49th IEEE Conference on Decision and Control
December 15-17, 2010
Hilton Atlanta Hotel, Atlanta, GA, USA

The role of social feedback in steady-state performance of human decision making for two-alternative choice tasks

Andrew Stewart and Naomi Ehrich Leonard

***Abstract*— With an eye towards design of human-in-the-loop systems, we investigate human decision making in a social context for tasks that require the human to make repeated choices among finite alternatives. We consider a human decision maker who receives feedback on his/her own performance as well as on the choices of others performing the same task. We use a drift-diffusion, decision-making model that has been fitted to human neural and behavioral data in sequential, two-alternative, forced-choice tasks and recently extended to the social context with an empirically derived feedback term that depends on choices of other decision makers. We show conditions for this model to be a Markov process, and we derive the steady-state probability distribution for choice sequences and individual performance as a function of the strength of the social feedback. It has recently been shown in behavioral experiments that human decision-making performance for a relatively easy task is decreased with this social feedback; we show that our analytic predictions agree with this finding.**

I. INTRODUCTION

There is increasing interest in design of human-in-the-loop systems to address complex problems that demand the combined strengths of humans and machines. In many contexts, human decision makers will be faced with making repeated choices among finite alternatives in response to priorities among objectives and observations of evolving performance. Examples include human flight control operators who choose between approving or grounding flights in order to maximize throughput and ensure safety in air traffic [1] and human supervisors of unmanned air vehicles who must choose between attending to targets and ensuring safe return of vehicles [2]. In [3] a human repeatedly chooses one of two robotic oxygen extraction systems operating on Mars; the investigation focuses on the well-known difficulty that humans have with making long-term optimal decisions when short-term performance is high.

In human-in-the-loop systems such as these there may be multiple human decision makers carrying out tasks in parallel; a challenging problem is the design of social feedback, that is how to pass decision-making information among decision makers, in order to improve performance. Consider, for example, taskable human sensors in information-gathering missions who decide their next move in order to best contribute to situational awareness in the field. Under what circumstances will it be helpful for an individual to learn what choices others are making and/or how well others are performing?

A. Stewart and N. E. Leonard are with Department of Mechanical and Aerospace Engineering, Princeton University, USA ({arstewar,naomi}@princeton.edu).

This research was supported by AFOSR grant FA9550-07-1-0528.

Models that reliably predict how humans behave under relevant circumstances are critical to developing principled methodology for design of human-in-the-loop systems with social feedback. We leverage experimental and theoretical studies of human decision making from cognitive and social psychology, where the *Two-Alternative Forced-Choice* (TAFC) task has been used to investigate human decision-making in problems that require sequential choices among finite alternatives [4].

The successful fitting of both behavioral and neural data taken during TAFC task experiments provides strong justification for use of the *Drift Diffusion Model* (DDM) to describe human decision making in TAFC tasks [4], [5]. Recently authors of [6], [7] have extended studies to investigate multiple human subjects in social TAFC tasks; using experimental data they derive a model that couples multiple DDMs with a social feedback term.

In this paper we use the coupled DDM model [6], [7] and our previous analysis of an individual decision maker [8] to derive an analytic prediction of the role of social feedback in the steady-state performance of a decision maker who sees choices of M others (who are not receiving social feedback). Our results apply to a subset of TAFC tasks studied in [6], [7]. It is observed in [7] that a human subject in the TAFC task with the relatively easy *converging gaussians* reward structure experiences a performance decrease when given feedback on choices of other decision makers. We show formally that our prediction agrees with this observation.

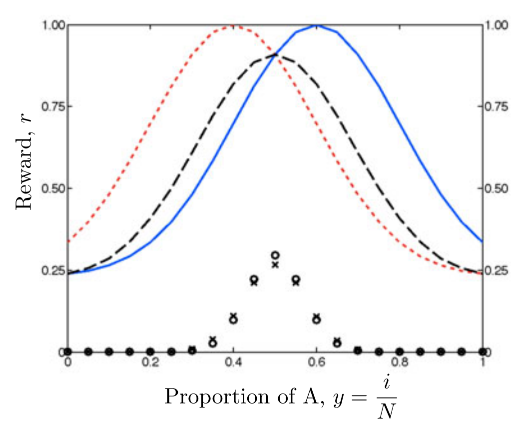

Fig. 1. The *converging gaussians* reward structure [6]: dotted curve r_A is the reward for choice A and solid curve r_B the reward for choice B. The dashed line is the average value of reward. The limiting distribution π_0, given by (11), is shown for $N = 20$ and $\mu = 2.5$ by the circular points. The limiting distribution π_0 for the model without Assumption 2 is shown by the x's. Each component π_{i0} is plotted against $y = \frac{i}{N}$.

The TAFC task in a social context is modeled in Section II and the DDM for decision making with social feedback is described in Section III. In Section IV we validate assumptions and in Section V we prove that our model is Markov. The steady-state choice distribution for an individual with feedback on choices of others is derived in Section VI and used in Section VII to prove performance in the converging gaussians task as a function of parameters defining the social feedback. We make final remarks in Section VIII.

