Reciprocity and Retaliation in Social Games With Adaptive Agents

Derrik E. Asher, Andrew Zaldivar, Brian Barton, Alyssa A. Brewer, and Jeffrey L. Krichmar

***Abstract*—Game theory has been useful for understanding risk-taking and cooperative behavior. However, in studies of the neural basis of decision-making during games of conflict, subjects typically play against opponents with predetermined strategies. The present study introduces a neurobiologically plausible model of action selection and neuromodulation, which adapts to its opponent's strategy and environmental conditions. The model is based on the assumption that dopaminergic and serotonergic systems track expected rewards and costs, respectively. The model controlled both simulated and robotic agents playing Hawk–Dove and Chicken games against subjects. When playing against an aggressive version of the model, there was a significant shift in the subjects' strategy from Win-Stay-Lose-Shift to Tit-For-Tat. Subjects became retaliatory when confronted with agents that tended towards risky behavior. These results highlight the important interactions between subjects and agents utilizing adaptive behavior. Moreover, they reveal neuromodulatory mechanisms that give rise to cooperative and competitive behaviors.**

***Index Terms*—Adaptive systems, cognition, cognitive robotics, human robot interaction, neurotransmitters.**

I. Introduction

ELUCIDATING the neurobiological basis for decision-making under competitive and conflicting situations is an important step towards understanding reciprocity, social cognition, cooperation, and competition [1], [2]. Game theory has been successful in describing such social behaviors [3]–[5] and has been applied to the investigation of their neural bases [1], [6]–[8].

The vertebrate neuromodulatory systems play a key role in the regulation of social behavior [9], [10]. In particular, the serotonergic neuromodulatory system is involved in a wide variety of emotional, cognitive, and social responses [11]. The raphe nucleus, which is the source of serotonin in the central nervous system (CNS), may underlie cognitive control of stress, social interactions, and risk-taking behavior [12].

Of particular interest to the present work are studies in which levels of neuromodulators are depleted or altered while subjects play cooperative and competitive games. In one such study, subjects, who were serotonin-depleted through dietary changes, cooperated less than control subjects in an iterated version of Prisoner's Dilemma against a computer-generated Tit-For-Tat opponent [13]. Another study showed similar results with a "one-shot" version of the Ultimatum Game, where serotonin-depleted subjects tended to reject monetary offers more than control subjects when they deemed these offers to be unfair [14]. Conversely, increasing serotonin levels through the drug citalopram, a selective serotonin reuptake inhibitor, resulted in a decrease in the rejection of such offers [15]. In the same study, citalopram was shown to increase subjects' aversion to personally harming others when faced with moral dilemmas.

It would be illuminating, however, to advance these "one-shot" games by examining human subject responses against opponents in iterative games, during which their opponents can adapt their strategies to aspects of the game, environment, and subject responses. In addition, subjects' decisions in games against physical agents that are embedded in the real world may be quite different from decisions in games against a computer simulation. Indeed, such embodied models have been shown to elicit strong reactions in humans [16], [17] and to exhibit more natural and complex behavior than pure simulations [18], [19]. The goal of the present study is to explore these issues by investigating human subject interactions with artificial neural agents, both embodied and simulated, that can adapt their behavior to their opponent's strategy and to environmental change.

Previously, we developed a computational model of neuromodulation and action selection based on the assumptions that dopamine levels are related to the expected reward of an action, and serotonin levels are related to the expected cost of an action [20], [21]. That is, serotonergic activity might track the expected cost of an action in a similar way that dopaminergic activity is thought to track the expected reward of an action [22], [23]. The neural agent adapted its behavior appropriately to changes in environmental conditions and to changes in its opponent's strategy [20], [21]. The present study embedded our neural model of neuromodulation and action selection in both simulated and embodied neural agents to investigate reciprocal social interactions in games of cooperation and conflict with human subjects. Subjects played a series of competitive games (i.e., Hawk–Dove and Chicken) with tradeoffs between high risk, high payoff actions and low risk, low payoff actions, against robotic and simulated

Manuscript received December 09, 2011; revised March 26, 2012; accepted May 27, 2012. Date of publication June 05, 2012; date of current version September 07, 2012. This work was supported by the National Science Foundation (EMT/BSSE Award 0829752), the Office of Naval Research (Award N000140910036), and the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH) and the NIH Roadmap for Medical Research, under Grant UL1 RR031985.

D. E. Asher, A. Zaldivar, B. Barton, and A. A. Brewer are with the Cognitive Sciences Department, University of California, Irvine, CA 92697-5100 USA (e-mail: dasher@uci.edu; azaldiva@uci.edu; bbarton@uci.edu; aabrewer@uci.edu).

J. L. Krichmar is with the Cognitive Sciences and Computer Science Departments, University of California, Irvine, CA 92697-5100 USA (e-mail: jkrichma@uci.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAMD.2012.2202658

1943-0604/$31.00 © 2012 IEEE

agents. The effects of serotonergic levels on adaptive behavior in these games were further tested by: 1) simulating serotonergic lesions in the neural agent; and/or 2) lowering the CNS serotonin levels of human subjects through a dietary manipulation called acute tryptophan depletion (ATD), which has been shown to decrease cooperation and lower harm-aversion [13], [14], [24].

II. Methods

A. Subjects

Eight subjects (three female; mean age: 26.6 ± 3.8 years) participated in this study. The Institutional Review Board at University of California, Irvine, approved the experimental protocol and informed consent was obtained from all subjects.

Prior to enrollment in the study, all potential participants were screened for psychiatric and neurological disorders using the Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID-I Research Version, Biometrics Research; [25]). Potential participants were excluded for a history of cardiac, hepatic, renal, pulmonary, neurological, psychiatric or gastrointestinal disorders, pregnancy, psychiatric medication, drug use, or a personal or family history of mood disorders. Because serotonin levels can be affected by estrogen fluctuations, female subjects participated in the study only during the first two weeks of their menstrual cycles [26]–[28].

B. Acute Tryptophan Depletion

Several studies of social behavior have used a dietary manipulation, called the acute tryptophan depletion procedure (ATD), to investigate the short-term effects of a decline in serotonin levels on mood in humans [13], [24]. The goal of ATD is to temporarily alter the levels of serotonin in the brain via a decrease in blood plasma tryptophan, the amino acid precursor to serotonin. Because free blood plasma tryptophan levels, and the corresponding serotonin levels in the brain, vary with the amount of dietary tryptophan and the rate of protein synthesis, these levels can be altered by a low protein diet in combination with a specially prepared "protein shake." This "protein shake" contains an amino acid load (lacking tryptophan), which has two effects. First, it stimulates protein synthesis in the liver, which uses up blood plasma tryptophan. Second, the amino acids that are given in the "protein shake" compete with tryptophan for transport across the blood-brain barrier, which restricts entry of tryptophan into the brain and leads to lower levels of serotonin in the brain [29], [30]; for reviews see [31] and [32].

The ATD "protein shake" (our Trp- mixture) contains 15 amino acids (NutraBio, www.nutrabio.com) mixed with ∼400 ml water and flavoring (Crystal Light, Kraft Foods, Inc.). The amino acids are used in proportions approximating human milk, except for three amino acids (see Table I). The ATD Trp- mixture lacks tryptophan, the amino acid under study, as well as aspartic acid and glutamic acid, which are omitted because of concern about their toxicity at high doses [13], [24]. Our Trp- mixture had a total protein content of ∼100 g. Our control mixture (Trp+ condition) used the same ratio of amino acids, but additionally included 2.3 g of tryptophan. For female participants, the same ratios of amino acids were used, but with a ∼17% reduction in quantity to take into account average lower body weight [24].

TABLE I
"Protein Shake" Amino Acid Levels for the Acute Tryptophan Depletion (ATD) Procedure (Males)

Amino Acid	Control/Depleted Mixture
L-alanine	5.5g
L-arginine	4.9g
L-cystine	2.7g
glycine	3.2g
histidine	3.2g
L-isoleucine	8.0g
L-leucine	13.5g
L-lysine monohydrochloride	11.0g
L-methionine	3.0g
L-proline	12.2g
L-phenylalanine	5.7g
L-serine	6.9g
L-threonine	6.5g
L-tyrosine	6.9g
L-valine	8.9g
L-tryptophan	**2.3g/0.0g**
Total	**104.4g/102.1g**

Subjects followed a low protein (<20 g) diet for 24 h before each experimental day. Additionally, they fasted 12 h prior to the administration of the amino acid drink and followed a low-protein diet (<5 g) throughout each experimental day.

C. General Procedures

In a double-blind study, human subjects were randomly assigned on the first experimental day to receive either the Tryp+ control mixture or the Tryp- mixture. Each subject then returned to participate in the other condition at least seven days later to ensure the return to baseline blood plasma tryptophan levels between experimental days. On the morning of each experimental day, a blood sample was drawn to determine baseline blood plasma tryptophan levels. Following the blood draw, subjects ingested one of the amino acid drinks (either Tryp+ or Tryp-). A second blood sample was drawn approximately five hours after ingestion of the amino acid drink to confirm reduction (Trp- condition) or maintenance (Trp+ condition) of blood plasma tryptophan levels. Roughly 5.5 hours after consumption of the amino acid drink, human subjects then participated in a series of Hawk–Dove and Chicken games against a neural agent.

In order to: 1) track potential short-term mood effects of ATD; and 2) ensure that ATD had no long term effects on subjects' mood, the positive and negative affect scale (PANAS) was administered on two occasions during each experimental day [33], once before the amino acid drink was consumed (i.e., baseline) and once just prior the start of the interactive games (i.e., 5.5 h after consumption of the amino acid drink). A follow-up PANAS assessment was also performed at least seven days after the experiments and compared with baseline measurements.

