PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Experiments with sensorimotor games in dynamic human/machine interaction

Benjamin Chasnov, Momona Yamagami, Behnoosh Parsa, Lillian J. Ratliff, Samuel A. Burden

Benjamin Chasnov, Momona Yamagami, Behnoosh Parsa, Lillian J. Ratliff, Samuel A. Burden, "Experiments with sensorimotor games in dynamic human/machine interaction," Proc. SPIE 10982, Micro- and Nanotechnology Sensors, Systems, and Applications XI, 109822A (13 May 2019); doi: 10.1117/12.2519258

SPIE.

Event: SPIE Defense + Commercial Sensing, 2019, Baltimore, Maryland, United States

Invited Paper

Experiments with sensorimotor games in dynamic human/machine interaction

Benjamin Chasnov[e], Momona Yamagami[e], Behnoosh Parsa[m], Lillian J. Ratliff[e], and Samuel A. Burden[e]

[e]Electrical and Computer Engineering, University of Washington, Seattle, WA, USA
[m]Mechanical Engineering, University of Washington, Seattle, WA, USA

ABSTRACT

While interacting with a machine, humans will naturally formulate beliefs about the machine's behavior, and these beliefs will affect the interaction. Since humans and machines have imperfect information about each other and their environment, a natural model for their interaction is a *game*. Such games have been investigated from the perspective of economic game theory, and some results on discrete decision-making have been translated to the neuromechanical setting, but there is little work on continuous sensorimotor games that arise when humans interact in a dynamic closed loop with machines. We study these games both theoretically and experimentally, deriving predictive models for steady-state (i.e. equilibrium) and transient (i.e. learning) behaviors of humans interacting with other agents (humans and machines). Specifically, we consider experiments wherein agents are instructed to control a linear system so as to minimize a given quadratic cost functional, i.e. the agents play a Linear-Quadratic game. Using our recent results on gradient-based learning in continuous games, we derive predictions regarding steady-state and transient play. These predictions are compared with empirical observations of human sensorimotor learning using a teleoperation testbed.

Keywords: sensorimotor learning, control, non-cooperative games, dynamical systems, multi-agent learning

1. INTRODUCTION AND BACKGROUND

We seek to experimentally characterize steady-state and transient behavior as humans learn to control dynamic systems, with an eye toward future applications in human/robot interaction (HRI) and human-cyber-physical systems (HCPS). These applications motivate us to combine principles from theories of decision-making and motor control to study *sensorimotor games*, that is, dynamic games involving agents that use physical bodies to produce control actions. We will briefly review key concepts from these areas in the remainder of this section, then continue in Section 2 with a the problem formulation and theoretical results, followed by simulation and experimental results in Section 3.

1.1 Decision-making and games

When an agent's preferences can be quantified with a single *objective function*, it is natural to model their decision-making in an optimization framework. The optimization framework for decision-making predicts that rational agents will seek *stationary* strategies, that is, they will vary their action until it is no longer possible to decrease their cost.* Under these premises, participating agents play a *game* wherein the conditions for

Further author information: (Send correspondence to S.A.B.)
B.C.: E-mail: bchasnov@uw.edu
M.Y.: E-mail: my13@uw.edu
B.P.: E-mail: behnoosh@uw.edu
L.J.R.: E-mail: ratliffl@uw.edu
S.A.B.: E-mail: sburden@uw.edu, Telephone: 1 206 221 3545

*Here and in what follows, we assume objective functions are *costs* that agents seek to minimize. This choice is without loss of generality since agents that seek to maximize *reward* also seek to minimize the negative of reward.

