogma is evident when only a single particle and detector are considered. The ingenuity of Einstein, Podolsky, and Rosen lay in discovering a situation involving a *pair* of particles and detectors, where the quantum dogma continued to deny the existence of 1-, 2-, and 3-colors, while, at the same time, quantum theory predicted correlations (RR and GG but never RG or GR) that seemed to require their existence. Einstein concluded that, if the quantum theory were correct, i.e., if the correlations were, as predicted, perfect, then the dogma on the nonexistence of complementary properties—essentially Bohr's doctrine of complementarity—had to be rejected.

Pauli's attitude toward this in his letter to Born is typical of the position taken by many physicists: since there is no known way to determine all three *n*-colors of a particle, why waste your time arguing about whether or not they exist? To deny their existence has a certain powerful economy—why encumber the theory with inaccessible entities? More importantly, the denial is supported by the formal structure of the quantum theory which completely fails to allow for any consideration of the simultaneous 1-, 2-, and 3-colors of a particle. Einstein preferred to conclude that all three *n*-colors did exist, and that the quantum theory was incomplete. I suspect that many physicists, though not challenging the completeness of the quantum theory, managed to live with the Einstein, Podolsky, Rosen argument by observing that though there was no way to establish the existence of all three *n*-colors, there was also no way to establish their nonexistence. Let the angels sit, even if they cannot be counted.

Bell changed all this by bringing into consideration the case *b* runs and pointing out that the quantitative numerical predictions of the quantum theory ($\frac{1}{4}$ vs. $\frac{1}{3}$) unambiguously ruled out the existence of all three *n*-colors. Experiments done since Bell's paper confirm the quantum-theoretic predictions.[12] Einstein's attack, were he to maintain it today, would be more than an attack on the metaphysical underpinnings of the quantum theory—more, even, than an attack on the quantitative numerical predictions of the quantum theory. Einstein's position now appears to be contradicted by nature itself. The device behaves as it behaves, and no mention of wave-functions, reduction hypotheses, measurement theory, superposition principles, wave-particle duality, incompatible observables, complementarity, or the uncertainty principle is needed to bring home its peculiarity. It is not the Copenhagen interpretation of quantum mechanics that is strange, but the world itself.

As far as I can tell, physicists live with the existence of the device by implicitly (or even explicitly) denying the absence of connections between its pieces. References are made to the "wholeness" of nature: particles, detectors, and box can be considered only in their totality; the triggering and flashing of detector *A* cannot be considered in isolation from the triggering and flashing of detector *B*—both are part of a single indivisible process. This attitude is sometimes tinged with Eastern mysticism, sometimes with Western know-nothingism, but, common to either point of view, as well as to the less

[12]Clauser and Shimony (1978). For a less technical summary, see d'Espagnat (1979).

trivial but considerably more obscure position of Bohr, is the sense that strange connections are there. The connections are strange because they play no explicit role in the theory: they are associated with no particles or fields and cannot be used to send any kinds of signals. They are there for one and only one reason: to relieve the perplexity engendered by the insistence that there are no connections. Whether or not this is a satisfactory state of affairs is, I suspect, a question better addressed by philosophers than by physicists.

I conclude with the recipe for making the device, which, I emphasize again, can be ignored

> The device exploits Bohm's version (1951, 614–619) of the Einstein, Podolsky, Rosen experiment. The two particles emerging from the box are spin $\frac{1}{2}$ particles in the singlet state. The two detectors contain Stern-Gerlach magnets, and the three switch positions determine whether the orientations of the magnets are vertical or at $\pm 120°$ to the vertical in the plane perpendicular to the line of flight of the particles. When the switches have the same settings the magnets have the same orientations. One detector flashes red or green according to whether the measured spin is along or opposite to the field; the other uses the opposite color convention. Thus when the same colors flash the measured spin components are different.
>
> It is a well-known elementary result that, when the orientations of the magnets differ by an angle Θ, then the probability of spin measurements on each particle yielding opposite values is $\cos^2(\Theta/2)$. This probability is unity when $\Theta = 0°$ (case *a*) and $\frac{1}{4}$ when $\Theta = \pm 120°$ (case *b*).
>
> If the subsidiary detectors verifying the passage of the particles from the box to the magnets are entirely nonmagnetic they will not interfere with this behavior.