Human-Robot Collaboration in Manufacturing: Quantitative Evaluation of Predictable, Convergent Joint Action

StefanosNikolaidis[1*], Przemyslaw Lasota[1], Gregory Rossano[2], Carlos Martinez[2], Thomas Fuhlbrigge[2] and Julie Shah[1]
[1]Massachusetts Institute of Technology, USA
[2]ABB, USA
*email: snikol@mit.edu

***Abstract*--New industrial robotic systems that operate in the same physical space as people highlight the emerging need for robots that can integrate seamlessly into human group dynamics. In this paper we build on our prior investigation, which evaluates the convergence of a robot computational teaming model and a human teammate's mental model, by computing the entropy rate of the Markov chain. We present and analyze the six out of thirty-six human trials where the human participant switched execution strategies while working with the robot. We conduct a post-hoc analysis of this dataset and show that the entropy rate appears to be sensitive to changes in the human strategy and reflects the resulting increase in uncertainty about the human next actions. We propose that these results provide first support that entropy rate may be used as a component of dynamic risk assessment, to generate risk-aware robot motions and action selections.**

***Index Term*—entropy rate, human-robot joint action, robot teaming model**

I. Introduction

When humans work in teams, it is crucial for the members to develop fluent team behavior. We believe that the same holds for robot teammates, if they are to perform in a similarly fluent manner as members of a human-robot team. New industrial robotic systems that operate in the same physical space as people highlight the emerging need for robots that can integrate seamlessly into human group dynamics. Learning from demonstration [3] is one technique for robot training that has received significant attention. In this approach, the human explicitly teaches the robot a skill or specific task [4], [1], [11], [6], [2]. However, the focus is on one-way skill transfer from a human to a robot, rather than a mutual adaptation process for learning fluency in joint action. In many other works, the human interacts with the robot by providing high-level feedback or guidance [5], [9], [7], [16], but this kind of interaction does not resemble the teamwork processes naturally observed when human teams train together on interdependent tasks [10].

In this paper we build on our prior investigation, which presents a human-inspired technique for programming flexible human-robot coordinated work, and validates the objective and subjective performance benefits of this approach through large-scale human subject experimentation [12]. The contribution of this prior art is a computational teaming model that is empirically validated and shown to be quantitatively comparable to the human mental model using standard human factors elicitation techniques [10]. In this work, we present a post-hoc analysis of these experiments to provide support that these computationally-derived teaming models may be used to quantify a robot's uncertainty in its human teammates' next actions and generate risk-aware robot behavior.

Fig. 1. (Left) Snapshop of human-robot task execution from human subject experiments, (Right) RobotStudio point-and-click simulation environment for robot training used in our prior investigation.

II. Quantitative Models for Collaborative Physical Interaction

In our ongoing research we utilize a Markov Decision Process (MDP) to computationally encode a teaming model that captures knowledge about the roles of the robot and the human team member [12]. The computational teaming model is generated using a human-robot interactive planning method. In our prior work, we perform interactive planning through cross-training, a training strategy widely used in human teams [10].

Human-robot cross-training has been compared to a prior state-of-the-art interactive reinforcement learning algorithm [16] through large-scale experimentation with 36 human subjects. Results indicated that the human-inspired training technique improved quantitative measures of team model convergence ($p = 0.04$) and mental model similarity ($p < 0.01$). Additionally, a post-experimental survey indicated statistically significant improvements in subjective measures of

human-robot team performance; participants agreed more strongly that the robot performed its role effectively, and agreed more strongly that they trusted the robot ($p < 0.01$). Finally, significant improvements in team fluency metrics were reported, including an increase of 71% in concurrent motion ($p = 0.02$) and a decrease of 41% in human idle time ($p = 0.04$), during the actual human-robot task execution phase that succeeded the human-robot interactive training process. These prior results provide the first evidence that human-robot teamwork is improved when a human and robot train together in a manner similar to effective human team training practices [12].

In this paper, we present a post-hoc analysis of these experiments indicating that a quantitative assessment of the computational teaming model may be used to generate risk-aware robot motions and action selections.