II. The TAFC task in a social context

The two-alternative forced-choice (TAFC) task, introduced by Montague and co-authors [5], [4], has become a valuable tool for studying decision making in a variety of contexts. In the TAFC task a decision maker is required to choose between two alternatives (denoted A and B), sequentially in time, and a reward (performance measure) is received after each choice is made. The goal is to maximize accumulated reward over the duration of the task (optimize performance over the long run). The reward is a function not only of the immediate choice but also of the subject's recent history of choices [4], [9], [5]; this dependence on past decisions is highly relevant for real-world human-in-the-loop decision-making problems.

Figure 1 shows an example reward schedule called *converging gaussians* (CG) that is used in the human behavioral studies of [6], [7]. The reward r_A for choosing A (resp. r_B for B) is plotted as a dotted curve (resp. solid curve) as a function of $y = i/N$, where i is the number of times A is chosen in the past N decision trials. E.g., when $y = 0.4$, a choice of A yields a large reward of value 1, whereas a choice of B yields a significantly lower reward. The average value of reward $r_A(y)y + r_B(y)(1-y)$ is the dashed curve. This curve shows that for the CG task, choice sequences corresponding to $y = .5$ maximize average reward and thus accumulated reward. Note that the CG structure of Figure 1 is symmetric about $y = 0.5$, e.g., if $y = 0.6$, then the situation is the same as above with the roles of A and B swapped.

The point at which the reward curves intersect is called the *matching point*; it corresponds to the point at which the decision maker gets the same reward whether A or B is chosen. The CG task is one of several well studied tasks for which there is a matching point; others are the *matching shoulders* (MS) and the *rising optimum* (RO) reward structures. There is extensive empirical evidence that human decision makers converge in aggregate to choice sequences y that correspond to the matching point [5], [4], [6]. Convergence of human decision making to the matching point has been analyzed using decision-making models in [10], [8], [4], [11].

Interestingly, for the MS and RO tasks, decision making at the matching point does *not* typically correspond to the maximum average reward. For example, the RO structure has a matching point that is separated from the point of maximum reward by a range of decision sequences that yield very low rewards; this a "difficult" task since the decision maker must move away from the matching point and forego higher rewards in order to find the global optimum. The CG task, on the other hand, is a relatively "easy" task, since the matching point is coincident with the maximum reward.

The authors of [6], [7] have run extensive experiments with multiple human subjects to investigate decision making in TAFC tasks with social feedback. A series of TAFC experiments using CG, RO and *diverging gaussians* (DG) reward structures were run with groups of five human subjects who each received feedback on choices, rewards or both choices and rewards of the other four subjects in the group. One of the reported observations concerns the case of the CG task with choice feedback: decision-making performance deteriorates with choice feedback. The feedback seems to trigger increased exploration that takes the decision maker away from the otherwise easy-to-find optimal solution.

In this paper we derive probabilistic predictions for performance in the social TAFC task in the case of choice feedback. Our results apply to the CG, DG and MS tasks where the assumptions we make are validated. We consider the case of directed information passing; i.e., we study an individual given feedback on the choices of others, who themselves are not receiving any feedback. This differs from experiments in [6], [7] where feedback is all-to-all; however, we find our predictions agree qualitatively with the finding of decreased performance in the CG task with choice feedback.

We consider a focal individual in the social TAFC receiving feedback on choices of M other decision makers. Let $x(t) = (x_1(t), x_2(t), \dots, x_N(t))$ denote the past N choices of a focal decision maker ordered sequentially with $x_1(t) \in \{A, B\}$ the most recent decision at time t, $x_2(t) \in \{A, B\}$ the most recent decision at time $t-1$, etc. We have

$$x_k(t{+}1) = x_{k-1}(t), \qquad k = 2, \dots, N,\ t = 0, 1, 2, \dots \quad (1)$$

The proportion of choice A in the last N trials at time t is

$$y(t) = \frac{1}{N} \sum_{k=1}^{N} \delta_{kA}(t) \quad (2)$$

where $\delta_{kA}(t) = 1$ if $x_k(t) = A$ and $\delta_{ka}(t) = 0$ if $x_k(t) = B$. Note that y takes values from a discrete set $\mathcal{Y} = \left\{ \frac{i}{N},\ i = 0, 1, \dots, N \right\}$. The reward at time t is

$$r(t) = \begin{cases} r_A(y(t)) & \text{if } x_1(t) = A \\ r_B(y(t)) & \text{if } x_1(t) = B\,. \end{cases} \quad (3)$$

We define the difference in reward as

$$\Delta r(y(t)) := r_B(y(t)) - r_A(y(t)). \quad (4)$$

III. Drift Diffusion Model with Social Feedback

The Drift Diffusion Model (DDM) for decision making is described by a scalar drift diffusion process given by the following stochastic differential equation [6], [12], [13]:

$$dz = \alpha dt + \sigma dW, \qquad z(0) = 0. \quad (5)$$

Here z represents accumulated evidence in favor of a candidate choice, α is the drift rate representing signal intensity of stimulus acting on z and σdW is a Wiener process with standard deviation σ, which is the diffusion rate representing the effect of white noise.

