Chapter 8

The Elitzur-Vaidman Bomb Test

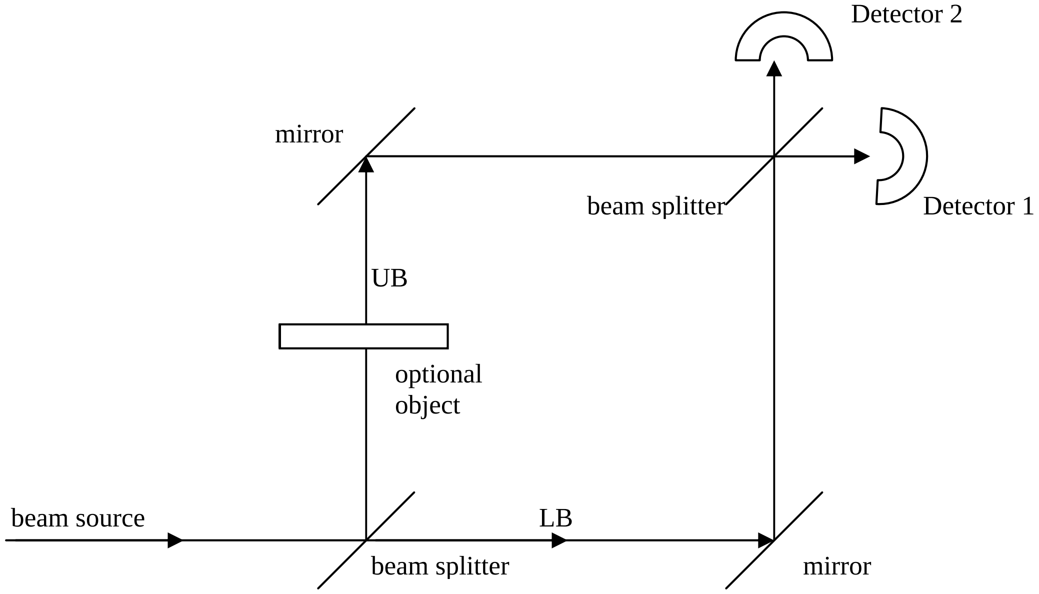
The Observability of Counterfactuals

Last Update: 31/12/11

1. Counterfactuals

Suppose something could have happened, but actually did not happen. In classical physics the fact that an event *could have* happened makes no difference to any future outcome if, in actual fact, it did not happen. Only those things which actually happen can influence the future evolution of the world. But in quantum mechanics it is otherwise. The potential for an event to happen can influence future outcomes even if the event does not happen. Something that could have happened but actually did not is called as *counterfactual*. In quantum mechanics *counterfactuals are observable* – they have measurable consequences. The Elitzur-Vaidman bomb test provides a striking illustration of this. But first we must revise how interferometers work – specifically Mach-Zehnder interferometers.

Figure 1 The Mach-Zehnder Interferometer
LB = Lower Beam; UB = Upper Beam



Referring to Figure 1, the incoming beam of particles (perhaps photons perhaps something else) hits a beam splitter which divides the beam into a lower beam (LB), which is simply transmitted, and an upper beam (UB) which is reflected. Ignoring any optional objects in the beam path for the moment, the two beams are reflected off mirrors at the top left and bottom right. The sole function of these mirrors is to focus the two beams onto the same spot of the beam splitter (or, if you will, the beam reuniter) at the top right. Thence any particle travelling through the interferometer may be detected in one or other of the two detectors.

A convenient simplification is to assume that both beam splitters split the beam 50/50, and that the mirrors are 100% efficient. This is merely to simplify the explanation and is not crucial to the principles of the device.

The key to the behaviour of the interferometer is the phase change suffered by the reflected and transmitted waves at the mirrors/beam splitters. There is potential confusion over this since the phase change depends upon the type of object used as a "mirror". For simplicity we shall assume plane glass plates are used as both mirrors and beam splitters here. This makes a small simplification in the phase change rules compared with real mirrors, though the outcome would be the same. Real mirrors would be used experimentally, of course, and half-silvered mirrors for the beam splitters.