III. Robot Teaming Model Formulated as MDP

We describe how the robot teaming model is computationally encoded as a Markov Decision Process. A Markov decision process is a tuple $\{S, A, T, R\}$, where:

- S is a finite set of states of the world; it models the set of world environment configurations
- A is a finite set of actions; this is the set of actions the robot can execute
- $T : S \times A \rightarrow \Pi(S)$ is the state-transition function, which, for each world state and action, gives a probability distribution over world states; the state transition function models the variability in human action. For a given robot action a, the human's next choice of action yields a stochastic transition from state s to a state s'. We write the probability of this transition as $T(s, a, s')$. In this formulation, human behavior is the cause of randomness in our model, although this can be extended to include stochasticity from the environment or the robot actions, as well.
- $R : S \times A \rightarrow \mathrm{R}$ is the reward function, giving the expected immediate reward gained by taking each action in each state. We write $R(s, a)$ for the expected reward of taking action a in state s.

The policy π of the robot is the assignment of an action $\pi(s)$ at every state s. The optimal policy π^* can be calculated using dynamic programming [14]. Under this formulation, the role of the robot is represented by the optimal policy π^*, whereas the robot's knowledge of the role of the human co-worker is represented by the transition probabilities T.

IV. Quantitative Evaluation of Predictable, Convergent Joint Action

As the mental model of human and robots converge, we describe the human and robot to perform similar patterns of actions. This means that the same states will be visited frequently and the robot's uncertainty about the human's action selection will decrease.

To evaluate the convergence of the robot's computational teaming model and the human mental model, we assume a uniform prior and compute the *entropy rate* [8] of the Markov chain (Eq. 1). The Markov chain is induced by specifying a policy π in the MDP framework. For the policy π we use the robot actions that match human preference, as it is elicited by the human after training with the robot. Additionally, we use the states s in S that match the preferred sequence of configurations to task completion. For a finite state Markov chain X with initial state s_0 and transition probability matrix T the entropy rate is always well defined [8]. It is equal to the sum of the entropies of the transition probabilities $T(s, \pi(s), s')$, for all s in S, weighted by the probability of occurrence of each state according to the stationary distribution μ of the chain:

$$H(X) = -\sum_{s \in S} \mu(s) \sum_{s' \in S} T(s, \pi(s), s') \log[T(s, \pi(s), s')] \quad (1)$$

The entropy rate measure has also been shown to produce different results (of statistical significance) for various interactive planning techniques, and to correlate to objective and subjective measures of team performance [12]. This measure can be generalized to encode situations where the human has multiple preferences or acts stochastically. In this work, we present and analyze the six out of thirty-six human trials where the human participant switched execution strategies while working with the robot. We conduct a post-hoc experiment analysis on this small data set and show that the entropy rate appears to be a sensitive to changes in the human's strategy and reflects the resulting increase in uncertainty about the human's next actions. We propose that these results provide intriguing first support that entropy rate may be used as a component of a dynamic assessment of risk, and may be used to generate risk-aware robot motions and action selections. Interestingly, the conditional entropy, given by Eq. (1), also represents the robot's uncertainty about the human's action selection. Post-hoc analysis of the human subject experiments verifies that this measure decreases as the human and robot train together, and increases when the human deviates from the robot's probabilistic model of human action-intent (Fig.2).

V. Experiment Protocol

A. Experiment Setting

In each experiment, one team of one human and one robot were tasked to perform a simple place-and-drill task. The human's role was to place screws in one of three available positions. The robot's role was to drill each screw. Although this task is simple, we observed a sufficient variety of different preferences for accomplishing the task. For example, some participants preferred to place all screws in a sequence from right-to-left and then drill them in the same sequence. Others preferred to place and drill each screw before moving on

to the next. The participants consisted of 36 subjects recruited from MIT. Videos of the experiment can be found at: **http://tinyurl.com/9prt3hb**

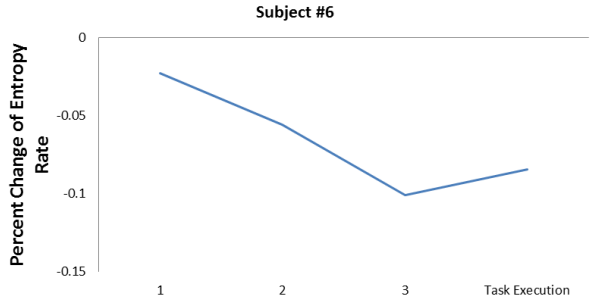

Fig. 2. From prior human subject experiments [12] - in this trial the participant changed strategies for working with the robot from training to execution. The entropy rate decreases over all three rounds of interactive training, and then sharply increases at execution as the person acts out a different strategy than planned.