Micro- and Nanotechnology Sensors, Systems, and Applications XI, edited by Thomas George, M. Saif Islam, Proc. of SPIE Vol. 10982, 109822A · © 2019 SPIE
CCC code: 0277-786X/19/$18 · doi: 10.1117/12.2519258

stationarity depend on the information structure governing agent interactions. When a single rational agent interacts with an environment that has known statistics,

$$\min_u c(u), \tag{1}$$

then *(local) minimizers* are stationary. When multiple rational agents interact with a known environment and are able to cooperate,

$$\min_{u_i, u_{-i}} \{c_i(u_i, u_{-i})\}_{i \in \mathcal{I}}, \tag{2}$$

then *(local) Pareto minimizers* are stationary, that is, collections of decision variables $\{u_i{}^\star\}_{i \in \mathcal{I}}$ which cannot be (locally) modified without increasing one or more agents' cost. When multiple rational agents interact with a known environment and are unable to cooperate,

$$\min_{u_i} c_i(u_i, u_{-i}), \ \forall i \in \mathcal{I}, \tag{3}$$

then *(local) Nash minimizers* are stationary, that is, collections of decision variables $\{u_i{}^\star\}_{i \in \mathcal{I}}$ which cannot be (locally) modified unilaterally without increasing the agent's cost. Although the preceding notation is concise, appropriately choices of the set of actions $u \in \mathcal{U}$ and corresponding cost function c can be defined to encompass *dynamic games*,[1] that is, games wherein costs are related to actions through their effect on the state of a dynamic system,

$$\dot{x} \text{ or } x^+ = F(x, u). \tag{4}$$

1.2 Sensorimotor learning and control

Multiple scientific theories have been proposed to describe and predict how humans learn to control their bodies and interact with their environment. In the present context, we are interested in understanding how agents adapt their actions to guide an external system to follow a specified reference; we will neglect the internal learning and control processes that produce these actions. This shift in focus from the internals of the human body to their *feedback* and *feedforward* actions on an external control system simplifies and unifies predictions from leading theories of human motor control. For instance, pilots control linear vehicle models using linear feedback and feedforward control,[2–4] and this observation is consistent with the theories that humans internalize dynamic models[5–7] or motion primitives,[8,9] as well as the theory that humans employ optimal control.[10,11] Recent work has experimentally corroborated theoretical predictions for the feedforward action;[12–15] we are not aware of work that probes feedback action analogously.

1.3 Notation and mathematical preliminaries

Key notation used in this paper is summarized in Table 1. We work in the framework of dynamic games,[1] and assume dynamics and costs are twice continuously differentiable so that stationary points can be characterized using first- and second-order approximations.[16]

$\mathcal{I} \quad (\lvert\mathcal{I}\rvert = n)$	set of n agents
$x \in \mathcal{X}$	dynamic system state
$u_i \in \mathcal{U}_i$	agent i's action set
$\mathcal{U} = \Pi_{i \in \mathcal{I}} \mathcal{U}_i$	joint action space
$u = (u_i)_{i \in \mathcal{I}} \in \mathcal{U}$	element of joint action space
$c_i : \ \mathcal{X} \times \mathcal{U} \to \mathbb{R}$	agent i's objective function
$G = (F, (c_i)_{i \in \mathcal{I}})$	sensorimotor game
$D_i c_i$	derivative of agent i's cost with respect to its own action
$D_j c_i$	derivative of agent i's cost with respect to agent j's action

Table 1. Notation used in this paper

2. SENSORIMOTOR GAMES

For concreteness, we will focus in what follows on a *trajectory tracking* task wherein agents are instructed to choose control actions that steer the system output y to follow a *reference* r. Multiple such agents $i \in \mathcal{I}$ may be tasked with facilitating trajectory tracking in the same system simultaneously, in which case the dynamics and output are given by

$$\dot{x} \text{ or } x^+ = F(x,u),\ y = H(x,u),\ u = (u_i)_{i\in\mathcal{I}} = \left(u_1, \ldots, u_{|\mathcal{I}|}\right). \tag{5}$$

In practical terms, such a scenario may arise when a human pilots a vehicle or a co-robot assists a human partner.

2.1 General sensorimotor games

When multiple agents interact with the same dynamic system simultaneously, they play a *game* in the sense discussed in Section 1.1. Thus, predicting how agents behave requires consideration of the game's *information structure*, that is, what knowledge each agent has about the system and the other agents, and whether agents have channels of communication other than that provided through interaction with the system. In scenarios where either multiple humans or multiple autonomous agents interact, the information structure may be complex: if humans are permitted to communicate through language (verbal, written, or postural), they may be able to collude and coordinate their actions; similarly, if autonomous agents are permitted centralized communication, they could effectively act as a single agent.

