Handout #2: Derivation of the Turbo Bell's Inequality (less formally known as the Clauser-Horne-Holt-Shimony inequality)

Suppose we have two apparatuses, one of which can be set to perform either experiment a_1 or a_2, and the other of which can be set to perform either experiment b_1 or b_2. (If it helps, you can think of these experiments as being measurements performed on each of a pair of particles.) Suppose that each of these four experiments has just two possible outcomes; in each case, we will label one of these outcomes "up" and the other "down". Thus, a_1 and a_2 might consist in sending one of a pair of particles through a Stern-Gerlach magnet with orientation -120° and 0°, respectively; while b_1 and b_2 might consist in sending the other one of the pair of particles through a Stern-Gerlach magnet with orientation +120° and 0°, respectively. So there are four possible combinations of experiments, and in each case the outcomes will either be correlated (both "up" or both "down") or anti-correlated (one "up" and the other "down"). Let us finally define a random variable Cor that has value 1 if the outcomes are correlated and value –1 if they are anti-correlated.

Then for each combination of experiments, we can define the expected correlation E(x,y) (where x = either a_1 or a_2 and y = either b_1 or b_2) to be the expected value of Cor, given that we are performing the combination of experiments (x,y). Thus,

$$
\begin{aligned}
E(x,y) = \; & \mathrm{Prob}(\mathrm{up},\mathrm{up} \mid x,y) + \mathrm{Prob}(\mathrm{down}, \mathrm{down} \mid x,y) \\
& - \mathrm{Prob}(\mathrm{down},\mathrm{up} \mid x,y) - \mathrm{Prob}(\mathrm{up}, \mathrm{down} \mid x,y).
\end{aligned}
$$

Note that E(x,y) is rather easy to measure experimentally: We just perform the combination of experiments x and y a large number of times N, count up the number of times the outcomes are the same (both "up" or both "down"), subtract the number of times they are different (one "up" and the other "down"), and divide by N.

Here is the Turbo Bell's Inequality:

$$| E(a_1, b_1) + E(a_1, b_2) + E(a_2, b_1) - E(a_2, b_2) | \leq 2.$$

Before proceeding to the derivation of this inequality, let us just note two things about it. First—provided the assumptions needed in its derivation hold—it is absolutely universal in scope. That is, it has nothing in particular to do with quantum mechanics, or with any other theory for that matter. Second, it is clearly violated by our correlated spin experiment, given the foregoing choice of a_1, a_2, etc. For $E(a_1, b_1)$ is just the expected value of Cor, when the magnet on the left has orientation –120° and the magnet on the right has orientation +120°; in such a case, we know that the probability that the outcomes are different is .25, and so the expected value of Cor is .75 - .25 = .5. Likewise, $E(a_1, b_2) = E(a_2, b_1) = .5$. Finally, since the outcomes are certain to be different if both magnets have orientation 0°, $E(a_2, b_2) = -1$. Plugging these values into the equation, we get the result that $| .5 + .5 + .5 - (-1) | = 2.5 \leq 2$, a contradiction.

On to the derivation. We begin by assuming that we are equipped with *some* theory—maybe quantum mechanics, maybe something entirely different—that provides us with the following three things:

(i) A set of possible initial **physical states** λ of the system upon which we are performing the combined experiments.

(ii) For each such state λ, and each choice of experiment combination (x,y), a **value** for each of the probabilities Prob(up,up | x,y,λ), Prob(up, down | x,y,λ), Prob(down,up | x,y,λ), Prob(down,down | x,y,λ).

(iii) For each experiment combination (x,y), a **probability density** $\rho(\lambda \mid x,y)$ that encapsulates how likely it is that the physical state λ has one value or another, given that we are performing experiments x and y.

We will assume that there is some way to parametrize the physical states λ so that the integral $\int\rho(\lambda \mid x,y)d\lambda$ is well-defined. Assuming it is, it must of course have value 1.

We now make two crucial assumptions about the probabilities that our theory introduces:

No conspiracy: The probability density distribution for λ is *independent* of our choice of experiment combination (x,y). Thus, $\rho(\lambda \mid a_1,b_1) = \rho(\lambda \mid a_2,b_1) = \rho(\lambda \mid a_1,b_2) = \rho(\lambda \mid a_2,b_2)$, in which case we can simplify notation by replacing all of these by $\rho(\lambda)$.

Locality: For any pair of outcomes (u,v), any choice of experiment combination (x,y), and any physical state λ, $\mathrm{Prob}(u,v \mid x,y,\lambda) = \mathrm{Prob}(u \mid x,\lambda) \cdot \mathrm{Prob}(v \mid y,\lambda)$. This condition reflects the thought that the physical state λ contains enough information to render *irrelevant* to the outcome of one experiment what is going on with the other experiment (i.e., either what the other experiment *is*, or what its outcome is).