D. Hawk–Dove Game

The Hawk–Dove game consisted of a human and a neural agent choosing a single action in response to a territory of interest (TOI). The Hawk–Dove game, which is similar to Pris-

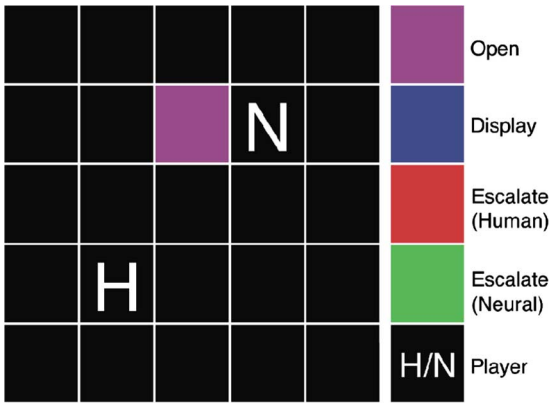

Fig. 1. Schematic representation of the Hawk–Dove game. The objective is for the human subject (H) and the neural agent (N) to reach the territory of interest (TOI). The TOI is denoted with magenta when it is an *Open* resource. Upon reaching the TOI, players decide whether to *Display*, a cooperative action to share the resource, which changes the TOI to blue, or to *Escalate*, a confrontational action to solely obtain the resource, which changes the TOI to green if chosen by the agent or red if chosen by the subject.

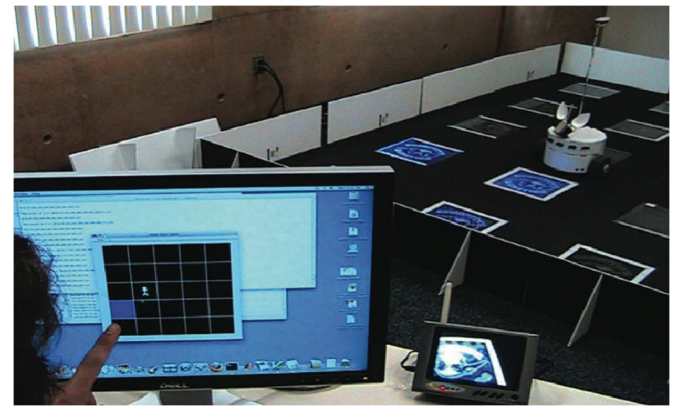

Fig. 2. Experimental apparatus for playing the Hawk–Dove game. A subject is playing the Hawk–Dove game with the CARL robot [34]. The GUI viewed by the subject reflects the state of the interactive floor and allows the subject to move the icon and to change the TOI state. A neural agent guides the robot's behavior. Note that the TOI on the GUI and CARL's interactive floor are in the same location with the same color.

TABLE II
PAYOFF MATRIX FOR HAWK–DOVE GAME BETWEEN SUBJECTS (H) AND NEURAL AGENTS (N)

	N. Escalate	**N. Display**
H. Escalate	H: (V-D)/2, N: (V-D)/2	H: V, N: 0
H. Display	H: 0, N: V	H: V/2, N: V/2

V is the value of the resource and is set to 0.60. D is the damage incurred when both players choose to *Escalate*. D is set to 1.60 for serious injury and 0.62 for low injury. The probability of a serious injury is 0.25 or 0.75.

oner's Dilemma, was chosen because it is amenable to a physical instantiation with a robot. Moreover, it has an additional strategic element since choices are different depending on who arrives at the TOI first. The agent that arrives at the TOI first can signal its intention to compete or cooperate. However, the agent that arrives second to the TOI has the advantage of seeing its opponent's move and responding appropriately.

At the start of the game, the TOI and the human subject's location were randomly placed on a 5×5 playing grid (see Fig. 1). If the neural agent was simulated, its location was randomly chosen as well. Otherwise, the current location of the robot was used as a starting position. The player that arrived at the neutral TOI first had the opportunity to take one of two possible actions: *Escalate* (i.e., an aggressive, confrontational tactic) or *Display* (i.e., a nonviolent, cooperative tactic). The player that arrived second responded with one of the two aforementioned actions. After each game, a payoff was calculated based on the matrix given in Table II. If both players chose *Escalate*, they received a penalty that was either a serious injury (large penalty) or just a scratch (small penalty). The probability of serious injury was set to 0.25 or 0.75 at the start of the game. If both players chose *Display*, they split the value of the TOI resource. If one player chose *Escalate* and the other chose *Display*, the player that chose *Escalate* received the entire value of the resource.

In order to carry out the embodied version of the Hawk–Dove game, we modified the CARL robot and its interactive floor apparatus, which was originally developed for a conditioning paradigm [34]. The robot consisted of a two wheeled mobile base equipped with IR sensors for obstacle avoidance, compass for orienting and navigation, a WiFi device server (http://www.sena.com) for communication between the robot and a computer workstation, and a CCD video camera with a RF transmitter for vision. The model for the neural agent ran on a computer workstation, which received sensor input through RF and RS-232 communication and sent motor commands to the CARL robot through RS-232 communication. The pan and tilt position of the camera was controlled by commands to a pair of servomotors. The base of the robot was 10 in in diameter and 8.5 in high. Visual processing was carried out on the workstation using open source Computer Vision (OpenCV) libraries (http://opencv.willowgarage.com/wiki/). A color histogram method was run across image frames to classify salient features [34].

The robot's environment consisted of a 10-foot by 10-foot enclosure that contained 25 light panels arranged in a 5-by-5 grid (see Fig. 2). The panel color was set to magenta, red, blue or green through RS-232 communication from the workstation to electronics controlling the panels. All 25 panels had IR transceivers that could communicate position information to the robot when it was directly above the panel. Robot navigation was achieved by combining heading information with visual tracking.

The game proceeded with the human subject and neural agent approaching the TOI and then upon arriving at the TOI, making a decision to choose *Escalate* or *Display*. A human subject sat at a computer workstation with a visual representation that reflected the state of the interactive floor (see Fig. 2). At the start of each game, the TOI was set to the *Open* state by displaying one panel on the human subject's user interface as magenta, and setting the corresponding four panels on CARL's interactive floor to magenta. After the TOI was presented to both players, the human subject and robot moved toward the *Open* resource. The human moved his or her icon by clicking on one adjacent panel at a time using a mouse. A ten-second delay between moves was used to prevent the human from moving toward the TOI faster than the robot. If the human subject's icon was adjacent to the TOI, the human was allowed to make a decision. The

human could *Display* by turning the panel to blue, or *Escalate* by turning the panel to Red. When the robot was at the TOI location, it would make a decision by visually recognizing the light panel's color (i.e., magenta for *Open*, red for *Escalate*, and blue for *Display*). The robot chose *Display* by turning the panel to blue, or *Escalate* by turning the panel to green. The change of state was reflected on the human subject's interactive screen.

We also developed a simulated variant of the Hawk–Dove game in which subjects used the same interactive screen, but instead of a real robot, there was a robot icon on the screen. The same neural model used for the robot dictated the control of the robot's icon and its decision-making. This simulated setup allowed us to judge whether playing against a robot had an effect on human behavior.

In the Hawk–Dove game there were three experimental conditions per subject per experimental day (i.e., Trp- and Trp+): 1) simulation versus Robot—subjects would play games against a computer agent or against the robot; 2) control versus Raphe—subjects would play games against a neural agent with an intact simulated neuromodulatory system (Control), or against a neural agent with a simulated lesion of its serotonergic system (Raphe); and 3) friendly versus Harsh environment—in the friendly environment, the chance of a serious injury was 25%, and in the harsh environment, the chance of a serious injury was 75%. The neural agents underwent the same three experimental conditions except that the neural agent's Control and Raphe corresponded to the human subjects' Tryp- and Tryp+ days. Subjects played 20 games of Hawk–Dove per condition. The average time of completion for all the conditions of the Hawk–Dove paradigm was roughly 60 min.

E. Chicken Game

In the chicken game, two cars approach each other on a collision course, and players must decide to drive *Straight* for a high-risk, high reward or *Swerve* away. While somewhat similar to Hawk–Dove and Prisoner's Dilemma, the Chicken game forces players to decide on an action quickly without knowledge of the opponent's choice. Therefore, the players must rely solely on prior game experience when making their decisions.

In our instantiation of the Chicken game, the human subject and neural agent drove their cars along a single lane from opposite directions (see Figs. 3 and 4). Both players started simultaneously at the same speed. The human subject had less than one second to decide to *Swerve* or to continue *Straight* and risk a crash. After each game, a payoff was calculated based on the outcome of the game (see Table III). If both players drove *Straight*, the result was a head-on collision with a heavy penalty. If one player *Swerved* (thereby deemed the "chicken"), then the player that continued moving *Straight* on the lane received a high payoff. If both players *Swerved*, both players received a small payoff.

In the embodied version of the Chicken game, human subjects and the neural agent controlled racecars from a digital slot car racing set (see Fig. 4). We modified the Carrera Digital 1/24 23602 Classic Legends racing car set (http://us.carrera-toys.com) such that two cars moved in opposite directions toward each other. Slot cars were placed on opposite ends of

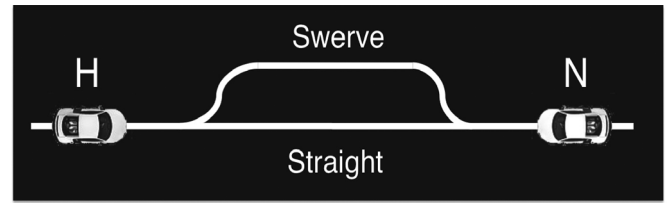

Fig. 3. Schematic representation of the Chicken game. In this game, two racecars approach each other on a collision course. The human subject (H) controlled one car and the neural agent (N) controlled the other. Players decide whether to *Swerve* onto another lane or stay *Straight*.