B. *Human-Robot Interactive Planning*

Before starting the training, all participants were asked to describe both verbally and in written form their preferred way of executing the task. We then initialized the robot policy from a set of prespecified policies so that it was clearly different from the participant's preference. For example, if the user preferred to "have the robot drill all screws as soon as they are placed, starting from left to right," we initialized the MDP teaming model so that the starting robot policy was to wait until all screws were placed before drilling. We did this to avoid the trivial case where the initial policy of the robot matches the preferred policy of the user, and to evaluate mental model convergence starting from different human and robot mental models.

The participants were randomly assigned to two groups, Group A and Group B. Each participant then did a training session in the ABB RobotStudio virtual environment with an industrial robot which we call "Abbie" (Figure 3). Depending on the assigned group, the participant participated in the following training session:

1) Cross-training session (Group A): The participant iteratively switches positions with the virtual robot, placing the screws at the forward phase and drilling at the rotation phase.
2) Reinforcement learning with human reward assignment session (Group B): This is the standard reinforcement learning approach, where the participant places screws and the robot drills at all iterations, with the participant assigning a positive, zero, or negative reward after each robot action [7].

For the cross-training session, the MDP policy update was performed using value iteration with a discount factor of 0.9, as described in [12]. The policy update for the reinforcement learning condition was performed using the Sarsa(λ) algorithms, where parameters in the standard notation of Sarsa [15] were empirically tuned ($\lambda = 0.9$, $\mu = 0.9$, $\alpha = 0.3$) for best performance on this task.

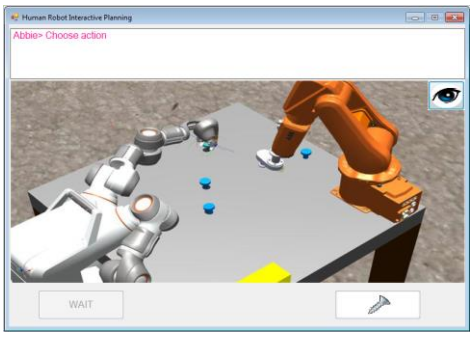

Fig. 3. Human-Robot Interactive Planning Using ABB RobotStudio Virtual Environment. The human controls the white anthropomorphic "Frida" robot on the left, to work with the orange industrial robot, "Abbie," on the right.

After the training session, the mental model of all participants was assessed as follows: for each workbench configuration through task completion, participants were asked to choose a human placing action and their preference for an accompanying robot drilling action, based on the training they had together (Figure 4).

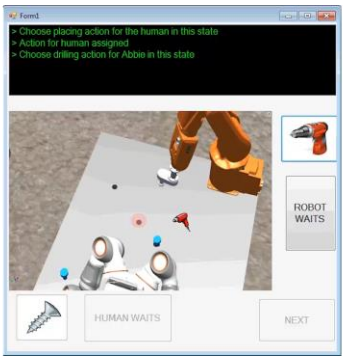

Fig. 4. Human-Robot Mental Model Elicitation Tool

C. *Human-Robot Task Execution*

We then asked all participants to perform the place-and-drill task with the actual robot, Abbie. To recognize the actions of the human we used a PhaseSpace motion capture system of eight cameras [13], which tracked the motion of a Phasespace glove worn by the participant (Figure 5). Abbie executed the policy learned from the training sessions. The task execution was videotaped and later analyzed for team fluency metrics. Finally, all participants were asked to answer a post-experiment survey.

VI. Experiment Analysis

In this section, we present and analyze the six out of thirty-six human trials where the human participants changed strategies or otherwise demonstrated inconsistencies in execution. We conduct a post-hoc experiment analysis on this small data set and show that the entropy rate appears to be sensitive to changes in the human's strategy.