Any transparent plate with a refractive index, *n*, greater than air will generally cause a mixture of transmission and reflection. The phase change rule is simply that the reflected wave acquires a phase factor of i compared to the transmitted wave. This relative phase is derived as follows. Consider transmission through a finite thickness, *a*, of transparent material. The incident plus reflected wave in the region $x < 0$ is $e^{ikx} + Be^{-ikx}$. Within the plate material, $0 < x < a$, the right plus left going waves are $Ee^{ik'x} + Fe^{-ik'x}$ where $k' = nk$ is the wavenumber in glass. In the region $x > a$ the transmitted wave is Ce^{ikx}. By applying the boundary conditions (continuity of the wavefunction and its x-derivative) at both boundaries $x = 0$ and $x = a$, the coefficients B, E, F, C are found. In particular the ratio of the reflection (B) and transmission (C) coefficients is,

$$\frac{B}{C} = i \cdot \left(\frac{n^2 - 1}{2n} \right) \sin k'a \cdot e^{ika} \qquad (1)$$

Now the factor of e^{ika} just accounts for the fact that the phase of the reflected wave has been reference to position $x = 0$ whereas the transmitted wave phase is referenced to $x = a$. Referencing them both to the same point, as required, leaves the phase factor from (1) as just i. This is all that is needed to analyse the interferometer.

2. The Mach-Zehnder Interferometer – No Specimen

Initially we analyse what signals are expected at the detectors 1 and 2 shown in Figure 1 if there is no "specimen", nor anything else, put in the way of the beams (e.g., any measuring device). Consider the phase changes of the upper beam (UB) with respect to the lower beam (LB) for exit into detector 2. Beam LB undergoes just one reflection (at the second plate) in order to enter detector 2, whereas beam UB requires three reflections, one at each of the three plates. Their relative phase entering detector 2 is therefore $i^3 / i = -1$. Consequently the waves travelling via paths LB and UB interfere destructively and hence no particles are recorded by detector 2.

Repeating this analysis for detector 1 we see that both the LB and UB paths involve two reflections. The two beams therefore enter detector 1 with the same relative phase and so there is constructive interference. Hence, all the particles are recorded by detector 1 and none by detector 2. This is the signature of interference in this device. From the classical perspective, a particle would travel by one or other of the two paths and have a 50/50 chance of ending up in either detector regardless of the path followed.

3. The Mach-Zehnder Interferometer With Measurements

From Chapter 4 we know that, in the context of a double slit interference pattern, obtaining “which path” information will destroy the interference. Exactly the same is true for a Mach-Zehnder Interferometer, and for identical algebraic reasons. However the matter is so fundamental that it bears repeating in this context.

Suppose the state of a particle which follows path LB into detector j is written $|LB\rangle_j$. Then the state of a particle following path UB into detector *j* is $|UB\rangle_j = e^{i\delta_j}|LB\rangle_j$. The preceding analysis has shown that $e^{i\delta_1} = 1$ and $e^{i\delta_2} = -1$. The state at the detectors is thus,

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|LB\rangle_j + e^{i\delta_j}|LB\rangle_j\right) \tag{2}$$

Hence the flux of particles into the detectors is proportional to,

$$\langle\psi|\psi\rangle = 1 + \cos\delta_j \tag{3}$$

thus giving constructive interference, $\langle\psi|\psi\rangle = 2$, at detector 1 and destructive interference, $\langle\psi|\psi\rangle = 0$, at detector 2.

How does this change if a measuring device is inserted into either beam path? A device to detect a particle in path UB must have two states, one of which corresponds to “a particle was detected in path UB” and the other corresponding to “a particle was not detected in path UB”. These states of the measuring device will be written $|M:UB\rangle$ and $|M:LB\rangle$ respectively. The notation reflects the assumption of perfectly efficient detection, and no particle losses, for which it follows that not detecting the particle in path UB implies that it followed path LB.

