Arm A	Arm B
100	60
-50	60
100	60
-50	60

21-301 Assignment 05
Karan Sikka
ksikka@cmu.edu
November 20, 2013

1a

The main difference in exploration/exploitation and UCB is that exploration affords us an opportunity to discover unknown properties of the arms, and UCB goes straight to exploitation.

Consider the case where an arm A's reward increases linearly but arm B's reward increases exponentially (as a function of the number of pulls). Note that initially, the rewards look quite similar. In UCB, you'd initially see arm A have higher reward than arm B, so you'd pull arm A more frequently and eventually you'd be so confident that A is the better arm so you'd only pull arm A. Note that here you made a premature assumption about arm B, which is that it's current performance reflects its future performance.

But if you had explored first, you'd see arm B beat arm A's reward, and it would get higher and higher rewards to outweight arm A. So in the exploitation phase you'd pull arm B and this strategy clearly beats that of pulling arm A.

The tradeoff is that exploration allows us to learn more about the true properties of the arms at the expense of time or pulls.

1b

and so on.

Arm A will make more money, but it has high variance. Arm B makes less money but has no variance.