# UG Project weekly task: due Sept 4

Hypothesis a: TB is more helpful in games with varying reward scales. The intuition of TB is to transform rewards to a smaller range without losing the relative differences, enabling using raw rewards in Atari games.

Methods: log all types of rewards and returns, compare the differences. Use MsPacman for now (each experiment will take ~3 days, all line numbers below refer to script "a3c_training_thread.py"):

a. Run A3C for 50 million steps. Log the following parameter:
   - *rewards* and batch_*cumsum_reward*
   - Name them "a3c_clipped_rewards.pkl" line206 and "a3c_clipped_returns.pkl line256" respectively.
b. Run A3CTB for 50 million steps. Log the following parameter:
   - *rewards*, *batch_cumsum_reward*, and *batch_raw_reward*
   - Name them "tb_raw_rewards.pkl" line204, "tb_transformed_returns,pkl" line253 and "raw_returns.pkl" line254
c. Generate **one** histogram of: a3c_clipped_returns vs. tb_transformed_returns vs. raw_returns
d. Generate **one** histogram of: a3c_clipped_rewards vs. tb_raw_rewards

Study task: revisit the concept of "reward" vs. "return".

a. Explain how they differs and why do we want to generate two histogram as in c and d.
b. What's the difference between the two "rewards"? What about the three"returns"?