Stochastic pairwise preference convergence in Bayesian agents

Jordan Kemp [1,2,*] Max-Olivier Hongler, [3,*] and Olivier Gallay [4]
[1] *Department of Physics, University of Chicago, Chicago, Illinois 60637, USA*
[2] *Mansueto Institute for Urban Innovation, University of Chicago, Chicago, Illinois 60637, USA*
[3] *STI, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland*
[4] *Département des Opérations, Université de, Lausanne 1015, Switzerland*

(Received 8 January 2024; accepted 14 March 2024; published 6 May 2024)

Beliefs inform the behavior of forward-thinking agents in complex environments. Recently, sequential Bayesian inference has emerged as a mechanism to study belief formation among agents adapting to dynamical conditions. However, we lack critical theory to explain how preferences evolve in cases of simple agent interactions. In this paper, we derive a Gaussian, pairwise agent interaction model to study how preferences converge when driven by observation of each other's behaviors. We show that the dynamics of convergence resemble an Ornstein-Uhlenbeck process, a common model in nonequilibrium stochastic dynamics. Using standard analytical and computational techniques, we find that the hyperprior magnitudes, representing the learning time, determine the convergence value and the asymptotic entropy of the preferences across pairs of agents. We also show that the dynamical variance in preferences is characterized by a relaxation time $t^{\star}$ and compute its asymptotic upper bound. This formulation enhances the existing toolkit for modeling stochastic, interactive agents by formalizing leading theories in learning theory, and builds towards more comprehensive models of open problems in principal-agent and market theory.

DOI: 10.1103/PhysRevE.109.054106

I. INTRODUCTION

Belief formation is essential for studying behavior in the social and cognitive sciences. In noisy environments, empirical beliefs are formed through observation [1] and probabilistically predict future states to optimize energy costs [2]. Belief dynamics are critical for modeling how agents interact strategically in varying socioeconomic contexts (through games), navigate uncertainty, and make decisions under imperfect information. However, there remain questions about how beliefs evolve in complex social environments such as networks [3] and markets [4], where the fluctuating beliefs (or perception of others') of asset values subject markets to intense volatility [5] and divergent valuations [6].

Several learning models have emerged to explain the formation of beliefs in stochastic multiagent games [7], including frequentist and regression approaches [8,9]. Reinforcement learning (RL) models are widely used and have intuitive descriptions [10,11], but they do not produce closed-form solutions to dynamics of agent preferences [12], hampering the search for generalizable results. These models are generally outperformed by learning frameworks based on Bayesian inference (BI) [13–15], where agents process information to inform history-dependent, optimally predictive, and (in some cases) analytically tractable models of their environment. BI has thus become foundational in human cognition [16–19], and in studying adaptive agent behavior in models of wealth and inequality [20,21], social dynamics [22], and coordinated action [23,24]. Additionally, Bayesian reversal learning has emerged as a more efficient alternative to RL in more realistic, nonstationary environments [25] where discerning signal dynamics from noise is difficult [26,27].

Solutions to closed-form belief dynamics in stationary environments have contributed to a growing literature [13,20,28]. However, they are not suitable for studying convergence in interacting models where signals are dynamic [29,30]. Studying pairwise dynamics in Gaussian models, for which analytical descriptions of distribution parameters exist [31], closes this theoretical gap while opening the door towards characterizing emergent population preference dynamics [6,32]. We can accomplish this using established methods in nonequilibrium statistical physics, where the relationship between BI and Ornstein-Uhlenbeck (OU) processes as noisy, mean-reverting processes with memory is well explored [33–35]. In the case of sequential Bayesian estimation, this analysis can be used to study how convergence time relates to behavioral properties.

In this paper, we propose a model for the statistical dynamics of two agents' preferences under Bayesian adaptation to another's behavior. By treating behaviors as a Gaussian-distributed quantity, we can study the dynamics of preferences through the coupled Markov dynamics of its first-order moments. We first show that in the absence of noise, the asymptotic preferences of the agents converge to one another both to a relative value and on a timescale set by the relative strength of their priors. Later, we introduce noise and show how the dynamics resemble an OU process with time-rescaling noise. Using the Fokker-Planck equation (FPE), we then show that the preferences converge to a stationary

*These authors contributed equally to this work.

distribution with a width set by the uncertainty in their behavior and with dynamics governed by a relaxation time, $t^{\star}$. We conclude by discussing how convergence can be broken by introducing unpredictable behavioral shocks and the model's implication for studying belief formation in a host of game-theoretical and principal-agent problems.

II. BAYESIAN PREFERENCE DYNAMICS

Consider agents A and B, who at time-step i exhibit a statistically distributed, real-valued behavior $x_i \in X$ and $y_i \in Y$. We denote the normalized distribution of their decisions $P_i(X|\theta_A)$ and $P_i(Y|\theta_B)$, parameterized by behavioral parameters θ_A, θ_B. Consider that the agents can learn each other's behavior and are motivated to align their decisions (e.g., $[x_i - y_i]^2$ is minimized) but cannot directly coordinate their actions before observation. While coordination can be accomplished by conditioning behavior on some shared signal [23], this would not change the general dynamics and is excluded for brevity.

Each agent infers the other agent's preferences by observing their cumulative noisy behavior and adjusting their preferences to match. By preferences, we mean the first moment of the distribution of behaviors that spans the agent's set of choices. This particular setup is motivated by open questions in principal-agent problems, where agents must coordinate their behavior through adaptation [36].

History-dependent learning is accomplished optimally through BI [20]. As such, the distribution of agent A's behaviors at $i = 0$ forms a prior for their guess of $B's$, $P(X = x) \equiv P_0(\tilde{Y} = x)$, for approximated behavior $\tilde{Y}$ (and $\tilde{X}$ for B). The distribution of decisions at later interactions is given by a posterior $P_i(\tilde{Y}|\{y_i\})$, where the decision is conditioned on the history of B's behavior [37]. After n steps, agent A's posterior is given by (and B by analogy)

$$P_n(\tilde{y}|\{y_i\}, \theta_A) = \left[\prod_{i=1}^{n} \frac{P(y_i|\tilde{y}_i)}{P(y_i)}\right] P(x|\theta_A). \qquad (1)$$

In sequential BI, an agent's behavior at step n follows a Markov process and is sampled from P_{n-1}. This process is illustrated in Fig. 1, where $\mathcal{L}$ denotes the likelihood given the evidence.

In this paper, we assume the behaviors are instantaneously described by Gaussian distributions with gamma-distributed priors, $x \sim \mathcal{N}(\mu_x, \sigma_x|\boldsymbol{\theta}_x)$ and $y \sim \mathcal{N}(\mu_y, \sigma_y|\boldsymbol{\theta}_y)$, where $\boldsymbol{\theta}$ is the gamma prior vector. The means μ_x, μ_y, describe the agents' preferences, whereas the fluctuation in true behavior is given by the Gaussian standard deviations σ_x, σ_y.

BI on this choice of distribution results in preference dynamics that are linear [31]. Therefore, we first study the dynamics of the preference averages, then later consider how noisy behavior couples into the preference variances. The following analysis gives a first-order approximation of the complete behavior (*vis a vis* the preferences) under BI, whereas dynamics of higher order naturally come from higher-order moments and their couplings. In this paper, we will assume $\sigma_x = \sigma_y$ and leave the dynamics of the standard deviations under BI for future work.

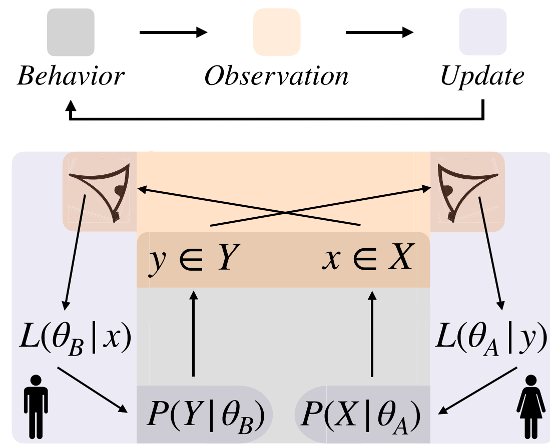

FIG. 1. Diagram of the interaction model. Agents A and B sample behaviors $x \in X$ and $y \in Y$ from respective distributions. Agent A updates their prior θ_A with the evidence $\mathcal{L}(y|\theta_a)$ from B's behavior, and vice versa.

A. Deterministic dynamics

First, we will study the dynamics of the preference parameter in the absence of noise. The rule for updating the mean parameter of a Gaussian-gamma model under BI is described recursively after n steps as (Appendix A) [31]

$$\mu_x^n = \frac{\mu_x^{n-1}\left(\frac{n-1}{\omega} + \alpha\right) + \mu_y^{n-1}}{\frac{n}{\omega} + \alpha}, \qquad \mu_x^1 = \frac{\alpha x_0 + y_0}{1 + \alpha},$$

where $\omega = n/t$ is the interaction rate. In the continuous limit $\omega \to \infty$, μ_x and μ_y become coupled by the linear differential equations

$$\frac{\partial \mu_x}{\partial t} = \frac{\mu_y(t) - \mu_x(t)}{t + \alpha}, \qquad \frac{\partial \mu_y}{\partial t} = \frac{\mu_x(t) - \mu_y(t)}{t + \beta}, \qquad (2)$$

where x_0, y_0 are the initial preferences and α, β are the hyperprior magnitudes with units t. Denoted the learning times, these parameters measure how resilient the preferences are to new evidence. These equations say that the dynamics of the preference parameters μ_x, μ_y decrease as the quantities converge in time. We demonstrate this by constructing the ODE for the difference measure $\Delta(t) = \mu_x(t) - \mu_y(t)$ [correspondingly $\Sigma(t) = \mu_x(t) + \mu_y(t)$], with solution (Appendix A 1)

$$\Delta(t) = \frac{\Delta_0 \alpha\beta}{R(t)}, \qquad (3)$$

where $\Delta_0 = x_0 - y_0$, and $R(t) = (\alpha + t)(\beta + t)$ is the time rescaling coefficient. This shows intuitively that the agents' preferences converge with power law -2 in time that increases symmetrically as $\alpha, \beta \to \infty$, and agent learning times increase.