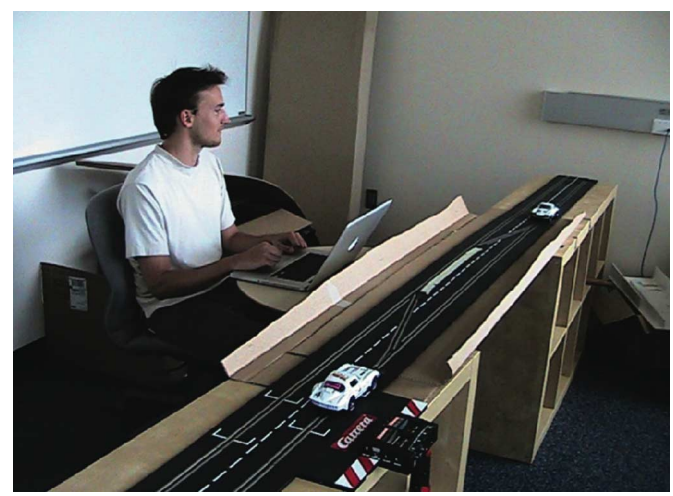

Fig. 4. Experimental apparatus for the Chicken game. Subjects have control of one racecar, and the agent has control of the other. The subject may choose to *Swerve* by clicking on a button shown on the interactive screen.

TABLE III
PAYOFF MATRIX FOR THE CHICKEN GAME BETWEEN HUMAN SUBJECTS (H) AND NEURAL AGENTS (N)

	N. Straight	**N. Swerve**
H. Straight	H: -4, N: -4	H: 3, N: 0
H. Swerve	H: 0, N: 3	H: 1, N: 1

a $12'6''$ long straightaway. The control of both cars was handled through RS-232 serial communication from a computer. The racecar controller consisted of a peripheral interface controller (PIC) microcontroller with two digital potentiometers used to control the speed of the cars, and a serial line level converter necessary for serial communication between the software and racing car set. The speed was fixed to insure that both cars reached the *Swerve/Straight* decision point at the same time. We developed a graphical user interface to allow the human subject's car to switch lanes with a mouse click. The output of the model for the neural agent dictated whether the other car would *Swerve* or not. If both cars chose *Straight*, they crashed in the middle of the track. Rubber foam bumpers were placed on the racecars to prevent damage from collisions. If both cars chose *Swerve*, they both switched lanes and stopped before hitting each other. In the case where one car chose *Swerve* and the other car chose *Straight*, the car that chose *Straight* traveled down the track, and the car that chose *Swerve* switched to the other lane and stopped.

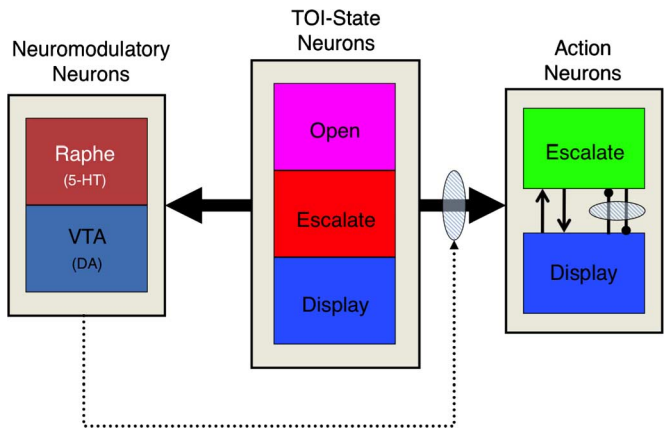

Fig. 5A. Neural network architectures and initial connection weights for the neural agents. The thick arrows represent plastic pathways. The dotted arrows and shaded ovals represent neuromodulatory pathways. Within the Action area, neurons with excitatory reciprocal connections are denoted with the lines with an arrow at the end, and neurons with reciprocal inhibitory connections are denoted with lines with a dot at the end. (a). Architecture for the Hawk–Dove neural model. The solid arrows extending from the TOI state neurons represent all-to-all connections.

From	To	Initial Weight	Plastic	Phasic Neuromodulation
TOI-State	*Action*	0.1	Y	Y
TOI-State	*Neuromodulatory*	0.1	Y	N
Action-Escalate	*Action-Display*	0.1	N	N
Action-Escalate	*Action-Display*	-0.1	N	Y
Action-Display	*Action-Escalate*	0.1	N	N
Action-Display	*Action-Escalate*	-0.1	N	Y

Fig. 5B. Synaptic connections between neural areas in the Hawk–Dove neural network that controlled the neural agent.

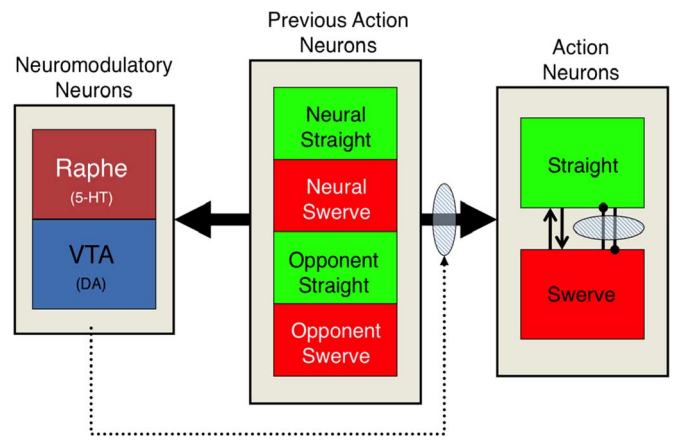

Fig. 5C. Architecture for the Chicken game neural model. The solid arrows extending from the Previous Action neurons represent all-to-all connections.

From	To	Initial	Plastic	Phasic
Previous Action	*Action*	0.1	Y	Y
Previous Action	*Neuromodulatory*	0.1	Y	N
Action-Straight	*Action-Swerve*	0.1	N	N
Action-Straight	*Action-Swerve*	-0.1	N	Y
Action-Swerve	*Action-Straight*	0.1	N	N
Action-Swerve	*Action-Straight*	-0.1	N	Y

Fig. 5D. Synaptic connections between neural areas in the Chicken neural network that controlled the neural agent.

Similar to the Hawk–Dove paradigm, human subjects played a pure simulation version of this game. The interactive representation was identical to controlling the racecars. Instead of seeing the cars move on the track, the outcome of the game was shown on the subject's computer screen.

For the Chicken game there were two experimental conditions per subject per experimental day: 1) simulation versus robot; and 2) control versus raphe. Similar to the Hawk–Dove game, the neural agents played against subjects in the Tryp- and Tryp+ states. Human subjects played 20 games of Chicken per condition. The average time of completion for all the conditions of the Chicken paradigm was roughly 40 min.

F. Neural Agent

Throughout these games, a neural network controlled the behavior of the neural agent. The neural agent was rigorously tested in a previous modeling study against simple opponents with fixed strategies [20], [21]. The focus of this study was to move past using simple opponents with fixed strategies and introduce adaptive neural agents. For both games, the neural network had three neural areas (see Fig. 5): 1) TOI-State for Hawk–Dove, and Previous Action for Chicken; 2) action; and 3) neuromodulatory. For the Hawk–Dove games, the TOI-State included three neurons that corresponded to the possible states the neural agent may observe [see Fig. 5(a)]: 1) *open*—the neural agent reached the TOI first; 2) *escalate*—the human player reached the TOI first and chose to create a conflict; or 3) *display*—the human player reached the TOI first but did not start a conflict. At the start of each condition, the weights of the neural network were set to their initial values [see Fig. 5(b)]. For the Chicken games, the Previous Action area featured four neurons that provided information on possible outcomes performed in the prior game for both itself and the human subject [see Fig. 5(c)]: 1) neural *straight*; 2) neural *swerve*; 3) opponent *straight*; or 4) opponent *swerve*. Congruent with the Hawk–Dove games, at the start of each condition, the weights of the neural network were set to their initial values [see Fig. 5(d)].

In both games, the equations for the activity of each of these neurons $(\mathrm{n_i})$ were computed similarly. In Hawk–Dove, the activity of TOI-State neurons was based on the current state of the TOI

$$n_i = \begin{cases} 0.75 + \mathrm{rnd}(0.0, 0.25); & i = \mathrm{TOIState} \\ \mathrm{rnd}(0.0, 0.25); & \mathrm{Otherwise} \end{cases} \quad (1)$$

where rnd (0.0, 0.25) was a random number uniformly distributed between 0.0 and 0.25.

In Chicken, the activity of Previous Action neurons was based on what occurred in the previous game

$$n_i = \begin{cases} 0.45 + \mathrm{rnd}(0.0, 0.5); & i = \mathrm{PreviousAction} \\ \mathrm{rnd}(0.0, 0.5); & \mathrm{Otherwise} \end{cases} \quad (2)$$

where rnd (0.0, 0.05) was a random number uniformly distributed between 0.0 and 0.05.

The Action area in both games included two neurons. For Hawk–Dove, these neurons were labeled: 1) *escalate*—the neural agent created a conflict; or 2) *display*—the neural agent did not start a conflict. For Chicken, the neurons in this area represented: 1) *straight*—the neural agent remained in the same lane; or 2) *swerve*—the neural agent moved away from the lane. The neural activity was based on input from TOI-State/Previous Action neurons and neuromodulation.