Although complex information structures are worthy of study, for simplicity we will restrict our attention to the simplest case involving one human and one autonomous agent that cannot directly communicate except through the effect of their actions on the state of the shared dynamic system. We will say that such games *have no side information*, and specify them by a tuple $G = \left(F, (c_i)_{i\in\mathcal{I}}\right)$ where:

- F specifies a differential or difference equation : $\mathcal{X} \times \mathcal{U} \to T\mathcal{X}$ or difference equation : $\mathcal{X} \times \mathcal{U} \to \mathcal{X}$;
- c_i specifies a cost function $c_i : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$ for each agent $i \in \mathcal{I}$.

This restriction leads us to focus on *Nash equilibria* as the stationarity concept in what follows.

Hypothesis 1. *In sensorimotor games that have no side information, humans learn to play Nash equilibria.*

The determination of Nash equilibria for a given game is a hard problem in general. We will assume the functions that define the game G are twice continuously differentiable so that Nash equilibria can be characterized locally using first- and second-order approximations of cost functions.[16] Thus, we tacitly assume agents have *bounded rationality*[17] in the sense that they utilize local information about the cost landscape to make small adjustments to their actions that are expected to decrease their costs.

2.2 Linear-quadratic sensorimotor games

In this section we restrict our attention to the class of sensorimotor games with linear dynamics,

$$F(x,u) = Ax + \sum_{i\in\mathcal{I}} B_i u_i, \tag{6}$$

and infinite-horizon quadratic costs,

$$c_i(x, u_i, u_{-i}) = \int_0^\infty x(t)^\top Q_i x(t) + \sum_{j\in\mathcal{I}} u_{-i}(t)^\top R_{ij} u_{-i}(t)\, dt. \tag{7}$$

Of course, our actual simulations and experiments will terminate on finite time horizons, so we tacitly assume that games are played sufficiently long that the infinite-horizon game solutions provide useful predictions for agent behaviors.

This special class of dynamic games has been extensively studied, producing a complete characterization of their Nash equilibria that we will leverage in our simulations and experiments. For instance, it is known that the

stationary cost-minimizing policies are linear state feedback, that the feedback matrices are determined from the positive-definite matrices that define the quadratic minimal cost-to-go for each agent, and that these cost-to-go matrices satisfy coupled Riccati equations. We refer the interested reader to [1, Sec. 6.2.3, 6.5.3] for details in the n-player case, and summarize these results in the 2-player case in the following example.

Example 1. *For concreteness, consider the 2-player, discrete-time game, so that the dynamics are given by*

$$x^+ = Ax + B_1u_1 + B_2u_2,$$

and the costs are given by

$$c_1(u_1, u_2) = \sum_{t=0}^{T} x(t)^\top Q_1 x(t) + u_1(t)^\top R_{11} u_1(t) + u_2(t)^\top R_{12} u_2(t)$$

and

$$c_2(u_1, u_2) = \sum_{t=0}^{T} x(t)^\top Q_2 x(t) + u_1(t)^\top R_{21} u_1(t) + u_2(t)^\top R_{22} u_2(t).$$

Observe that in general $Q_1 \neq Q_2$, i.e. the agents do not need to weight tracking error equally. Note also that although R_{11}, R_{22} must be positive-definite, there are no such restrictions on R_{12}, R_{21} (indeed, these matrices may be nonsquare if the agents' actions differ in dimension). If $Q_1 = Q_2$ and $R_{ij} = 0$, then we recover the single agent cooperative LQR solution. Indeed, the linear state feedback $u_i = -K_i x_i$ yields closed-loop dynamics

$$x(t+1) = (A - B_1K_1 - B_2K_2)x(t) = \mathbf{A}x(t) = \mathbf{A}^t x(0) \tag{8}$$