With intuition for the coupled system established, we can now study the dynamics of the full system. There exist two solutions to Eq. (2) given by the equality of the learning times. First, when $\alpha = \beta$, the dynamics have the asymptotically symmetric solution $f(x_0, y_0, t) = \mu_x(t)$ and $f(y_0, x_0, t) = \mu_y(t)$, where f is defined as

$$f(x_0, y_0, t) = \frac{2\alpha^2 x_0 + (2\alpha t + t^2)(x_0 + y_0)}{2(\alpha + t)^2}. \qquad (4)$$

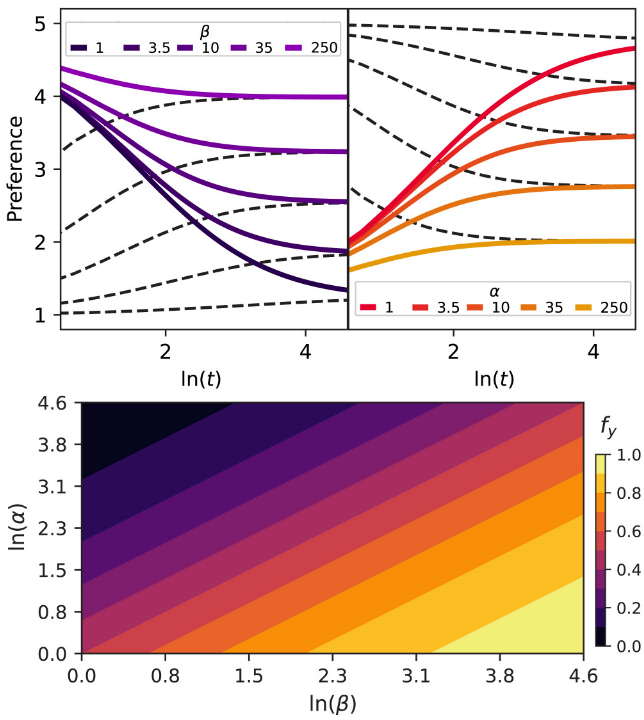

FIG. 2. Behavior of the noiseless model in Eq. (2) with $x_0 = 1$, $y_0 = 5$. Top: Convergence values computed from Eq. (5) for variable β (left) and variable α (right) for constant reference agent ($\alpha, \beta = 5$), represented by dashed lines. Bottom: Asymptotic fractional drift computed from (6) on a logarithmic parameter scale.

It follows that $\lim_{t\to\infty} f = (x_0 + y_0)/2$, and both agents' preferences converge to the average of their initial preferences asymptotically at times $t \gg 2\alpha$.

In the case $\alpha \neq \beta$, the solution for μ_x is given by

$$\mu_x(t) = \frac{\alpha x_0}{\alpha + t} + \frac{\alpha\beta(x_0 - y_0)}{(\alpha - \beta)^2} K(t) + \frac{t(\alpha x_0 - \beta y_0)}{(\alpha - \beta)(\alpha + t)}, \quad (5)$$

where $K(t) = \ln\left[\frac{\alpha\beta + \beta t}{\alpha\beta + \alpha t}\right]$ is a dynamical value with $\lim_{t\to\infty} K(t) = \ln[\beta/a]$. As we would expect, $\mu_x(0) = x_0$ and at long times, $\mu_x(t) \to s$, where s is the weighted average between the initial values:

$$s \equiv \frac{x_0[\alpha/\beta + \ln[\beta/\alpha] - 1] + y_0[\beta/\alpha - \ln[\beta/\alpha] - 1]}{(\alpha - \beta)^2/\alpha\beta}.$$

The solution for $\mu_y(t)$ is given in the Appendixes, with $\mu_y(t) \to s$ asymptotically. These results are demonstrated at the top of Fig. 2 for various learning times, with $y_0 = 5$ and $x_0 = 1$. In matrix form, these dynamics are given by $\mathbb{M}[x_0, y_0] \equiv [x(t), y(t)]$, where the drift matrix is

$$\mathbb{M} = \frac{\alpha\beta}{\alpha - \beta}\begin{pmatrix} M_2(t) - M_1(0) & M_2(0) - M_2(t) \\ M_1(t) - M_1(0) & M_2(0) - M_1(t) \end{pmatrix},$$

$$M_1(t) = K(t) - \tfrac{1}{(t+\alpha)}, \quad M_2(t) = K(t) - \tfrac{1}{(t+\beta)}.$$

This invertible matrix has a nonzero determinant $\det[\mathbb{M}(t)] = \alpha\beta/R(t)$. As we will see, this gives the constant of motion for constructing exact solutions for the dynamics of the system with noise [38].

1. Asymptotic preference behavior

Conveniently, the asymptotic preference value can be expressed independently of the initial condition, allowing us to compute the relative shift in preferences as a function of learning times. Consider the initial parameter difference Δ_0, and the difference in asymptotic value from the initial parameter $\delta_x = x_0 - s$. The fractional similarity of X is given by $f_x = 1 - \delta_x/\Delta_0$. This expresses how close X has remained to x_0 relative to y_0, and is useful for measuring the change in preferences of an agent represented by X (and Y by analogy). It is given by

$$f_x = 1 - \alpha\beta\frac{\beta/\alpha - \ln(\beta/\alpha) - 1}{(\alpha - \beta)^2}, \quad f_y = 1 - f_x. \quad (6)$$

These fractional limits are demonstrated in the bottom of Fig. 2 over various learning times.

So far, we have explored the dynamics of this model without noise, and have shown that both preference parameters converge to a value set by the relative magnitude of the learning times. We have shown that the deterministic dynamics are isomorphic and that we glean useful information about the relative change in preference between the agents without knowledge of the initial conditions. These results establish intuition for how, on average, agent characteristics determine the convergence process. In the following section, we will introduce noise to the inference process and demonstrate a procedure for constructing exact solutions using the linear and isomorphic properties of the dynamics. While this procedure results in lengthy analytical solutions that are not explored, we will demonstrate some key insights from the coupled dynamics, $\Delta(t)$, $\Sigma(t)$.

B. Full dynamics under noisy sampling

We introduce noise by rewriting Eq. (2) as the stochastic differential equations on quantities X_t, Y_t,

$$dX_t = \frac{\sigma_y}{t + \alpha} dW_{2,t} - \frac{\Delta_t}{t + \alpha} dt, \;\; dY_t = \frac{\sigma_x}{t + \beta} dW_{1,t} + \frac{\Delta_t}{t + \beta} dt,$$

with boundary conditions $X_0 = x_0, Y_0 = y_0$. We have introduced white Gaussian noise (WGN) processes, $dW_{1,t}$, $dW_{2,t}$ with magnitudes σ_x, σ_y that describe i.i.d fluctuations in agent behavior. Recalling previously that the asymptotic preferences depend on the initial conditions, we note that while the dynamics of the SDEs are Markovian, they cannot be ergodic. The dynamics of both preferences behave like OU processes, as the magnitude of the attractive drifts increases with the magnitude of the difference. However, this OU process is time inhomogeneous, as the magnitude of all dynamics decay with a power law in time. We interpret these dynamics in terms of the underlying BI process. The rate of parameter convergence slows as the agents converge in parameter value, and the effect of each interaction decreases in time as the agent weighs cumulatively larger sums of evidence. At long times, when preferences have nearly converged and have accumulated lengthy histories, small fluctuations dominate the dynamics.

To explore the statistics of the two-dimensional process, we define the bivariate transition probability distribution (TPD) as $P(x, y, t|x_0, y_0)$. The evolution for this distribution is given by the FPE, $\partial_t P(x,y,t|x_0,y_0) = \mathcal{F}[P]$, where $\mathcal{F}[\cdot]$ is defined:

$$\mathcal{F}[\cdot] = \partial_x\left(\frac{y-x}{t+\alpha}[\cdot]\right) + \partial_y\left(\frac{x-y}{t+\beta}[\cdot]\right) + \frac{\sigma_y^2}{2(t+\alpha)^2}\partial_{xx}[\cdot] + \frac{\sigma_x^2}{2(t+\beta)^2}\partial_{yy}[\cdot]. \quad (7)$$

One can marginalize the distribution for x, $P_M(x,t|x_0) = \int_{\mathbb{R}} P(x,y,t|x_0,y_0)dy$, and by analogy, y. To solve these equations exactly, we transform the set of equations into the frame of constant motion, defined by $\mathbb{M}(t)$, in which the process is purely diffusive and described by a Gaussian. In this frame, solutions for the dynamics of $P_M(x,t)$ and $P_M(y,t)$ are exactly solvable [38]. However, this procedure does not lead to concise results and is detailed only in the Appendixes.

As in the deterministic case, we glean tractable insights into the dynamics by solving the FPE for the coupled system, $X_t \to \Delta_t = X_t - Y_t$, $Y_t \to \Sigma_t = X_t + Y_t$. In the following section, we will use an exact solution of the FPE to show how the mean and variance of the TPD of Δ_t converges to zero, encoding the system's entropy into Σ_t. We will conclude this work by approximating an upper bound for the asymptotically stationary variance of $\Sigma(t)$.