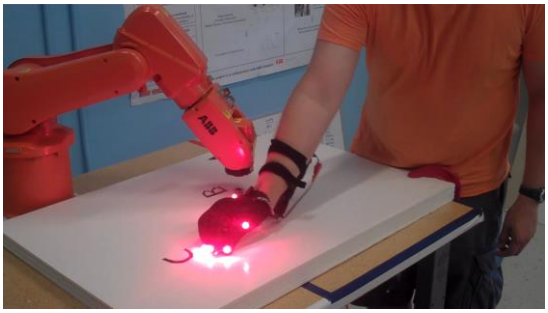

Fig. 5. Human-Robot Task Execution

A. *Calculation of Entropy Rate*

We calculate entropy rate taking into account only the states that appear in the sequence annotated by the user as his or her preferred action sequence. As the robot learns the user's preference (for example, "drill as soon as a screw is placed"), some of these states, but not necessarily all, appear in the sequence executed during training and task execution. A change in the observed sequence results in an increase in the entropy rate. For example, if the users preference is to drill as soon as a screw is placed in the order A-B-C, the states that matter for the calculation of the entropy are: "no screw placed", "screw A placed", "screw A drilled and screw B placed," etc. However, if the robot has not yet learned that it should drill after the user places a screw, most of these states are not reached. That is, after the state "screw A placed", the robot waits and the state "screw A drilled and screw B placed" is not reached during training with the robot. Instead, the state evolves to "screw A placed and screw B placed," since the robot does not drill and so instead the user places another screw. Note that even if the user changes the sequence of placement from A-B-C to, say, B-C-A, the change mostly affects states that are irrelevant to the users initial preference, and therefore affects states irrelevant to the entropy calculation.

The entropy rate measure has been shown to produce different results (of statistical significance) for various interactive planning techniques, and to correlate to objective and subjective measures of team performance [12]. Here we investigate more closely the six out of thirty-six trials where the participants changed strategies for working with the robot, and map those events to changes in the entropy rate.

1) *Subject 1, Group A*

The user's preference at the beginning of the experiment was to: "place the screws down in the order B-A-C and for Abbie to drill them in that order during placement." In the first two rounds, the user followed this preference. However, at the third and final round, the user changed the sequence from B-A-C to A-C-B, which caused the increase in the entropy rate. At the task execution, the user followed the predefined sequence B-A-C, and the entropy decreased again.

2) *Subject 2, Group A*

This subject follows a pattern of action that is similar to Subject 1. The user's preference was to place screws in the sequence A-C-B, and have the robot drill after each placement. The user followed his preference for the first two rounds, but then at the third training round he changed the sequence to C-A-B. The user did this consciously, saying he "wanted to see the response of the system to bimodal preferences."

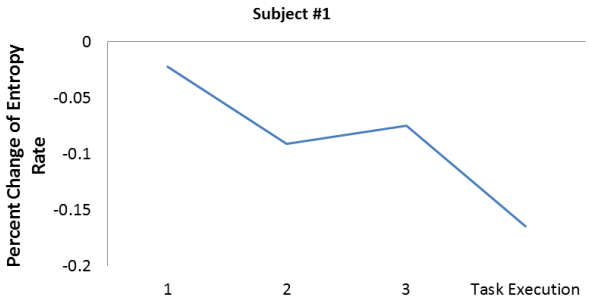

Fig. 6. Subject 1

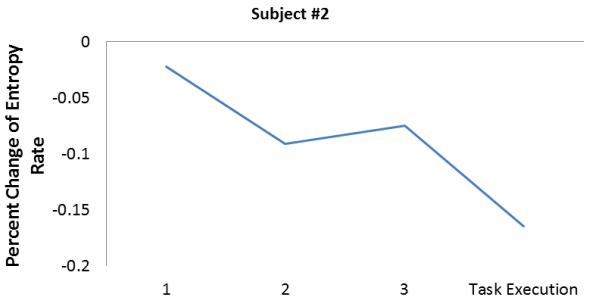

Fig. 7. Subject 2

3) *Subject 3, Group A*

The user followed the preferred sequence of A-C-B for the first two rounds. Then, at the final training round the user placed screws according to the sequence B-A-C. Finally, at the task execution the user followed the sequence A-B-C, a strategy inconsistent with all the previous training rounds. The entropy decreased from the first training round to execution, but not significantly.