and corresponding costs for the agent 1,

$$\begin{aligned} c_1(x(0), K_1, K_2) &= \sum_{t=0}^{\infty} x(t)^\top \left(Q_1 + K_1^\top R_{11} K_1 + K_2^\top R_{12} K_2\right) x(t) \\ &= \sum_{t=0}^{\infty} x(0)(A - B_1K_1 - B_2K_2)^{t\top} \left(Q_1 + K_1^\top R_{11} K_1 + K_2^\top R_{12} K_2\right) (A - B_1K_1 - B_2K_2)^t x(0) \\ &= \sum_{t=0}^{\infty} x(0)(\mathbf{A}^t)^\top \mathbf{Q}_{\mathbf{A}}^t x(0), \end{aligned}$$

and similar for agent 2.To determine the optimal (Nash) feedback gains for these non-cooperative agents, we compute each agent's unique positive definite cost-to-go matrix P_i via coupled Riccati equations and subsequently compute each agent's stationary feedback matrix $(K_i^\star,\ K_{-i}^\star)$. For the infinite time horizon case, we have

$$P_{i,k-1} = \mathbf{A}^\top P_{i,k} \mathbf{A} + K_1^\top R_{i1} K_i + K_2^\top R_{i2} K_2 + Q_i$$

and

$$K_{i,k-1} = (R_{ii} + B_i^\top P_{i,k} B_i)^{-1} B_i^\top P_i (\mathbf{A} + B_i K_{i,k})$$

for each $i \in \mathcal{I}$ for k backwards in time until convergence.

2.3 Predictions for steady-state and transient play

As indicated by the theoretical results summarized in the preceding sections, the behaviors exhibited when multiple agents seek to minimize their individual costs by controlling a shared dynamic system can be complex. Even if the game admits only a single stationary joint action, it is not obvious whether or how the agents will learn to play this action. In the remainder of this section we discuss a range of learning processes that may theoretically be observed, and conclude with a discussion of the behaviors we expect human agents to exhibit.

Consider a mathematically simple model of learning where each agent myopically descends the gradient of their own cost with respect to their own action. In the single-agent case, this is simply gradient descent, whose

convergence is well-established [18, Ch. 1]. However, when multiple agents perform this update simultaneously in the same game, counter-intuitive transient behaviors can arise.[19] For instance, agents' costs may in fact increase after each iteration due to the coupling of cost functions. Alternatively, the agents' learning process may converge to a limit cycle[20] or a more complicated attractor,[21] and therefore never approach a stationary strategy. Finally, some games have attractive stationary points that are local maxima for each agent.[22] These observations indicate the complexity of behaviors that can arise in sensorimotor games involving multiple agents.

In our experiments, the human agents are not informed in advance about the dynamic systems they are tasked with controlling – anything they learn about the systems must be obtained through experience. It is possible that the human subjects learn the system's dynamics (indeed, we have previously reported experimental evidence that human subjects learn to invert the system's dynamics[12]); in principle, the human subjects could then choose actions that are stationary for their learned model, and update their actions as they learn a better model. However, it is also possible that the subjects directly adjust their actions to decrease cost without relying on an internal model estimate. Regardless of the strategy actually employed, we hypothesize that human sensorimotor learning can be modeled as a stochastic process that converges in expectation to stationary play.

3. RESULTS

In this section, we provide preliminary results that illustrate steady-state and transient learning in sensorimotor games. We provide simulation results in Section 3.1 that demonstrate complex learning processes, and preliminary experimental results in Section 3.2 where human subjects learn to control simple linear systems.