C. Solutions of the FPE for coupled dynamics

In terms of the original model parameters, the new SDEs are

$$d\Delta_t = \frac{\sqrt{\sigma_y^2(t+\beta)+\sigma_x^2(t+\alpha)}}{R(t)}dW'_{1,t} - \frac{2t+\alpha+\beta}{R(t)}\Delta_t dt,$$

$$d\Sigma_t = \frac{\sqrt{\sigma_y^2(t+\beta)+\sigma_x^2(t+\alpha)}}{R(t)}dW'_{2,t} + \frac{\alpha-\beta}{R(t)}\Delta_t dt, \quad (8)$$

where the dW' terms are now correlated WGN processes. Again, we see that the difference equation behaves like an OU process, where drift is set by the difference in preferences, with time-rescaling noise. In this sense, Σ_t does not couple into the dynamics of Δ_t, permitting us to solve for the statistics of Δ_t first, then Σ_t.

1. The difference equation

In these coordinates, the statistics of Δ_t are fully described by the TPD $P_\Delta(z,t|z_0,0) = \text{Prob}\{z \leqslant \Delta_t \leqslant (z+dz)|z_0\}$, which solves the FPE

$$\partial_t P_\Delta = \partial_z\left[\frac{2t+\alpha+\beta}{R(t)}zP_\Delta\right] + D(t)\partial_{zz}P_\Delta, \quad (9)$$

where the diffusivity $D(t) = \frac{\sigma_y^2(t+\beta)^2+\sigma_x^2(t+\alpha)}{2[R(t)]^2}$. To solve this partial differential equation, we seek the reference frame where the process becomes purely diffusive. Consider the change of variables $z \mapsto z' \equiv z\frac{R(t)}{\alpha\beta}$ and $t \mapsto \tau \equiv t$. The differential operators transform as $\partial_z \mapsto \frac{R(t)}{\alpha\beta}\partial_{z'}$, $\partial_t \mapsto \frac{2t+\alpha+\beta}{R(t)}z'\partial_{z'} + \partial_t$, where we used the equivalence $t = \tau \to \partial_t = \partial_\tau$, yielding $\partial_t P_\Delta = \frac{2t+\alpha+\beta}{R(t)}P_\Delta + D(t)\frac{R(t)^2}{\alpha^2\beta^2}\partial_{z'z'}P_\Delta$ [Eq. (C5) in Appendix C]. Introducing the rescaling $P_\Delta \equiv R(t)Q_\Delta$, diffusion absorbs the drift term and reduces the dynamics to time inhomogeneous diffusion $\partial_t Q_\Delta = D'(t)\partial_{z'z'}Q_\Delta$, where $D'(t) = \frac{\sigma_x^2(t+\alpha)^2+\sigma_y^2(t+\beta)^2}{2\alpha^2\beta^2}$. To solve this equation, we introduce the time rescaling, $t \to s(t) = \int_0^t D(\xi)d\xi$, giving

$$s(t) = \frac{\sigma_x^2(t+\alpha)^3+\sigma_y^2(t+\beta)^3}{3\alpha^2\beta^2} - s_0,$$

where $s_0 = \frac{\sigma_x^2\alpha}{\beta^2} + \frac{\sigma_y^2\beta}{\alpha^2}$. Equation (9) has a Gaussian solution $Q_\Delta \sim \mathcal{N}(z'_0, \sqrt{s(t)})$. This Gaussian transforms back to the moving frame, $z' \to z = z'\alpha\beta/R(t)$, to get the full solution for P_Δ,

$$P_\Delta(z,t|z_0,0) = \frac{R(t)}{\sqrt{2\pi\sigma_\Delta^2(t)}}\exp\left[-\frac{\left(z - z_0\frac{\alpha\beta}{R(t)}\right)^2}{2\sigma_\Delta^2(t)}\right],$$

where we have introduced the difference variance as

$$\sigma_\Delta^2(t) \equiv \frac{\sigma_x^2(t+\alpha)^3+\sigma_y^2(t+\beta)^3}{3R^2(t)}. \quad (10)$$

This probability density has a few key features. First, $\lim_{t\to\infty}\langle\Delta_t\rangle = 0$ and $\lim_{t\to\infty}\sigma_\Delta^2(t) = 0$, so the distribution asymptotically converges to a delta function at $\Delta = 0$. By construction, $\sigma_\Delta(0) = 0$ and $\sigma_\Delta^2(t) \geqslant 0$, meaning the variance evolves nonmonotonously, and maximizes at a relaxation time $t^\star$.

While cumbersome, these results are intuitive in terms of the underlying inference process. Agents' preferences converge to one another almost certainly in a time that increases with the magnitude of fluctuations but decreases in the strength of the agents' learning times. Therefore, the strength of attraction is asymptotically stronger than noise fluctuations. These convergence dynamics are demonstrated by Monte Carlo (MC) simulations in Fig. 3 with $N = 1000$, $\alpha = \beta = 15$, where we see that $\Delta \to 0$ on a logarithmic scale, in agreement with theory. We can define the observables $m(t) = \Sigma(t)/2$, $\Delta_m(t) = m(t) - m(0)$, where Δ_m measures how the mean of the dynamics change over time, and $\Delta_{y,m} = m(t) - y(t)$ demonstrates how y differs from the mean. We see that $\Delta_{y,m}$ ($\Delta_{x,m}$) converges to zero with asymptotically vanishing noise, indicating that the preferences are converging to the mean almost certainly, while the entropy is increasingly expressed through the convergence value.

However, convergence in realistic agents are subject to social, environmental, and biological constraints to their interaction time (such as life expectancies). So, while convergence is asymptotically guaranteed, it is not guaranteed for particularly noisy or stubborn agents that have interaction time horizons shorter than the convergence timescale.

In the following section, we will use these insights to solve for the TPD of the Σ_t process.

2. The sum equation

The cumbersome dynamics of Δ_t lead to an even more complex analytical description for Σ_t. However, we know that $\lim_{t\to\infty}\langle\Delta\rangle(t) = 0$ and $\lim_{t\to\infty}\sigma_\Delta(t) = 0$. It follows that the initially bivariate diffusion process asymptotically collapses into a univariate pure diffusion centered at $\mu_{\Sigma,f}$. Hence, for

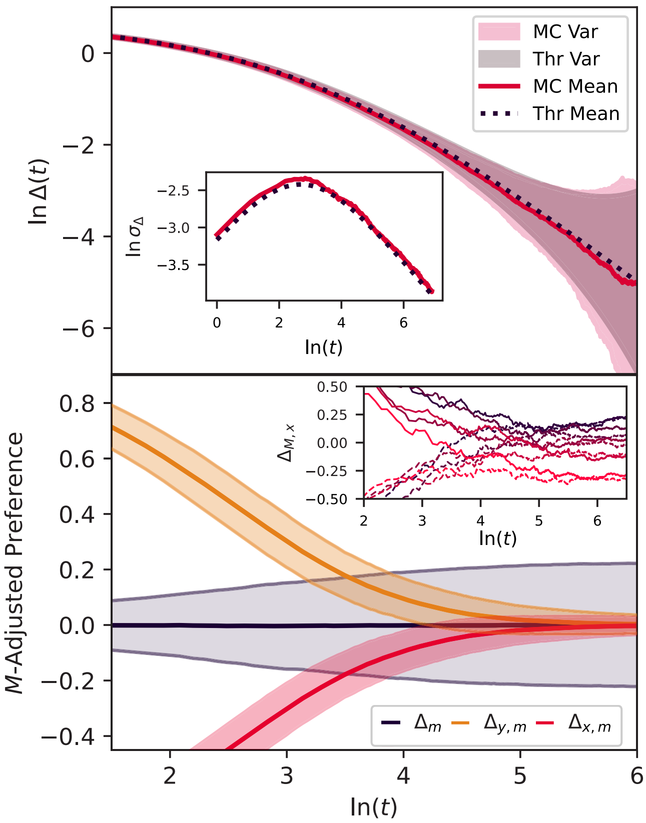

FIG. 3. Coupled dynamics of stochastic agent preferences with $x_0 = 0$, $y_0 = 2$, $\alpha = \beta = 25$ and 95% CI shaded regions. Top: MC dynamics of Δ_t match theory. Inset: The variance initially increases, reaches a maximum at $t^\star$, and then decreases. Bottom: The mean-adjusted dynamics ($\Delta_m = \Sigma/2$) is constant for this choice of parameters, with asymptotically constant noise. The difference in agent parameters from the mean, $\Delta_m - y$, $\Delta_m - x$, converge to 0 with time vanishing noise. Inset: Selected trajectories demonstrating different asymptotic values.

