INVITED PAPER

A Decision Task in a Social Context: Human Experiments, Models, and Analyses of Behavioral Data

Studies of decision making in small human groups reveal behavior that illustrates both advantages and disadvantages to social feedback.

By Andrea Nedic, Damon Tomlin, Philip Holmes, Deborah A. Prentice, and J. D. Cohen

ABSTRACT | To investigate the influence of information about fellow group members in a constrained decision-making context, we develop four two-armed bandit tasks in which subjects freely select one of two options (*A* or *B*) and are informed of the resulting reward following each choice. Rewards are determined by the fraction x of past *A* choices by two functions $f_A(x), f_B(x)$ (unknown to the subject) which intersect at a matching point $\bar{x}$ that does not generally represent globally optimal behavior. Playing individually, subjects typically remain close to the matching point, although some discover the optimum. Each task is designed to probe a different type of behavior, and subjects work in parallel in groups of five with feedback of other group members' choices, of their rewards, of both, or with no knowledge of others' behavior. We employ a soft-max choice model that emerges from a drift-diffusion process, commonly used to model perceptual decision making with noisy stimuli. Here the stimuli are replaced by estimates of expected rewards produced by a temporal-difference reinforcement-learning algorithm, augmented to include appropriate feedback terms. Models are fitted for each task and feedback condition, and we compare choice allocations averaged across subjects and individual choice sequences to highlight differences between tasks and intersubject differences. The most complex model, involving both choice and reward feedback, contains only four parameters, but nonetheless reveals significant differences in individual strategies. Strikingly, we find that rewards feedback can be either detrimental or advantageous to performance, depending upon the task.

KEYWORDS | Decision making; drift-diffusion model; exploitation; exploration; group dynamics; human behavior; reinforcement learning; social information; two-armed bandit task

Manuscript received August 18, 2010; revised June 9, 2011; accepted July 29, 2011. Date of publication November 2, 2011; date of current version February 17, 2012. This work was supported by the Air Force Office of Scientific Research (AFOSR) under Grant FA9550-07-1-0528 under the Multidisciplinary University Research Initiative. A preliminary account of part of this work appeared in J. D. Cohen *et al.*, "Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration," *Phil. Trans. Roy. Soc. Lond. B*, vol. 362, pp. 933–942, 2007.

A. Nedic is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: nedic@Princeton.edu).

D. Tomlin and **J. D. Cohen** are with the Department of Psychology and Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544 USA (e-mail: dtomlin@Princeton.edu; jdc@Princeton.edu).

P. Holmes is with the Princeton Neuroscience Institute, Department of Mechanical and Aerospace Engineering, and Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544 USA (e-mail: pholmes@math.princeton.edu).

D. A. Prentice is with the Department of Psychology, Princeton University, Princeton, NJ 08544 USA (e-mail: predebb@Princeton.edu).

Digital Object Identifier: 10.1109/JPROC.2011.2166437

I. INTRODUCTION

In order to understand and model decision making in small human groups, we have designed and executed a highly constrained experiment that probes the manner in which limited sharing of information among the group influences individual choices. We have adapted and generalized an experimental paradigm of Egelman *et al.* [12], cf. [26], to a social context, allowing different types of feedback about group members to subjects performing a simple task: a two-armed bandit in which one of two alternatives is

0018-9219/$26.00 ©2011 IEEE

selected on each trial and a reward is delivered. Performing individually, subjects presumably base their current choices on previous rewards. We wish to understand how limited information about others' rewards, choices, or both, modifies individual behaviors, and to determine neural correlates of, and mechanisms determining, this process.

Rewards are computed by a deterministic rule, based on the subject's choice history, but unknown to him or her. In the task of [12] and [26], described in Section II-A as the simple rising optimum, two different types of behavior were observed. A majority of subjects settled near a matching point, at which both choices result in the same (moderate) reward, while a minority of "explorers" endured runs of low rewards and discovered a global optimum that is substantially better than the matching strategy. This type of rule permits examination of whether subjects exploit a particular, easily discovered strategy, or explore different ones in seeking to maximize their rewards.[1] The *explore versus exploit* question, central to studies of foraging in animal communities [23], is of increasing interest in cognitive neuroscience [8].

We generalize the model of [12] and [26] to include bias terms, representing feedback about other group members, in the expression for choice probabilities. We show that the models fit behaviors of subjects performing alone, and of subjects with access to information about the choices and/or rewards of other group members. We also show that these probabilities have the same form as those predicted by the widely used drift-diffusion (DD) model of perceptual choices, in which evidence in favor of one choice over the other is integrated until a predetermined threshold is reached. The DD model and extensions of it have been fitted to accuracy and reaction time data in numerous two-alternative forced-choice tasks [30], [31], [37]. For a recent review and derivations of DD processes from other, more complex, neurally based models, see [2]. Related articles map a robot foraging task onto the two-armed bandit task and provide rigorous results on performance of simplified choice models [6], [7], [39], [41].

In Section II, we describe the four tasks, review the reinforcement-learning model of [12] and [26], explain the model's dynamics in the limiting cases of completely random and purely deterministic choices, and note implications for convergence toward specific choice behaviors. In Section III, we extend the model to the group context by introducing biasing terms driven by information on choices and rewards of other group members; several alternative models are proposed. Section IV presents analyses of data and fits of these models to data, starting with choice behaviors averaged across subjects and over time (Section IV-A), and moving to individual choice sequences (Section IV-B). We identify the most promising models, and use them to analyze group and individual behaviors in Sections IV-C and D. A discussion ensues in Section V. Details of experimental methods and data analysis are provided in the Appendix. Functional magnetic resonance imaging (fMRI) data from the same groups is described in a forthcoming paper [44], and analyses of related models appear in a companion paper in this special issue [40].

[1] In [26], these behaviors are called "conservative" and "risky."

II. FOUR GAMBLING TASKS AND A SIMPLE CHOICE MODEL

Here we describe the tasks and reward schedules, review the model for individual choices, and provide brief analyses of its behavior.

A. Four Gambling Tasks

Upon pressing buttons A or B our two-armed bandits deliver rewards determined by schedules $r = f_A(x)$ and $r = f_B(x)$, respectively, where $x \in [0,1]$ is the fraction of A choices or *allocation to* A made over the past N trials. (In the present study $N = 20$.) If f_A lies entirely above (or below) f_B, subjects will rapidly deduce the better option and thereafter always choose A (or B); interesting cases occur when the functions f_A, f_B cross at a *matching point*, $\bar{x}$, so called because the rewards are equal there.

At any *constant allocation* $\boldsymbol{x}$ *to* $\boldsymbol{A}$ *the expected* reward is

$$R(x) = xf_A(x) + (1-x)f_B(x). \tag{1}$$

Since local maxima of (1) are stable fixed points of a gradient dynamical system [16] with potential function $R(x)$

$$\dot{x} = R'(x) = f_A(x) - f_B(x) + f_B'(x) + x\left[f_A'(x) - f_B'(x)\right] \tag{2}$$

rewards could be maximized by following a hill-climbing algorithm. However, subjects do not know that their rewards depend on their choice histories, nor that they are deterministic, and in any case they cannot easily estimate the functions $f_{A,B}(x)$ from choice-to-choice observations, especially if N is large. This has motivated the neurobiologically plausible model described in Section II-B.

The reward schedules used in this work generalize a "matching shoulders" task with linear functions $f_A(x) = a_1 - b_1 x$ and $f_B(x) = a_2 + b_2 x$, where $a_j, b_j > 0$ and $b_1 + b_2 > a_1 - a_2 > 0$. These cross at $\bar{x} = (a_1 - a_2)/(b_1 + b_2)$ and the global maximum lies at $x_{\max} = (a_1 - a_2 - b_2)/[2(b_1 + b_2)]$, showing that maxima need not lie at matching points: see Fig. 1. When these points do not coincide, experiments have shown that subjects typically tend to match rather than maximize, with allocations hovering around the matching point (Herrnstein's matching law) [19]–[21].

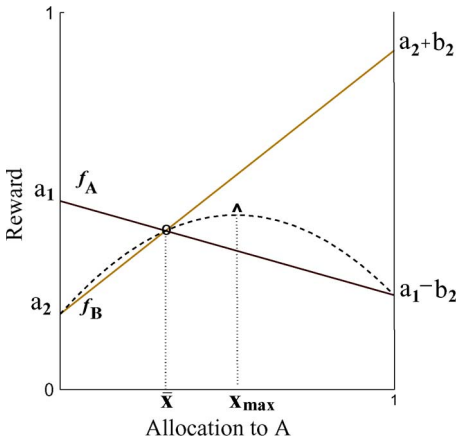


Fig. 1. ***The matching shoulders task with linear reward schedules, showing that optimal behavior (triangle) need not coincide with matching (circle); coincidence occurs if and only if*** $f_A(\mathbf{0}) = f_B(\mathbf{1})$ ***[26, App. B].***

In order to probe a wide range of behaviors, and in particular to investigate exploitation versus exploration, we have developed four tasks with the reward schedules shown in Fig. 2. The converging Gaussians (CG) and diverging Gaussians (DG) both have symmetric average reward curves $R(x)$ and unique global optima at 50% *A*-allocation, which are also their matching points. The optimum in the CG task is easy to find and maintain (as noted below, it is dynamically stable), so that the distance that subjects move from optimum may be used to gauge baseline levels of their exploratory tendencies. In contrast, the optimum in the DG task is unstable, pushing subjects to the right and left with equal probabilities, so that the extent to which they move together in one direction or the other may be used to gauge baseline levels of herding tendencies.

The complex rising optimum (CRO) and simple rising optimum (SRO) tasks are modifications of the task of [12] and [26]. In those papers, the net rewards curve rises monotonically with x and there is no local maximum at or near the matching point. (It was nonetheless found in [12] that a majority of subjects remained near this point.) The present CRO and SRO schedules have local maxima at 0% *A*'s, below the matching points, and global optima at 75% *A*'s (CRO) and 100% *A*'s (SRO), respectively. These were designed to make the global optima more difficult to discover (fewer than one in five subjects playing alone find them), but to draw exploratory and herding behaviors from other participants, given appropriate feedback. These two nonsymmetric schedules are also administered in versions reflected about 50% *A*'s, with local maxima at 100% *A*'s.

Each of the resulting four tasks is performed under each of four feedback conditions regarding the previous trial: no-feedback (subjects see only their own rewards), choice (subjects see other group members' choices), reward (subjects see others' rewards), and both (subjects see others' choices and rewards). The experimental procedure is described in part A of the Appendix.

In the CG task, the optimum coincides with the matching point $\bar{x}$, and so we expect all subjects, whether performing with or without feedback, to hover around this point. Near it the reward schedules are well approximated by the linear case of Fig. 1, and this "stable" behavior can be intuitively understood by noting that, if a subject is at the matching point and chooses *A*, his *A*-allocation typically rises and his reward drops ($f_A(x) < f_A(\bar{x})$ for $x > \bar{x}$), prompting him to choose *B* on the next trial. Choice of *B* and a probable reduction in *A*-allocation also results in a lower reward ($f_B(x) < f_B(\bar{x})$ for $x < \bar{x}$), again prompting reversal. Thus, choices tend to cycle around $\bar{x}$. As we will see in Section IV-A, our subjects do exhibit this behavior.

The optimum and matching point $\bar{x}$ also coincide at 50% for the DG schedule, but this point is typically *unstable*, as a similar argument suggests. The slopes of f_A

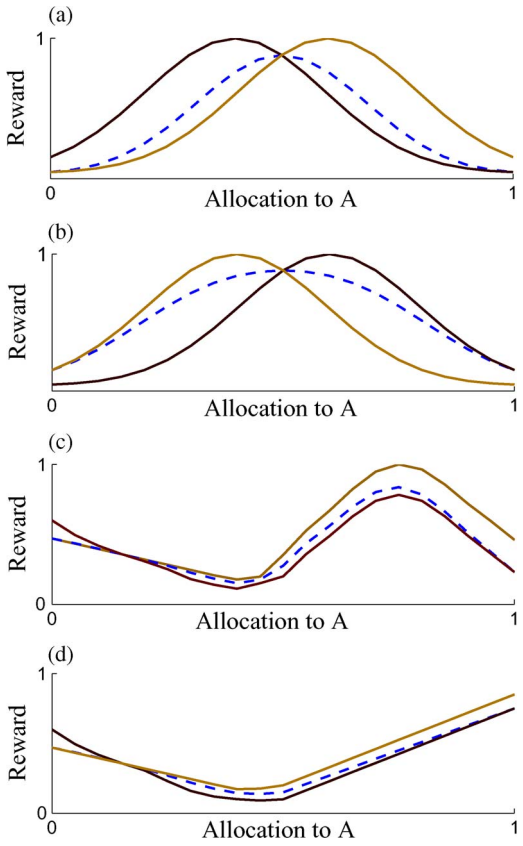


Fig. 2. ***Reward schedules*** f_A ***(dark solid) and*** f_B ***(light solid) for the four tasks: (a) CGs; (b) DGs; (c) CRO; and (d) SRO. Dashed curves denote average rewards*** $R(x)$***. There are two versions of each rising optimum task, one as shown, and one with schedules reflected about the midpoint 50%*** *A*.

and f_B are now reversed and, starting from $\bar{x}$, repeated A *and* B choices both initially lead to *higher* rewards and further divergence, suggesting that subjects will not match in this task.