The neuromodulatory areas were identical in both games and included two neurons: 1) Raphe—a simulated raphe nucleus, which was the source of serotonergic neuromodulation; and 2) VTA—a simulated ventral tegmental area, which was the source of dopaminergic neuromodulation. The synaptic connectivity of the network is shown in Fig. 5. Some of these connections were subject to synaptic plasticity and phasic neuromodulation, where the activity of neuromodulatory neurons affected the synaptic efficacy.

The neural activities in both games were simulated by a mean firing rate neuron model, where the firing rate of each neuron ranged continuously from 0 (quiescent) to 1 (maximal firing). The equation for the mean firing rate neuron model was

$$s_i(t) = \rho_i s_i(t-1) + (1-\rho_i)\left(\frac{1}{1+\exp\left(-5I_i(t)\right)}\right) \quad (3)$$

where t was the current time step, s_i was the activation level of neuron i, ρ_i was a constant set to 0.1 and denoted the persistence of the neuron, and I_i was the synaptic input. The synaptic input of the neuron was based on presynaptic neural activity, the connection strength of the synapse, and the amount of neuromodulatory activity

$$I_i(t) = \mathrm{rmrnd}(-0.5, 0.0) + \sum_j (t-1) w_{ij}(t-1) s_j(t-1) \quad (4)$$

where w_{ij} was the synaptic weight from neuron j to neuron i, and nm was the level of neuromodulation, which was the combined average activity of Raphe and VTA.

Phasic neuromodulation can have a strong effect on action selection and learning [10]. During phasic neuromodulation, synaptic projections from sensory systems and inhibitory neurons are amplified relative to recurrent or associational connections. In our model, the input (TOI-State/Previous Action) to Action neurons represented sensory connections and the excitatory Action-to-Action neurons represented the associational connections. To simulate the effect of phasic neuromodulation, inhibitory and sensory connections were amplified by setting nm to ten times the combined average activity of the simulated Raphe and VTA neurons. Otherwise, nm was set to 1 for association connections. The last columns of Fig. 5(b) and (d) list connections amplified by phasic neuromodulation. In previous simulation studies and robotic experiments, this mechanism was shown to be effective in making the network exploitive when neuromodulation levels were high and exploratory when neuromodulation levels were low [10], [34].

Action selection depended on the activity of the Action neurons when the neural agent reached the TOI in Hawk–Dove, or just prior to the racecar moving in Chicken. At this time, neural activities of all neurons were calculated for ten time-steps [see(1)–(4)]. The Action neuron with the largest cumulative activity after ten time-steps dictated the action taken.

After both the neural agent and human player chose an action, a learning rule was applied to the plastic connections of the neural model [see Fig. 5(b)(d)]. The learning rule depended on the current activity of the presynaptic neuron, the postsynaptic neuron, the overall activity of the neuromodulatory systems, and the payoff from the game

$$\Delta w_{ij} = nm(t-1) s_j(t-1)\left(s_j(t-1)\right) * R \quad (5)$$

where s_j was the presynaptic neuron activity level, s_i was the postsynaptic neuron activity level, nm was the average activity of all neuromodulatory neurons, and R was the level of reinforcement based on payoff and cost (6). The presynaptic neuron (s_j) in (5) was either the most active TOI-State neuron for Hawk–Dove or the most active Previous Action neurons for Chicken. The postsynaptic neuron (s_i) could either be the most active Action neuron, the Raphe neuron, or the VTA neuron. Weights were normalized by the square root of sum of squared weights.

The level of reinforcement was given by [see (6), shown at bottom of page], where the *Reward* and *Cost* were the neural agent's payoff from Tables II and III were divided by the maximum positive payoff and maximum negative payoff, respectively, [see (7)]

$$\text{Reward} = \left\{ \begin{array}{ll} \frac{\text{Payoff}}{0.60} & \text{for HawkDove} \\ \frac{\text{Payoff}}{3} & \text{for Chicken} \end{array} \right\}$$

$$\text{Cost} = \left\{ \begin{array}{ll} \frac{\text{Payoff}}{-0.50} & \text{for HawkDove} \\ \frac{\text{Payoff}}{-4} & \text{for Chicken} \end{array} \right\}$$

Equations (6) and (7) were based on the assumption that dopaminergic activity predicted the reward resulting from an action and serotonergic activity predicted the cost of an action. If the prediction were accurate, there would be little change in synaptic plasticity, whereas if the prediction were inaccurate, synaptic plasticity would occur [see (5)–(7)].

III. Results

We first report on the performance of the neural agent, both in simulations and in interactions with human subjects. We then

$$R = \left\{ \begin{array}{ll} (\text{Reward} - \text{VTA}) - (\text{Cost} - \text{Raphe}); & \text{TOI-State/Previous Action} \rightarrow \text{Action Connection} \\ \text{Reward} - \text{VTA}; & \text{TOI-State/Previous Action} \rightarrow \text{VTA Connection} \\ \text{Cost} - \text{Raphe}; & \text{TOI-State/Previous Action} \rightarrow \text{Raphe Connection} \end{array} \right\} \quad (6)$$

TABLE IV
PERCENTAGE OF ESCALATIONS BY 100 NEURAL AGENTS PLAYING 100 GAMES EACH IN SIMULATIONS AGAINST OPPONENTS WITH PRESET STRATEGIES

Opponents	Control		Raphe	
	Safe	Harsh	Safe	Harsh
Statistical	97.65%	10.00%	99.06%	92.86%
T4T	34.15%	13.64%	81.82%	81.82%
WSLS	93.22%	9.09%	96.88%	96.88%

describe the subjects' performance against the neural agents in select conditions.

A. Neural Agent's Performance

1) Neural Agent's Performance in Simulations: In prior work to assess the neural agent's performance [20], [21], we analyzed the responses of the neural agent in the Hawk–Dove game against three specific opponent strategies: 1) statistical—its opponent chose to *Escalate* either 25% or 75% of the time; 2) tit-for-tat (T4T)—its opponent copied the most recent move of the neural agent; or 3) win–stay, lose–shift (WSLS)—its opponent selected the same action that led to a positive payoff in the previous game (win–stay), or selected a different action from the previous game if that action led to zero or negative payoff (Lose-Shift). The purpose of these initial simulations was to develop the model of neuromodulatory decision-making against simpler models and opponents. This allowed us to develop and analyze the model guiding the neural agent's behavior under more controlled conditions.

The neural agent adapted its behavior appropriately to the environmental conditions (see Table IV) and its opponent's strategy [20], [21]. The neural agent tended to be less aggressive (i.e., chose *Display* more frequently) when playing in a harsh environment where the probability of a serious injury was high, or against an aggressive opponent that retaliates (T4T). Conversely, when the neural agent was in a cooperative, forgiving environment, it tended to take advantage of the situation by choosing to *Escalate* more (see Table IV).

An intact neuromodulatory system was necessary for appropriate behavior [20], [21]. When serotonin was removed from the system by simulated lesions to the neural agents' raphe nucleus (Raphe), the neural agent's behavior became more Hawk-like, with a high probability of choosing to *Escalate*, even when the chance of serious injury was high (see *Harsh* columns in Table IV). This aggressive behavior was due to its inability to assess the cost of its actions. Even when the chance of serious injury was high, the Raphe lesioned agent still tended to Escalate, thus incurring lower payoffs.

The activity of the simulated neuromodulatory systems drove the agent's behavior and adapted to environmental context. For example, the *Neural* agent adapted to a T4T opponent by oscillating between escalating and displaying in successive games. In essence, the *Neural* agent learned to adopt a T4T strategy against this opponent, which yielded approximately equal reward to both agents [see Fig. 6(a): games 84–92]. Similarly, the *Neural* agent adjusted appropriately to a WSLS opponent, by creating opportunities for high payoffs. High cost and reward were expected when both agents escalated [see Fig. 6(b): games 79, 82, or 86]. In these examples, the *Neural* agent escalated first and its *Opponent* escalated second. The *Neural* agent learned that this tactic caused the *Opponent* agent to "lose-shift'" towards *Display* in the following game, which could be taken advantage of by escalating [see Fig. 6(b): games 80, 83, or 87)]. This tactic resulted in a maximal reward to the *Neural* agent, but caused the *Opponent* agent to "lose-shift" back to *Escalate* in the following game [see Fig. 6(b): games 81, 84, or 88].

The neural response of the simulated neuromodulators governed the *Neural* agent's actions. Dopaminergic and serotonergic activity tracked the expected rewards and costs respectively. When the VTA (dopamine) activity dropped below the Raphe (serotonin) activity, the neural agent chose a *Display* action, and when the VTA activity was greater than the Raphe, the neural agent chose to *Escalate*. Raphe activity may be acting as a threshold for the expected cost of upcoming actions, whereas the VTA activity rises and falls based on the expected reward. For example, when a *Neural* agent behaved *Dove-like*, its serotonin activity was high relative to the dopamine activity due to the low expected reward from displaying [see Fig. 6(a): games 78–80]. In addition, the oscillatory actions taken by the *Neural* agent are exactly matched by the oscillatory VTA neuromodulatory activity [see Fig. 6(a): games 84–99] rising above and falling below the Raphe neuromodulatory activity. The low fluctuation in Raphe values from one game to the next in Fig. 6(a) result from the predictable cost estimates when playing an opponent using the *T4T* strategy. Predicted cost was not as regular for the *Neural* agent when playing against the *WSLS* opponent, which may explain why the Raphe neuromodulatory activity fluctuated more in Fig. 6(b). Although the Raphe activity fluctuated more when playing against the *WSLS* opponent, the actions taken by the *Neural* agent were consistent with the VTA and Raphe neuromodulatory activity.