$t \to \infty$, we shall approximately have (see Appendix C 2)

$$d(\Sigma_t - \mu_{\Sigma,f}) = d\Sigma_t \approx \frac{\sqrt{\sigma_y^2(t+\beta) + \sigma_x^2(t+\alpha)}}{(t+\beta)(t+\alpha)} dW'_{2,t},$$

where the constant $\mu_{\Sigma,f}$ is given in Eq. (A12). The corresponding TPD $P_\Sigma(z, t|\mu_{\Sigma,f})dz = \text{Prob}\{z \leqslant \Sigma_t \leqslant (z + dz)|\sigma_0\}$ solves the FPE:

$$\partial_t P_\Sigma = \frac{\sigma_x^2(t+\beta)^2 + \sigma_y^2(t+\alpha)^2}{2R(t)} \partial_{zz} P_\Sigma.$$

By inspection, P_Σ is a Gaussian law with mean $\mu_{\Sigma,f}$ and, by an *ad hoc* time re-scaling (see Appendix C 2), we obtain the time-dependent variance as

$$\langle \Sigma_t^2 \rangle = \sigma_x^2 \left[\frac{1}{\alpha} - \frac{1}{\alpha + t} \right] + \sigma_y^2 \left[\frac{1}{\beta} - \frac{1}{\beta + t} \right]. \quad (11)$$

We now see that the second moment (and hence the variance) converges in the limit $t \to \infty$ to a stationary value $\langle \Sigma_t^2 \rangle = \frac{\sigma_x^2}{\alpha} + \frac{\sigma_y^2}{\beta}$. Similarly to the case of σ_Δ^2, the variance of this distribution increases with the fluctuations in behavior and decreases with hyperprior strength. We note, though, that this is an upper-bound estimation for variance, as

$$\sigma_\Sigma^2(t) = \langle (\Sigma_t - \mu_{\Sigma,f})^2 \rangle = \langle \Sigma_t^2 \rangle - \mu_{\Sigma,f}^2 \leqslant \langle \Sigma_t^2 \rangle.$$

Furthermore, from the convergence of $\Delta_t \to 0$, we know preferences converge to $X_t = Y_t$ and the variances converge to $\langle X_t^2 \rangle = \langle Y_t^2 \rangle = \frac{\langle \Sigma_t^2 \rangle}{2}$ as $t \to \infty$.

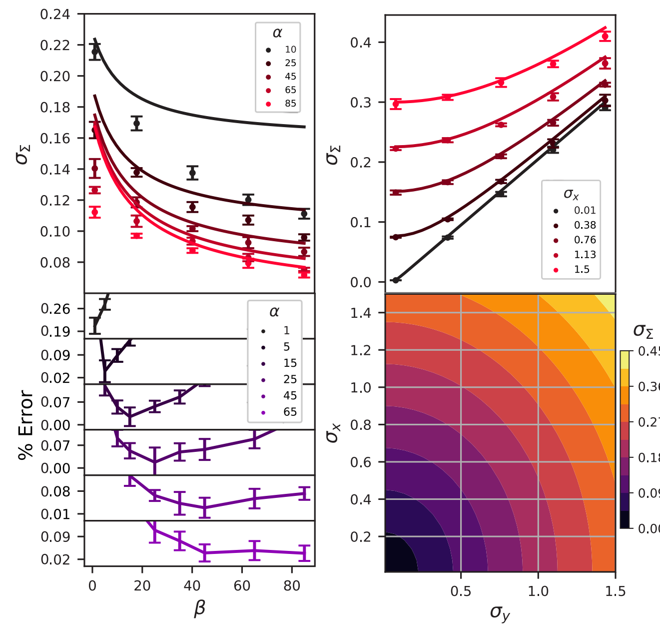

FIG. 4. Asymptotic variance upper bound $\langle \Sigma_t^2 \rangle$ under various parameters. Top: $\langle \Sigma_t^2 \rangle$ increases with agent noise with agreement between MC simulations and theory. Bottom: Variance decreases with agent learning time. Left: Variance upper bound $\langle \Sigma_t^2 \rangle$, diverges from empirical results as β and α diverge. Right: Deviation between the upper bound and MC experiments as a fraction of theoretical prediction.

The asymptotic preference variance demonstrated in Eq. (11) shows that more noisy agents who learn quickly will converge to more entropic states. This is because the fluctuations of one agent are recorded and immediately reciprocated by the partner agent (on average), biasing future preferences towards early fluctuations. When noise is strong or learning fast, agents weigh fluctuations more heavily than weak noise or slow learning. In both cases, these characteristics more strongly ossify early, noisy shifts in preferences before mean behavior is fully resolved, coupling more noise into the asymptotic behavior of the system. MC simulations demonstrate the relationships between these quantities and the variance in Fig. 4. We also see through simulations that the percent error in the upper bound estimate and the true variance are closest when $\beta = \alpha$ and increases as the learning times diverge. However, this difference decreases as both learning times approach large values.

III. DISCUSSION

In this paper, we studied a simple model for pairwise belief formation in Bayesian agents who adapt to each other's behaviors. We showed that preferences converge on a timescale and to a value given by the agents' relative learning times. Using the FPE, we then explored the convergence characteristics of the Gaussian PDF for preferences in the combined frame. We showed that while agents' preferences invariably converge to one another, the relative value is noisy, is characterized by a relaxation time $t^\star$, and is bounded above by a sum of the standard deviations of agent behaviors weighted by the learning times.

There remain several challenges concerning the full characterization of this system. First, deriving the full dynamics of Σ_t would be useful for attaining better bounds on the asymptotic coupled behavior. Second, solving the nonlinear dynamics for the covariance matrix of the inference process gives the full dynamics of the interaction, although this would likely require numerical treatment. Once we understand the full dynamics of this interaction, we can scale this model to include agents with multidimensional, covarying preferences. This builds towards a Bayesian analog of a self-other model [11], wherein agents coordinate decisions by approximating the other agent's behavior and serve as a microfoundation for more robust statistical mechanical models of network belief formation [3,32]. However, doing so also requires mechanisms for polarization, such as biased assimilation [39] where agents become resistant to preferences that diverge strongly from their own. While this behavior has been explored in DeGroot models of opinion formation [40], we must still explore how these interactions compete with convergence dynamics in a Bayesian context.

Furthermore, we can extend this analysis by studying preference dynamics in agents that must balance learning each other's signals with some additional, external signals. When only one agent observes an additional, stationary signal, it is natural that the agents' preferences would converge to a value biased by the external signal. However, when one agent observes an additional, nonstationary signal, as a form of unpredictable shock, or both agents observe separate, stationary signals as a form of reality check, their preferences are not guaranteed to converge [6]. The existence of a phase transition would depend on whether the external signal alters their preferences on a timescale comparable to the relaxation time $t^\star$, and can be applied at scale to study many-body preference dynamics.

Although this work can already be applied to game theoretic models of dynamical persuasion [41], dynamical prisoner's dilemma [42], and other games of trust and coordination [43–45], where the transmission of preferences through behavior determines asymptotic Nash equilibria. This formalism can also be adapted to study how preferences evolve in principal-agent models. In cases where the agent serves as an information channel for the principal, there remain questions of how the value and variance of asymptotic preferences behave as agents adapt to post-contract disagreements [36]. Models of information-driven resources dynamics [20] can be used to study how the convergence rate affects agent resources and how the entropy of convergence values affects the quality of information transmitted to the principal. Generally, these results constitute a step towards more robust quantitative models of interagent and market interactions that incorporate findings from the cognition community.

ACKNOWLEDGMENTS

We thank A. Kline, A. Stier, and L. Bettencourt for their discussions and comments on the paper. This work is supported by the Department of Physics at the University of Chicago, the National Science Foundation Graduate Research Fellowship (Grant No. DGE 1746045 to J.K.), and the Swiss State Secretariat for Education, Research and Innovation SERI (to J.K.).

APPENDIX A: DEFINING THE DETERMINISTIC ODE

The equation for the mean of the Gaussian Gamma for variable x is given by

$$\mu_x^n = \frac{\sum_{i=0}^{n} y_i + \mu_0\alpha}{n+1+\alpha} = \frac{y_{n-1} + \sum_{i=0}^{n-1} y_i + \mu_0\alpha}{n+1+\alpha}.$$

The sample y_n corresponds to the mean of $P(y)$, μ_y, plus some noise ξ_n. In the deterministic case, $\xi_n = 0$, and the remaining two terms constitute $\mu_x^{n-1}(n+\alpha)$. We can therefore redefine this quantity as Eqs. (2) from the main text,

$$\mu_x^n = \frac{\mu_x^{n-1}[(n-1)/\omega + \alpha] + \mu_y^{n-1}}{\frac{n}{\omega} + \alpha}, \tag{A1}$$

where we have written $n \to \frac{n}{\omega}$ to enable a conversion to continuous time variables later on. In the deterministic case, we define the difference operator

$$\Delta\mu_x = \mu_x^n - \mu_x^{n-1} = \frac{\mu_x^{n-1}[(n-1)/\omega + \alpha] + \mu_y^{n-1}}{\frac{n}{\omega} + \alpha} - \frac{\mu_x^{n-1}\left(\frac{n}{\omega} + \alpha\right)}{\frac{n}{\omega} + \alpha} = \frac{\mu_y^n - \mu_x^n}{\frac{n}{\omega} + \alpha}, \tag{A2}$$

which as $\omega \to \infty$ converges to the continuous time ODE used in the main text. Finally, Gaussian noise will be linearly added to the deterministic dynamics.