B. A Model for Individual Choices Without Group Feedback

Egleman *et al.* [12] model the probability of choosing A on the nth trial by the softmax function

$$P(A) = \frac{1}{1 + \exp(-\mu[w_A(n) - w_B(n)])} \tag{3}$$

where $w_A(n)$ and $w_B(n)$ are weights or *expected rewards* accorded to choices A and B and the gain parameter $\mu \in (0, \infty)$ controls the slope of $P(A)$, and hence the degree of randomness in choosing. (The probability of choosing B is simply $1 - P(A)$.) Weights are updated as follows: if A is chosen on the nth trial, resulting in a reward $r(n)$, one sets

$$\begin{aligned} w_A(n+1) &= (1-\lambda)w_A(n) + \lambda r(n) \\ w_B(n+1) &= w_B(n) \end{aligned} \tag{4}$$

and if B is chosen, the roles of A and B in (4) are reversed and $w_A(n+1)$ remains unchanged. This rule is motivated by the role of dopamine neurons in coding for reward prediction error [27], [32] and by temporal difference reinforcement learning (TDRL) theory [42], [43].

The *learning rate* $\lambda \in [0, 1]$ determines the time scale on which the memory of previous choices decays: for $\lambda = 0$, no learning occurs; for $\lambda = 1$, the expected rewards for A and B are determined solely by the most recent rewards for those choices, previous choices being forgotten. Both weights are initialized at $w_A(0) = w_B(0) = 0$, except as noted in Sections IV-C1 and C2. In [3], a further time scale was incorporated in decaying eligibility traces to better describe self- or fast-paced versions of the task. Since this did not improve fits of allocation distributions for the present data (for which the task was neither self- nor fast-paced), we do not use it here.

The two-armed bandit demands a choice between two alternatives. In the case of perceptual decisions, such tasks are frequently modeled by a DD process, the simplest form of which is

$$dy = \alpha\, dt + \sigma\, dW, \qquad y(0) = 0 \tag{5}$$

where the drift rate α quantifies the difference between mean incoming evidence for A and B, $\sigma\, dW$ denote increments drawn from a Wiener process with standard deviation σ, and $y(t)$ represents integrated evidence in favor of A over B (the logarithmic likelihood ratio [15]). On each trial the choice A or B is made when $y(t)$ first crosses one of the predetermined thresholds $\pm y_{th}$, and the probability of choosing A is given by [5], [13]

$$P_{\mathrm{DD}}(A) = \frac{1}{1 + \exp(-2\alpha Z/\sigma^2)}. \tag{6}$$

Identifying the product of the gain parameter and weight difference $\mu[w_A - w_B]$ with $2\alpha Z/\sigma^2$, (3) and (6) coincide and we can therefore map the choice component of Egleman *et al.*'s model onto a well-understood decision making model [30], [31], [37]. Specifically, interpreting μ as the ratio of threshold separation to noise variance, the difference between expected rewards plays a role analogous to drift rate. Moreover, Simen *et al.* [36] show that, with thresholds inversely proportional to reward rates and $\alpha/\sigma = 0$, or with fixed thresholds and adjustable drift rate, the DD model (5) can reproduce Herrnstein's matching law (cf. Fig. 1), and that reward rates can be estimated by TDRL.

As described in [2] and [15], (5) is a continuum limit of the sequential probability ratio test of statistical decision theory [46], [47], which is optimal in the sense that it identifies a sequence of noisy stimuli, with guaranteed average accuracy, in the shortest possible time. However, we cannot predict maximization of rewards even if subjects follow this procedure: that would require optimal learning of the weights $w_A(n), w_B(n)$, and we are not aware of general results in TDRL theory that imply this across tasks such as those used here. We discuss the relationship between the models further in Section V.

The stability argument sketched in Section II-A for the CG and matching shoulders tasks is formalized for the model of (3) and (4) in [26, App.] where it is shown that, if the weight difference $w_A - w_B$ approximates the difference in rewards $\Delta r(x) = f_A(x) - f_B(x)$ in a neighborhood of $\bar{x}$ and $f_A'(\bar{x}) < 0 < f_B'(\bar{x})$, then allocations will remain near $\bar{x}$. More extensive and rigorous results on stability and equilibrium distributions for simplified versions of this and a related "win-stay lose-switch" model, both with and without group feedback, appear in [6], [7], and [45].

C. Dynamical Analyses of the Choice Model

Complete stability proofs for the choice model of Section II-B (3), (4) seem problematic due to the decaying memory, but its behavior is easily understood in the random and deterministic limits $\mu = 0$ and $\mu \to \infty$. In the former case neither previous choices made nor rewards accrued influence the choice probability $P(A) \equiv 1/2$, and the allocation sequence is an unbiased random walk on the interval $x \in [0, 1]$ with reflecting boundary conditions.

This generates a uniform allocation distribution regardless of the underlying reward schedules.

For $\mu = \infty$, the current value of $\Delta w(n) = w_A(n) - w_B(n)$ determines each choice: A if $\Delta w(n) > 0$ and B if $\Delta w(n) < 0$. Supposing that $\Delta w(n) > 0$ and the last B choice occurred at trial $n-k$ $(k \geq 1)$, the update algorithm (4) implies that A is chosen repeatedly on trials $n, n+1, \ldots$ until $\Delta w(n+j)$ changes sign, which cannot occur until $f_A(x_{n+j}) < w_B(n-k)$. (For $\lambda = 1$ $w_A(n+j) = f_A(x_{n+j})$, but for $\lambda \in (0,1)$ $w_A(n+j)$ can remain above $f_A(x_{n+j})$ for one or more further choices.) A similar conclusion holds with A and B interchanged if $\Delta w(n) < 0$. For the matching shoulders and CG schedules, initialized at the matching point $\bar{x}$ with $w_A(0) = w_B(0) = f_A(\bar{x}) = f_B(\bar{x})$, this algorithm predicts growing strings of A's and B's as the A-allocation describes unstable oscillations around $\bar{x}$ until it settles at $x = 1$ if $f_A(1) > f_B(0)$ or $x = 0$ if $f_A(1) < f_B(0)$. Applied to the diverging Gaussians schedule, it predicts divergence towards $x = 1$ (resp., $x = 0$) if the first choice is A (resp., B), with one or more B (resp., A) choices interposed when the current rewards fall below $f_A(\bar{x}) = f_B(\bar{x})$.

The rising optimum schedules may be partitioned into segments containing the matching points in which $f_A(x)$ and $f_B(x)$ both decrease with increasing x, and segments in which $f_B(x) > f_A(x)$ [on the left and right, respectively, in Fig. 2(c) and (d)]. Initialized at the matching point as above, the deterministic limit first chooses A or B (with equal probability), and the choice repeats until the reward falls below $f_A(\bar{x}) = f_B(\bar{x})$. This occurs in the case of A's as soon as $x > \bar{x}$, when choices switch to B's; then, since $f'(B) < 0$, a string of B's ensues and $x \to 0$. If B is chosen first, an uninterrupted string of B's leads directly to $x = 0$. On the right, an initial B choice will lead to a string of B's, and decreasing x, until $f_B(x) < w_A(0)$, but the resulting switch to A will produce a lower reward and choices will therefore revert to B's, possibly interrupted by occasional A's, passage into the left-hand side region, and convergence to $x = 0$. Only an initial choice of A with $w_B(0) < f_A(x)$ can produce a string of A's (with increasing rewards) and convergence to the global optimum.

Initialization of $w_A(0)$ and $w_B(0)$ is critical: employing $w_A(0) = w_B(0) = 0$ as in the model runs reported below, the deterministic limit predicts that the initial (randomly determined) choice persists without change for every task, since rewards for either choice are always strictly positive. This and the analyses above show that behavior is typically unstable, resulting in growing oscillations or monotonic changes in A-allocation, followed by convergence to repeated A's or B's. An element of random, exploratory behavior is evidently key to the simple reinforcement learning algorithm's success. This fact is used in recent studies of a Markov approximation to the model, for which equilibrium distributions of A-allocations can be explicitly computed [39]–[41].

III. MODELS FOR CHOICES WITH GROUP FEEDBACK

In the no-feedback condition subjects can still be modeled by independent soft-max/DD processes, although the knowledge that they are in a group situation may require parameter modifications. Given explicit information regarding other group members' choices and/or rewards on the previous trial, however, we *expect* updates of the weights w_A, w_B to differ. When only choice feedback is provided it is not clear that other subjects, whose choices deviate significantly from one's own, are doing better or worse, while if only reward scores are provided, the strategies by which these are achieved remain unknown. With this in mind, we model parameter updates as follows.

For choice feedback, we propose a majority rule: each individual increases his or her probability of choosing A (or B) by an amount determined by the fraction of A's (or B's) chosen by the other group members on the previous trial. Specifically, on the $(n+1)$th trial we add the quantity

$$\nu_c f(n) \cdot \begin{cases} +1, & \text{if } \#A\text{'s} > \#B\text{'s in } n\text{th choice} \\ -1, & \text{if } \#B\text{'s} > \#A\text{'s in } n\text{th choice} \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

to the difference $w_A(n+1) - w_B(n+1)$ between the weights, so that the softmax function of (3) becomes

$$P(A) = \frac{1}{1+\exp(-\mu[w_A(n+1) - w_B(n+1) + \nu_c f(n) \cdot \{\ldots\}])}. \tag{8}$$

Here $\{\ldots\}$ denotes the three alternatives of (7) and $f(n)$ is a ternary function that takes the value $+1$ if the individual followed the majority choice on trial $n-1$, -1 if (s)he went against it, and 0 if there was no majority (e.g., *AABB*). This range of values allows for "contrarians" who tend to oppose the majority $(f(n) < 0)$, and $f(n)$ can also be smoothed by averaging over windows of length $k > 1$. The parameter ν_c, which remains fixed for a given subject and task, describes the influence accorded to information about the group; it will provide a third fitting parameter in the studies of Section IV below. Examination of typical variations of $|w_A(n) - w_B(n)|$, which remained less than 0.5 in preliminary fitting, led us to restrict $\nu_c \in [0,1]$ to avoid overlarge biases.

In preliminary work, we considered a simpler rule, omitting the correlation function $f(n)$ in (7) and ignoring the fact that a subject may tend to follow the lead of others (herding behavior), or actively counter it; $f(n)$ was introduced to allow for such subject- and/or time-dependent

behavior. We also investigated a rule that updates the gain of the choice function (3) according to

$$\mu(n+1) = \mu(n) + \nu_c f(n) \cdot \begin{cases} +1, & \text{if one's choice matches the majority} \\ -1, & \text{if one's choice opposes the majority} \\ 0, & \text{if no majority exists.} \end{cases} \tag{9}$$

This makes a noncontrarian subject's behavior more deterministic when choices are validated, and more exploratory when the opposite occurs.

Given feedback on others' rewards without knowledge of their choices, we suppose that, when another group member's reward exceeds his own, an individual will tend to explore different strategies than those previously pursued. A rule analogous to (7) achieves this, in which the individual increases his probability of choosing one of the alternatives, selected at random on each trial, by an amount scaled to the absolute difference between his own reward and the maximum reward of the entire group. Specifically, to the preference weight difference $w_A(n+1) - w_B(n+1)$ we add the quantity $u(n+1)$, determined by

$$u(n+1) = u(n) + \nu_r b(n) |r(n) - \max(\text{all rewards on trial } n)| \tag{10}$$

so that the exponent in (3) becomes $-\mu[w_A(n+1) - w_B(n+1) + u(n+1)]$. Here $u(0) = 0$ and $b(n)$ is a binary random variable taking the values ± 1 with equal probability. No augmentation occurs if no other group member's reward exceeds the individual's. The quantity $r(n) - \max(\ldots)$ in (10) can be replaced by $1 - rr(n)$, where rr is the individual's *reward rank*, equaling 1 if he or she has the highest reward and 0 if the lowest, in decrements of 0.25 (cf. [44], and recall that subjects play in groups of five).

