Figure 2.5: Preference gradient no baseline, bandit