1. Deterministic evolution of preferences

We start by solving the deterministic motion, namely, the set of ODEs:

$$\frac{d\mu_x(t)}{dt} = \frac{\mu_y(t) - \mu_x(t)}{t+\alpha}, \qquad \mu_x(0) = x_0$$
$$\frac{d\mu_y(t)}{dt} = \frac{\mu_x(t) - \mu_y(t)}{t+\beta}, \qquad \mu_y(0) = y_0. \tag{A3}$$

To process further, we introduce the new variables:

$$\mu_\Delta(t) = [\mu_x(t) - \mu_y(t)] \quad \text{and} \quad \mu_\Sigma(t) = [\mu_x(t) + \mu_y(t)]. \tag{A4}$$

In terms of the new variables, Eq. (A3) reads

$$[(t+\alpha)(t+\beta)]\frac{d\mu_\Delta(t)}{dt} = -[2t+\alpha+\beta]\mu_\Delta(t)$$
$$[(t+\alpha)(t+\beta)]\frac{d\mu_\Sigma(t)}{dt} = [\alpha-\beta]\mu_\Delta(t). \tag{A5}$$

From Eq. (A5), we immediately have

$$\frac{d\ln(\mu_\Delta(t))}{dt} = -\frac{[2t+\alpha+\beta]}{[(t+\alpha)(t+\beta)]} = -\frac{d\ln[(t+\alpha)(t+\beta)]}{dt} \Rightarrow \mu_\Delta(t) = \frac{\Delta_0\alpha\beta}{[(t+\alpha)(t+\beta)]} := \frac{\Delta_0\alpha\beta}{R(t)}$$
$$[(t+\alpha)(t+\beta)]\frac{d\mu_\Sigma(t)}{dt} = [\alpha-\beta]\mu_\Delta(t) \quad \Rightarrow \quad \frac{d\mu_\Sigma(t)}{dt} = \frac{(\alpha-\beta)\alpha\beta\Delta_0}{[(t+\alpha)(t+\beta)]^2} = \Delta_0\frac{\alpha\beta(\alpha-\beta)}{R^2(t)}, \tag{A6}$$

with

$$R(t) := [(t+\alpha)(t+\beta)]. \tag{A7}$$

By direct calculation, one may verify the identities,

$$\int \frac{1}{R(t)}dt = \frac{1}{\alpha-\beta}\ln\left[\frac{t+\beta}{t+\alpha}\right],$$
$$\int \frac{1}{R^2(t)}dt = -\frac{1}{(\alpha-\beta)^2}\left\{\frac{d\ln[R(t)]}{dt} + 2\int\frac{1}{R(t)}dt\right\} = -\frac{1}{(\alpha-\beta)^2}\left\{\frac{d\ln[R(t)]}{dt} + \frac{2}{\alpha-\beta}\ln\left[\frac{t+\beta}{t+\alpha}\right]\right\}$$
$$= -\frac{1}{(\alpha-\beta)^2}\left\{\frac{2t+\alpha+\beta}{t^2+(\alpha+\beta)t+\alpha\beta} + \frac{2}{\alpha-\beta}\ln\left[\frac{t+\beta}{t+\alpha}\right]\right\}. \tag{A8}$$

Accordingly, from Eqs. (A6) and (A8), we obtain

$$\mu_\Sigma(t) = \int \frac{(\alpha-\beta)\alpha\beta\Delta_0}{R^2(t)}dt = -\alpha\beta\Delta_0\left\{\frac{2t+\alpha+\beta}{(\alpha-\beta)R(t)} + \frac{2}{(\alpha-\beta)^2}\ln\left[\frac{t+\beta}{t+\alpha}\right]\right\} + C, \tag{A9}$$

where C is an integration constant to be determined by the initial condition. At time $t = 0$, we have

$$C = \Sigma_0 + \Delta_0\left\{\frac{\alpha+\beta}{(\alpha-\beta)} + \frac{2\alpha\beta[\ln\beta - \ln\alpha]}{(\alpha-\beta)^2}\right\}. \tag{A10}$$

Hence, we can write

$$\mu_\Sigma(t) = \Sigma_0 + \frac{\Delta_0}{(\alpha-\beta)}\left[\alpha+\beta - \frac{(2t+\alpha+\beta)\alpha\beta}{(t+\alpha)(t+\beta)} + \frac{2\alpha\beta}{(\alpha-\beta)}\ln\left(\frac{\beta t+\alpha\beta}{\alpha t+\alpha\beta}\right)\right]. \tag{A11}$$

Note that we have

$$\lim_{t\to 0^+}\mu_\Sigma(t) = \Sigma_0$$
$$\lim_{t\to\infty}\mu_\Sigma(t) := \mu_{\Sigma,f} = \Sigma_0 + \Delta_0\left[\frac{\alpha+\beta}{\alpha-\beta}\right] = 2\left[\frac{\alpha x_0 - \beta y_0}{\alpha-\beta}\right] + 2\Delta_0\frac{\alpha\beta\ln[\beta/\alpha]}{(\alpha-\beta)^2}, \qquad \alpha \neq \beta$$
$$\lim_{t\to\infty}\mu_\Sigma(t) := \Sigma_0, \qquad \alpha = \beta, \tag{A12}$$

In terms of $\mu_x(t)$ and $\mu_y(t)$, we have

$$\mu_x(t) = \frac{\mu_\Delta(t) + \mu_\Sigma(t)}{2} = \left[\frac{\alpha x_0 - \beta y_0}{\alpha - \beta}\right] + (x_0 - y_0)\left\{\frac{\alpha\beta}{\alpha - \beta}\ln\left[\frac{\alpha\beta + \beta t}{\alpha\beta + \alpha t}\right] - \frac{\alpha\beta}{(\alpha - \beta)(t + \alpha)}\right\}$$
$$\mu_y(t) = \frac{\mu_\Delta(t) - \mu_\Sigma(t)}{2} = \left[\frac{\alpha x_0 - \beta y_0}{\alpha - \beta}\right] + (x_0 - y_0)\left\{\frac{\alpha\beta}{\alpha - \beta}\ln\left[\frac{\alpha\beta + \beta t}{\alpha\beta + \alpha t}\right] - \frac{\alpha\beta}{(\alpha - \beta)(t + \beta)}\right\}, \tag{A13}$$

which can be summarized as

$$\begin{pmatrix}\mu_x(t)\\ \mu_y(t)\end{pmatrix} = \mathbb{M}(t)\begin{pmatrix}x_0\\ y_0\end{pmatrix}, \tag{A14}$$

where the matrix $\mathbb{M}(t)$ reads

$$\mathbb{M} = \begin{pmatrix}-M_1(0) + M_2(t) & M_2(0) - M_2(t)\\ -M_1(0) + M_1(t) & M_2(0) - M_1(t)\end{pmatrix},$$
$$M_1(t) = \frac{\alpha\beta}{\alpha - \beta}\ln\left[\frac{\alpha\beta + \beta t}{\alpha\beta + \alpha t}\right] - \frac{\alpha\beta}{(\alpha - \beta)(t + \alpha)},$$
$$M_2(t) = \frac{\alpha\beta}{\alpha - \beta}\ln\left[\frac{\alpha\beta + \beta t}{\alpha\beta + \alpha t}\right] - \frac{\alpha\beta}{(\alpha - \beta)(t + \beta)}. \tag{A15}$$

a. Stationary regime

From Eqs. (A14) and (A15), one immediately concludes that the final state $(\mu_x(\infty), \mu_x(\infty))$ is given by

$$\begin{pmatrix}\mu_x(\infty)\\ \mu_y(\infty)\end{pmatrix} = \begin{pmatrix}\frac{\alpha}{\alpha-\beta} & \frac{\beta}{\beta-\alpha}\\ \frac{\alpha}{\alpha-\beta} & \frac{\beta}{\beta-\alpha}\end{pmatrix}\begin{pmatrix}x_0\\ y_0\end{pmatrix} := \mathbb{M}_\infty\begin{pmatrix}x_0\\ y_0\end{pmatrix}. \tag{A16}$$

We observe that $(\mu_x(\infty), \mu_y(\infty))$ depends on the initial condition (x_0, x_0). We note that $\alpha = \beta$ is a singular situation. In addition, observe also that for $\alpha = 0$, we obtain $\mu_x(\infty) = \mu_y(\infty) = y_0$ and, conversely for $\beta = 0$, we have $\mu_x(\infty) = \mu_y(\infty) = x_0$, thus showing that in both of these limiting cases, the evolution affects a single variable.

APPENDIX B: SOLVING THE TPD USING LIOUVILLE COORDINATES

This Appendix shows that an *ad hoc* time-dependent change of coordinates $(X_t, Y_t) \mapsto (U_t, V_t)$ transforms the nominal drifted process into a pure bivariate diffusion for which the analytical probability density can be calculated. First, one observes

$$\mathrm{Det}[\mathbb{M}(t)] = \frac{\alpha\beta}{R(t)} \neq 0 \ \text{ for } \ 0 < t < \infty, \tag{B1}$$

with $R(t)$ as defined in Eq. (A7). Hence, the inverse matrix $\mathcal{M}(t)$ exists and reads

$$\mathbb{M}(t)\mathcal{M}(t) = \mathcal{M}(t)\mathbb{M}(t) = \mathbb{I}\mathrm{d}$$
$$\mathcal{M}(t) = \frac{R(t)}{\alpha\beta}\begin{pmatrix}M_2(0) - M_1(t) & -M_2(0) + M_2(t)\\ M_1(0) - M_1(t) & -M_1(0) + M_2(t)\end{pmatrix} := \begin{pmatrix}m_{11}(t) & m_{12}(t)\\ m_{21}(t) & m_{22}(t)\end{pmatrix}. \tag{B2}$$

In particular, using Eq. (A14), we can write

$$\begin{pmatrix}x_0\\ y_0\end{pmatrix} = \mathcal{M}(t)\begin{pmatrix}\mu_x(t)\\ \mu_y(t)\end{pmatrix}, \tag{B3}$$

where the initial values (x_0, y_0) are constants of the motion. This suggests introducing the time-dependent change of coordinates (i.e., Liouville coordinates) defined by

$$\mathbf{z} := \begin{pmatrix}x\\ y\end{pmatrix} \mapsto \mathbf{w} := \begin{pmatrix}u\\ v\end{pmatrix}$$
$$\begin{pmatrix}u\\ v\end{pmatrix} = \mathcal{M}(t)\begin{pmatrix}x\\ y\end{pmatrix} := \begin{pmatrix}m_{11}(t) & m_{12}(t)\\ m_{21}(t) & m_{22}(t)\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}. \tag{B4}$$