3.1 Simulation results

We consider synthetic agents that employ a local gradient update of a parameterization of their action strategy in an effort to find Nash minimizers of (3). Specifically, the agents adjust their action strategy using a *policy gradient*[23] learning rule

$$K_i^+ = K_i - \gamma_i D_i c_i(K_i, K_{-i}), \quad \forall i \in \mathcal{I},$$

where γ_i is the ith agent's (possibly variable) learning rate. When learning rates are small, the discrete-time learning process in (3.1) is approximated by the continuous-time vector field $\dot{K} = -\omega(K) = -[D_i c_i(K)]_{i\in\mathcal{I}}$. To analyze the asymptotic behavior of the gradient flow, we compute the Jacobian of this vector field,

$$J(K) = \begin{bmatrix} D_1^2 c_1(K) & \cdots & D_{1n} c_1(K) \\ \vdots & \ddots & \vdots \\ D_{n1} c_n(K) & \dots & D_n^2 c_n(K) \end{bmatrix}.$$

From dynamical systems theory, we know that the discrete time updates will converge locally to a stationary point $K^\star$ where $\omega(K^\star) = 0$ and the eigenvalues of $-\Gamma J$ are negative, where $\Gamma = \mathrm{diag}(\gamma_i)$ is a diagonal matrix with agents' learning rates on its diagonal. The stationary point $K^{\star}*$ will be a local Nash minimizer if the block diagonal terms of the Jacobian $(D_i^2 c_i)$ are positive-definite.

3.1.1 A stable Nash equilibrium becomes unstable

Agents may learn at different rates: some agents may wish to learn quickly to gain an advantage, while others may have a variable learning rate. In this section, we show that variable learning rates can change the stability properties of a Nash equilibrium under the learning rule (3.1).

Equilibria of the learning process (3.1) are invariant under change of learning rates, i.e. the solutions to $\omega(K) = 0$ and $\Gamma\omega = 0$ are the same for any non-zero diagonal matrix γ. However, the eigenstructure of $\Gamma\omega(K) = 0$ need not stay constant. The following example illustrates a counter-intuitive but non-degenerate situation in which changes to one agent's learning rate causes a stable Nash equilibrium to become unstable.

Consider a three-player game where the Jacobian at a fixed point has positive-definite block diagonals and strictly positive eigenvalues. This implies the equilibrium is a Nash equilibrium and that the dynamics $\dot{x} = -\omega(x)$ are stable in a neighborhood around the fixed point under uniform learning rate, i.e. $\Gamma = I$, the identity. Now suppose agent 3 slows down its learning by a factor of 5, from γ_3 to $\gamma_3/5$. We show in what follows that this change can cause the learning dynamics to go unstable.

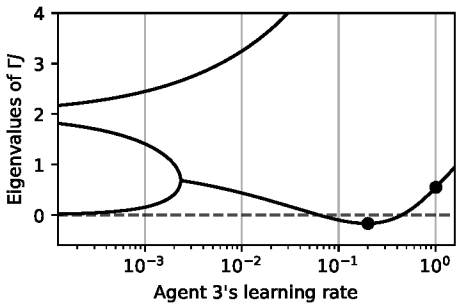

Figure 1. The eigenvalues of the Jacobian of the vector field $\dot{K} = -\Gamma\omega(K)$ as agent 3 varies their learning rate. When the real part of any eigenvalue drops below zero, the stationary Nash equilibrium becomes unstable under the learning process (3.1).

Example 2. *Suppose the Jacobian of the dynamics at this equilibrium is*

$$J = \begin{bmatrix} 2 & 9 & 0 \\ 0 & 2 & 6 \\ 9 & 0 & 12 \end{bmatrix}$$

whose spectrum lies on the right half complex plane with eigenvalues $14.9,\ 0.5 \pm 6.0i$. *However, premultiplying the Jacobian by* $\Gamma = diag(1,\ 1,\ 1/5)$, *the eigenvalues of* ΓJ *become* $6.7,\ -0.2 \pm 4.0i$, *which indicate a saddle point. Interestingly and quite counter-intuitively, the first player's cost (locally) does not depend on the choice variable of the last player: that is,* $J_{1,3} = 0$. *However, under the learning process in (3.1), a decrease in the third players' learning rate will destabilize the first player's learning process. This phenomenon is illustrated in Figure 1 as a bifurcation diagram where we adjust* γ_3 *while keeping all other agents' learning rates constant.*