We considered several other rules for reward feedback. Exploration can also be promoted by reducing the gain parameter μ of (3) to promote random choices, or, if current rewards exceed those of all other group members, increasing it to favor deterministic choices and exploitation, according to

$$\mu(n+1) = \mu(n) \cdot \begin{cases} (1+\nu_r), & \text{if one's reward is maximum} \\ (1-\nu_r), & \text{if another's reward is maximum} \\ 1, & \text{otherwise.} \end{cases} \tag{11}$$

Here $\mu(0)$ is set at the value estimated with no-feedback, and we take $\nu_r \in [0,1]$, to avoid changing the sign of $\mu(n)$. We also examined additive updates $\mu(n+1) = \mu(n) \pm \nu_r$, as well as versions of the rules (10) and (11) in which $u(n)$ and $\mu(n)$ are not cumulatively updated, but modified from fixed values on each trial. A deterministic rule analogous to the choice feedback of (7) was also considered, in which $w_A(n+1) - w_B(n+1)$ was augmented by adding

$$\nu_r f(n) \cdot \begin{cases} +1, & \text{if chose } A \text{ and own reward maximum or} \\ & \text{chose } B \text{ and own reward not maximum} \\ -1, & \text{if chose } B \text{ and own reward maximum or} \\ & \text{chose } A \text{ and own reward not maximum.} \end{cases} \tag{12}$$

In all these cases, one additional parameter (ν_c or ν_r), to be fitted to data, describes the susceptibility to feedback.

More complex rules can be envisaged for both choice and reward feedback, but here we will consider only linear superposition of the separate choice and reward updating rules [see (7)–(10)] proposed above. As noted in Section II-C, analyses of approximations to some of the above models, including computation of equilibrium allocation distributions, appear in [39]–[41].

IV. MODEL FITS TO BEHAVIORAL DATA

In this section, we present analyses of behavioral data from the five-member groups. We start by comparing block- and subject-averaged behaviors for each task, expressed as distributions of A-allocations. We then fit individual subjects' data using a maximum-likelihood criterion, and using these fits, investigate the striking findings that rewards feedback can hurt performance in a simple task (converging Gaussians), while it can assist in finding global optima in the more difficult rising optimum tasks, by promoting exploration. We finally return to examine individual parameter differences across subjects and the corresponding individual allocation distributions.

A. A-Allocations With No-Feedback: Averaged Behaviors

We start by matching the choice model (3), without feedback, to data from individuals performing all the tasks of Fig. 2 in the no-feedback condition, weights updated according to (4). Fig. 3 shows histograms of choice allocations from the data pooled across subjects for each of the four tasks. To determine initial A-allocations and hence rewards, subjects were all given "seed" histories of $N = 20$ trials (unknown to them). All subjects performed all six variants of the four tasks, including both versions of the

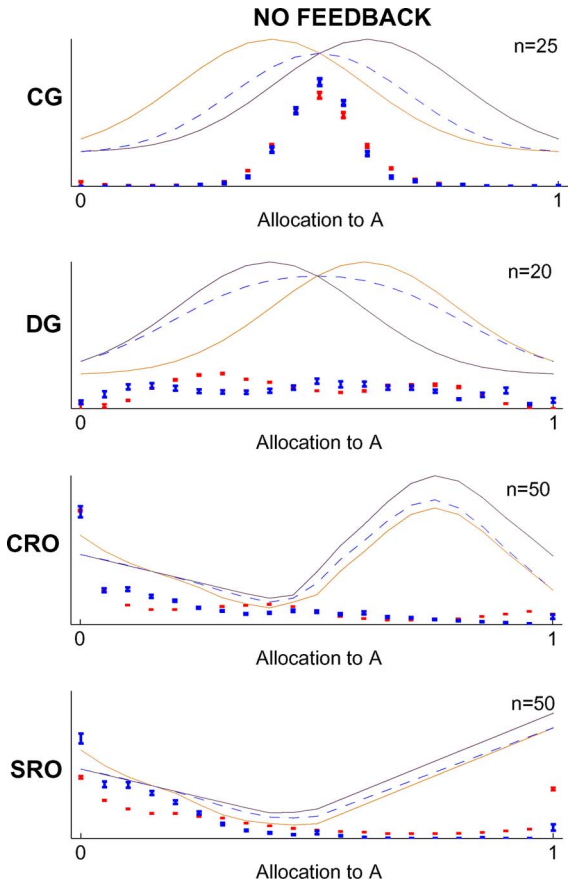


Fig. 3. ***Allocation distributions with standard error bars for subjects performing the four tasks of Fig. 2 (blue) in the no-feedback condition, compared with fits of the choice model (3) with no-feedback (red). Reward schedules are shown as faint curves with average rewards dashed. Top to bottom: CG, DG, CRO, and SRO. Subject numbers for each task are given by n and data from both versions of the rising optimum tasks are combined and presented as if all subjects had performed the version shown.***

CRO and SRO, and also experienced all four feedback conditions, although a given subject performed each task under only one feedback condition.

Choice distributions were also obtained from model simulations of 150 trials, initialized by the starting *A*-allocations supplied to the individuals performing each task. A learning rate λ and gain parameter μ were estimated for each task by attempting to minimize the mean square difference between the allocation distribution averaged across all individuals who performed that task without feedback, and a predicted distribution obtained by averaging across the same number of model runs. To constrain the search, bounds on μ were established for each task as described in part B of the Appendix (the fact that $\lambda \in [0, 1]$ was already noted in Section II-B). Table 1 (top panel) lists the resulting parameter values and fit errors. Constraining λ and μ to common values for both Gaussian tasks (CG and DG column) shows that different reward schedules, and the resulting choice-to-choice feedback of individual performance, can lead to markedly different choice allocations *without* changes in DD parameters, although fit errors modestly increase.

Fig. 3 shows that the model captures both the "stable" behavior of subjects in the CG task, whose *A*-allocations remain close to optimal at 50%, as well as their much more diffuse exploratory behavior in the DG task. The model captures the allocation distribution of the CRO slightly better than that of the SRO. In both cases, it reproduces the peak at 0% *A*'s (the local maximum) and approximates the small upticks near 100% *A*'s, but it overestimates *A*-allocations immediately to the right of the matching point, where rewards are lower, and underestimates them to its left, where rewards are higher.

We note that gain values are significantly higher for the rising optimum tasks ($\mu \approx 11$) than for the Gaussian tasks ($\mu \in [2.5, 3]$). The learning rates also differ ($\lambda = 0.9$–1 for the CG and CG and DG fits, and $\lambda \approx 0.1$ for rising optima). This suggests that subjects adopt different strategies for the "easy" stable matching tasks than for the rising optimum tasks. Specifically, for the simpler CG task, they rely largely on current reward information (implying larger learning parameters) and employ more exploration (smaller gain parameters). In contrast, for the more difficult rising

Table 1 Parameter Values and Mean Square *A*-Allocation Errors for the Choice Model With No-Feedback, Fitted to Subject Data From No-Feedback (Top), Choice (Middle), and Reward (Bottom) Conditions. Recall That $\lambda \approx 0$ Implies Little Learning, and $\lambda \approx 1$ Implies Reliance on the Most Recent Reward

		CG	DG	CG & DG	CRO	SRO
NO FEEDBACK	λ	0.98	0.14	0.91	0.11	0.06
	μ	2.50	2.91	2.50	11.3	11.0
	error	0.061	0.110	0.070; 0.119	0.085	0.188
CHOICE	λ	1.00	0.04	0.99	0.09	0.08
	μ	2.60	2.90	2.50	11.2	11.1
	error	0.049	0.103	0.067; 0.112	0.097	0.272
REWARD	λ	0.99	0.06	0.89	0.10	0.09
	μ	2.50	3.00	2.60	11.0	11.5
	error	0.119	0.091	0.121; 0.105	0.099	0.152

optimum tasks subjects draw on a longer history (smaller learning parameters) and display more exploitation (higher gain parameters).

The model also produces choice-to-choice dynamics similar to those of typical subjects. *A*-allocations as a function of trial number for the model show that it qualitatively reproduces rapid cycling around the matching point in the CG task, and considerably slower, larger amplitude cycling characteristic of the DG task. That this also holds for the common fit to both Gaussian tasks (Table 1) shows that learning dominated by current rewards can yield very different behaviors on different tasks without parameter changes (indeed, behavior on the DG task depends only weakly on λ). The model also captures some features of the rising optimum choice sequences, occasionally reproducing the discovery, and subsequent abandonment of, the global optimum (data not shown).

To further assess the need for the reinforcement learning and probabilistic components of the model, we fitted a simple deterministic "win-stay, lose-switch" model that starts at random with two *A* or two *B* choices and thereafter switches when rewards decrease, and otherwise stays with the same choice

$$\begin{aligned} \text{choice}(n+1) &= \text{choice}(n), && \text{if } r(n) \geq r(n-1) \\ \text{choice}(n+1) &\neq \text{choice}(n), && \text{if } r(n) < r(n-1). \end{aligned} \quad (13)$$

Fits were significantly poorer, with mean square errors more than double those of Table 1 (data not shown).

We also fitted the model of Section II-B, still with no-feedback, to allocation distributions of groups performing under the choice and reward conditions, again finding reasonable fits with mean square errors similar to those above: see Fig. 4 and the lower panels of Table 1. As expected, fits improved slightly upon including the additional feedback parameters ν_c and ν_r of Section III (data not shown). However, as Fig. 4 shows, with the exception of the CG task under reward feedback (note local maxima

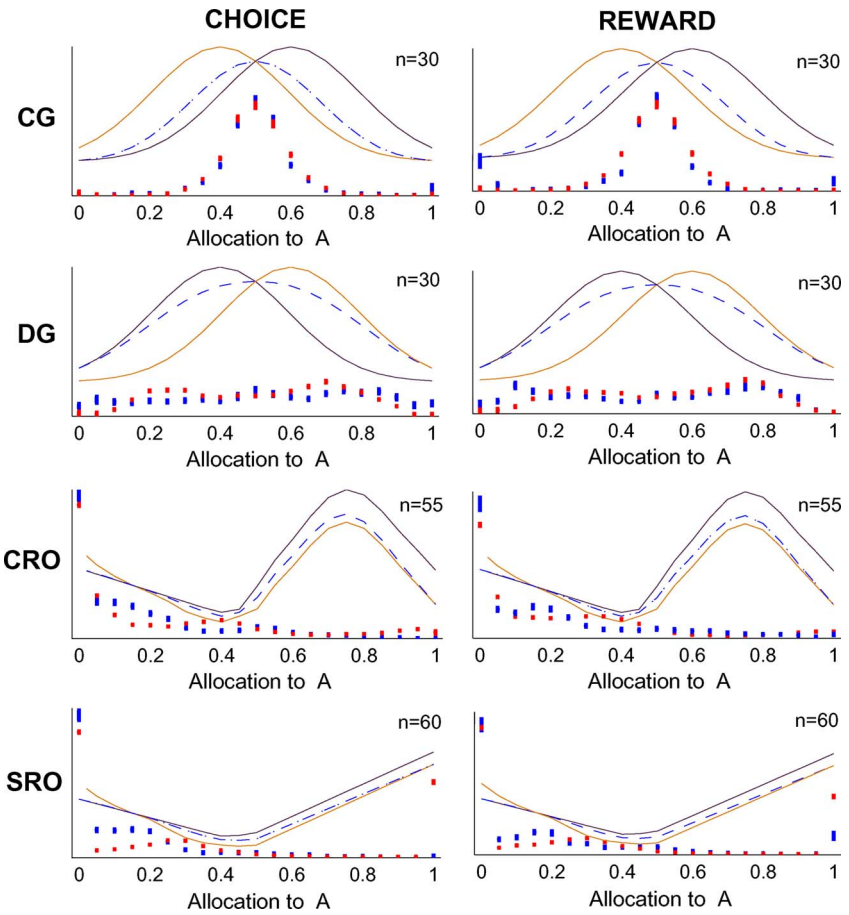


Fig. 4. ***Allocation distributions with standard error bars for subjects performing the four tasks of Fig. 2 (blue), compared with fits of the choice model (3) with no-feedback (red). Reward schedules are shown as faint curves with average rewards dashed. Data from choice and rewards feedback conditions shown in left and right columns, respectively, in the same format as Fig. 3, CG (top) to SRO (bottom). Note that distributions are similar to those of the subjects in Fig. 3, with no-feedback.***

at 0 and 100% *A*'s and see Section IV-C), averaged *A*-allocations of subjects with both types of feedback are very similar to those with no-feedback. It is thus not surprising that they can be reproduced by a model with averaged feedback, or even one with no-feedback. Nonetheless, behaviors may vary significantly over time and/or across individual subjects, and comparison of averaged allocation distributions is inadequate to test for these phenomena. Therefore, we next describe a method that assesses the models' capacities to generate realistic choice sequences. Using it, we then show that both time-dependent and cross-subject differences occur.