Now a perfect measurement requires the two states of the measuring device to be absolutely distinguishable. This means that the two measurement states must be orthogonal, $\langle M:UB|M:LB\rangle = 0$. However, it is possible to envisage a poor measurement which might indicate a probability of the particle being in path LB, but with some residual possibility of being in path UB. For such an imperfect measurement we would have $\langle M:UB|M:LB\rangle \neq 0$.

The combined state of the particle and the measuring device, at either detector, is thus,

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|LB\rangle_j|M:LB\rangle + e^{i\delta_j}|LB\rangle_j|M:UB\rangle\right) \tag{4}$$

The flux of particles into the detectors is thus,

$$\langle\psi|\psi\rangle = 1 + \Re\left\{e^{i\delta_j}\langle M:UB|M:LB\rangle\right\} \tag{5}$$

Consequently, if we have performed a perfect measurement which definitely detects which path the particle took, then, because $\langle M:UB|M:LB\rangle = 0$, we have from (5) that the particle flux is unity, $\langle\psi|\psi\rangle = 1$, into *both* detectors. All interference has been lost.

This establishes quite generally that any means of detecting which path the particle takes around the interferometer will destroy the interference (in the sense that both detectors will detect equal numbers of particles).

However, the occurrence or not of interference is not an all-or-nothing affair. An imperfect measurement, which has $0 < \left| \langle M:UB | M:LB \rangle \right| < 1$, will still leave residual interference and hence more particles into detector 1 than into detector 2.

4. The Elitzur-Vaidman Bomb Test

Everything is now set up for this lovely example of the observability of counterfactuals. Imagine that you have manufactured a large number of bombs. The bombs are so sensitive that the slightest interaction, say with a single very low energy photon, would make the bomb explode. The trouble is that you know that some bombs are duds but you need to identify one which is definitely not a dud without it exploding.

How on earth can you do this? The problem defines the bombs to be so sensitive that the slightest physical interaction with any given bomb will make it explode. So you need to determine if a bomb is not a dud without interacting with it at all! Of course you can easily determine which are the duds by poking all the bombs. The ones that don't explode are the duds. But unfortunately that leaves you with no live bombs.

In classical physics the problem is insoluble. But in quantum physics, amazingly, it can be solved.

A bomb is placed[1] within the UB beam path of the interferometer in such a way that a passing particle may or may not interact with it. For example, this might be accomplished by attaching a rod to the mirror at the bottom right such that there was the tiniest gap between the rod and the bomb. If the mirror is free to move when struck by a particle, the rod would then strike the bomb and set if off. (The mechanism is impractical, of course, but the principle is what matters – a truly practical device is perfectly feasible).

An active bomb has thus been made into a measuring device regarding which of the paths, UB or LB, the particle takes. If it takes path UB, the bomb explodes. If the bomb does not explode either the particle took path LB or the bomb is a dud. The very sensitivity of an active bomb makes it a perfect measuring device.

Now if the bomb is a dud, then the bomb does not constitute a measuring device. A dud bomb might as well not be there. So, with a dud bomb, the particles will always emerge into detector 1, never into detector 2.

But if the bomb is *not* a dud, and assuming it does not explode, then the measurement implemented by the bomb destroys the interference and the particle could be detected in either detector 1 or 2. But if it is detected in detector 2 the bomb cannot be a dud! So we have successfully identified a bomb which is definitely not a dud but without it exploding. Miraculous! Note that in terms of the state algebra, and assuming a live bomb, the absence of an explosion means that the measurement has selected (collapsed to) the first term in the state,

---

[1] Of course, you can't actually pick the bomb up and move it, or it would go off. So this is really shorthand for "a Mach-Zehnder interferometer is constructed around a given bomb". You might also be wondering what use these bombs could possibly be, since you could never move them to where you might want to destroy something. Hey, don't take this example so literally!