3.1.2 Stationary points of LQR games under policy gradient learning updates

Motivated by the preceding example, we find it important to study the stationary points of a game and whether the stationary points are stable under the learning process in (3.1). To visualize the game's vector field $\dot{K} = -\omega(K)$, we construct a simple scalar LQR game with fixed constants $A = B_i = 1$ and agent control cost structure

$$\begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = \begin{bmatrix} 5 & -5 \\ -5 & 4 \end{bmatrix}.$$

We plot three stationary points of the game in Figure 2 with K_1 and K_2 on the x and y axes. In this setup, the only stable stationary point is the Nash equilibrium. This is not true in general, as some learning processes produce stable stationary points that are not Nash minimizers.

3.2 Experimental results

We developed a one-dimensional trajectory tracking task to quantify an agent's ability to learn to control a dynamic system. In this task, a single agent operates a slider joystick to provide a scalar input u to a linear system. The system is a second-order integrator

$$\ddot{y} = u + d$$

with disturbance d and an internal state $x \in \mathbb{R}^2$ as velocity and position of the system. The position and reference signals are displayed on a screen with some preview (in the case of the reference) and time history (for both position and reference), and subjects are instructed to minimize reference tracking error; we refer the reader to[12] for specifics of the experiment.

Assuming the subjects have a perfect model of the feedforward component of the controller, our investigation of this paper is to determine how subjects learn to perform feedback to correct for the disturbances in the system. In other words, we assume agents control the system by $u = Lr + K(r - x)$, where L is the optimal feedforward gain and K may vary over each trial as the subject improves on the task. We solve for K as a least squares problem $\min_K \sum_{t=0}^{T} \|Lr_t + K(r_t - x_t) - u_t\|_2^2$ for a trial of T sample points. The solution of this least squares problem is an estimate of the subject's feedback gain averaged over a single trial. We record the subject's performance over 30 trials and plot the resulting Ks for each of 10 subjects and each trial in Figure 3.

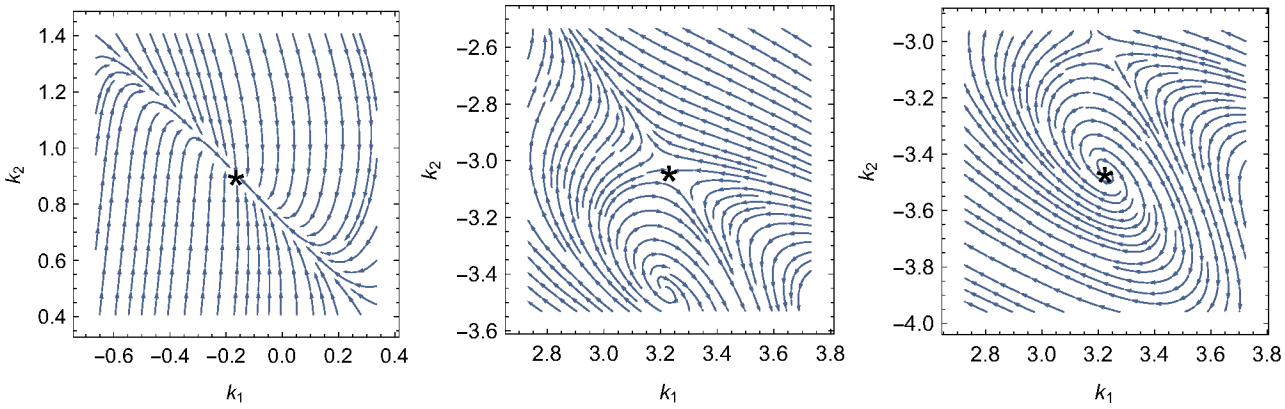

Figure 2. Several types of equilibria arise in a linear-quadratic game under policy gradient learning dynamics: a stable equilibrium, a saddle point, and an unstable equilibrium. In this example, the stable equilibrium is a Nash equilibrium of the game, but this is not guaranteed to be the cased in general.