B. Maximum-Likelihood Assessment of Choice Sequences

We adopt the maximum-likelihood method used by Corrado *et al.* [9] to assess similar binary choice data. We ask how likely each individual's choice sequence is, given the basic probabilistic model of Section II-B supplemented by the appropriate feedback model of Section III, and supplied with the actual history of choice and feedback data preceding each new trial. The likelihood that the model predicts a specific ordered sequence $\{d(n)\}_{n=0}^{K}$ of choices *A* and *B* is computed by multiplying by $P_n(A)$ on all instances in which $d(n) = A$ and by $1 - P_n(A)$ on all instances in which $d(n) = B$, where $P_n(A)$ [see (3)] denotes the probability that the model chose *A* on the *n*th trial

$$L(d|p) = \prod_{\{n|d(n)=A\}} P_n(A) \times \prod_{\{n|d(n)=B\}} [1 - P_n(A)]. \quad (14)$$

It is convenient to express (14) in terms of logarithmic likelihoods so that the products become sums, to normalize by the number of trials *K* (150, here), and finally, to exponentiate the resulting sum

$$\text{Avg } L(d|p) = \exp\left\{\frac{1}{K}\left[\sum_{\{n|d(n)=A\}} \log P_n(A) + \sum_{\{n|d(n)=B\}} \log[1 - P_n(A)]\right]\right\}. \quad (15)$$

The average likelihood expression (15) is interpreted as follows: prediction at chance level (e.g., if $P_n(A) \equiv 0.5$) yields the value 0.5, perfect prediction yields 1.0, and prefect antiprediction 0.0.

To fit each model we attempt to maximize (15) over a suitable range of $\mu, \lambda \in [0, 1]$ and for data from the feedback conditions, over the relevant parameters ν_c and/or ν_r. Since ν_c and ν_r quantify an individual's susceptibility to information about other group members (Section III), we initially estimated common (μ, λ) values for each group, and then sought individual estimates for ν_c and/or ν_r. However, many individuals exhibited behaviors very different from those captured by the average μ and λ values of their group, which necessitated fitting an *individual* (μ, λ, ν) *combination* for each subject. For gain feedback via (11), the resulting $\mu = \mu(0)$ value is used to initialize the sequence $\mu(n)$. Further details are supplied in part B of the Appendix.

We start by reexamining the model of Section II-B with no-feedback, but unlike the group-averaged *A*-allocations of Section IV-A, we fit individual μ and λ values for each subject who performed a given task without feedback. Fig. 5(a) shows the resulting average likelihood values computed from (15). Fits for the rising optimum tasks are clearly better than those for the Gaussian tasks, for which many subjects are predicted only at chance ($\text{Avg } L(d|p) = 0.5$). This is not unreasonable, in that optimal performances are obtained by choosing *A* and *B* with equal probabilities in each run of $N = 20$ trials, specific sequences being irrelevant; thus one does not expect past actions to determine a subject's current choice with high probability. The rising optimum reward schedules promote longer runs of *A*'s or *B*'s as subjects settle near local or global maxima, so predictions are more reliable.

In Fig. 5, we also compare average likelihood values for the models with choice, rewards, and both types of feedback, fitted to the appropriate data, with average likelihood values for the model with no-feedback fitted to the same data. Fig. 5(b) shows that the choice feedback model of (7) provides substantial improvements in predictions for 158/175 individuals, especially in the rising optimum tasks. In the case of reward feedback [Fig. 5(c)], the model of (10) offers improvement for 145/175 individuals, although the increases in average likelihood are not as striking as those obtained for choice feedback. Finally, Fig. 5(d) shows improvements for the majority of individual fits (149/165) under the superposed choice and reward feedback conditions of (7) and (10).

Improvements in fits for each feedback condition can be quantified by computing the ratios of feedback models' likelihood values, averaged over all subjects in that condition, to the models with no-feedback. The choice feedback model of (7) shows a ratio of 1.26, while the reward feedback model [see (10)] yields a lower ratio of 1.21 and combined choice and rewards yields 1.19. Computations of the Akaike information criterion (AIC) [1] and paired t-tests of these values, described in part B of the Appendix, confirm that these improved fits outweigh the additional feedback parameters. Summing the $\text{Avg } L(d|p)$ values for each feedback condition and dividing by the number of subjects performing in that condition yields the following overall measures of goodness of fit: $\langle \text{Avg } L(d|p) \rangle = 0.88$, 0.72, and 0.73 for choices, rewards, and both feedbacks, respectively. Paired t-tests performed for individuals in each feedback-task combination confirm that these are

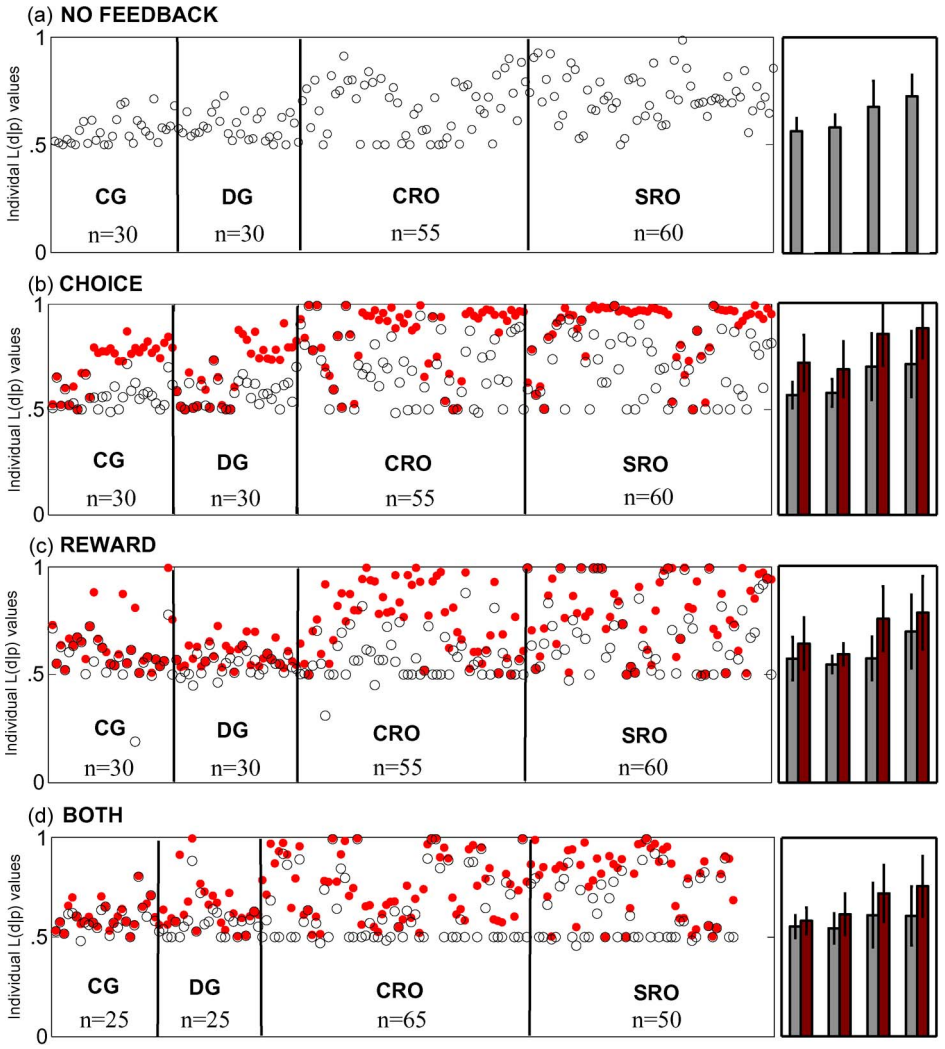


Fig. 5. ***Average likelihood values of the (b) choice, (c) reward, and (d) both feedback models of (7), (10), and (12) (solid red circles), compared to model with no-feedback (empty black circles) for all subjects performing under the appropriate feedback conditions. Panel (a) shows likelihood values for model with no-feedback fitted to subjects performing with no-feedback. Vertical lines from left to right divide subjects, respectively, performing CG, DG, CRO, and SRO tasks. Bar graphs on the right-hand side show*** Avg $L(d|p)$ ***values for each task in the same order, comparing model with (red) and without (gray) feedback;*** n ***denotes subject numbers; thin lines denote standard errors.***

significantly higher than the corresponding values for the models with no-feedback. The p-values are less than 10^{-4} for the rising optima and DG tasks in all feedback conditions, and $10^{-8}, 0.0137$, and 0.001 for the CG task for choice, reward, and both feedbacks, respectively; also see bar graphs on the right-hand side in Fig. 5. These may reflect higher quality of the choice model, but may also be due to the fact that subjects are less likely to process and thereby be influenced by feedback when it indicates no clear course of action (reward), or contains too much information (reward and both conditions). Paired t-tests of the respective Akaike values confirm these findings in part B of the Appendix.

We also fitted the alternative model of (9) to the choice feedback data, but found poorer fits than those for the model with no-feedback: most subjects' choices being predicted at chance level Avg $L(d|p) = 0.5$ [cf. Fig. 5(a)]. In addition to (7), which determines the additional bias applied to $w_A(n) - w_B(n)$ from the immediately preceding choice, we considered functions $f(n)$ that average over windows of up to ten preceding choices, and over windows of lengths 1–10 individually fitted to each subject. We also tried applying the bias of (7) *only* when an individual's rewards do not increase from trial $n-2$ to trial $n-1$ [resulting in $\langle \text{Avg } L(d|p) \rangle = 0.77$] and only when rewards do not increase *and* all other group members make the same choice on trial $n-1$ ($\langle \text{Avg } L(d|p) \rangle = 0.69$), and also only when the other four chose in unison ($\langle \text{Avg } L(d|p) \rangle = 0.73$). Finally, we considered a simpler rule that excludes the function $f(n)$ in (7), and applies

a constant weight ν_c to the choices of others throughout the block of trials ($\langle \text{Avg } L(d|p) \rangle = 0.75$). The model (7) significantly outperforms all of these, with $\langle \text{Avg } L(d|p) \rangle = 0.88$.

For the rewards data we found that the iterated additive rule (10) yielded the best fits, and that using reward rank in place of the term $r(n) - \max(\ldots)$ in (10) yielded fits that were almost as good. The alternate feedback rules (11) and (12) produced significantly poorer fits. Specifically, multiplicative updating of the gain parameter via (11) improved fits for 84/145 subjects, but with an average likelihood ratio of only 1.02 (results not shown), and (12) slightly improved Avg $L(d|p)$ values for only seven individuals, while all others remained at the values with no-feedback.

We did not test the alternate rules of Section III on data from the both condition, since they performed poorly on choice and rewards data separately and the model employs linear superposition of the rules.

C. The Perils and Pleasures of Feedback

In this section, we provide a sample of results from the behavioral data, specifically describing how reward feedback can be detrimental in the simple CG task, while exploratory tendencies and reward and/or choice feedback can improve performance in the more difficult SRO task.

1) Rewards Feedback Degrades Performance in the CG Task: The top left panel of Fig. 6 shows group allocation variances, as functions of trial number, averaged across the reward, choice, and no-feedback conditions in blue, red, and black, respectively. Higher variance on a given trial implies greater exploratory behavior of the group as a whole at that point in time. The reward feedback data clearly indicate greater exploration than in the no- and choice-feedback conditions; indeed, group variances grow over the block of trials. We believe that this is due to the fact that, on any given trial, at least one other group member gains a higher reward than one's own with probability 0.61 (estimated by averaging across subjects). This evidently promotes exploration in pursuit of higher rewards, albeit leading to poorer performance. Indeed, the top right panel of Fig. 6 shows that, after a brief initial transient, the mean rewards obtained per trial consistently lie *below* those for the no- and choice-feedback conditions.