Consider the TAFC task with choices A and B. The drift rate α, as described in [5], [14], is determined by a subject's anticipated rewards (denoted w_A for a decision of A and w_B for a decision of B). Let z be the accumulated evidence for choice A relative to choice B. Then on each trial a choice is made when $z(t)$ *first* crosses one of the predetermined thresholds $\pm\xi$. If $+\xi$ is crossed then choice A is made, and if $-\xi$ is crossed then choice B is made.

As pointed out in [6], using tools in [15] it can be shown that the probability of choosing A is

$$p_A(t+1) = \frac{1}{1+e^{-\mu(w_A(t)-w_B(t))}}, \tag{6}$$

where $\mu(w_A - w_B)$ is identified with $2(\alpha/\sigma)^2(\xi/\alpha)$. The right side of (6) is a sigmoidal function of $w_A - w_B$ with slope μ. Larger μ implies more certainty in decision making, sometimes interpreted as less of a tendency to explore.

We follow the approach in [6], [7] to model choice feedback by biasing anticipated rewards with a feedback parameter ν so that the probability of choosing A in the next time step is

$$p_A(t+1,\nu) = \frac{1}{1+e^{-\mu(w_A(t)-w_B(t)+\nu u(t))}} \tag{7}$$

$$u(t) = \begin{cases} 1 & \text{if } \#A\text{'s} \geq \lceil \frac{M+1}{2} \rceil \\ -1 & \text{if } \#B\text{'s} \geq \lceil \frac{M+1}{2} \rceil \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where $\#A$'s refers to the number of others (not receiving feedback) who choose A at time t, and $\lceil\cdot\rceil$ gives the smallest integer greater than its argument. The no-feedback case (6) is equivalent to $p_A(t+1,0)$ in (7).

Studies of the role of dopamine neurons in coding for reward prediction [16] motivate the use of temporal difference learning theory [17] to describe the dynamics of w_A and w_B. Let $Z \in \{A, B\}$ be the choice made at time t, then

$$\begin{aligned} w_Z(t+1) &= (1-\lambda)\omega_Z(t) + \lambda r(t) && (9) \\ w_{\bar{Z}}(t+1) &= w_{\bar{Z}}(t) \qquad t = 0,1,2,\ldots && (10) \end{aligned}$$

where $\bar{\cdot}$ denotes the "not" operator. Here, $\lambda \in [0,1]$ acts as a learning rate, controlling how the anticipated reward of choice Z at $t+1$ is affected by its value at t.

IV. Assumptions

In Section VI we analyze the focal DDM decision maker receiving choice feedback from M others in the TAFC task as modeled in Sections II and III by (1)-(3) and (7)-(10). Our approach is to use $p_A(t+1,0)$ to model the probability that each of the M others makes choice A at time $t+1$. The focal decision maker has the N-element decision history $x(t)$, the expected rewards $w_A(t)$ and $w_B(t)$ and the choices of M others at time t as the state. We make use of the following assumptions in Sections V and VI:

Assumption 1: $\Pr\{x_k(t) = A | x(t)\} = y(t)$

Assumption 2: $w_B(t) - w_A(t) = \Delta r(y(t))$.

Assumption 1 implies that the yN A's and $(1-yN)$ B's in $x(t)$ are uniformly distributed in the finite history. Assumption 2 sets the difference in anticipated rewards at time t equal to the difference in rewards evaluated at $y(t)$; according to Montague and Berns [4] this assumption is true "on average" in experiments.

We further examine Assumption 2 in this section by comparing equilibrium distributions for the model computed with and without Assumption 2 and showing they are close.

Let π_{i0} denote the long-run probability that $y(t) = \frac{i}{N}, i = 0, 1, \ldots, N$ for a decision maker without social feedback. In [8] we use Assumptions 1 and 2 to show that $\pi_{i0} = \pi_i(\mu,\nu,N,\Delta r)|_{\nu=0}$ is given by

$$\pi_{i0} = \frac{\alpha_i(1+e^{\mu\Delta r(\frac{i}{N})})e^{-\mu\beta_i}}{\sum_{j=0}^{N}\alpha_j e^{-\mu\beta_j}(1+e^{\mu\Delta r(\frac{j}{N})})} \tag{11}$$

where $\alpha_i = \frac{N!}{(N-i)!i!}$ and $\beta_i = \sum_{j=1}^{i}\Delta r(\frac{j}{N})$. This distribution for the CG reward structure is plotted with circular points in Figure 1 with $\mu = 2.5$ (the fitted value of μ from experimental data in the no-feedback condition [7]).