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|LB\rangle|M:LB\rangle + |UB\rangle|M:UB\rangle bang!\right) \quad (6)$$

So the particle state hitting the final beam splitter is just $|LB\rangle$, which thus has a 50/50 chance of ending up at detector 1 or 2.

The fact that the bomb *might* have gone off, despite the fact that it did not, is crucial to the bomb constituting a measuring device and hence to the identification of the unexploded bomb as not being dud. The counterfactual has had an observable consequence, namely that the bomb is now known with certainty to be live.

This curious phenomenon, and the bomb scenario described above, was originally described by Elitzur and Vaidman (1993). Do not think that it is too theoretical to be demonstrated experimentally. On the contrary, this was done almost as soon as the effect was discovered, by Anton Zeilinger's group in Vienna, Kwiat (1995).

As described, the bomb test is terribly inefficient. Of the live bombs, half are exploded, and of the remaining 50% only half of these result in a particle at detector 2 – and hence are identified definitely as live bombs. Thus, a single application of the bomb tester identifies ¼ of the live bombs. But another ¼ still remain unexploded and unidentified. Running these through the bomb tester again results in a further ¼ of this ¼ being identified as live. Hence, repeated applications identify,

$$\tfrac{1}{4} + \tfrac{1}{4} \times \tfrac{1}{4} + \tfrac{1}{4} \times \tfrac{1}{4} \times \tfrac{1}{4} \ldots = {}^{1}/_{3}$$

of the live bombs. But the remaining ${}^{2}/_{3}$ are destroyed – terribly inefficient. However, this is avoidable. In principle virtually all the live bombs can successfully be identified, see for example Kwiat (1999).

5. Forward to Magic

The startling realisation that information can be obtained without interacting with the object opens up many possibilities. One is to replace the bomb with a quantum computer. The claim is sometimes loosely made that this can provide a means of obtaining the result of a quantum computation without even switching the computer on. This is rather misleading because we are apt to interpret the "switching on" as a classical process, i.e., involving a real macroscopic switch and the initiation of large currents or voltages. In contrast what is really envisaged is that the quantum computer includes an "activation" qubit, such that the computer does nothing if the qubit is set to 0 but runs its computation if the qubit is switch to 1. The difference between the two scenarios is that the qubit is quantum mechanical and will in general be in a superposition of states.

This is crucial to the claim, which is more correctly stated as being that it is possible to extract the result of the computation whilst also confirming that the activation qubit remains at 0, the "off position". This is referred to as *counterfactual computation*, see for example Mitchison and Jozsa (2001). It is directly analogous to obtaining the information that a bomb is live whilst not exploding it (i.e., whilst not "switching it on").

The great curse of quantum computers is decoherence. The possibility of obtaining the result of a quantum computation without the computer actually having to run might appear to be a very neat way of avoiding the problem of decoherence. Unfortunately this is not so. Obtaining counterfactual information depends crucially on unitary evolution. Just as the *possibility* of the computer running is essential in obtaining its result counterfactually, so the *possibility* of its decoherence if it did run is sufficient to undermine a successful outcome. The magic does not extend to getting something for nothing.

6. References

Elitzur A. C. and Vaidman L. (1993). Quantum mechanical interaction-free measurements. *Found. Phys.* **23**, 987-97. arxiv:hep-th/9305002.

P. G. Kwiat, H. Weinfurter, T. Herzog, A. Zeilinger, and M. A. Kasevich (1995). "Interaction-free Measurement". *Phys. Rev. Lett.* **74** (24): 4763.

P.G.Kwiat, A.G.White, J.R.Mitchell, O.Nairz, G.Weihs, H.Weinfurter, A. Zeilinger (1999), "High-efficiency quantum interrogation measurements via the quantum Zeno effect", Physical Review Letters **83**, no. 23, pp. 4725-4728 (1999). See also Paul Kwiat, http://physics.illinois.edu/people/kwiat/interaction-free-measurements.asp.

G.Mitchison and R.Jozsa (2001), "Counterfactual Computation", Proc.Roy.Soc.Lond. A457 (2001) 1175-1194.