The bottom panels of Fig. 6 show analogous variance and reward histories obtained from model simulations with no-feedback, and with the additional terms (7) and (10) modeling choice and reward feedback. To compute these we used the parameters estimated for each individual by maximizing likelihoods as in Section IV-B, and averaged over models representing the same subject groups as in the data. The resulting variance and reward histories are similar to those of the data after an initial transient of ≈20 trials, and average rewards are predicted well, although group variances under choice feedback are underestimated.

In the case of reward feedback, the model yielded long runs of identical choices for three subjects, due to the zero weights initially assigned to each choice and (10)'s additional biasing of the weight difference. We overcame this by setting both $w_A(n)$ and $w_B(n)$ equal to the highest of the two values predicted by (4) for the first four trials and thereafter reverting to the original update rule. This

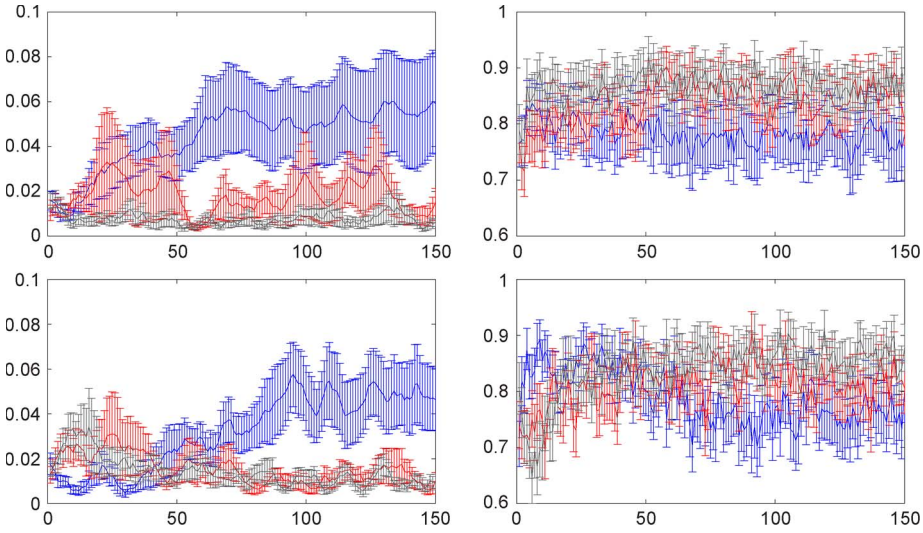


Fig. 6. *Comparison of averaged group A-allocation variances (left column) and average rewards obtained (right column) over 150-trial blocks obtained from behavioral data (top row) and from model simulations (bottom row) for the CG task under no-feedback (black), choice feedback (red), and reward feedback (blue) conditions. Data and model simulations are averaged over all subjects performing under each feedback condition. Vertical bars indicate standard errors.*

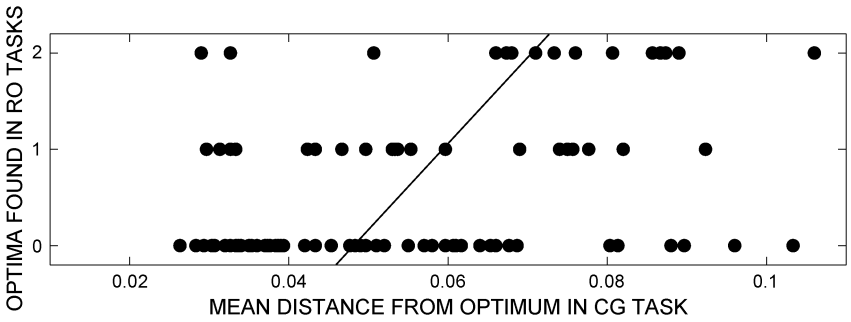


Fig. 7. ***Exploratory behavior in the CG task is positively correlated with discovery of global optima in RO tasks. Each filled circle represents a subject, with abscissa denoting average distance from 50% A-allocation in the CG task, and ordinate, number of optima found in CRO and SRO tasks, averaged across all feedback conditions. Line indicates best linear fit*** $(r = 0.44, p < 10^{-6})$***, and axes are clipped to show trends clearly: some subjects found up to four optima and some deviated from 50% A-allocation more than those shown here.***

eliminates long runs of the initial choice and consequent underestimation of rewards by preventing the first (randomly selected) choice from gaining an advantage due to the other weight's initialization at zero, and thereby precludes reinforcement of this choice by (10). This stratagem was only needed here and in producing Fig. 9 in the next subsection.

2) Exploratory Tendencies and Feedback Improve Performance in RO Tasks: While exploration is detrimental in the CG task, it can be beneficial in more complex tasks. To quantify this, the mean absolute difference between each subject's *A*-allocation and the optimum value of 50% was calculated for the CG task, irrespective of feedback condition, and the resulting value correlated with the number of global optima found by the subject on both rising optimum tasks. (Recall from Section II-A that the CG task was designed to gauge exploratory behavior.) Here "finding the optimum" was defined as attaining an *A*-allocation of $\geq 80\%$ in the SRO and $\geq 60\%$ in the CRO; percentages selected because they represent the first points at which the reward experienced is higher than that obtained at 0% *A*: the local maximum.

Fig. 7 illustrates the results. The correlation coefficient ($r = 0.44$, $p < 10^{-6}$ with Spearman test) indicates that exploratory behavior on the CG task predicts behavior on the rising optimum tasks (significant correlations were also found when the data are separated by feedback condition). However, rewards were not significantly higher for more exploratory subjects playing SRO or CRO ($r = -0.001$, $p = 0.99$ with Spearman test), primarily because reward schedules were designed such that maintaining a global maximum is difficult, and many subjects did not find it early enough in the block of trials, or, having found it, did not remain there. For example, the bottom two panels in the third column of Fig. 8 correspond to subjects who reach 95% *A*'s but do not remain there.

We also asked how feedback affects the probability that subjects find global optima, thereby influencing the group's performance as a whole. The numbers of individuals finding the global maximum (as defined above) in the SRO task under the four feedback conditions are as follows. With no-feedback, 4/60 subjects found it (6.7%, each alone in their group in doing so); with choices feedback, 5/60 found it (8.3%: two pairs in two groups, and one singleton); with rewards feedback, 9/60 found it (15%: a triplet in one group, two pairs in two groups, and two singletons), and in the both condition 11/50 found it (22%: one quadruplet, one pair, and five singletons). Passing through these feedback modes, not only does the fraction of individuals reaching the global optimum rise, but the size of subgroups displaying this behavior also increases, as illustrated in Fig. 8. Data from the CRO revealed no consistent patterns, other than occurrence of only singletons in the no-feedback condition.

Our models, with parameter values fitted to the appropriate subjects, yield similar trends. Simulating each subject's behavior 100 times yielded the following: with no-feedback, most frequently 3/60 subjects found the optimum, with a high likelihood of singletons and rare appearances of doubles. With choice feedback, most frequently 8/60 subjects found it, most often in pairs, with occasional singletons and triplets. Under rewards feedback, most frequently 11/60 subjects found it, with singletons, pairs, and triplets almost equally likely, and a few quadruplets. In both condition, most frequently 18/50 subjects found it, with common occurrences of all group sizes, but a preponderance of doubles.

Fig. 9 (top) probes this further by showing the distributions of subject subgroup sizes that found the optimum under the four feedback conditions, in comparison with model simulations (bottom). The distributions are strikingly similar: the model captures the greater likelihood of doubles under choices feedback and the progressive flattening of the distributions as one passes though choices and rewards to both feedbacks, although it underestimates the fraction of singletons and overestimates the incidence of larger subgroups under all conditions.

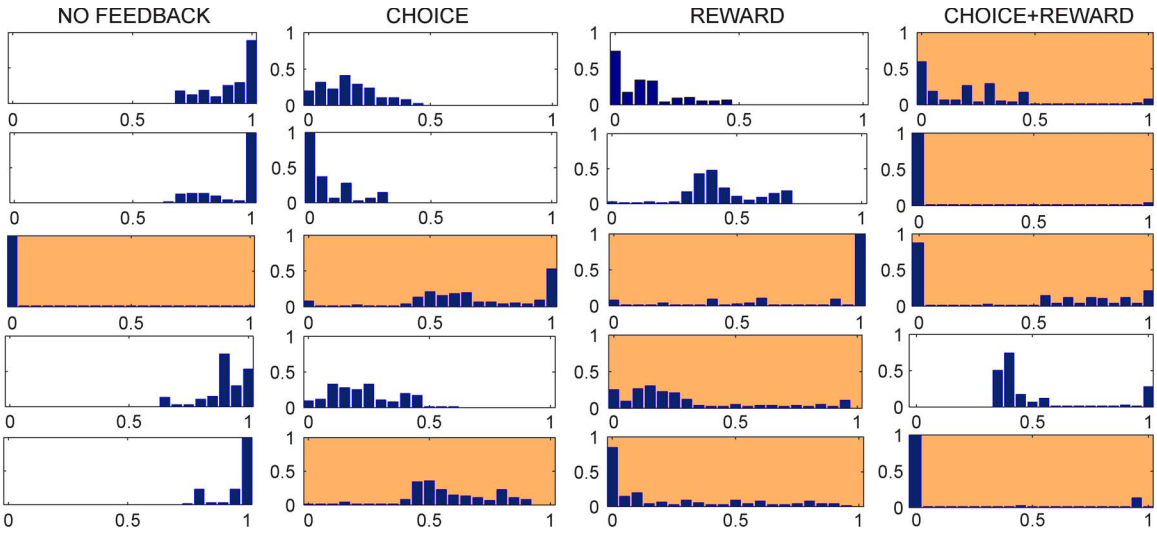


Fig. 8. ***Allocation distributions for the SRO task showing typical individual behaviors in high-performing groups under each feedback condition. Highlighted backgrounds identify subjects within each group who reach the vicinity of the global optimum (on the left for outer columns, on the right for inner columns). Subjects differ from group to group.***

In addition, trial-by-trial behavioral data from the same task shows that under choice and both feedbacks, an individual's probability of choosing the option previously picked by all four other group members increases on average by 10% and 17%, respectively, compared to the no-feedback condition. The DG task confirms such behavior with increases of 22% for choice feedback and 28% for both feedback.

These results and those of Section IV-C1 support the following interpretation. Reward feedback has the potential to stimulate exploration; this typically degrades performance in a simple task, but raises the probability that

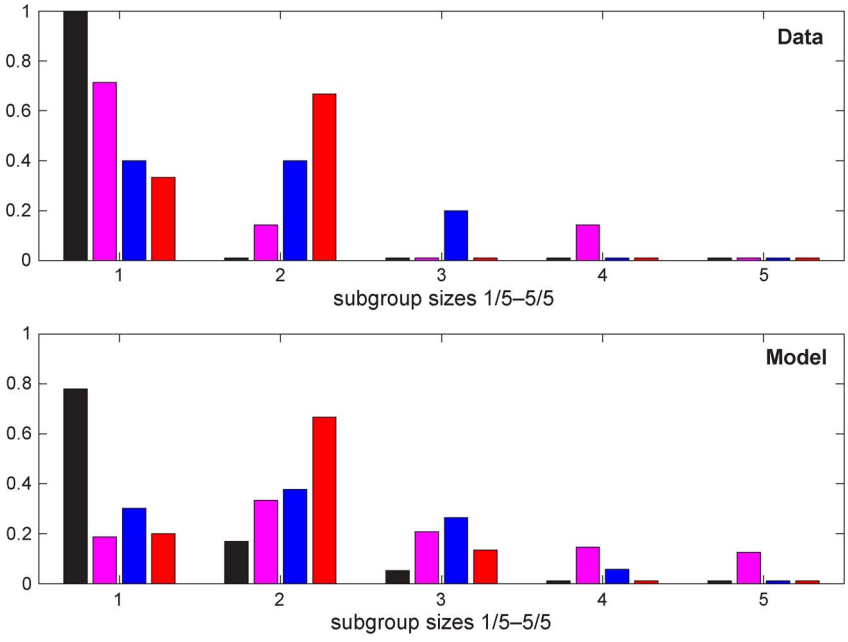


Fig. 9. ***Distributions of subgroup sizes (1/5–5/5) within all groups in which at least one subject reached the global optimum in the SRO task for data (top) and model (bottom). No-feedback, choices, rewards, and both conditions are indicated by black, red, blue, and magenta boxes, respectively; ordinates denote the fraction of each subgroup size. See text for discussion.***

individuals find the global optimum in more complex tasks. Moreover, an individual's probability of finding this optimum increases further if another group member finds it [cf. (10)]. Choice feedback increases the probability that individuals will find the global optimum if another group member finds it, and it does so by promoting mimicry [see (7)]. Related findings are reported in [25]. Hence, choice feedback should increase the size of subgroups finding the global optimum, but not the number of groups in which someone finds it, while reward feedback should increase the number of individuals who find it, both between and within groups. When both are provided, reward feedback stimulates exploration and choice feedback directs it, further increasing individual numbers and group sizes. Our models consistently capture these feedback-dependent group dynamics seen in behavioral data, albeit slightly overestimating the incidence of larger subgroups.