With these conditions in place, the derivation is fairly straightforward. We first note (making implicit use of **No conspiracy**) that

$$E(x,y) =$$

$$\int[\mathrm{Prob}(u,u \mid x,y,\lambda) + \mathrm{Prob}(d,d \mid x,y,\lambda) - \mathrm{Prob}(d,u \mid x,y,\lambda) - \mathrm{Prob}(u,d \mid x,y,\lambda)]\rho(\lambda)d\lambda.$$

Using **Locality**, and rearranging, this becomes

$$\int\{[\mathrm{Prob}(u \mid x,\lambda) - \mathrm{Prob}(d \mid x,\lambda)] \cdot [\mathrm{Prob}(u \mid y,\lambda) - \mathrm{Prob}(d \mid y,\lambda)]\}\rho(\lambda)d\lambda.$$

Let us now introduce the following useful abbreviations:

$A_1(\lambda) = \mathrm{Prob}(u \mid a_1,\lambda) - \mathrm{Prob}(d \mid a_1,\lambda)$

$A_2(\lambda) = \mathrm{Prob}(u \mid a_2,\lambda) - \mathrm{Prob}(d \mid a_2,\lambda)$

$B_1(\lambda) = \mathrm{Prob}(u \mid b_1,\lambda) - \mathrm{Prob}(d \mid b_1,\lambda)$

$B_2(\lambda) = \mathrm{Prob}(u \mid b_2,\lambda) - \mathrm{Prob}(d \mid b_2,\lambda)$

Observe that for every value of λ, $-1 \leq A_1(\lambda) \leq 1$; likewise for the other three functions.

Putting all this together, we get

$| E(a_1, b_1) + E(a_1, b_2) + E(a_2, b_1) - E(a_2, b_2) |$

$= | \int\{A_1(\lambda)\bullet B_1(\lambda) + A_1(\lambda)\bullet B_2(\lambda) + A_2(\lambda)\bullet B_1(\lambda) - A_2(\lambda)\bullet B_2(\lambda)\}\rho(\lambda)d\lambda |$

$\leq \int | \{A_1(\lambda)\bullet B_1(\lambda) + A_1(\lambda)\bullet B_2(\lambda) + A_2(\lambda)\bullet B_1(\lambda) - A_2(\lambda)\bullet B_2(\lambda)\}\rho(\lambda) | d\lambda$

$= \int | \{A_1(\lambda)\bullet B_1(\lambda) + A_1(\lambda)\bullet B_2(\lambda) + A_2(\lambda)\bullet B_1(\lambda) - A_2(\lambda)\bullet B_2(\lambda)\} | \rho(\lambda) d\lambda$

$\leq \int\{ | A_1(\lambda)\bullet B_1(\lambda) + A_1(\lambda)\bullet B_2(\lambda) | + | A_2(\lambda)\bullet B_1(\lambda) - A_2(\lambda)\bullet B_2(\lambda) | \}\rho(\lambda) d\lambda$

$= \int\{ | A_1(\lambda) | \bullet | B_1(\lambda) + B_2(\lambda) | + | A_2(\lambda) | \bullet | B_1(\lambda) - B_2(\lambda) | \}\rho(\lambda) d\lambda$

$\leq \int\{ | B_1(\lambda) + B_2(\lambda) | + | B_1(\lambda) - B_2(\lambda) | \}\rho(\lambda) d\lambda$

$\leq \int 2\rho(\lambda) d\lambda$

$= 2.$

Why are we so sure that $\{ | B_1(\lambda) + B_2(\lambda) | + | B_1(\lambda) - B_2(\lambda) | \} \leq 2$? Here is one way to see this: First, remember that $B_1(\lambda)$ and $B_2(\lambda)$ have values between –1 and 1 (inclusive). For notational convenience, replace them by "X" and "Y". Then if we take the square of the quantity $\{ | X + Y| + | X – Y| \}$, we get

$X^2 + 2XY + Y^2 + 2|X^2 – Y^2| + X^2 – 2XY + Y^2,$

which simplifies to

$4X^2$, if $X^2 \geq Y^2$, and

$4Y^2$, if $X^2 \leq Y^2$.

Either way, since $X^2 \leq 1$ and $Y^2 \leq 1$, we know that this quantity is ≤ 4. But that means that its positive square root $\{ | X + Y | + | X – Y | \}$ is ≤ 2.