

Probe weight cosine (random_direction_1) steer_0 probe1

Weight index