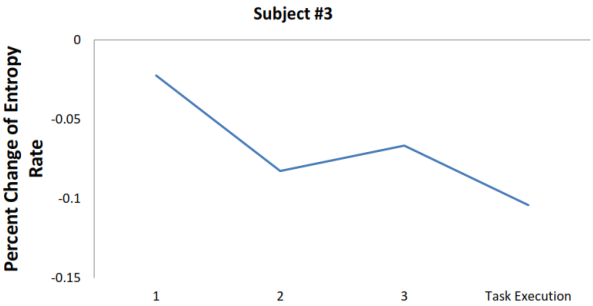

Fig. 8. Subject 3

4) *Subject 4, Group B*

The participant stated their preference as "B-C-A, Abbie drills while I place the next screw", and trained with the robot using reinforcement learning with reward assignment. The participant was consistent in all training rounds, however, the robot did not converge to the users preferences. In particular, the robot learnt to drill screw B when the user was placing screw A, but then waited for the participant to finish placing the rest

of the screws, before drilling them. Therefore, although the entropy rate decreases at each iteration, the slope is less steep than for subjects (e.g. Subject #7), whose mental model converged with the robot teaming model during training.

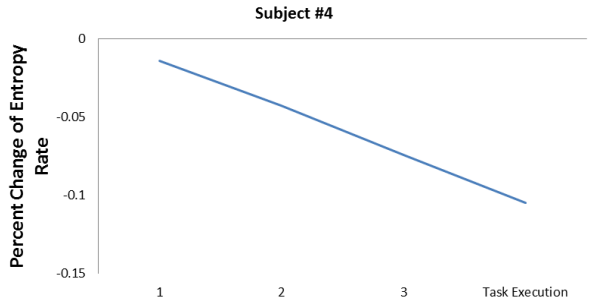

Fig. 9. Subject 4

5) *Subject 5, Group B*

The user started with a preference of placing screws in the order of C-B-A, with Abbie "drilling them as they are put in place". At the second round, the participant changed the sequence and placed the screws in the order A-B-C. We see that the entropy remained nearly constant rather than decreasing. It may seem counter-intuitive that the entropy did not increase, even though the user changed sequences. The explanation is that the user changed his preferred sequence early in the training process, when the entropy of the Markov-chain was still high. Furthermore, when we calculate the entropy of the Markov-chain, we use the states reached when following the preferred policy of the user. Since the robot has not learned that it should drill the screws after placement, following the participants preference, these states are not reached and their entropy remains constant. Therefore, the effect on the entropy rate when the participant changed the sequence is small in this case. During task execution, the user expressed confusion that the robot did not follow his preference of drilling the screw upon placement. The user waited for the robot to drill before giving up and placing all the screws by himself. This is illustrated by the change in slope of the entropy rate at task execution in Figure 10.

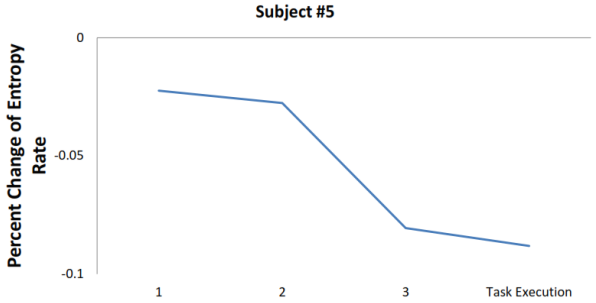

Fig. 10 Subject 5

6) *Subject 6, Group A*

The participant followed her preference of placing the screws in the order C-B-A, but then at the task execution switched to the sequence A-B-C without realizing it. The robot had learned her preference of C-B-A during training, and the result of the change of strategies at execution was a sharp increase in the entropy, as illustrated.