In terms of the **w** coordinates, the motion is purely diffusive (i.e., the drift components in the FPE cancel out) and we have

$$\partial_t[\cdot] \mapsto \partial_t[\cdot] + \frac{\partial_u}{\partial_t}\partial_u[\cdot] + \frac{\partial_v}{\partial_t}\partial_v[\cdot]$$
$$\begin{pmatrix}\partial_x[\cdot]\\ \partial_y[\cdot]\end{pmatrix} \mapsto \mathcal{M}^\dagger(t)\begin{pmatrix}\partial_u[\cdot]\\ \partial_v[\cdot]\end{pmatrix}$$
$$\frac{1}{2}[\partial_{xx} + \partial_{yy}][\cdot] \mapsto \triangle_{uv}[\cdot] := \frac{1}{2}(\partial_u[\cdot], \quad \partial_v[\cdot])\mathcal{M}(t)\mathcal{M}^\dagger(t)\begin{pmatrix}\partial_u[\cdot]\\ \partial_v[\cdot]\end{pmatrix}$$
$$\partial_t P = \triangle_{uv}[P]$$
$$P = \frac{1}{2\pi\sqrt{\mathrm{Det}[\Pi(t)]}}e^{-\mathbf{w}^\dagger \frac{\Pi(t)}{\mathrm{Det}(\Pi(t))}\mathbf{w}}. \tag{B5}$$

Equation (B5) describes the TPD evolution of the pure diffusion process:

$$\begin{pmatrix}dU_t\\ dV_t\end{pmatrix} = \mathcal{M}(t)\begin{pmatrix}dB_{1,t}\\ dB_{2,t}\end{pmatrix}, \tag{B6}$$

where $dB_{1,t}$ and $dB_{2,t}$ are independent WGN processes. The FPE in Eq. (B5) describes the TPD of the bivariate (U_t, V_t) pure Gaussian process and we have

$$\Pi(t) := \begin{pmatrix}\mathbb{E}\{V_t^2\} & \mathbb{E}\{V_t U_t\}\\ \mathbb{E}\{U_t V_t\} & \mathbb{E}\{U_t^2\}\end{pmatrix} := \begin{pmatrix}\pi_{11}(t) & \pi_{12}(t)\\ \pi_{21}(t) & \pi_{22}(t)\end{pmatrix}, \tag{B7}$$

$$\begin{aligned}
\mathbb{E}\{dU_t; dU_\tau\} &= \mathbb{E}\{[m_{11}(t)dB_{1,t} + m_{12}(t)dB_{2,t}]; [m_{11}(\tau)dB_{1,\tau} + m_{12}(\tau)dB_{2,\tau}]\}\\
&= [m_{11}(t)m_{11}(\tau) + m_{12}(t)m_{12}(\tau)]\delta(t-\tau)\\
\mathbb{E}\{dU_t; dV_\tau\} &= [m_{11}(t)m_{12}(\tau) + m_{22}(t)m_{21}(\tau)]\delta(t-\tau)\\
\mathbb{E}\{dV_t; dU_\tau\} &= [m_{11}(t)m_{21}(\tau) + m_{22}(t)m_{12}(\tau)]\delta(t-\tau)\\
\mathbb{E}\{dV_t; dV_\tau\} &= [m_{22}(t)m_{22}(\tau) + m_{21}(t)m_{21}(\tau)]\delta(t-\tau).
\end{aligned} \tag{B8}$$

Invoking Theorem 3.6 of Jazwinsky [46], Eq. (B8) leads to

$$\begin{aligned}
\mathbb{E}\{U_t^2\} &= \int_0^t ds \int_0^s d\tau[m_{11}(s)m_{11}(\tau) + m_{12}(s)m_{12}(\tau)]\delta(s-\tau)\\
&= \int_0^t dt[m_{11}^2(s) + m_{12}^2(s)]ds\\
\mathbb{E}\{U_t V_t\} &= \int_0^t [m_{11}(s)m_{12}(s) + m_{22}(s)m_{21}(s)]ds\\
\mathbb{E}\{V_t^2\} &= \int_0^t [m_{22}^2(s) + m_{21}^2(s)]ds.
\end{aligned} \tag{B9}$$

Going back to the nominal variables $\mathbf{z}^\dagger = (x, y)$, Eqs. (B4) and (B5) imply

$$\begin{aligned}
P &= \frac{1}{2\pi\sqrt{\mathrm{Det}(\mathbb{W})}}e^{-\frac{\mathbf{z}^\dagger \mathbb{W}(t)\mathbf{z}}{2\mathrm{Det}(\mathbb{W}(t))}}\\
\mathbb{W}(t) &= \mathcal{M}^\dagger(t)\Pi(t)\mathcal{M}(t) = \begin{pmatrix}A(t) & -H(t)\\ -H(t) & B(t)\end{pmatrix}\\
\mathbb{E}\{X_t^2\} &:= \int_{\mathbb{R}}\int_{\mathbb{R}} x^2 P\, dxdy = B(t)\\
\mathbb{E}\{Y_t^2\} &:= \int_{\mathbb{R}}\int_{\mathbb{R}} y^2 P\, dxdy = A(t),
\end{aligned} \tag{B10}$$

with

$$\begin{aligned}
B(t) &= [m_{12}^2\Pi_{11} + m_{12}\Pi_{12}m_{22} + m_{22}\Pi_{21}m_{12} + m_{22}^2\Pi_{22}](t)\\
A(t) &= [m_{11}^2\Pi_{11} + m_{11}\Pi_{12}m_{21} + m_{21}\Pi_{21}m_{11} + m_{21}^2\Pi_{22}](t),
\end{aligned} \tag{B11}$$

where the matrix elements $m_{ij}(t)$ and $\pi_{ij}(t)$ are explicitly given in Eqs. (B2) and (B7). While the present procedure is exact, it leads to cumbersome algebra.

Remark. Note that a simpler case of the above general scheme was explored by Chandrasekhar [38] (see Lemma II), for the simpler case:

$$\mathcal{M} = \begin{pmatrix} m_{11}(t) & m_{11}(t) \\ m_{22}(t) & m_{22}(t) \end{pmatrix}$$
$$\Delta_{uv} = m_{11}^2(t)\partial_{uu} + 2m_{11}(t)m_{22}(t)\partial_{uv} + m_{22}^2(t)\partial_{vv}$$
$$\Pi(t) = \begin{pmatrix} 2\int_0^t m_{22}^2(s)ds & -\int_0^t m_{11}(s)m_{22}(s)ds \\ -\int_0^t m_{11}(s)m_{22}(s)ds & 2\int_0^t m_{11}^2(s)ds \end{pmatrix}.$$

APPENDIX C: THE (X_t, Y_t) STOCHASTIC PROCESS USING COUPLED DYNAMICS

Consider the stochastic process $(X_t, Y_t) \in \mathbb{R}^2$:

$$\begin{aligned} dX_t &= \frac{-\Delta_t dt + \sigma_y dW_{1,t}}{t+\alpha} = \frac{(y-x)dt + \sigma_y dW_{1,t}}{t+\alpha}, \qquad X_0 = x_0 \\ dY_t &= \frac{+\Delta_t dt + \sigma_x dW_{2,t}}{t+\beta} = \frac{(x-y)dt + \sigma_x dW_{2,t}}{t+\beta}, \qquad Y_0 = y_0, \end{aligned} \tag{C1}$$

where $dW_{1,t}$ and $dW_{2,t}$ are independent WGN processes. To study the (X_t, Y_t) bivariate Gaussian [47] and Markovian diffusion process, it is advantageous to proceed with the change of variables:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} \mapsto \begin{pmatrix} \Delta_t := X_t - Y_t \\ \Sigma_t := X_t + Y_t \end{pmatrix}$$
$$\begin{aligned} d\Delta_t &= -\frac{[2t+\alpha+\beta]}{[(t+\alpha)(t+\beta)]}\Delta_t dt + \frac{\sqrt{\left[\sigma_x^2(t+\beta)^2 + \sigma_y^2(t+\alpha)^2\right]}dB_{1,t}}{[(t+\alpha)(t+\beta)]} \\ d\Sigma_t &= +\frac{(\alpha-\beta)}{[(t+\alpha)(t+\beta)]}\Delta_t dt + \frac{\sqrt{\left[\sigma_x^2(t+\beta)^2 + \sigma_y^2(t+\alpha)^2\right]}dB_{2,t}}{[(t+\alpha)(t+\beta)]}, \end{aligned} \tag{C2}$$

where we have used the property

$$\begin{aligned} \frac{\sigma_y}{(t+\alpha)}dW_{1,t} - \frac{\sigma_x}{(t+\alpha)}dW_{2,t} &= \frac{\sqrt{\left[\sigma_x^2(t+\beta)^2 + \sigma_y^2(t+\alpha)^2\right]}dB_{1,t}}{[(t+\alpha)(t+\beta)]} \\ \frac{\sigma_y}{(t+\alpha)}dW_{1,t} + \frac{\sigma_x}{(t+\alpha)}dW_{2,t} &= \frac{\sqrt{\left[\sigma_x^2(t+\beta)^2 + \sigma_y^2(t+\alpha)^2\right]}dB_{2,t}}{[(t+\alpha)(t+\beta)]}, \end{aligned} \tag{C3}$$

with $dB_{1,t}$ and $dB_{2,t}$ being now *correlated* WGNs. Since the Δ_t process is actually decoupled from the Σ_t, we shall proceed in two steps.