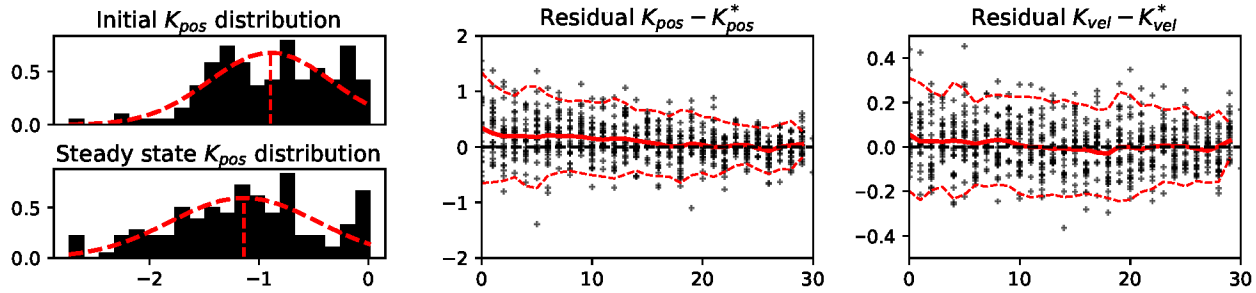

Figure 3. Learning the feedback gain of a second-order system trajectory tracking task. The gain $K \in \mathbb{R}^2$ consists of a position and velocity gain. Left: Comparison of the distribution of position gains from initial trials to final trials. The mean of the "steady-state" position gains become more negative over time. Middle and right: Understanding this phenomenon by plotting the residual $K - K^\star$ over the 30 trials, with 2 sigma intervals plotted as dashed lines. We observe a slight decrease in variance over the trials as well as a slight decrease in the residuals of the position gain.

We observe some general trends in the results of our experiments. Firstly, subjects' feedback gain seem to converge to a stationary distribution as shown in Figure 3 (left). However, this distribution has large variance which possibly indicates variable subject-specific skill levels. Secondly, there is some indication of transient learning dynamics of the K gains over the trials, specifically for the position gain K_{pos}, as illustrated in Figure 3 (middle). Further investigation is required to determine the statistical significance of these preliminary findings. These insights will guide our experiment design for future work in trajectory tracking task for multiple agents.

4. DISCUSSION

Building on established results from decision theory and sensorimotor control, we formulated predictions for steady-state and transient behaviors of humans learning to play linear-quadratic dynamic games. These predictions were investigated using simulations involving multiple synthetic agents and experiments involving a single human agent. The simulation results demonstrated that simple mechanisms can lead to complex learning processes when multiple agents learn to control a shared dynamic system. The preliminary experimental results indicated that humans modulate their feedback over time to converge on stationary play. We intend to continue this line of inquiry in future work by more fully characterizing how humans learn to play sensorimotor games, and how their play is affected by the behavior of other agents, the system being controlled, and the sensory and motor modalities employed in the sensorimotor loop.

ACKNOWLEDGMENTS

This work was supported in part by Award #1836819 from the Cyber-Physical Systems (CPS) Program in the the National Science Foundation (NSF) Directorate for Computer & Information Science & Engineering (CISE), by the Center for Amplifying Motion and Performance (AMP Center) Strategic Research Initiative (SRI) in the University of Washingtons College of Engineering (UW-CoE), the Washington Research Foundation Funds for Innovation in Neuroengineering, and by the Computational Neuroscience Graduate Training Program at the University of Washington. The human subjects data reported herein were collected with the approval of the University of Washington Human Subjects Division (UW-HSD) under Study #909.

REFERENCES

[1] Başar, T. and Olsder, G. J., [*Dynamic Noncooperative Game Theory*], Society for Industrial and Applied Mathematics (1998).

[2] Allen, R. W. and McRuer, D., "The man/machine control interface—pursuit control," *Automatica: the journal of IFAC, the International Federation of Automatic Control* **15**(6), 683–686 (1979).

[3] McRuer, D. T. and Jex, H. R., "A review of Quasi–Linear pilot models," *IEEE Transactions on Human Factors in Electronics* **HFE–8**(3), 231–249 (1967).