2) Neural Agent's Performance Against Subjects: Similar to its performance against simulated opponents, the neural agent needed an intact neuromodulatory system to appropriately adapt its performance to a subject's strategy and the environmental conditions. We assessed the percentage of times the neural agents chose to *Escalate* in the Hawk–Dove game with a four-way repeated-measures analysis of variance (ANOVA; alpha $= 0.0125$, Bonferroni corrected) that included the factors of Neural State (Control and Raphe), Embodiment (Robot and Sim), Probability of Serious Injury (0.25 and 0.75), and Experimental Day (Tryp- and Tryp+). We found a significant main effect for Neural State [$\mathrm{F}(1,7) = 254.085$, $\mathrm{p} < 1 \times 10^{-4}$; Fig. 6(a)], driven by a higher percentage of choices to *Escalate* for Raphe (mean $= 66.56\% \pm 2.22\%$ standard error of the mean (SEM)) than for Control (mean $= 35.47\% \pm 2.64\%$ SEM). We found no significant effect for Embodiment ($\mathrm{F}(1,7) = 0.050$, $\mathrm{p} = 0.829$), Probability of Serious Injury ($\mathrm{F}(1,7) = 3.651$, $\mathrm{p} = 0.097$), or Experimental Day ($\mathrm{F}(1,7) = 0.116$, $\mathrm{p} = 0.743$). There were no significant interactions ($\mathrm{p} > 0.05$).

Similar to the Hawk–Dove game statistical analysis, we assessed the percentage of times the neural agents chose to drive *Straight* in the Chicken game with a three-way repeated-measures analysis of variance (ANOVA; $\mathrm{alpha} = 0.017$, Bonferroni corrected) that included factors Neural State (Control and Raphe), Embodiment (Robot and Sim), and Experimental Day (Tryp- and Tryp+). We found a significant main effect of Neural

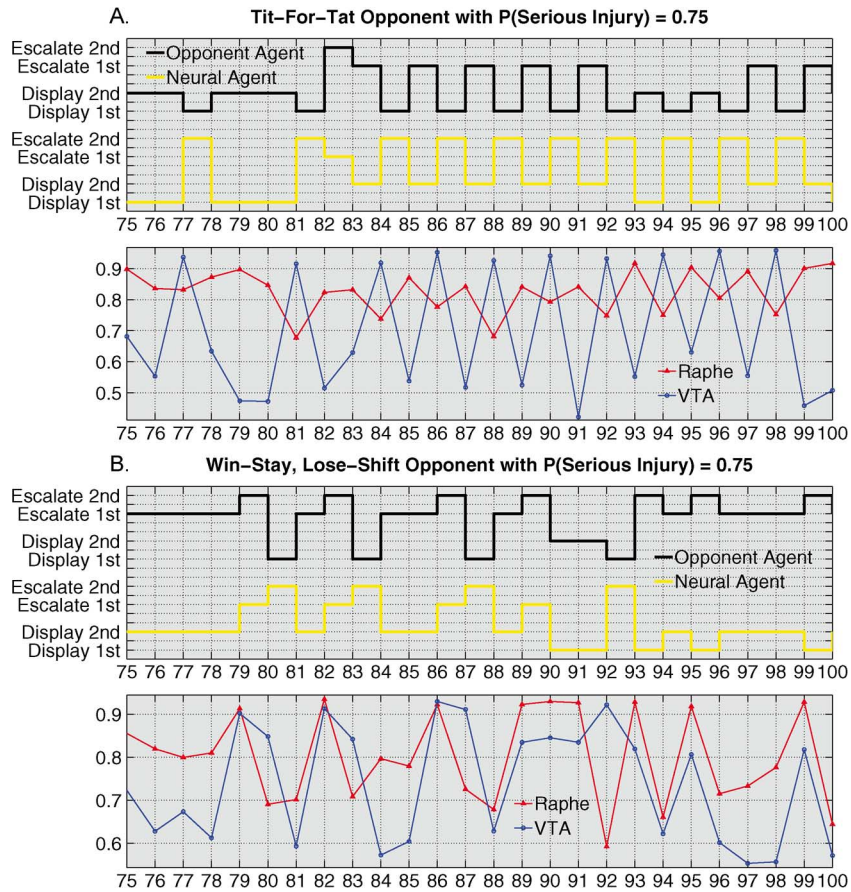

Fig. 6. Actions taken by both a *Neural* and *Opponent* agent during the last 25 games of a 100 game series, and the corresponding neuromodulatory activity for the *Neural* agent. The stair plots located on the top half of A and B show the actions taken by both the *Neural* (yellow) and *Opponent* (black) agents. The line plots located in the bottom of A and B show neuromodulatory activity for the *Neural* agent during the same 25 games. The red line represents the Raphe activity, and the blue line represents the VTA activity. (a). *Control* agent versus the *Tit-For-Tat* opponent. (b). *Control* agent versus *Win-Stay*, *Lose-Shift* opponent.

State [$F(1,7) = 116.069$, $p < 1 \times 10^{-4}$; Fig. 6(b)], driven by a higher percentage of choices to drive *Straight* for Raphe (mean $= 54.69\% \pm 1.63\%$ SEM) than for Control (mean $= 32.66\% \pm 1.71\%$ SEM). We found no significant effect of Embodiment ($F(1,7) = 1.252$, $p = 0.300$) or Experimental Day ($F(1,7) = 0.101$, $p = 0.760$). There were no significant interactions ($p > 0.100$).

In both the Hawk–Dove and Chicken games, the neural agent became more aggressive when its simulated serotonergic system was lesioned by increasing the number of times it to chose to *Escalate* or to drive *Straight* (see Fig. 7). These results from games against subjects were consistent with the behavior of the neural agent against simulated opponents.

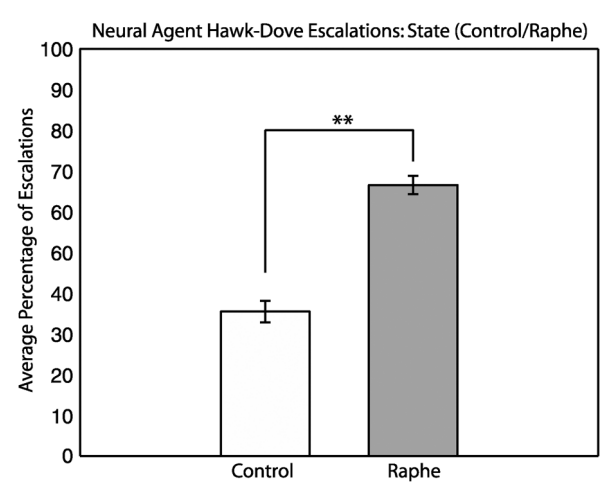

Fig. 7A. Responses of neural agents against human subjects. The bar plots show the mean and SEM for each level (Control and Raphe) within the main factor Neural State for the dependent variable *%Escalations*. The double asterisks indicate significance (a). Percentage of choices to *Escalate* in the Hawk–Dove game for an intact neural agent (Control) and for a neural agent with a lesion to its simulated serotonergic system (Raphe-lesioned). There was an increase in the percentage of choices to *Escalate* for Raphe-lesioned neural agents (four-way repeated-measures ANOVA; $F(1,7) = 254.085$, $p < 1 \times 10^{-4}$; $\text{alpha} = 0.0125$, Bonferroni corrected).

B. Subjects' Performance

1) Acute Tryptophan Depletion: The ATD procedure effectively altered subjects' blood plasma tryptophan levels. The ratio between total blood plasma tryptophan levels at the two time points (baseline $T = 0$ h and experimental $T = 5.5$ h) for each day resulted in a highly significant difference when comparing Tryp- with Tryp+. ($p < 0.0005$, Wilcoxon rank-sum test; Fig. 8). At baseline, the total blood plasma tryptophan levels for both days ranged from 49–69 $\mu\text{mol/L}$. At five hours after the amino acid drink, tryptophan levels ranged from 5–8 $\mu\text{mol/L}$ for the Tryp- condition, and 51–182 $\mu\text{mol/L}$ for the Tryp+ condition.

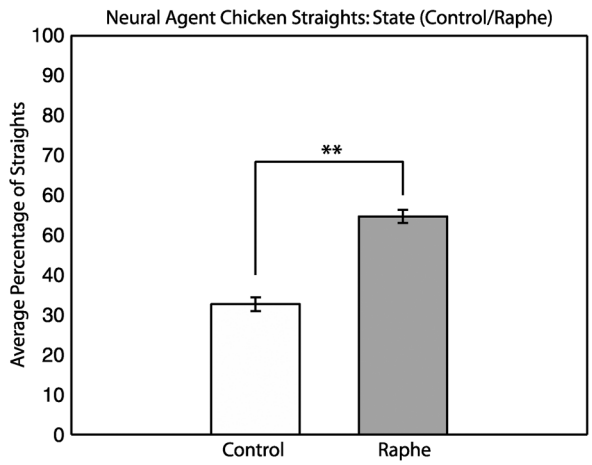

Fig. 7B. There was an increase in the percentage of choices to stay *Straight* for Raphe-lesioned neural agents in the Chicken game (three-way repeated-measures ANOVA; $F(1,7) = 116.069$, $p < 1 \times 10^{-4}$; $alpha = 0.017$, Bonferroni corrected).