We also compute the steady-state distribution for the CG reward structure without using Assumption 2 in the case that $\lambda = 1$, which is typical of the fitted data for the CG, DG and MS reward structures [7]. When $\lambda = 1$ it can be seen in (9) and (10) that w_A and w_B are each restricted to a finite set defined by the values of r_A and r_B, respectively. Inclusion of w_A and w_B in the state space for $\lambda = 1$ therefore results in a state transition matrix with $(N+1)^3 \times (N+1)^3$ elements. Numerical computation of the steady-state distribution for the CG reward structure without using Assumption 2 yields the distribution plotted with x's in Figure 1. The closeness of the two distributions suggests Assumption 2 is valid for the CG task. We have similarly validated Assumption 2 in the DG and MS cases.

We can use (11) to determine $p_A(\infty, 0)$, the long-run probability that any of the M others (not receiving feedback) chooses A:

$$p_A(\infty,0) = \sum_{i=0}^{N}\frac{\pi_{i0}}{1+e^{\mu\Delta r(\frac{i}{N})}}. \tag{12}$$

Assumption 2 allows us to compute the probability that a focal decision maker will make choice A simply as a function of choice history, $y(t)$. By conditioning on the value of $u(t)$ we can compute the expectation of $p_A(t+1)$ in (7). Transition probabilities for a focal decision maker's proportion of choice A are determined by computing the probability that $x_N(t)$ is an A or B via Assumption 1. This reduces the state space to one dimension. We next show that the task and decision-making model in the social context is Markov under these assumptions.

V. Markov model of social decision making

Consider again the focal DDM decision maker in the social TAFC task. We derive the probability transition function for $y(t)$ and build a state transition matrix which is used in Section VI to compute the steady-state distribution for the process.

Proposition 1: Suppose Assumptions 1 and 2 hold. Then, the DDM with social feedback (7) for the TAFC task (1)-(3) is a Markov Process with state $y(t)$ and transition probabilities given by

$$\Pr\{y(t+1) = y(t) - \frac{1}{N}\} = \big[1 - \bar{p}_A\big(y(t)\big)\big]y(t) \quad (13)$$

$$\Pr\{y(t+1) = y(t)\} = \big[1 - \bar{p}_A\big(y(t)\big)\big]\big(1 - y(t)\big) + \bar{p}_A\big(y(t)\big)y(t) \quad (14)$$

$$\Pr\{y(t+1) = y(t) + \frac{1}{N}\} = \bar{p}_A\big(y(t)\big)\big(1 - y(t)\big) \quad (15)$$

where $\Delta r = \Delta r(y(t))$ is given by (4) and $\bar{p}_A(y(t))$ is

$$\bar{p}_A(y(t), \nu) = \frac{\Pr\{u(t) = 1\}}{1 + e^{\mu(\Delta r - \nu)}} + \frac{\Pr\{u(t) = -1\}}{1 + e^{\mu(\Delta r + \nu)}} + \frac{\Pr\{u(t) = 0\}}{1 + e^{\mu \Delta r}}. \quad (16)$$

The conditional probabilities on $u(t)$ are given by

$$\Pr\{u(t) = 1\} = \sum_{k=\lceil \frac{M+1}{2} \rceil}^{M} \begin{pmatrix} M \\ k \end{pmatrix} p_A(\infty, 0)^k (1 - p_A(\infty, 0))^{M-k}, \quad (17)$$

$$\Pr\{u(t) = -1\} = \sum_{k=\lceil \frac{M+1}{2} \rceil}^{M} \begin{pmatrix} M \\ k \end{pmatrix} (1 - p_A(\infty, 0))^k p_A(\infty, 0)^{M-k}, \quad (18)$$

$$\Pr\{u(t) = 0\} = 1 - \big(\Pr\{u(t) = -1\} + \Pr\{u(t) = 1\}\big) \quad (19)$$

and $\begin{pmatrix} M \\ k \end{pmatrix} = \frac{M!}{k!(M-k)!}$.

Proof of Proposition 1:

Since for a given choice $x_1(t+1)$ at time $t+1$, $y(t+1)$ can only change from $y(t)$ to $y(t) + \frac{1}{N}$, $y(t) - \frac{1}{N}$ or stay at $y(t)$, we need only compute the probability of each event for all $y(t) \in \mathcal{Y}$. Each of these depends upon the current value of $y(t)$ as well as $x_1(t+1)$ and $x_N(t)$ since $y(t+1)$ will only differ from $y(t)$ if $x_1(t+1)$ also differs from $x_N(t)$.

The event that $y(t+1) = y(t) - \frac{1}{N}$ requires $x_1(t+1) = B$ and $x_N(t) = A$. Treating these as independent events and using (7) yields

$$\Pr\{y(t+1) = y(t) - \frac{1}{N}\} = \frac{e^{\mu(w_B(t) - w_A(t) - \nu u(t))} y(t)}{1 + e^{\mu(w_B(t) - w_A(t) - \nu u(t))}}.$$

Substituting in Assumption 2, and treating the M peer decisions as independent events, we condition on the value of $u(t)$ and get $\Pr\{x_1(t+1) = B\} = 1 - \bar{p}_A(y(t), \nu)$ which with Assumption 1 gives us (13).