a. The Δ_t process

The probabilistic properties of the Δ_t stochastic process in Eq. (C2) are fully described by the TPD $P_\Delta(x, t|x_0, 0)dx :=$ Prob$\{x \leqslant \Delta_t \leqslant (x+dx)|x_0\}$ which solves the FPE:

$$\begin{aligned} \partial_t P_\Delta &= \partial_x\left\{\left[\frac{[2t+\alpha+\beta]}{[(t+\alpha)(t+\beta)]}\right]x\right\} + D(t)\partial_{xx}P_\Delta \\ D(t) &:= \frac{\sigma_y^2(t+\beta)^2 + \sigma_x^2(t+\alpha)^2}{2(t+\alpha)^2(t+\beta)^2}. \end{aligned} \tag{C4}$$

To solve Eq. (C4), similarly to Appendix B, we express the evolution in terms of the constant of the motion Δ_o. Accordingly, we introduce the change of variables:

$$\begin{aligned} t \mapsto \tau = t \Rightarrow \quad \partial_t &\mapsto \frac{\partial x'}{\partial t}\partial_{x'} + \frac{\partial \tau}{\partial t}\partial_\tau = \left[\frac{2t+\alpha+\beta}{\alpha\beta}\right]x\partial_{x'} + \partial_\tau = \left[\frac{2t+\alpha+\beta}{(t+\alpha)(t+\beta)}\right]x'\partial_{x'} + \partial_t \\ x \mapsto x' := x\frac{(t+\alpha)(t+\beta)}{\alpha\beta} \quad \Rightarrow \quad \partial_x &\mapsto \frac{\partial x'}{\partial x}\partial_{x'} + \frac{\partial \tau}{\partial x}\partial_\tau = \frac{(t+\alpha)(t+\beta)}{\alpha\beta}\partial_{x'}. \end{aligned} \tag{C5}$$

In terms of the (t, x'), Eq. (C4) is transformed into a pure diffusion process. This can be seen as follows (omitting the arguments of P):

$$\partial_t P_\Delta + \left[\frac{2t+\alpha+\beta}{(t+\alpha)(t+\beta)}\right] x'\partial_{x'}P_\Delta = \frac{(\alpha+t)(\beta+t)}{\alpha\beta}\partial_{x'}\left\{\frac{[2t+\alpha+\beta]}{[(t+\alpha)(t+\beta)]}\frac{\alpha\beta}{(t+\alpha)(t+\beta)}x'P\right\} + D(t)\frac{(\alpha+t)^2(\beta+t)^2}{\alpha^2\beta^2}\partial_{x'x'}P_\Delta$$
$$= \frac{[2t+\alpha+\beta]}{[(t+\alpha)(t+\beta)]}\partial_{x'}(x'P_\Delta) + D(t)\frac{(\alpha+t)^2(\beta+t)^2}{\alpha^2\beta^2}\partial_{x'x'}P_\Delta,$$

yielding

$$\partial_t P_\Delta = \frac{[2t+\alpha+\beta]}{[(t+\alpha)(t+\beta)]}P_\Delta + D(t)\frac{(\alpha+t)^2(\beta+t)^2}{\alpha^2\beta^2}\partial_{x'x'}P_\Delta. \tag{C6}$$

Writing $P_\Delta := (t+\alpha)(t+\beta)Q$, Eq. (C6) reduces to the pure time inhomogeneous diffusion:

$$\partial_t Q_\Delta = \left[D(t)\frac{(\alpha+t)(\beta+t)}{\alpha^2\beta^2}\right]\partial_{x'x'}Q_\Delta = \left[\frac{\sigma_x^2(t+\alpha)^2+\sigma_y^2(t+\beta)^2}{2\alpha^2\beta^2(\alpha+t)(\beta+t)}\right]\partial_{x'x'}Q_\Delta. \tag{C7}$$

Finally, we introduce the time rescaling:

$$t \mapsto s(t) := \int_0^t \left[\frac{\sigma_x^2(\alpha+\xi)^2+\sigma_y^2(\beta+\xi)^2}{\alpha^2\beta^2(\alpha+\xi)(\beta+\xi)}\right]d\xi = \frac{(\sigma_x^2+\sigma_y^2)t + (\alpha-\beta)\left(\sigma_x^2\ln\frac{\beta+t}{\beta} - \sigma_y^2\ln\frac{\alpha+t}{\alpha}\right)}{\alpha^2\beta^2}. \tag{C8}$$

This enables us to rewrite Eq. (C7) as

$$\partial_s Q_\Delta = \frac{1}{2}\partial_{yy}Q_\Delta \quad \Rightarrow \quad Q_\Delta = \frac{1}{\sqrt{2\pi s(\tau)}}\exp\left[-\frac{(y-y_0)^2}{2s(\tau)}\right]. \tag{C9}$$

Proceeding backwards to the nominal (x, t) variables, one ends with

$$lP_\Delta(x,t|x_0,0)dx = \frac{(t+\alpha)(t+\beta)}{\alpha\beta\sqrt{2\pi s(t)}}\exp\left[-\frac{\left(x\frac{(t+\alpha)(t+\beta)}{\alpha\beta}-x_0\right)^2}{2s(\tau)}\right]dx$$
$$= \frac{1}{\sqrt{2\pi\sigma_\Delta^2(t)}}\exp\left[-\frac{\left(x-\frac{\alpha\beta x_0}{(t+\alpha)(t+\beta)}\right)^2}{2\sigma_\Delta^2(t)}\right], \tag{C10}$$

where we used the notation

$$\sigma_\Delta^2(t) := \frac{\alpha^2\beta^2 s(t)}{(t+\alpha)^2(t+\beta)^2} = \frac{(\sigma_x^2+\sigma_y^2)t + (\alpha-\beta)\left(\sigma_x^2\ln\frac{\beta+t}{\beta} - \sigma_y^2\ln\frac{\alpha+t}{\alpha}\right)}{(t+\alpha)^2(t+\beta)^2}$$
$$\lim_{t\to\infty}\sigma_\Delta^2(t) = 0. \tag{C11}$$

1. Nonmonotonous relaxation of the variance $\sigma_\Delta^2(t)$

As $X_0 = x_0$ and $Y_0 = y_0$ are fixed and deterministic, we obviously have $\lim_{t\to\infty}\sigma_\Delta^2(t) = 0$, In parallel, from Eq. (C11) we have $\lim_{t\to\infty}\Delta_t = 0$. Since $\sigma_\Delta^2(t) \geqslant 0$, we conclude that $\sigma_\Delta^2(t)$ follows a nonmonotonous evolution reaching a maximum at a relaxation time t^* such that $\frac{d}{dt}\sigma_\Delta^2(t)\mid_{t=t^*}= 0$. On physical grounds, this nonmonotonous evolution describes the underlying trade-off between two distinct mechanisms: a disorganizing mechanism generated by the noisy driving forces versus the organising mechanism generated by the mutual interactions. During the early stage $0 < t < t^*$, the fluctuations dominate while later for $t > t_c$ the learning mechanism overcomes the underlying noise to ultimately drive $\sigma_\Delta(t)$ towards zero. Accordingly, it is legitimate to interpret t_c as a relaxation time.

2. The Σ_t process: Approximation for the time asymptotic regime

Since the full transient evolution of variances as given in Appendix B leads to cumbersome expressions, let us focus on the time asymptotic development. We already know exactly that $\lim_{t\to\infty}\mathbb{E}\{\Delta_t\} = \lim_{t\to\infty}\mu_\Delta(t) = 0$, and from the last section we have $\lim_{t\to\infty}\sigma_\Delta(t) = 0$. Accordingly, in the time asymptotic regime, the initially bivariate diffusion collapses to a scalar (i.e., univariate) pure diffusion Σ_t process centered at the constant final value $\mu_{\Sigma,f}$ given in Eq. (A12). Hence, for asymptotic times Eq. (C2), the

Σ_t evolution can be approximately written as

$$d(\Sigma_t - \mu_{\Sigma,f}) = d\Sigma_t \approx \frac{\sqrt{\left[\sigma_x^2(t+\beta)^2 + \sigma_y^2(t+\alpha)^2\right]}}{[(t+\alpha)(t+\beta)]} dB_{2,t}. \tag{C12}$$

The associated TPD $P_\Sigma(x, t|\Sigma_0, 0)dx = \text{Prob}\{x \leqslant \Sigma_t \leqslant (x+dx)|\Sigma_0\}dx := P_\Sigma dx$ obeys to the FPE:

$$\partial_t P_\Sigma = \frac{\left[\sigma_x^2(t+\beta)^2 + \sigma_y^2(t+\alpha)^2\right]}{(t+\alpha)^2(t+\beta)^2} \frac{1}{2} \partial_{xx} P_\Sigma. \tag{C13}$$

To solve Eq. (C13), as usual we introduce the rescaling,

$$t \mapsto s(t) := \int_0^t \frac{\left[\sigma_x^2(\xi+\beta)^2 + \sigma_y^2(\xi+\alpha)^2\right]}{(\xi+\alpha)^2(\xi+\beta)^2} d\xi = \sigma_x^2\left[\frac{1}{\alpha} - \frac{1}{\alpha+t}\right] + \sigma_y^2\left[\frac{1}{\beta} - \frac{1}{\beta+t}\right], \tag{C14}$$

yielding

$$P_\Sigma(x, t|\Sigma_0, 0) = \frac{1}{\sqrt{2\pi s(t)}} e^{-\frac{[x-\mu_{\Sigma,f}]^2}{2s(t)}}$$

$$\lim_{t\to\infty} P_\Sigma(x, t|\mu_{\Sigma,f}, 0) = \frac{1}{\sqrt{2\pi s_\infty}} e^{-\frac{[x-\mu_{\Sigma,f}]^2}{2s_\infty}}, \quad s_\infty = \left[\frac{\sigma_x^2}{\alpha} + \frac{\sigma_y^2}{\beta}\right]$$

$$\lim_{t\to\infty} \sigma_\Sigma^2(t) = \mathbb{E}\{(\Sigma_t - \mu_{\Sigma,f})^2\} = s_\infty. \tag{C15}$$

Finally, for asymptotic times, we have that $\lim_{t\to\infty} X_t = \lim_{t\to\infty} Y_t$, we can conclude that the asymptotic variances of X_t and Y_t converge to

$$\lim_{t\to\infty} \sigma_X^2(t) = \lim_{t\to\infty} \sigma_Y^2(t). \tag{C16}$$

[1] R. J. Seitz and H.-F. Angel, Belief formation–a driving force for brain evolution, Brain Cognit. **140**, 105548 (2020).