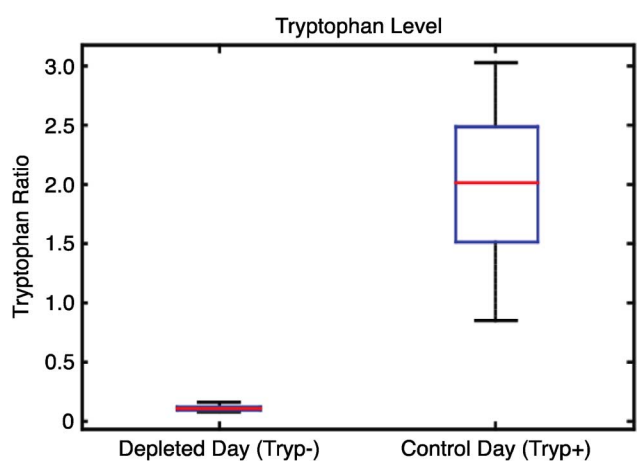

Fig. 8. The y-axis shows the ratio of total blood plasma tryptophan levels at the experimental time point (5.5 h) to the baseline time point (before ATD shake consumption). The left column represents the ATD day and the right column represents the control day. The red lines represent the median value for each distribution, and the whiskers are the spread of each distribution. The horizontal bars represent the upper and lower quartile values for each distribution. The two distributions of baseline and postshake blood tryptophan levels were significantly different ($p < 0.0005$, Wilcoxon rank-sum test).

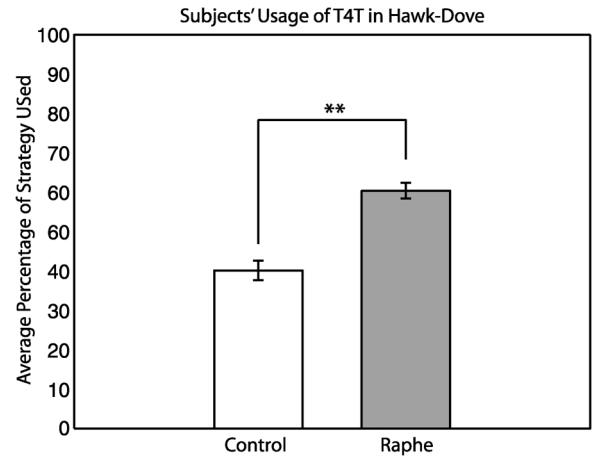

Fig. 9A. Subjects' performance while playing the Hawk–Dove game against neural agents. The bar plots show the mean and SEM for each level (Control and Raphe) within the main factor Neural State for the respective dependent variables, *%T4T* and *%WSLS*. (a). The percentage of choices by subjects to use the T4T strategy while playing the Hawk–Dove game. The double asterisks indicate that there was a significant increase in the percentage of choices to use T4T against Raphe-lesioned neural agents (four-way repeated-measures ANOVA; $F(1,7) = 38.949$, $p < 0.001$; $alpha = 0.0125$, Bonferroni corrected).

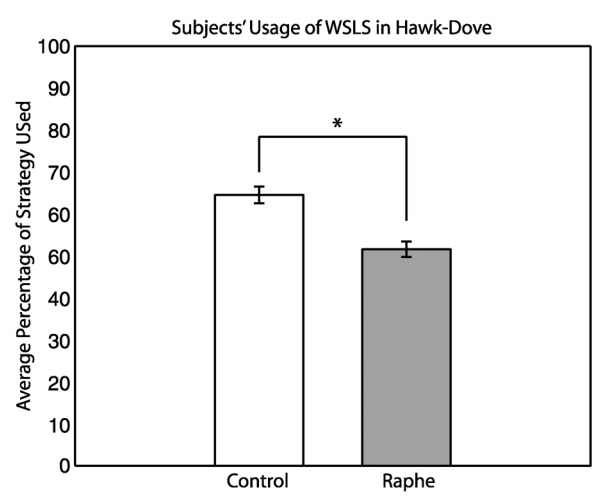

Fig. 9B. Percentage of choices by subjects to use the WSLS strategy while playing the Hawk–Dove game. The single asterisk indicates that there was a marginally significant decrease in the percentage of choices to use WSLS against Raphe-lesioned neural agents (four-way repeated-measures ANOVA; $F(1,7) = 5.731$, $p < 0.05$; $alpha = 0.0125$, Bonferroni corrected).

ATD did not have an effect on mood assessment. No significant differences were found through analysis of the human subjects' responses to the PANAS immediately before drink consumption and immediately before human robot interaction for each experimental day. We assessed positive affect and negative affect with 2 separate two-way repeated-measures analysis of variance (ANOVA; alpha $= 0.025$, Bonferroni corrected) that included factors Time of Day (Morning and Afternoon) and Experimental Day (Tryp- and Tryp+). Consistent with other studies we found no significant main effects (positive affect: $p > 0.400$; negative affect: $p > 0.200$) or interactions (positive affect: $p > 0.800$; negative affect: $p > 0.100$) when comparing the two time points within an experimental day or across days, or when comparing negative affect within a day or across days [35].

2) Subjects' Responses in the Hawk–Dove Game: Subjects tended to change their strategies depending on the type of neural agent they were playing against. Specifically, they tended to adopt a WSLS strategy against control neural agents with intact neuromodulatory systems and a T4T strategy against simulated Raphe-lesioned neural agents (see Fig. 9).

We assessed the percentage of choices by the subject to use each strategy (T4T or WSLS) and the percentage of choices to *Escalate* with two separate four-way repeated-measures analysis of variance (ANOVA; alpha $= 0.0125$, Bonferroni corrected) that included factors Neural State (Control and Raphe), Embodiment (Robot and Sim), Probability of Serious Injury (0.25 and 0.75), and Experimental Day (Tryp- and Tryp+). We found a significant main effect of Neural State ($F(1,7) = 38.949$, $p < 0.001$; Fig. 9(a)), driven by

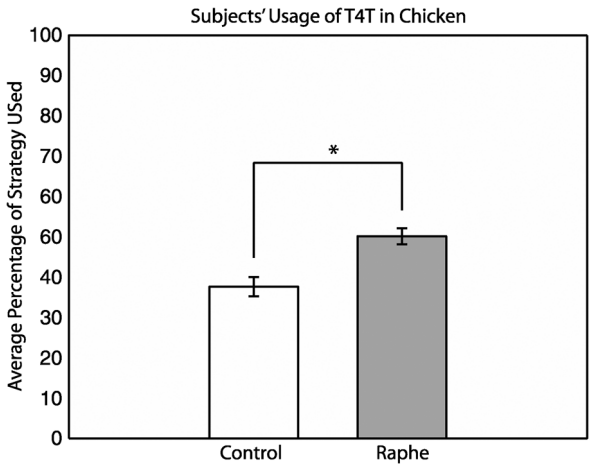

Fig. 10A. Subjects' performance while playing the Chicken game against neural agents. The bar plots show the mean and SEM for each level (Control and Raphe) within the main factor Neural State for the respective dependent variables, *%T4T* and *%WSLS*. The single asterisk indicates marginal significance. (a). The percentage of choices by subjects to use the T4T strategy while playing the Chicken game. There was an increase in the percentage of choices to use T4T against Raphe-lesioned neural agents (three-way repeated-measures ANOVA; $\mathrm{F}(1,7) = 8.537$, $\mathrm{p} < 0.025$; alpha $= 0.017$, Bonferroni corrected).

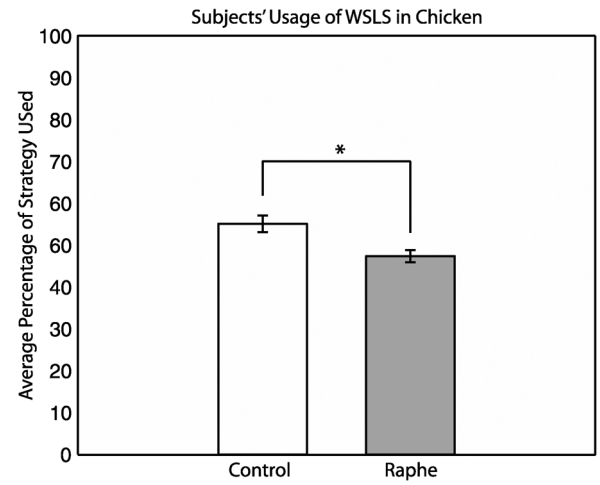

Fig. 10B. Percentage of choices by subjects to use the WSLS strategy while playing the Chicken game. There was a decrease in the percentage of choices to use WSLS against Raphe-lesioned neural agents (three-way repeated-measures ANOVA; $\mathrm{F}(1,7) = 5.240$, $\mathrm{p} < 0.06$; alpha $= 0.017$, Bonferroni corrected).

a higher percentage of choices to use the T4T strategy for Raphe (mean $= 60.36\% \pm$ of 1.98% SEM) than for Control (mean $= 40.13\% \pm 2.49\%$ SEM). Additionally, we found a marginally significant main effect of Neural State ($\mathrm{F}(1,7) = 5.731$, $\mathrm{p} < 0.05$; Fig. 9(b)), driven by a higher percentage of choices to use the WSLS strategy for the Control (mean $= 64.56\% \pm 1.99\%$ SEM) than for Raphe (mean $= 51.56\% \pm 1.83\%$ SEM). We found no significant effects for Embodiment (T4T: $\mathrm{F}(1,7) = 0.310$, $\mathrm{p} = 0.595$; WSLS: $\mathrm{F}(1,7) = 0.455$, $\mathrm{p} = 0.522$), Probability of Serious Injury (T4T: $\mathrm{F}(1,7) = 3.309$, $\mathrm{p} = 0.112$; WSLS: $\mathrm{F}(1,7) = 0.133$, $\mathrm{p} = 0.726$), or Experimental Day (T4T: $\mathrm{F}(1,7) = 1.075$, $\mathrm{p} = 0.334$; WSLS: $\mathrm{F}(1,7) = 0.319$, $\mathrm{p} = 0.590$). There were no significant interactions (T4T: $\mathrm{p} > 0.100$; WSLS: $\mathrm{p} > 0.150$). We also found no significant effects when assessing the percentage of choices to *Escalate* for all the factors: Neural State ($\mathrm{F}(1,7) = 1.43$, $\mathrm{p} = 0.271$), Embodiment ($\mathrm{F}(1,7) = 0.704$, $\mathrm{p} = 0.429$), Probability of Serious Injury ($\mathrm{F}(1,7) = 0.178$, $\mathrm{p} = 0.686$), and Experimental Day ($\mathrm{F}(1,7) = 0.7103$, $\mathrm{p} = 0.427$). There were no significant interactions with the percentage of choices to *Escalate* ($\mathrm{p} > 0.213$).