Howard-Jones *et al.* [22], using a four-armed bandit task in which individuals competed with a computer, found that reward combined with choice feedback did *not* stimulate exploration. Nor did we observe increased exploration under both condition, but the experiment of [22] is not directly comparable to ours for several reasons: in [22], the other player was known to be a computer (which has significant effects on both behavior and neural activity of human decision makers [33], [34]), and was described as a "competitor" (a framing which we explicitly avoided). Also, Howard-Jones *et al.* [22] employed a single competitor (a nominal equal), while our tasks used groups of five coequals capable of exhibiting an actual consensus.

D. Individual Differences: Parameter Distributions

In Section IV-A, we noted striking differences in gain and learning rate parameters (μ, λ) across the four tasks; we now build on this by examining individual differences in parameter fits. Fig. 10 shows distributions of μ, λ and of the influence coefficients ν_c and ν_r for choice and rewards feedback obtained by seeking individual fits that maximize average likelihoods as in Section IV-B.

The overall finding of lower $\mu \in [1, 5]$ for the Gaussian tasks and higher $\mu \in [6, 12]$ for rising optima, noted for the averaged fits of Section IV-A, is preserved, suggesting that subjects approach the tasks in different ways, performing more randomly in the former pair, and more deterministically in the latter pair. The learning rate λ also exhibits strong bimodality across subjects, with the majority of individual fits lying in [0, 0.2] and [0.8, 1] in all cases except for reward feedback on the CG task. (Weaker bimodality is also apparent in μ for choice feedback in all but the DG task.) Under choice feedback, high values of ν_c dominate and average μ values for individuals are higher across all tasks than for rewards feedback. In contrast, low ν_r values dominate for rewards feedback, particularly for the CG task, in which subjects with lower ν_r's do better than those with higher values. For the remaining tasks, ν_r distributions exhibit stronger and longer tails, with a substantial fraction of ν_r's around 1 for the SRO task.

Bimodality in learning rates λ and in gain μ suggests that distinct subjects may approach the tasks in different ways. We sought, but failed to find, correlations between high/low μ and λ values across subjects, but for the 30 individuals who performed the DG and SRO tasks with no-feedback (the only sizable number of groups performing two tasks under the same feedback condition), we found that 16 had a correlation of $+0.6715$ between λ and μ values, six had an inverse correlation of -0.6790, and the remaining eight exhibited no significant correlation. These correlations within subjects across tasks, and the structures evident in the influence coefficient distributions of Fig. 10, support a potential separation of individuals into subgroups, including those less and more susceptible to group feedback. Such correlations have been questioned in fitting a reinforcement learning model by maximum-likelihood estimates [10], but we propose to examine neuroimaging data for subgroups, to seek neural correlates of susceptibility (cf. [44]).

Fig. 11 shows that parameters estimated from individual subjects' choice sequences also yield reasonable fits for their *A*-allocation distributions. Here we show sample groups performing the CG task without feedback and with choice and rewards feedback. To compare with *A*-allocation distributions obtained by averaging behavioral data, model runs were averaged over 100 blocks of trials.

Individual distributions and influence parameter values (given above each subject's histogram) show that members of a group can exhibit a variety of different behaviors. Some are highly susceptible to feedback (ν_r or $\nu_c \approx 1$), while others appear to pay less or no attention (ν_r or $\nu_c \approx 0$). Under rewards feedback (right column) the fourth member improves on the group average reward by maintaining a tight distribution while ignoring others $(\nu_r = 0.05)$, but the fifth $(\nu_r = 1.11)$ is evidently led to explore the entire range and linger at the left margin, significantly reducing net rewards and providing an example of the more prevalent behavior discussed in Section IV-C. Under choice feedback (center) the first member gains higher average rewards by maintaining a tight distribution around the optimum and ignoring the others $(\nu_c = 0.06)$, but the second, fourth, and fifth $(\nu_c \approx 1)$ explore away from the optimum, producing broader allocation distributions, and thereby receiving lower average rewards. The third $(\nu_c = 0.59)$ mediates between these extremes.

V. DISCUSSION

The behavioral data and mathematical models described in this paper provide a window into the dynamics of decision making in a group context. We have designed and parameterized two-armed bandit tasks that can be performed

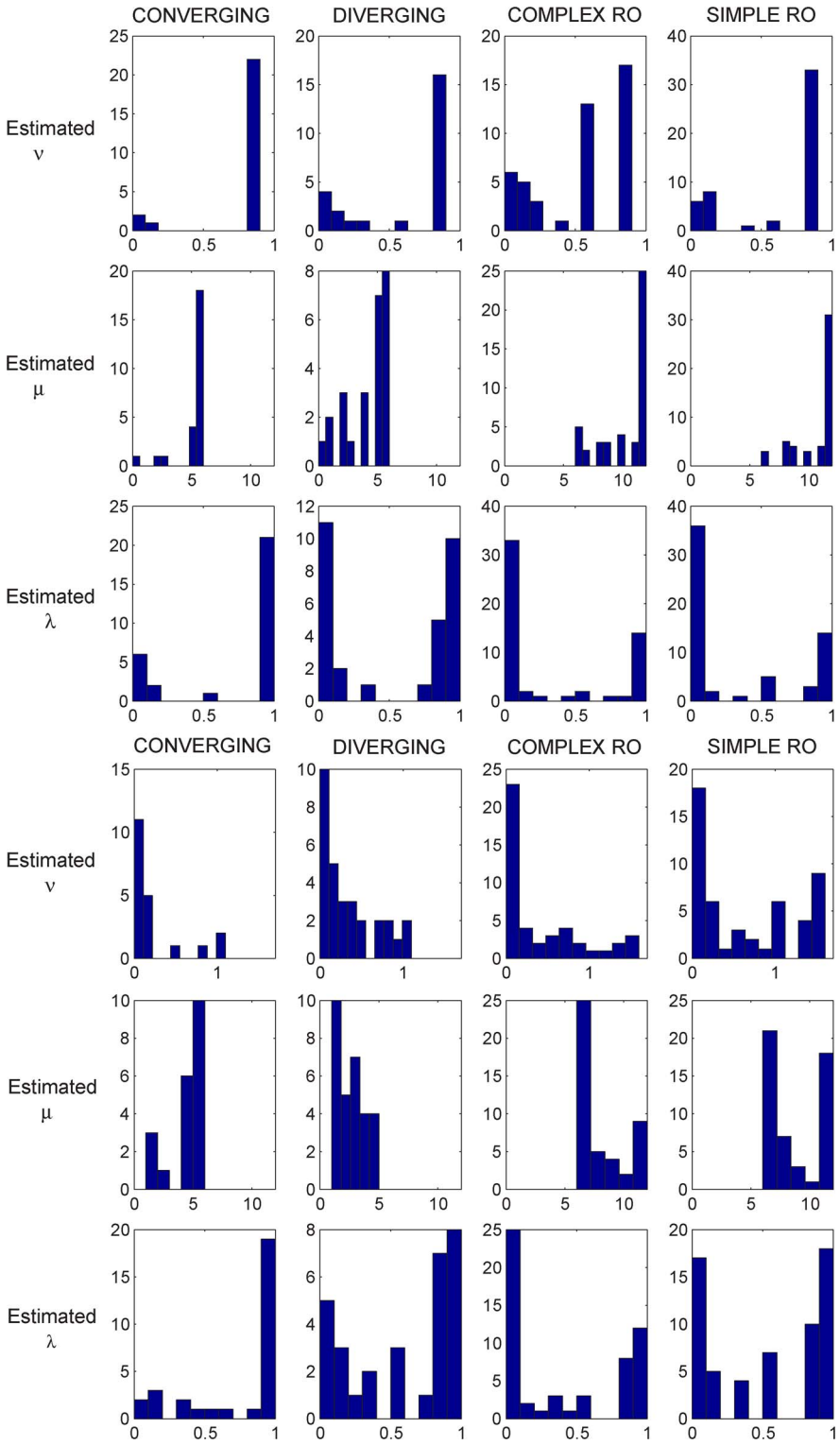


Fig. 10. ***Histograms showing distributions of parameter values for individual parameter fits for the four tasks under choice (top panel) and reward (bottom panel) feedback conditions. Note tendency to bimodality in μ and λ (middle and bottom rows of both panels), and substantial fraction of high values of ν_c and low values of ν_r (top rows of both panels). See text for discussion.***

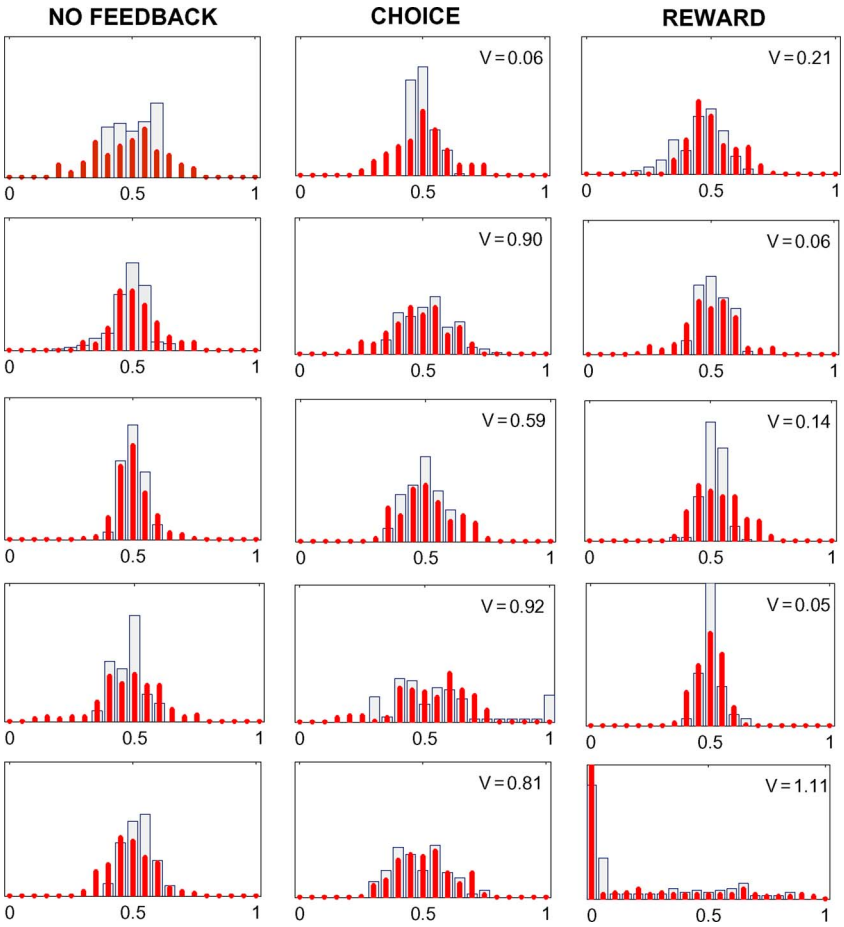


Fig. 11. *Model fits (red bars) to block-averaged A-allocation distributions (shaded boxes) of individuals performing the converging Gaussians task in groups of five with no-feedback (left), with choice feedback (7) (center) and with rewards feedback (10) (right), with ν_c and ν_r obtained by maximizing average likelihoods. Each panel depicts a different subject.*

both with and without feedback regarding other group members' choices and rewards, and proposed simple feedback rules coupled with previous individual reinforcement learning and choice models [12]. Our analyses of behavioral data, along with model fitting, confirm earlier work on the matching and rising optimum tasks [3], [26], and provide new insights on the effects of social information.

