L4 (Combined\_Dynamic\_Mutation): Final Population Beliefs (Sample Run) Benign **Neutral** 2.0 **Deceptive Benign Mean** 1.5 **Neutral Mean Deceptive Mean** Deceptive Region (v<0, a>0) 1.0 Alignment Signal (a) 0.5 0.0 -0.5 -1.0 -1.5 -2 2 True Value (v)