This shift from WSLS to T4T against a neural agent with a Raphe lesion suggested that subjects were retaliating against an aggressive opponent. Subjects tended to respond to cooperation with cooperation and aggression with aggression by adopting a T4T strategy.

3) Subjects' Responses in the Chicken Game: Consistent with the results of the Hawk–Dove game, subjects tended to change their strategies depending on the state of the neural agent they were playing against in the Chicken game. That is, they tended to adopt a WSLS strategy against a control neural agent with an intact neuromodulatory system and tended to adopt the T4T strategy against a simulated Raphe-lesioned neural agent (see Fig. 10).

We assessed the percentage of choices by subjects to use each strategy (T4T or WSLS) and the percentage of choices by subjects to drive *Straight* with two separate three-way repeated-measures analysis of variance (ANOVA; $\mathrm{alpha} = 0.017$, Bonferroni corrected) that included factors Neural State (Control and Raphe), Embodiment (Robot and Sim), and Experimental Day (Tryp- and Tryp+). We found a marginally significant main effect of Neural State ($\mathrm{F}(1,7) = 8.537$, $\mathrm{p} < 0.025$; Fig. 10(a)), driven by a higher percentage of choices to use the T4T strategy for Raphe ($\mathrm{mean} = 50.17\% \pm 2.01\%$ SEM) than for Control ($\mathrm{mean} = 37.67\% \pm 2.38\%$ SEM). Additionally, we found a marginally significant main effect of Neural State [$\mathrm{F}(1,7) = 5.240$, $\mathrm{p} < 0.06$; Fig. 10(b)], driven by a higher percentage of choices to use the WSLS strategy for the Control ($\mathrm{mean} = 55.10\% \pm 1.97\%$ SEM) than for Raphe ($\mathrm{mean} = 47.37\% \pm 1.43\%$ SEM). We found no significant effects for Embodiment (T4T: $F(1,7) = 0.492$, $\mathrm{p} = 0.506$; WSLS: $\mathrm{F}(1,7) = 3.943$, $\mathrm{p} = 0.088$) or Experimental Day (T4T: $\mathrm{F}(1,7) = 1.584$, $\mathrm{p} = 0.249$; WSLS: $\mathrm{F}(1,7) = 2.696$, $\mathrm{p} = 0.145$). There were no significant interactions (T4T: $\mathrm{p} > 0.340$; WSLS: $\mathrm{p} > 0.170$). We also found no significant effects when assessing the percentage of choices to drive *Straight* for all factors: Neural State ($\mathrm{F}(1,7) = 0.069$, $\mathrm{p} = 0.801$), Embodiment ($\mathrm{F}(1,7) = 0.003$, $\mathrm{p} = 0.957$), and Experimental Day ($\mathrm{F}(1,7) = 0.349$, $\mathrm{p} = 0.573$). No significant interactions were found between factors for choices to drive *Straight* ($\mathrm{p} > 0.4$).

4) Assessing Individual Subject Behavior Through Cognitive Modeling: We used hierarchical Bayesian cognitive modeling to analyze individual differences in behavior [36]. Bayesian cognitive models can illuminate individual differences between subjects that may otherwise get lost in population averages. Using the Hawk–Dove dataset, this approach to cognitive modeling found a subgroup within the subject population who tended to *Escalate* more frequently when tryptophan depleted, while there was another subgroup who tended to *Escalate* less

when tryptophan depleted. Similarly, there was a subgroup that *Escalated* more when playing the embodied neural agent (robot) and another that *Escalated* more against the simulated neural agent. The presence of two subgroups canceling each other out may have resulted in the insignificant effects of ATD and embodiment using population statistics. These cognitive modeling results point to the importance of taking individual differences into consideration in these types of studies. Moreover, these results show that subject decision-making is strongly influenced by the embodiment of the neural agent and the effect of tryptophan depletion.

IV. Discussion

The main contributions of the present study are: 1) playing games against opponents that are interactive and personified evoke strong responses in subjects; 2) an agent with the ability to adapt to contextual changes in the environment or its opponent's behavior is an important factor in evoking these responses; and 3) subjects tend to reciprocate and retaliate against adaptive agents when they believe they are being treated unfairly.

Game theory has had a long and productive history of predicting and describing human behavior in cooperative and competitive situations [3]–[5], [37]. The theory of games has also been used to illuminate the neural basis of economic and social decision-making [1], [2], [8]. These studies have identified brain areas and neural systems involved in social decision-making through brain imaging, as well as pharmacological and dietary manipulations. However, these studies typically have human subjects play against computer opponents with set strategies having predictable behavior.

The present work extends these studies by pitting human subjects against neural models that have the ability to adapt their behavior based on their opponent's strategy or changes in environmental conditions. Specifically, human subjects played Hawk–Dove and Chicken games against a simulated neural agent with the ability to assess the potential costs and rewards of its actions and adapt its behavior accordingly. The model for the neural agent was based on known interactions of the dopaminergic and serotonergic neuromodulatory systems with cortical areas [10], [20], [21], [34]. This double-blind study consisted of two experimental days during which the levels of serotonin were lowered in both humans and the neural model in half of the trials.

Similar to the idea of autonomous mental development (AMD), the appropriate behavior with which the neural agent should effectively play these games was unknown at the outset and the agent had to develop strategies based on its experience [38]–[40]. The adaptive agent did have innate values in the form of positive payoffs (reward related dopamine signals) and negative payoffs (cost related serotonin signals). From these inborn values, the neural agent learned when to act aggressively and when to act cooperatively based on the responses of its opponent and the environmental context. The plasticity in the model was initially driven by cost and rewards from payoff, but over time, the game cues and prior history shaped synaptic plasticity and behavioral responses. Both in simulations and in games with human subjects, the neural model developed the appropriate responses by adapting its decision-making to the game context.

The main finding of the study was that human subjects changed their overall strategies in response to changes in the neural agent's state (Control or Raphe). Specifically, subjects switched from a WSLS strategy when playing against a neural agent with an intact simulated nervous system to a T4T strategy when playing against a neural agent with a lesion to its serotonergic system (see Figs. 9 and 10). This change in strategy was independent of the embodiment of the neural agent and independent of tryptophan levels. A neural agent with a simulated lesion to its serotonergic system tended toward more aggressive behavior, because it lost its ability to assess the cost of an action (see Fig. 7). Subjects playing against such an opponent did not increase their levels of aggression; that is, there were not significant increases in their decisions to choose to *Escalate* or drive *Straight* (see Results for details). Rather, subjects responded to aggressive behavior with aggression and cooperative behavior with cooperation, through the adoption of the T4T strategy.

The shift to a T4T strategy may be similar to the rejection of unfair offers in the Ultimatum Game [4]. In both cases, subjects behaved irrationally by lowering their overall utility through aggressive behavior. That is, aggressive behavior by both resulted in lower payoffs (see Tables II and III). In the Ultimatum Game, a subject rejects what he deems to be unfair offers even if he is the only one penalized by the rejection, and even if the proposer of the offer is unaware of his actions [8]. A T4T strategy, which is strategically less advantageous than WSLS, could send a message to another player that the subject believes he is being treated unfairly. The neural agent, which was developed in simulations against simpler opponents, did not have the capacity to retaliate. It will be of interest to include this capacity in future versions of these human-robot interaction studies.

Playing an opponent that is interactive and personified has previously been observed to evoke strong responses in subjects. For example, in the Ultimatum Game, subjects reject more offers made by a human partner than those offers made by a computer, suggesting that participants have a stronger emotional reaction to unfair offers from humans than from a computer [41]. In our study, however, the physical instantiation of the neural agent in both games did not evoke stronger responses from subjects than did the simulated neural agent. We suggest that both the simulated and embodied versions of the neural agent evoked strong responses in subjects because of the neural agent's adaptive behavior. The neural agent demonstrated a variety of strategies and adjusted its behavior to environmental conditions and its opponent. Moreover, lesions to the neural agent's simulated serotonergic system resulted in additional classes of more aggressive opponents.

In previous studies, treatment with ATD has led to an increased number of defections in the Prisoner's Dilemma [13] and more rejections of offers in the Ultimatum Game [14]. In contrast, we did not observe a decrease of cooperativeness in our subjects due to ATD, but rather the emergence of a significant shift in strategies based on opponent type (Figs. 9 and 10). It may be that iterative interactions with a responsive, adaptive agent outweighed the effects of ATD in our human subjects.

Another possible reason for lack of significant effect due to embodiment and ATD was the small subject size in the study $(n = 8)$. The ATD protocol is invasive and time consuming for subjects. Therefore, it is difficult to run large-scale studies under this protocol. Moreover, the effects on behavior can be somewhat mild [13], [14].