Similarly,

$$\Pr\{y(t+1) = y(t) + \frac{1}{N}\} = \frac{1 - y(t)}{1 + e^{\mu(w_B(t) - w_A(t) - \nu u(t))}}.$$

Conditioning on the value of $u(t)$ and substituting in Assumption 2, we get (15).

The event that $y(t+1) = y(t)$ requires either $x_1(t+1) = A$ and $x_N(t) = A$ or $x_1(t+1) = B$ and $x_N(t) = B$. The probability of the union of these events is

$$\Pr\{y(t+1) = y(t)\} = \frac{y(t) + (1 - y(t)) e^{\mu(w_B(t) - w_A(t) - \nu u(t))}}{1 + e^{\mu(w_B(t) - w_A(t) - \nu u(t))}}.$$

Conditioning on the value of $u(t)$ and substituting in Assumption 2, we get (14). Since the probabilities depend only upon $y(t)$, the current value of the state at time t, the process is Markov. $\square$

Equations (13)-(15) are used to build the $(N+1) \times (N+1)$ one-step state transition matrix $\mathbf{P}$ which has entries

$$P_{ij} = \Pr\{y(t+1) = \frac{j}{N} | y(t) = \frac{i}{N}\}, \quad (20)$$

$i, j \in \{0, 1, \ldots, N\}$.

VI. Steady-State Choice Distribution

Since the Markov process modeled in Section V is irreducible and aperiodic, it has a unique limiting distribution $\pi = (\pi_0, \pi_1, \ldots, \pi_N)$ describing the fraction of time the chain spends in each enumerated state ($y = \frac{i}{N}, i = 0, 1, 2 \ldots, N$) in the long run (as $t \to \infty$) [18]. The steady-state distribution is the solution to:

$$\pi \mathbf{P} = \pi \quad (21)$$

$$\sum_{i=0}^{N} \pi_i = 1. \quad (22)$$

Proposition 2: For the transition probabilities given by (13) - (15) the unique steady-state distribution is

$$\pi_i = \alpha_i \frac{\prod_{j=1}^{i} q(\frac{i}{N}, \nu)}{\sum_{j=0}^{N} \alpha_j \prod_{k=1}^{j} q(\frac{k}{N}, \nu)} \quad (23)$$

where $\alpha_i = \frac{N!}{(N-i)!i!}$ and $q(\frac{i}{N}, \nu) = \frac{\bar{p}_A(\frac{i-1}{N}, \nu)}{1 - \bar{p}_A(\frac{i}{N}, \nu)}$.

Proof of Proposition 2: Solving (21) alone yields a row vector v whose elements are given by

$$v_i = \frac{N!}{(N-i)!i!} \prod_{j=1}^{i} \frac{\bar{p}_A(\frac{j-1}{N}, \nu)}{1 - \bar{p}_A(\frac{j}{N}, \nu)}.$$

To solve (22) we normalize the vector v to get $\pi = v / \sum_{i=0}^{N} v_i$. The elements of π are then given by (23). $\square$

VII. Social feedback performance in Converging Gaussians task

A key measure of performance in the TAFC task with CG reward structure is variance in the decision making. Since the convergent matching point ($y = 0.5$) is coincident with the optimal allocation to choice A, better performance corresponds to minimizing variance about $y = 0.5$. In this section we prove for the predictive model that variance is minimal for $\nu = 0$; i.e. receiving social feedback in the CG task decreases performance. This is in direct agreement with experimental results of [7]. We also investigate the effect on performance of the strength of the feedback ν, the number of other decision makers M and the focal individual's own certainty parameter μ.

Let Σ denote the second moment of the steady-state distribution about $y = 0.5$. Then using π derived in Section VI, Σ can be written as a function of the feedback gain ν as

$$\Sigma(\nu) = \frac{\sum_{i=0}^{N} \alpha_i (\frac{i}{N} - \frac{1}{2})^2 Q_i(\nu)}{\sum_{j=0}^{N} \alpha_j Q_j(\nu)}, \tag{24}$$

where $Q_i(\nu) := \prod_{j=1}^{i} q(\frac{j}{N}, \nu)$.

The converging gaussians structure is given by

$$r_A(y) = e^{-\left(\frac{y-\bar{y}_A}{\sqrt{2}\sigma_A}\right)^2} + c_A, \;\; r_B(y) = e^{-\left(\frac{y-\bar{y}_B}{\sqrt{2}\sigma_B}\right)^2} + c_B. \tag{25}$$

In this work we consider symmetric CG structures in which $c_A = c_B$, $\bar{y}_A < \bar{y}_B$ and $|\bar{y}_A| = |\bar{y}_B|$ (as in Figure 1) so that $y = 0.5$ is the converging matching point.

A. *Effect of choice feedback on performance*

In Theorem 1 we prove that variance is minimized at $\nu = 0$ in the case $M = 4$. This implies that with choice feedback (corresponding to $\nu \neq 0$), the focal individual tends to do more exploring away from the optimal solution and performance deteriorates.

Theorem 1: For the CG reward structure of (25), in the case that after every choice the focal decision maker sees the most recent decision of each of $M = 4$ other decision makers performing the same task, the variance $\Sigma(\mu, \nu)$ of the focal individual is minimal for $\nu = 0$.