[2] K. Friston, The free-energy principle: A unified brain theory? Nat. Rev. Neurosci. **11**, 127 (2010).

[3] J. Dalege, M. Galesic, and H. Olsson, Networks of beliefs: An integrative theory of individual-and social-level belief dynamics, OSF Preprints (2023).

[4] S. Gjerstad and J. Dickhaut, Price formation in double auctions, Games Econ. Behav. **22**, 1 (1998).

[5] M. Kurz and M. Motolese, Endogenous uncertainty and market volatility, Econ. Theory **17**, 497 (2001).

[6] R. Farmer and J.-P. Bouchaud, Self-fulfilling prophecies, quasi non-ergodicity and wealth inequality, J. Political Economy, **131**, 947 (2023).

[7] L. S. Shapley, Stochastic games, Proc. Natl. Acad. Sci. USA **39**, 1095 (1953).

[8] D. Fudenberg and D. K. Levine, Whither game theory? Towards a theory of learning in games, J. Econ. Perspect. **30**, 151 (2016).

[9] G. W. Evans and S. Honkapohja, Learning and macroeconomics, Annu. Rev. Econ. **1**, 421 (2009).

[10] L. Buşoniu, R. Babuška, and B. De Schutter, Multi-agent reinforcement learning: An overview, *Innovations in Multi-agent Systems and Applications-1*, edited by D. Srinivasan and L. C. Jain (Springer, Berlin, Germany, 2010), Vol. 310, pp. 183–221.

[11] R. Raileanu, E. Denton, A. Szlam, and R. Fergus, Modeling others using oneself in multi-agent reinforcement learning, in *Proceedings of the 35 th International Conference on Machine Learning* (PMLR, Stockholm, Sweden, 2018), pp. 4257–4266.

[12] R. S. Sutton and A. G. Barto, *Reinforcement lLearning: An Introduction*, (MIT Press, Cambridge Massachusetts, 2018).

[13] D. Acuna and P. Schrater, Bayesian modeling of human sequential decision-making on the multi-armed bandit problem, in *Proceedings of the 30th Annual Conference of the Gognitive Science Society* (Cognitive Science Society, Washington, DC, 2008), Vol. 100, pp. 200–300.

[14] S. J. Gershman and N. Uchida, Believing in dopamine, Nat. Rev. Neurosci. **20**, 703 (2019).

[15] V. D. Costa, V. L. Tran, J. Turchi, and B. B. Averbeck, Reversal learning and dopamine: A Bayesian perspective, J. Neurosci. **35**, 2407 (2015).

[16] T. E. Behrens, M. W. Woolrich, M. E. Walton, and M. F. Rushworth, Learning the value of information in an uncertain world, Nat. Neurosci. **10**, 1214 (2007).

[17] M. D. Lee and E.-J. Wagenmakers, *Bayesian Cognitive Modeling: A Practical Course* (Cambridge University Press, Cambridge, England, UK, 2014).

[18] A. Gopnik, S. O'Grady, C. G. Lucas, T. L. Griffiths, A. Wente, S. Bridgers, R. Aboody, H. Fung, and R. E. Dahl, Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood, Proc. Natl. Acad. Sci. USA **114**, 7892 (2017).

[19] K. B. Korb and A. E. Nicholson, *Bayesian Artificial Intelligence* (CRC Press, Boca Raton, Florida, USA, 2010).

[20] J. T. Kemp and L. M. Bettencourt, Learning increases growth and reduces inequality in shared noisy environments, PNAS Nexus **2**, pgad093 (2023).

[21] L. M. Bettencourt, Towards a statistical mechanics of cities, C. R. Phys. **20**, 308 (2019).
[22] J. Pallavicini, B. Hallsson, and K. Kappel, Polarization in groups of Bayesian agents, Synthese **198**, 1 (2021).
[23] J. T. Kemp, A. G. Kline, and L. MA. Bettencourt, Information synergy maximizes the growth rate of heterogeneous groups, arXiv:2307.01380.
[24] S. A. Wu, R. E. Wang, J. A. Evans, J. B. Tenenbaum, D. C. Parkes, and M. Kleiman-Weiner, Too many cooks: Bayesian inference for coordinating multi-agent collaboration, Topics Cognit. Sci. **13**, 414 (2021).
[25] A. Izquierdo, J. L. Brigman, A. K. Radke, P. H. Rudebeck, and A. Holmes, The neural basis of reversal learning: An updated perspective, Neuroscience **345**, 12 (2017).
[26] A.-M. D'Cruz, M. E. Ragozzino, M. W. Mosconi, M. N. Pavuluri, and J. A. Sweeney, Human reversal learning under conditions of certain versus uncertain outcomes, Neuroimage **56**, 315 (2011).
[27] M. K. Eckstein, S. L. Master, R. E. Dahl, L. Wilbrecht, and A. G. Collins, Reinforcement learning and Bayesian inference provide complementary models for the unique advantage of adolescents in stochastic reversal, Dev. Cognit. Neurosci. **55**, 101106 (2022).
[28] F. Milani, Expectations, learning and macroeconomic persistence, J. Monetary Econ. **54**, 2065 (2007).
[29] M. E. Khan, Y. J. Ko, and M. Seeger, Scalable collaborative Bayesian preference learning, in *Artificial Intelligence and Statistics* (PMLR, Reykjavik, Iceland, 2014), pp. 475–483.
[30] T. Ignatenko, K. Kondrashov, M. Cox, and B. de Vries, On preference learning based on sequential bayesian optimization with pairwise comparison, arXiv:2103.13192.
[31] K. P. Murphy, Conjugate Bayesian analysis of the Gaussian distribution, *def*, Vol. 1, p. 16, 2007.
[32] J. Dalege, D. Borsboom, F. van Harreveld, and H. L. van der Maas, The attitudinal entropy (AE) framework as a general theory of individual attitudes, Psycholog. Inquiry **29**, 175 (2018).
[33] G. O. Roberts, O. Papaspiliopoulos, and P. Dellaportas, Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes, J. Royal Stat. Soc. B **66**, 369 (2004).
[34] Z. Oravecz, F. Tuerlinckx, and J. Vandekerckhove, Bayesian data analysis with the bivariate hierarchical Ornstein-Uhlenbeck process model, Multivariate Behav. Res. **51**, 106 (2016).
[35] R. Singh, D. Ghosh, and R. Adhikari, Fast Bayesian inference of the multivariate Ornstein-Uhlenbeck process, Phys. Rev. E **98**, 012136 (2018).
[36] L. Schenker, Performance games: Simulating the effect of incentives, monitoring, and agent adaptation on institutional data manipulation, Annual Congress of the Swiss Political Science Association, Geneva, Switzerland, 2023.
[37] Through observation, $\tilde{Y}$ is conditioned on $\tilde{X}$, and agent A's past behavior, filtered through B, influences their future behavior. B is similarly conditioned by A's.
[38] S. Chandrasekhar, Stochastic problems in physics and astronomy, Rev. Mod. Phys. **15**, 1 (1943).
[39] C. G. Lord, L. Ross, and M. R. Lepper, Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence, J. Pers. Soc. Psychol. **37**, 2098 (1979).
[40] P. Dandekar, A. Goel, and D. T. Lee, Biased assimilation, homophily, and the dynamics of polarization, Proc. Natl. Acad. Sci. USA **110**, 5791 (2013).
[41] E. Kamenica and M. Gentzkow, Bayesian persuasion, Am. Econ. Rev. **101**, 2590 (2011).
[42] S. Le and R. Boyd, Evolutionary dynamics of the continuous iterated prisoner's dilemma, J. Theor. Biol. **245**, 258 (2007).
[43] S. Barrett, Coordination vs. voluntarism and enforcement in sustaining international environmental cooperation, Proc. Natl. Acad. Sci. USA **113**, 14515 (2016).
[44] A. Kumar, V. Capraro, and M. Perc, The evolution of trust and trustworthiness, J. R. Soc., Interface **17**, 20200491 (2020).
[45] T. Cimpeanu, F. C. Santos, L. M. Pereira, T. Lenaerts, and T. A. Han, Artificial intelligence development races in heterogeneous settings, Sci. Rep. **12**, 1723 (2022).
[46] A. H. Jazwinsky, *Stochastic Processes and Filtering Theory* (Academic Press, New York, 1970).
[47] The Gaussian property is ensured since the drifts are linear and linear transformations of Gaussian restitutes Gaussian.