

# COMP SCI 5001 FS2017 Final Project

Dalton Cole  
drcgy5@mst.edu

December 13, 2017

For this project, Q-Learning was used to compute the optimal policy for Tinny Tim to maximize his donuty reward.

The policy found along with the reward for each square is shown in Table 2. The configurations file used is shown in Table 1.

As can be seen by Table 2, each of the squares where tiles fall are avoided with no arrow leading to them. Each donut square, minus the lower right hand corner, has a square pointing to the donut square. There is also only one positive square on the entire grid. This is probably because of the number of evaluations performed. If a tile hit Tim, or if he accidentally ran into a wall, he lost a reward, thus it was easier to be in the negative than positive.

Table 1: Configuration File Used

Random Seed	0
Discount Rate	0.1
TD Step Size	0.1
Evaluations	123,456,789,012

Table 2: Resulting Grid

	↓ -0.09041	← -0.7826	← -0.05455	→ -1.139	→ -0.3364	→ -0.2377	→ -0.7371	← -0.207	
	↑ 1.018		↓ -0.3946				↓ -0.07519	↑ -0.1634	
	↓ -0.1095		↓ -0.24	↓ -0.08025	↓ -0.5231	→ -0.1269	← -0.2535	↑ -0.04677	
	↑ -0.177		↓ -0.07754	→ -0.238	↓ -0.07535		↓ -0.1368	↓ -0.0108	
	↓ -0.2259	← -0.0845	↓ -0.09152	← -0.1432	← -0.06724		↑ -0.263	↓ -0.09724	
	↓ -0.03103	← -0.8109	↑ -0.0801		↑ -0.1481		↓ -0.1945	↓ -0.127	
	↑ -0.1442	← -0.003676	↓ -0.2479		→ -0.5369	↓ -0.02842	→ -0.01054	↑ -0.05041	
	↑ -0.2627	← -1.195	↑ -0.2279		→ -0.2731	← -0.06344	← -0.07356	← -0.1539	