Our choice feedback model (7) shows that the addition of a time-dependent preference weight based on the majority actions of group members can capture some aspects of individual and group behavior and dynamics. Similarly, our comparison of several models of reward feedback suggests the importance of individual performance relative to that of the group as well as drift-directed randomness: (10) produces runs of identical choices, permitting effective exploration over the space of rewards.

Related studies of individuals performing a two-alternative choice task with monotonically rising reward schedules illustrate the use of spatial cues in promoting directed exploration [17], [18], [29]. Here such cues are absent but exploration increases when subjects are provided with information on the choices and rewards of others in their group. Specifically, in the rising optimum task, the number of individuals reaching the optimum within a group increases as the feedback mode passes through choices, to rewards, to both. Furthermore, the average exploratory tendency of individuals performing the converging Gaussians task without feedback correlates with a higher probability of reaching global optima. However, in the same task, reward feedback induces exploration, which leads to poorer performance. In contrast, choice feedback in this task does not significantly reduce rewards below those accrued without feedback.

After this paper was submitted we learned of recent work [4] in which signed prediction errors are invoked to explain observational learning. However, the experiment of [4] differs from the present one in several ways, and direct comparisons with the models of Sections II-B and III are problematic. While there are analogs of no-feedback, choice, and both conditions, in the latter cases participants

received information on a single "confederate's" choice or reward *before* choosing themselves, and the model of [4] (which is not described in detail) uses differences between actual and *predicted* choices and outcomes of the confederate, rather than differences between the participant's choices and rewards and those of four other group members.

In Section II-B, we observed that the probability of choices predicted by the present model coincides with that of a DD process, which is known to be optimal in both free response and timed (deadlined) perceptual decision tasks in which a sequence of noisy stimuli must be correctly identified to maximize rewards [2]. As noted there, this does not imply optimality in the present tasks, since that would require learning complex rules or statistical descriptions of the different reward schedules, in order to accurately estimate weights. However, sample paths of DD processes also resemble ramping firing rates in cortical areas involved in perceptual decisions [14], [24], [35], [37], suggesting neural substrates for the choice portion of each trial.

The link with statistical decision theory via the DD process also suggests that related methods in signal processing, such as signal detection and change detection theories, might be useful in modeling social behavior. It prompts more general questions of how optimality in an individual's behavior (maximizing one's net rewards, with or without feedback) might be related to strategies for maximizing group rewards. Parameter studies of the models could suggest strategies for sharing information that promote the latter. Signal detection theory has been used to investigate group decisions in target identification (binary decision) tasks, when individuals discuss and pool their judgements after initially reaching them independently [38]. In related work, decisions rendered by pattern classifiers, based on neural signals from different subjects who do not otherwise communicate during binary decisions, have been combined in various ways to improve accuracy by pooling individual data [11].

We propose to use the influence parameters of [see (7)–(12)] in analyses of fMRI data collected during the experiment. Variations in susceptibility to social pressure both between and within subjects may reveal brain structures that process social feedback or regulate its influence on behavior. In particular, the bimodal trends in Fig. 10 suggest dividing subjects into subgroups with high and low feedback susceptibility quantified by ν_c and ν_r, and asking if blood oxygenation level dependent (BOLD) signals in regions of interest correlate with these parameters. Averaging over all subjects performing a particular task under a given feedback condition, we have already found significant negative correlations between activity in the insula and both group alignment (the number of group members making the same choice as a given subject) and reward rank (the subject's order in the group based on reward magnitudes), under these feedback conditions [44].

In summary, when our models perform the tasks with information from others, they can reasonably capture changing feedback dynamics and the resulting A-allocations averaged across subjects, as well as individual choice behaviors. We have also verified that when the models receive information about other group members' choices, they produce cycling and "transfer of allegiance" behaviors similar to those of human groups (data not shown here). Hence, the simple feedback rules of Section III successfully capture aspects of both individual behaviors and group dynamics. ■

APPENDIX

A. Experimental Method

Participants were recruited at Baylor College of Medicine, Houston, TX, via e-mail and word of mouth, and informed consent was obtained according to protocols approved by Baylor College of Medicine and Princeton University's Institutional Review Boards. They were then instructed regarding the experiment and led to one of the facility's scanners, which consisted of two Siemens 3.0 Tesla Allegra scanners and three Siemens 3.0 Tesla Trio scanners. In total, 23 separate groups of five participants performed a two-alternative choice task while fMRI data were acquired ($n = 115$ individuals; 68 female, 47 male; ages 18–57). Four different reward-based decision-making tasks were performed (two of them in left- and right-handed variants, as described in Section II-A), in which each subject chose by pressing one of two buttons (A or B), receiving reward points after each choice. Rewards were determined as explained in the main text, but the rule was not explained to participants, who were simply told to accumulate as many points as possible. Each subject performed each task for a session containing 150 choices (2.5-s intertrial interval, synchronized across group members), after which a screen indicated the start of a new task, as shown in the top row of Fig. 12. Tasks were presented in a randomized order.

There were four information or feedback conditions in group tasks: *no-feedback*, in which subjects only received information on their own rewards; *choice*, in which subjects could see other group members' choices; *reward*, in which points earned by other group members after each choice were displayed, and *both*, in which points earned and choices were displayed. Subjects were shown the type of information being presented, and feedback conditions remained constant over each block of choices. Each group performed under each condition at least once, and the conditions performed by the various groups were balanced across the total number of groups. The bottom row of Fig. 12 shows typical screen displays during each feedback condition. After completion, subjects were debriefed and compensated according to their point totals, with payments ranging between \$30 and \$50, following guidelines

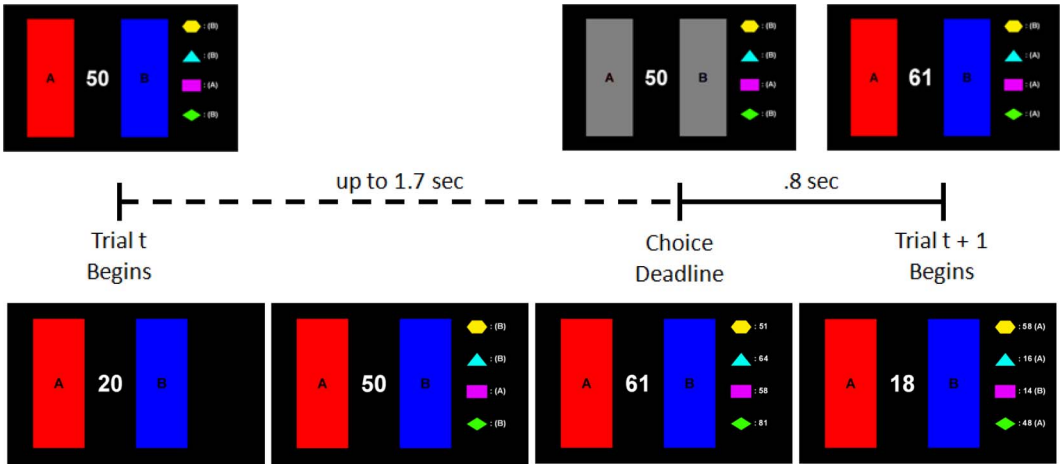


Fig. 12. *Top: The timeline of a single trial. Subjects had to respond within a 1.7-s deadline to register their choices, otherwise the computer selected the button chosen on the previous trial. Decisions or deadline passage were indicated by the "A" and "B" buttons turning gray. This state persisted until the total length of the trial was 2.5 s, resulting in a minimum intertrial interval of 0.8 s and synchronizing participants' choices. The buttons then regained their colors and the appropriate reward and social information were shown for the previous trial. Each session contained 150 trials. Bottom: Subjects performed each task under a particular social condition, viewing screens as in the samples for no-feedback, choices, rewards, and both conditions (left to right). In the no-feedback condition the task was performed without group information, only the participant's own reward being shown (between the buttons), although decisions were still synchronized across the group. In choices, the button previously chosen by each of the other group members was shown, updated synchronously following each trial. In rewards, the number of points earned by each of the other group members was shown following each trial, and in both conditions, choices and rewards were simultaneously displayed.*

set by the Princeton University and Baylor College of Medicine Institutional Review Panels.

B. Parameter Fitting Methods

Parameters were fitted by attempting to find the best matching A-allocation distribution produced by the model over the given parameter space using grid search in the Matlab environment.

To match allocation distributions in Section IV-A we used the mean square or L^2-norm

$$\text{error} = \sqrt{\sum_{i=0}^{20} [g_d(i) - g_m(i)]^2} \tag{16}$$

where $g_d(i)$ and $g_m(i)$, respectively, denote the number of entries falling into allocation bin $[i/20, (i+1)/20]$ in the data, and an analogous number predicted by the model (the 21 bins range from 20 B's to 20 A's). The number of entries in each bin was obtained by summing over all subjects performing a given task under the no-feedback condition, and the model was run 50 times for each subject over the 150-choice session, with independent drawings of $P(A)$, to obtain reliable estimates. The algorithm sought the global minimum of (16) over $\lambda \in [0, 1]$ and suitable ranges of parameter values μ. The following bounds, suggested by preliminary computations, were applied: CGs and DGs: $\mu \in [0, 6]$, CRO and SRO: $\mu \in [6, 12]$.

To assess prediction of choice sequences, as explained in Section IV-B, we estimated maximum likelihoods by computing the exponential of the logarithmic likelihood normalized for the number of trials in each block

$$\log L(d|p) = \frac{1}{150}\left[\sum_{\{n|d(n)=A\}} \log P_n(A) + \sum_{\{n|d(n)=B\}} \log[1 - P_n(A)]\right]. \tag{17}$$

Each subject performed each task only once, so no further averaging was necessary. The algorithm then attempted to maximize the average likelihood of (15) over $\lambda \in [0, 1]$, μ in the ranges specified above, $\nu_c \in [0, 1]$, and $\nu_r \geq 0$. We examined three fitting methods: 1) estimating common values (μ, λ) for each group with ν_c or $\nu_r = 0$, and then estimating ν_c, ν_r values for each individual with (μ, λ) fixed at the group values; 2) estimating common values (μ, λ, ν_c) or (μ, λ, ν_r) for each group and proceeding as in 1); and 3) estimating separate values (μ, λ, ν_c) and/or (μ, λ, ν_r) for each individual. We selected the third, because it offers

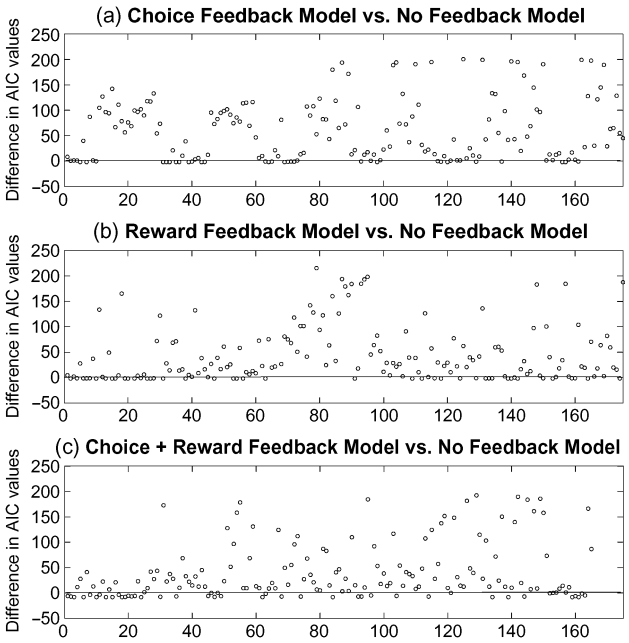


Fig. 13. ***Differences in AIC values for two-parameter fits of the model with no-feedback, three-parameter fits of the model with choice feedback (a) and rewards feedback (b) and four-parameter fits of the model with choice and rewards feedback (c) for all subjects performing under each of the three feedback conditions. Positive AIC differences imply that the models with feedback are superior.***

the best fit improvements on an individual basis, and accounts for differing individual behaviors within a group.