We prove Theorem 1 by first proving four lemmas.

Lemma 1: $\nu = 0$ is a critical point of $\Sigma(\mu, \nu)$

To prove Lemma 1 we introduce the following:

Lemma 2: $Q_i'(0) := \frac{\partial}{\partial \nu} \prod_{j=1}^{i} q(\frac{i}{N}, \nu)\big|_{\nu=0} = 0$.

Proof of Lemma 2: We compute

$$\frac{\partial q}{\partial \nu}(\tfrac{i}{N}, \nu) = \frac{\frac{\partial}{\partial \nu}\bar{p}_A\left(\frac{i-1}{N}, \nu\right)}{1 - \bar{p}_A\left(\frac{i}{N}, \nu\right)} - \frac{\frac{\partial}{\partial \nu}\bar{p}_A\left(\frac{i}{N}, \nu\right)\left(1 - \bar{p}_A\left(\frac{i-1}{N}, \nu\right)\right)}{\left(1 - \bar{p}_A\left(\frac{i}{N}\right), \nu\right)^2}. \tag{26}$$

In the case that $M = 4$, for the CG reward schedule $p_A(\infty, 0) = \frac{1}{2}$ and $\bar{p}_A$ in (7) becomes $\bar{p}_A(\frac{i}{N}, \nu) = \frac{3}{4}\frac{1}{(1+e^{\mu\Delta r})}\frac{1}{8}\left[\frac{1}{1+e^{\mu(\Delta r - \nu)}} + \frac{1}{1+e^{\mu(\Delta r + \nu)}}\right]$. Differentiating $\bar{p}_A$ with respect to ν yields

$$\frac{\partial}{\partial \nu}\bar{p}_A(\tfrac{i}{N}, \nu) = \frac{\mu e^{\mu\Delta r}}{8}\left[\frac{e^{-\mu\nu}}{(1+e^{\mu(\Delta r - \nu)})^2} - \frac{e^{\mu\nu}}{(1+e^{\mu(\Delta r + \nu)})^2}\right]. \tag{27}$$

Evaluating (27) at $\nu = 0$ we get $\frac{\partial}{\partial \nu}\bar{p}_A(\frac{i}{N}, \nu)|_{\nu=0} = 0, \forall i$. Therefore, in (26) we see that $\frac{\partial}{\partial \nu}q(\frac{i}{N}, \nu)|_{\nu=0} = 0$. From the definition of $Q_i'(\nu)$ we can write

$$Q_i'(\nu) = \sum_{k=1}^{i} \frac{\partial}{\partial \nu} q\left(\tfrac{k}{N}, \nu\right) \prod_{j=1, j\neq k}^{i} q\left(\tfrac{j}{N}, \nu\right). \tag{28}$$

Evaluating (28) at $\nu = 0$ with $\frac{\partial}{\partial \nu}q(\frac{i}{N}, \nu)|_{\nu=0} = 0$ gives $Q_i'(0) = 0$. □

Proof of Lemma 1: The derivative of $\Sigma(\mu, \nu)$ can be written

$$\frac{\partial}{\partial \nu}\Sigma(\mu, \nu) = \frac{\sum_{i=1}^{N} \alpha_i (\frac{i}{N} - \frac{1}{2})^2 Q_i'(\nu)}{\sum_{k=1}^{N} \alpha_k Q_k(\nu)} - \frac{\sum_{i=1}^{N} \alpha_i (\frac{i}{N} - \frac{1}{2})^2 Q_i(\nu) \sum_{k=1}^{N} \alpha_k Q_k'(\nu)}{\left(\sum_{k=1}^{N} \alpha_k Q_k(\nu)\right)^2}.$$

It follows from Lemma 2 that $\frac{\partial}{\partial \nu}\Sigma(\mu, \nu)|_{\nu=0} = 0$. □

It is now left to show that $\nu = 0$ is a minimum of $\Sigma(\mu, \nu)$.

Lemma 3: $\frac{\partial^2}{\partial \nu^2}\Sigma(\mu, \nu)\big|_{\nu=0} > 0$.

To prove Lemma 3 we introduce the following:

Lemma 4: $Q_i''(\nu) < 0$.

Proof of Lemma 4: Differentiating $Q_i'(\nu)$ with respect to ν, and making use of the fact that $\frac{\partial}{\partial \nu}\bar{p}_A(\frac{i}{N}, \nu)|_{\nu=0} = 0$ gives

$$Q_i''(0) = \frac{\frac{\partial^2}{\partial \nu^2}\bar{p}_A(\frac{i-1}{N}, \nu)|_{\nu=0}\left(1 - \bar{p}_A(\frac{i}{N}, 0)\right)}{\left(1 - \bar{p}_A(\frac{i}{N})\right)^2} + \frac{\frac{\partial^2}{\partial \nu^2}\bar{p}_A(\frac{i}{N}, \nu)|_{\nu=0}\left(\bar{p}_A(\frac{i-1}{N}, 0)\right)}{\left(1 - \bar{p}_A(\frac{i}{N})\right)^2}. \tag{29}$$