[4] McRuer, D. T. and Krendel, E. S., "The human operator as a servo system element," *Journal of the Franklin Institute* **267**(5), 381–403 (1959).

[5] Wolpert, D. M., Ghahramani, Z., and Jordan, M. I., "An internal model for sensorimotor integration," *Science* **269**(5232), 1880–1882 (1995).

[6] Kawato, M., "Internal models for motor control and trajectory planning," *Current opinion in neurobiology* **9**(6), 718–727 (1999).

[7] Bhushan, N. and Shadmehr, R., "Computational nature of human adaptive control during learning of reaching movements in force fields," *Biological cybernetics* **81**, 39–60 (1999).

[8] Hogan, N. and Sternad, D., "Dynamic primitives of motor behavior," *Biological cybernetics* **106**, 727–739 (Dec. 2012).

[9] Schaal, S., Ijspeert, A., and Billard, A., "Computational approaches to motor learning by imitation," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **358**(1431), 537–547 (2003).

[10] Todorov, E. and Jordan, M. I., "Optimal feedback control as a theory of motor coordination," *Nature neuroscience* **5**(11), 1226–1235 (2002).

[11] Diedrichsen, J., Shadmehr, R., and Ivry, R. B., "The coordination of movement: optimal feedback control and beyond," *Trends in cognitive sciences* **14**(1), 31–39 (2010).

[12] Yamagami, M., Howell, D. B., Roth, E., and Burden, S. A., "Contributions of feedforward and feedback control in a manual trajectory–tracking task," in [*Cyber–Physical–Human Systems (CPHS)*], (2018).

[13] Yu, B., Gillespie, R. B., Freudenberg, J. S., and Cook, J. A., "Human control strategies in pursuit tracking with a disturbance input," in [*Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*], 3795–3800, ieeexplore.ieee.org (2014).

[14] Drop, F. M., Pool, D. M., van Paassen, M. R. M., Mulder, M., and Bulthoff, H. H., "Objective model selection for identifying the human feedforward response in manual control," *IEEE transactions on cybernetics* **48**, 2–15 (Jan. 2018).

[15] Zhang, X., Wang, S., Hoagg, J. B., and Seigler, T. M., "The roles of feedback and feedforward as humans learn to control unknown dynamic systems," *IEEE transactions on cybernetics* **48**, 543–555 (Feb. 2018).

[16] Ratliff, L. J., Burden, S. A., and Sastry, S. S., "On the characterization of local nash equilibria in continuous games," *IEEE transactions on automatic control* **61**, 2301–2307 (Aug. 2016).

[17] Simon, H. A., [*Models of bounded rationality: Empirically grounded economic reason*], vol. 3, MIT press (1997).

[18] Bertsekas, D. P., [*Nonlinear Programming*], Athena Scientific, 2nd ed. (1999).

[19] Hart, S. and Mas-Colell, A., "Uncoupled dynamics do not lead to nash equilibrium," *American Economic Review* **93**(5), 1830–1836 (2003).

[20] Mertikopoulos, P., Papadimitriou, C., and Piliouras, G., "Cycles in adversarial regularized learning," in [*Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*], 2703–2717, SIAM (2018).

[21] Sato, Y., Akiyama, E., and Farmer, J. D., "Chaos in learning a simple two-person game," *Proceedings of the National Academy of Sciences* **99**(7), 4748–4751 (2002).

[22] Mazumdar, E. and Ratliff, L. J., "On the convergence of competitive, multi-agent gradient-based learning," *arXiv preprint arXiv:1804.05464* (2018).

[23] Fazel, M., Ge, R., Kakade, S., and Mesbahi, M., "Global convergence of policy gradient methods for the linear quadratic regulator," in [*Proceedings of the 35th International Conference on Machine Learning*], Dy, J. and Krause, A., eds., *Proceedings of Machine Learning Research* **80**, 1467–1476, PMLR, Stockholmsmssan, Stockholm Sweden (10–15 Jul 2018).