In comparing models with different numbers of fitting parameters, such as with and without feedback, we appeal to the AIC

$$\text{AIC} = 2(\text{number of parameters}) - 2\log L(d|p) \quad (18)$$

[1], which discounts the logarithmic likelihood by the number of fitting parameters, providing a quantitative measure of the advantage that additional complexity confers. Due to the sign of the $\log L(d|p)$ term, the smaller is AIC, the better is the fit. For example, in Fig. 13(a) and (b), we show the difference in AIC values computed from choice and rewards feedback data fitted with no-feedback (two parameters λ and μ) and, respectively, with the choice rule (7) (three parameters λ, μ, and ν_c) and the reward rule (10) (three parameters λ, μ, and ν_r). The fact that the vast majority of three-parameter fits (146/175 individuals in the choice condition and 131/175 individuals in the reward condition) show lower AIC values (differences in AIC for no-feedback and feedback values are positive) implies that the additional complexity of adding feedback is more than compensated by the improved data fits. In the choice and reward feedback condition, there is an improvement for 125/165 individuals, slightly lower than under the choice feedback condition, yet still representing a majority.

Table 2 Significance *p*-Values Resulting From Paired t-Tests of Individual AIC Values of Models With and Without Feedback

	Choice	Reward	Both
CG	$9.4289\text{x}10^{-9}$	0.0396	0.0297
DG	$4.0847\text{x}10^{-6}$	$2.2360\text{x}10^{-4}$	$5.7624\text{x}10^{-4}$
CRO	$3.7409\text{x}10^{-9}$	$4.4874\text{x}10^{-11}$	$4.8985\text{x}10^{-11}$
SRO	$1.3891\text{x}10^{-9}$	$5.1448\text{x}10^{-7}$	$6.3353\text{x}10^{-8}$

Paired t-tests of individual AIC values within each feedback-task combination and those corresponding to the models with no-feedback yield the *p*-values given in Table 2, confirming significant differences.

Acknowledgment

The authors would like to thank the three reviewers for numerous suggestions, and for drawing their attention to related studies. Experiments were conducted in the Human Neuroimaging Lab at Baylor College of Medicine, Houston, TX.

REFERENCES

[1] H. Akaike, "Likelihood of a model and information criteria," *J. Econometrics*, vol. 16, pp. 3–14, 1981.

[2] 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," *Psychol. Rev.*, vol. 113, no. 4, pp. 700–765, 2006.

[3] 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 Res.*, vol. 1153, pp. 111–121, 2007.

[4] C. J. Burke, P. N. Tobler, M. Baddesley, and W. Schultz, "Neural mechanisms of observational learning," in *Proc. Nat. Acad. Sci.*, 2010, vol. 107, no. 32, pp. 14431–14436.

[5] J. R. Busemeyer and J. T. Townsend, "Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment," *Psychol. Rev.*, vol. 100, pp. 432–459, 1993.

[6] M. Cao, A. Stewart, and N. E. Leonard, "Integrating human and robot decision-making dynamics with feedback: Models and convergence analysis," in *Proc. 47th IEEE Conf. Decision Control*, 2008, pp. 1127–1132, Session TuB15.3: Mixed Robot/Human Team Decision Dynamics.

[7] M. Cao, A. Stewart, and N. E. Leonard, "Convergence in human decision-making dynamics," *Syst. Control Lett.*, vol. 59, pp. 87–97, 2010.

[8] J. D. Cohen, S. M. McClure, and A. J. Yu, "Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration," *Phil. Trans. Roy. Soc. Lond. B.*, vol. 362, pp. 933–942, 2007.

[9] G. S. Corrado, L. P. Sugrue, S. Seung, and W. T. Newsome, "Linear-nonlinear-poisson models of primate choice dynamics," *J. Exp. Anal. Behav.*, vol. 84, pp. 581–617, 2005.

[10] N. D. Daw, "Trial-by-trial analysis using computational models," in *Affect, Learning and Decision Making. Attention and Performance, Vol XXIII*, E. A. Phelps, T. W. Robbins, and M. Delgado, Eds. Oxford, U.K.: Oxford Univ. Press, 2011.

[11] M. P. Eckstein, K. Das, B. T. Pham, M. F. Peterson, C. K. Abbey, J. L. Sy, and B. Giesbrecht, "Neural decoding of collective wisdom with multi-brain computing," *NeuroImage*, 2011, DOI:10.1016/j.neuroimage.2011.07.009.

[12] D. M. Egelman, C. Person, and P. R. Montague, "A computational role for dopamine delivery in human decision-making," *J. Cogn. Neurosci.*, vol. 10, pp. 623–630, 1998.

[13] C. W. Gardiner, *Handbook of Stochastic Methods*, 2nd ed. New York: Springer-Verlag, 1985.

[14] J. I. Gold and M. N. Shadlen, "Neural computations that underlie decisions about sensory stimuli," *Trends Cogn. Sci.*, vol. 5, no. 1, pp. 10–16, 2001.

[15] J. I. Gold and M. N. Shadlen, "Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward," *Neuron*, vol. 36, pp. 299–308, 2002.

[16] J. Guckenheimer and P. J. Holmes, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*. New York: Springer-Verlag, 1983.

[17] T. M. Gureckis and B. C. Love, "Learning in noise: Dynamic decision-making in a variable environment," *J. Math. Psychol.*, vol. 53, pp. 180–193, 2009.

[18] T. M. Gureckis and B. C. Love, "Short-term gains, long-term pains: How cues about state aid learning in dynamic environments," *Cognition*, vol. 113, pp. 293–313, 2009.

[19] R. J. Herrnstein, "Melioration as behavioral dynamism," in *Quantitative Analyses of Behavior. Matching and Maximizing Accounts, Vol II*, M. L. Commons, R. J. Herrnstein, and H. Rachlin, Eds. Cambridge, MA: Ballinger, 1982.

[20] R. J. Herrnstein, "Rational choice theory: Necessary but not sufficient," *Amer. Psychologist*, vol. 45, pp. 356–367, 1990.

[21] R. J. Herrnstein, "Experiments on stable suboptimality in individual behavior," *AEA Papers Proc.*, vol. 81, pp. 360–364, 1991.

[22] P. A. Howard-Jones, R. Bogacz, J. H. Yoo, U. Leonards, and S. Demetriou, "The neural mechanisms of learning from competitors," *Neuroimage*, vol. 53, pp. 790–799, 2010.

[23] J. R. Krebs, A. Kacelnik, and P. Taylor, "Tests of optimal sampling by for aging great tits," *Nature*, vol. 275, pp. 27–31, 1978.

[24] M. E. Mazurek, J. D. Roitman, J. Ditterich, and M. N. Shadlen, "A role for neural integrators in perceptual decision making," *Cerebral Cortex*, vol. 13, no. 11, pp. 891–898, 2003.

[25] R. McElreath, M. Lubell, P. J. Richardson, T. M. Waring, W. Braun, E. Edsten, C. Efferson, and B. Paciotti, "Applying evolutionary models to the laboratory study of social learning," *Evol. Human Behav.*, vol. 26, pp. 483–508, 2005.

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

[27] P. R. Montague, P. Dayan, and T. J. Sejnowski, "A framework for mesencephalic dopamine systems based on predictive Hebbian learning," *J. Neurosci.*, vol. 16, pp. 1936–1947, 1996.

[28] 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 analyses of behavioral data," in *Proc. 47th IEEE Conf. Decision Control*, 2008, pp. 1115–1120, Session TuB15.1: Mixed Robot/Human Team Decision Dynamics.

[29] A. R. Otto, T. M. Gureckis, A. B. Markham, and B. C. Love, "Navigating through abstract decision spaces: Evaluating the role of state generalization in a dynamic decision-making task," *Psychonomic Bull. Rev.*, vol. 16, no. 5, pp. 957–963, 2009.

[30] R. Ratcliff, "A theory of memory retrieval," *Psychol. Rev.*, vol. 85, pp. 59–108, 1978.

[31] R. Ratcliff, T. Van Zandt, and G. McKoon, "Connectionist and diffusion models of reaction time," *Psychol. Rev.*, vol. 106, no. 2, pp. 261–300, 1999.

[32] J. N. Reynolds and J. Wickens, "Dopamine-dependent plasticity of corticostriatal synapses," *Neural Netw.*, vol. 15, pp. 507–521, 2002.

[33] J. K. Rilling, D. A. Gutman, T. R. Zeh, G. Pagnoni, G. S. Berns, and C. D. Kilts, "A neural basis for social cooperation," *Neuron*, vol. 35, pp. 395–405, 2002.

[34] A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, and J. D. Cohen, "The neural basis of economic decision-making in the ultimatum game," *Science*, vol. 300, pp. 1755–1758, 2003.

[35] J. D. Schall, "Neural basis of deciding, choosing and acting," *Nature Rev. Neurosci.*, vol. 2, pp. 33–42, 2001.

[36] P. Simen and J. D. Cohen, "Explicit melioration by a neural diffusion model," *Brain Res.*, vol. 1299, pp. 95–117, 2009.

[37] P. L. Smith and R. Ratcliff, "Psychology and neurobiology of simple decisions," *Trends Neurosci.*, vol. 27, no. 3, pp. 161–168, 2004.

[38] R. D. Sorkin, C. J. Hays, and R. West, "Signal-detection analysis of group decision making," *Psychol. Rev.*, vol. 108, no. 4, pp. 183–203, 2001.

[39] A. Stewart, M. Cao, and N. E. Leonard, "Steady-state distributions for human decisions in two-alternative choice tasks," in *Proc. Amer. Control Conf.*, 2010, pp. 2378–2383.

[40] A. Stewart, M. Cao, A. Nedic, D. Tomlin, and N. E. Leonard, "Towards human-robot teams: Model-based analysis of human decision making in two-alternative choice tasks with social feedback," *Proc. IEEE*, vol. 100, no. 3, pp. 751–775, Mar. 2012.

[41] A. Stewart and N. E. Leonard, "The role of social feedback in steady-state performance of human decision making for two-alternative choice tasks," in *Proc. 49th IEEE Conf. Decision Control*, 2010, pp. 3796–3801.

[42] R. Sutton, "Learning to predict by the method of temporal differences," *Mach. Learn.*, vol. 3, pp. 9–44, 1988.

[43] R. S. Sutton and A. G. Barto, *Reinforcement Learning*. Cambridge, MA: MIT Press, 1998.

[44] D. Tomlin, A. Nedic, D. A. Prentice, P. Holmes, and J. D. Cohen, "Group foraging task reveals neural substrates of social influence," 2011, submitted to *PLoS One*.

[45] L. Vu and K. A. Morgansen, "Modeling and analysis of dynamic decision making in sequential two-choice tasks," in *Proc. 47th IEEE Conf. Decision Control*, 2008, pp. 1121–1126, Session TuB15.1: Mixed Robot/Human Team Decision Dynamics.

[46] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.

[47] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Stat.*, vol. 19, pp. 326–339, 1948.

ABOUT THE AUTHORS

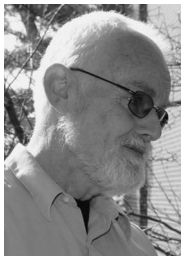
Andrea Nedic received the B.S. degree in electrical and computer engineering from Rutgers University, New Brunswick, NJ, in 2006 and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 2011.

She has been exploring social decision-making in association with the Princeton Neuroscience Institute.

Damon Tomlin received the Ph.D. degree in neuroscience from Baylor College of Medicine, Houston, TX, in 2006.

He has been a Postdoctoral Associate in the Princeton Neuroscience Institute, Princeton, NJ, since completing his degree in 2006. His research interests include reward-based decision making and social neuroscience.

Dr. Tomlin is a member of the Society for Neuroscience.



Philip Holmes received the Ph.D. degree in engineering from the Institute of Sound and Vibration Research, University of Southampton, Southampton, U.K., in 1974.

He is Eugene Higgins Professor of Mechanical and Aerospace Engineering, Professor of Applied and Computational Mathematics, and a member of the Neuroscience Institute at Princeton University, Princeton, NJ.

Dr. Holmes is a former Guggenheim Fellow, a member of the American Academy of Arts and Sciences, an Honorary member of the Hungarian Academy of Sciences, and a Fellow of the Society for Industrial and Applied Mathetics and the American Physical Society. He received the 2009 Liapunov and 2011 T. K. Caughey Awards from the American Society of Mechanical Engineers.

Deborah A. Prentice received the Ph.D. degree in psychology from Yale University, New Haven, CT, in 1989.

She is the Alexander Stewart 1886 Professor of Psychology at Princeton University, Princeton, NJ, and Chair of the Department of Psychology. She has recently held visiting appointments at the René Descartes University, Paris, France and the Institute for Advanced Study, Princeton. Her research interests include social norms, social learning, and group decision making.

J. D. Cohen, photograph and biography not available at the time of publication.