Since $\frac{\partial^2}{\partial \nu^2}\bar{p}_A(\frac{i}{N}, \nu)|_{\nu=0} = -\frac{\mu^2 e^{\mu\Delta r(\frac{i}{N})}(1+e^{2\mu\Delta r(\frac{i}{N})})}{(1+e^{\mu\Delta r(\frac{i}{N})})^4} < 0$, we can conclude that $Q_i''(0) < 0$. □

Proof of Lemma 3: Invoking Lemma 2 we can write

$$\frac{\partial^2}{\partial \nu^2}\Sigma(\mu, \nu)\big|_{\nu=0} = \frac{\sum_{i=0}^{N} \alpha_i \left(\frac{i}{N} - \frac{1}{2}\right)^2 Q_i''(0)}{\sum_{k=0}^{N} \alpha_k Q_k(0)} - \frac{\sum_{i=0}^{N} \alpha_i \left(\frac{i}{N} - \frac{1}{2}\right)^2 Q_i(0) \sum_{k=0}^{N} \alpha_k Q_k''(0)}{\left(\sum_{k=0}^{N} \alpha_k Q_k(0)\right)^2}. \tag{30}$$

Denote the numerator of $\frac{\partial^2}{\partial \nu^2}\big|_{\nu=0}\Sigma(\mu, \nu)$ by Γ. Then

$$\Gamma = \sum_{i=0}^{N}\sum_{k=0}^{N} \gamma_{i,k} \tag{31}$$

where $\gamma_{i,k} = \alpha_i Q_i''(0) \alpha_k Q_k(0) \left[\left(\frac{i}{N} - \frac{1}{2}\right)^2 - \left(\frac{k}{N} - \frac{1}{2}\right)^2\right]$. Lemma 4 tells us that $\gamma_{i,k} > 0$ for all i, k that satisfy

$$\left(\frac{i}{N} - \frac{1}{2}\right)^2 - \left(\frac{k}{N} - \frac{1}{2}\right)^2 < 0. \tag{32}$$

It is also true that $\gamma_{\frac{N}{2},\frac{N}{2}} = 0$. It can be shown that for all $i, k \neq \frac{N}{2}$, $\gamma_{i,k} > 0$ and $\gamma_{\frac{N}{2},\frac{N}{2}} = 0$. It therefore must be true that $\Gamma = \sum_{i=0}^{N}\sum_{k=0}^{N} \gamma_{i,k} > 0$. □

Proof of Theorem 1: Lemma 1 and Lemma 3 guarantee that $\nu = 0$ is a minimum of $\Sigma(\mu, \nu)$. □

B. *Sensitivity to ν, μ and M*

It is also of interest to examine sensitivity of performance in the social feedback case to decision-making parameters ν, μ and M. In Figure 2 the steady state distribution of y is plotted for $\mu = 2.6$ (the fitted value for an individual in the CG task with social feedback) without feedback and with feedback in the case $M = 4$ and $M = 2$. In Figure 3 the normalized standard deviation $100\sqrt{\Sigma(\mu, \nu, M)}$ is plotted as a function of ν for three different values of μ and M.

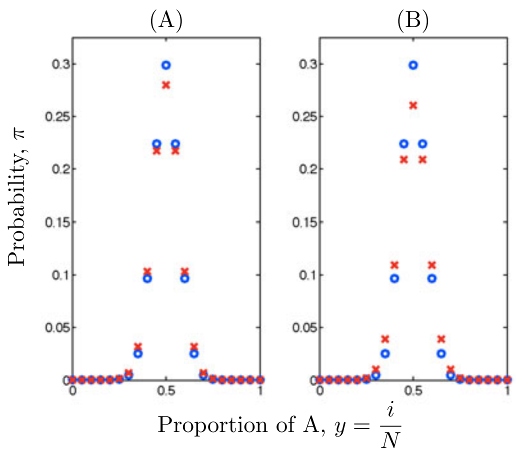

Fig. 2. Steady-state distribution of y for the CG task with social feedback. (A) $M = 4$ (B) $M = 2$. In each plot $\mu = 2.6$, the circular points correspond to $\nu = 0$, and the x's to $\nu = 1$. In each plot $N = 20$.

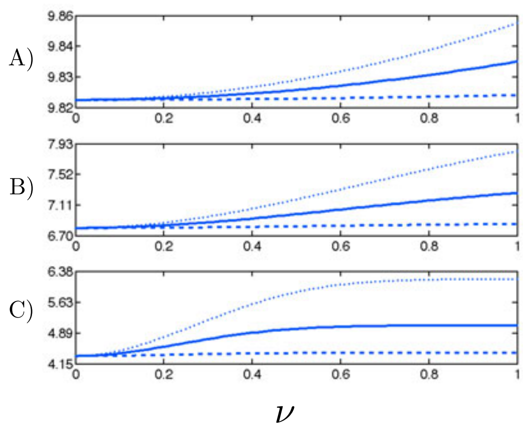

Fig. 3. Standard deviation of steady-state distribution of y from the mean $y = 0.5$ for the CG task as a function of feedback parameter ν (given by $100\sqrt{\Sigma(\mu, \nu, M)}$). A) $\mu = 0.5$ B) $\mu = 2.6$ C) $\mu = 10$. In each plot, the dotted curve corresponds to $M = 2$, the solid curve to $M = 4$ and the dashed curve to $M = 10$. In each plot $N = 20$.