Rather than attempting to expand the study, we turned to hierarchical Bayesian cognitive modeling to analyze individual differences in behavior during the Hawk–Dove games [36]. This approach to cognitive modeling found a subgroup within the subject population who tended to *Escalate* more frequently when tryptophan depleted, while there was another subgroup who tended to *Escalate* less when tryptophan depleted. Similarly, there was a subgroup that *Escalated* more when playing the embodied neural agent (robot) and another that *Escalated* more against the simulated neural agent. Taken together, these results: 1) point to the importance of taking individual differences into consideration in these types of studies; and 2) demonstrate that embodiment and serotonin levels had a strong effect on subject decision-making.

Game theory has revealed the mechanisms underlying cognitive behaviors, such as cooperation, competition, social contracts, and reciprocity. Our study, which included both embodied and simulated versions of adaptive agents, sheds light on how humans interact with others in conflicting situations. Moreover, this study will assist in the development of neural agents that can respond more naturally in human-robot interactions.

ACKNOWLEDGMENT

We would like to thank C. Guthrie for programming the graphical user interfaces. We would like to thank B. Cox and L. Bucci f or designing the robot and experimental apparatus. We would like to thank Jim at Nomad Slot Car Racing (http://www.nomadslotracingestore.com) for many useful discussions.

REFERENCES

[1] D. Lee, "Game theory and neural basis of social decision making," *Nat.Neurosci.*, vol. 11, pp. 404–409, Apr. 2008.
[2] J. K. Rilling and A. G. Sanfey, "The neuroscience of social decision-making," *Annu. Rev. Psychol.*, vol. 62, pp. 23–48, 2011.
[3] J. Maynard Smith, *Evolution and the Theory of Games*. Cambridge, U.K.: Cambridge Univ. Press, 1982.
[4] M. A. Nowak *et al.*, "Fairness versus reason in the ultimatum game," *Science*, vol. 289, pp. 1773–1775, Sep. 8, 2000.
[5] B. Skyrms and R. Pemantle, "A dynamic model of social network formation," *Proc. Nat. Acad. Sci. USA*, vol. 97, pp. 9340–9346, Aug. 1, 2000.
[6] D. J. Barraclough *et al.*, "Prefrontal cortex and decision making in a mixed-strategy game," *Nat. Neurosci.*, vol. 7, pp. 404–410, Apr. 2004.
[7] J. Rilling *et al.*, "A neural basis for social cooperation," *Neuron*, vol. 35, pp. 395–405, Jul. 18, 2002.
[8] T. Yamagishi *et al.*, "The private rejection of unfair offers and emotional commitment," *Proc. Nat. Acad. Sci. USA*, vol. 106, pp. 11520–11523, Jul. 14, 2009.
[9] L. A. Briand *et al.*, "Modulators in concert for cognition: Modulator interactions in the prefrontal cortex," *Prog. Neurobiol.*, vol. 83, pp. 69–91, Oct. 2007.
[10] J. L. Krichmar, "The neuromodulatory system—A framework for survival and adaptive behavior in a challenging world," *Adapt. Behav.*, vol. 16, pp. 385–399, 2008.
[11] R. Cools *et al.*, "Serotoninergic regulation of emotional and behavioural control processes," *Trends Cogn. Sci.*, vol. 12, pp. 31–40, Jan. 2008.
[12] M. J. Millan, "The neurobiology and control of anxious states," *Prog. Neurobiol.*, vol. 70, pp. 83–244, Jun. 2003.
[13] R. M. Wood *et al.*, "Effects of tryptophan depletion on the performance of an iterated Prisoner's Dilemma game in healthy adults," *Neuropsychopharmacology*, vol. 31, pp. 1075–1084, May 2006.
[14] M. J. Crockett *et al.*, "Serotonin modulates behavioral reactions to unfairness," *Science*, vol. 320, p. 1739, Jun. 27, 2008.
[15] M. J. Crockett *et al.*, "Serotonin selectively influences moral judgment and behavior through effects on harm aversion," *Proc. Nat. Acad. Sci. USA*, vol. 107, pp. 17433–17438, Oct. 5, 2010.
[16] C. Breazeal and B. Scassellati, "Robots that imitate humans," *Trends Cogn. Sci.*, vol. 6, pp. 481–487, Nov. 1, 2002.
[17] C. Kidd and C. Breazeal, E ect of a robot on user perceptions 2004, pp. 3559–3564.
[18] J. L. Krichmar and G. M. Edelman, "Machine psychology: Autonomous behavior, perceptual categorization and conditioning in a brain-based device," *Cereb. Cortex*, vol. 12, pp. 818–830, Aug. 2002.
[19] J. L. Krichmar and G. M. Edelman, "Brain-based devices for the study of nervous systems and the development of intelligent machines," *Artif. Life*, vol. 11, pp. 63–77, Winter-Spring, 2005.
[20] D. E. Asher *et al.*, "Effect of neuromodulation on performance in game playing: A modeling study," in *Proc. IEEE 9th Int. Conf. Develop. Learn.*, Ann Arbor, MI, 2010, pp. 155–160.
[21] A. Zaldivar *et al.*, "Simulation of How Neuromodulation Influences Cooperative Behavior," in *Simulation of Adaptive Behavior: From Animals to Animats*, S. Doncieux, Ed. Berlin, Germany: Springer-Verlag Lecture Notes Artif. Intell. (LNAI 6226), 2010, pp. 649–660, et al..
[22] W. Schultz, "Dopamine neurons and their role in reward mechanisms," *Current Opinion Neurobiol.*, vol. 7, pp. 191–197, Apr. 1997.
[23] W. Schultz *et al.*, "A neural substrate of prediction and reward," *Science*, vol. 275, pp. 1593–1599, Mar. 14, 1997.
[24] S. N. Young *et al.*, "Tryptophan depletion causes a rapid lowering of mood in normal males," *Psychopharmacology (Berl)*, vol. 87, pp. 173–177, 1985.
[25] J. B. W. Williams and M. Gibbon, "The structured clinical interview for DSM-III-R (SCID) II. Multisite test-retest reliability," *Archives General Psych.*, vol. 49, p. 630, 1992.
[26] A. Oldman *et al.*, "Biochemical and behavioural effects of acute tryptophan depletion in abstinent bulimic subjects: A pilot study," *Psychol. Med.*, vol. 25, pp. 995–1001, Sep. 1995.
[27] M. A. Ellenbogen *et al.*, "Mood response to acute tryptophan depletion in healthy volunteers: Sex differences and temporal stability," *Neuropsychopharmacology*, vol. 15, pp. 465–474, Nov. 1996.
[28] L. A. Jans *et al.*, "Serotonergic vulnerability and depression: Assumptions, experimental evidence and implications," *Mol. Psych.*, vol. 12, pp. 522–543, Jun. 2007.
[29] G. Biggio *et al.*, "Rapid depletion of serum tryptophan, brain tryptophan, serotonin and 5-hydroxyindoleacetic acid by a tryptophan-free diet," *Life Sci.*, vol. 14, pp. 1321–1329, Apr. 1, 1974.
[30] G. L. Gessa *et al.*, "Effect of the oral administration of tryptophan-free amino acid mixtures on serum tryptophan, brain tryptophan and serotonin metabolism," *J. Neurochem.*, vol. 22, pp. 869–870, May 1974.
[31] C. J. Bell *et al.*, "Acute tryptophan depletion. Part II: Clinical effects and implications," *Aust N Z J. Psych.*, vol. 39, pp. 565–574, Jul. 2005.
[32] S. D. Hood *et al.*, "Acute tryptophan depletion. Part I: Rationale and methodology," *Aust N Z J. Psych.*, vol. 39, pp. 558–564, Jul. 2005.
[33] D. Watson *et al.*, "Development and validation of brief measures of positive and negative affect: The PANAS scales," *J. Personality Soc. Psychol.*, vol. 54, p. 1063, 1988.
[34] B. R. Cox and J. L. Krichmar, "Neuromodulation as a robot controller: A brain inspired design strategy for controlling Autonomous robots," *IEEE Robot. Autom. Mag.*, vol. 16, pp. 72–80, 2009.
[35] M. J. Crockett *et al.*, "Reconciling the role of serotonin in behavioral inhibition and aversion: Acute tryptophan depletion abolishes punishment-induced inhibition in humans," *J. Neurosci.*, vol. 29, pp. 11993–11999, Sep. 23, 2009.
[36] D. E. Asher *et al.*, "Using bayesian graphical models to categorize subject variability during socioeconomic game playing," presented at the 34th Annu. Conf. Cogn. Sci. Soc. (CogSci 2012), Sapporo, Japan, 2012.
[37] M. Nowak and K. Sigmund, "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game," *Nature*, vol. 364, pp. 56–58, Jul. 1, 1993.

[38] M. Solgi and J. Y. Weng, "Developmental stereo: Emergence of disparity preference in models of the visual cortex," *IEEE Trans. Autonom. Mental Develop.*, vol. 1, no. 4, pp. 238–252, Dec. 2009.
[39] J. Y. Weng and W. S. Hwang, "From neural networks to the brain: Autonomous mental development," *IEEE Comput. Intell. Mag.*, vol. 1, pp. 15–31, Aug. 2006.
[40] J. Y. Weng and M. Luciw, "Dually optimal neuronal layers: Lobe component analysis," *IEEE Trans. Autonom. Mental Develop.*, vol. 1, no. 1, pp. 68–85, May 2009.
[41] A. G. Sanfey *et al.*, "The neural basis of economic decision-making in the ultimatum game," *Science*, vol. 300, pp. 1755–1758, Jun. 13, 2003.

Derrik E. Asher, photograph and biography not available at the time of publication.

Andrew Zaldivar, photograph and biography not available at the time of publication.

Brian Barton, photograph and biography not available at the time of publication.

Alyssa A. Brewer, photograph and biography not available at the time of publication.

Jeffrey L. Krichmar, photograph and biography not available at the time of publication.