Variance increases as a function of ν for all μ and M plotted; this is as predicted for $M = 4$ by Theorem 1. We also see variance is higher for smaller M. This implies that social feedback has a greater effect on performance in smaller groups of decision makers. As the number of decision makers in the group increases, so does the probability that among the M other choices there will be an equal number of $A's$ and $B's$, thereby decreasing the influence of social effects.

The results show that dependence of variance on μ is significant. In Figure 3 it can be seen that increasing μ (certainty in decision making) magnifies sensitivity to the feedback gain ν. In previous work [8] we determined that increasing μ in the CG task decreases variance for a single individual without social feedback. The "uncertainty" parameter μ and feedback gain ν have a coupling effect in the CG task with social feedback that causes a more substantial decrease in performance as ν increases for larger values of certainty μ.

VIII. Final Remarks

In ongoing work we are investigating the effect of social feedback on convergence rate and the role of the feedback interconnection topology. We are extending the current analysis to the undirected case in which each of the $M + 1$ decision makers receives feedback from M others as in experiments of [6], [7]. We are also considering the role of social feedback in tasks with more "difficult" reward structures such as the Rising Optimum; we predict that social feedback will improve performance there. A key goal of this research is to determine a principled means to design social interaction to improve performance and mitigate performance degradation in scenarios like the one studied in this paper.

References

[1] K. C. Campbell, Jr. W. W. Cooper, D. P. Greenbaum, and L. A. Wojcik. Modeling distributed human decision-making in traffic flow management operations. In *3rd USA/Europe Air Traffic Management R&D Seminar*, Napoli, June 2000.

[2] B. Donmez, M. L. Cummings, and H. D. Graham. Auditory decision aiding in supervisory control of multiple unmanned aerial vehicles. *Human Factors: The Journal of the Human Factors and Ergonomics*, 51(5):718–729, October 2009.

[3] A. R. Otto, T. M. Gureckis, A. B. Markman, and B. C. Love. Navigating through abstract decision spaces: Evaluating the role of state generalization in a dynamic decision-making task. *Psychonomic Bulletin and Review*, 16(957-963), 2009.

[4] P. R. Montague and G. S. Berns. Neural economics and the biological substrates of valuation. *Neuron*, 36:265–284, 2002.

[5] D. M. Egelman, C. Person, and P. R. Montague. A computational role for dopamine delivery in human decision-making. *Journal of Cognitive Neuroscience*, 10:623–630, 1998.

[6] A. Nedic, D. Tomlin, P. Holmes, D.A. Prentice, and J.D. Cohen. A simple decision task in a social context: experiments, a model, and preliminary analyses of behavioral data. In *Proc. of the 47th IEEE Conference on Decision and Control*, 2008.

[7] A. Nedic, D. Tomlin, P. Holmes, D.A. Prentice, and J.D. Cohen. A decision task in a social context: Behavioral experiments, models, and analyses of behavioral data. Submitted, 2010.

[8] A. Stewart, M. Cao, and N. E. Leonard. Steady-state distributions for human decisions in two-alternative choice tasks. In *Proc. of the American Control Conference*, 2010.

[9] T.M. Gureckis and B.C. Love. Learning in noise: Dynamic decision-making in a variable environment. *Journal of Mathematical Psychology*, 53:180–193, 2008.

[10] M. Cao, A. Stewart, and N. E. Leonard. Convergence in human decision-making dynamics. *Systems and Control Letters*, 59:87–97, 2010.

[11] L. Vu and K. Morgansen. Modeling and analysis of dynamic decision making in sequential two-choice tasks. In *Proc. of the 47th IEEE Conference on Decision and Control*, 2008.

[12] B. K. Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer-Verlag, Berlin, 2003.

[13] P. Simen and J. D. Cohen. Explicit melioration by a neural diffusion model. *Brain Research*, 1299:99–117, 2009. Submitted to *Brain Research*.

[14] R. Bogacz, S. M. McClure, J. Li, J. D. Cohen, and P. R. Montague. Short-term memory traces for action bias in human reinforcement learning. *Brain Research*, 1153:111–121, 2007.

[15] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen. The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113:700–765, 2006.

[16] P. R. Montague, P. Dayan, and T. J. Sejnowski. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, 16:1936–1947, 1996.

[17] R. S. Sutton and A. G. Barto. *Reinforcement learning*. MIT Press, Cambridge, MA, 1998.

[18] H. M. Taylor and S. Karlin. *An Introduction to Stochastic Modeling -3rd ed*. Academic Press, 1998.