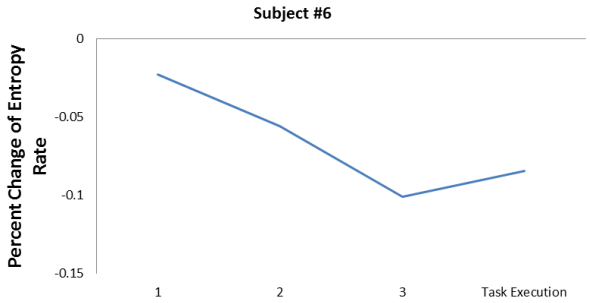

Fig. 11. Subject 6

7) *Subject 7, Group A*

This is an example of a participant that remained consistent with his preferences, during the training round and task execution. The robot learned his preference. Note the difference in the magnitude of the entropy rate decrease, compared to Subject #4, whose mental model did not converge with the robot teaming model.

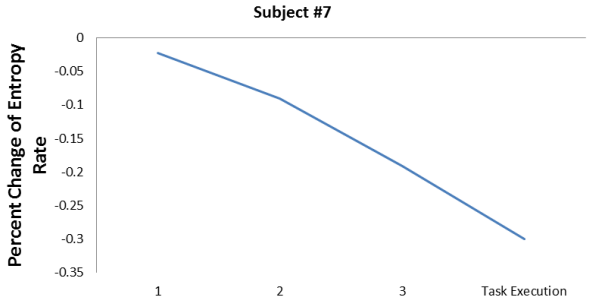

Fig. 12. Subject 7

B. Discussion

Relatively few of the thirty-six subjects changed strategies or otherwise demonstrated inconsistencies in execution. With the small sample size (six subjects) we are not able to demonstrate that the observed increases in entropy rate are of statistical significance. Nonetheless, each of the observed increases in entropy rate can be directly linked to changes or inconsistencies in human behavior. Our post-hoc analysis of this small data set provides support that the entropy rate is sensitive to changes in the human's strategy, and reflects the robot's increase in uncertainty about the human's next actions. We believe these initial results provide adequate support to justify large scale human subject experiments aimed at investigating entropy rate as a component of a dynamic assessment of risk in human-robot collaboration.

As a next step we will use changes in entropy rate to adapt robot motions. Given probability distributions of the human teammate's next actions, we can determine with what probability the human worker will occupy various locations in the workspace shared with the robot. If we know with high probability that the human worker will reach toward location C next, we can utilize

task and human motion models to determine what portion of the shared workspace will be occupied through space and time by the worker while he or she executes the task. The anticipated obstructed space can be reformulated as a cost function as input for a robot motion planner, which will change robot motion parameters (e.g. speed and path) to maneuver around the person. When the human teammate deviates from the robot's probabilistic model of human action-intent (manifested as a sudden increase in entropy rate), the robot will plan risk-aware motions and action selections (e.g. by slowing down, or choosing to execute actions that maintain a wider berth around the human).

There are several reasons why incorporating such adaptations would be beneficial. A robot which does not adapt to a human worker and simply performs a pre-set sequence of actions has to stop any time the human worker is in the way of the robot's next task. Additionally, precedence complications could arise when the human worker performs a task which needs to be done prior to a certain robot action. For example, if the human places screws to be drilled by the robot in a sequence other than the robots pre-programmed plan, the robot will have to sit idle until the human places a screw at the robot's anticipated drilling location. These problems could potentially lead to significant decreases in efficiency, especially if the pre-programmed sequence the robot is using is significantly different from the worker's preferred order of actions. A robot that adapts to the uncertainty inherent in working with a human will mitigate these problems and will result in a more efficient and human-friendly system.

VII. Conclusions

In this paper we build on our prior investigation, which evaluates the convergence of a robot's computational teaming model and a human teammate's mental model, by computing the entropy rate of the Markov chain. We present and analyze the six out of thirty-six human trials where the human participant switched execution strategies while working with the robot. We conduct a post-hoc analysis of this small data set and show that the entropy rate appears to be sensitive to changes in the human's strategy and reflects the resulting increase in uncertainty about the human's next actions. With the small sample size (six subjects) we are not able to demonstrate that the observed increases in entropy rate are of statistical significance. Nonetheless, the observed increases in entropy rate can be directly linked to changes or inconsistencies in human behavior. We believe these initial results provide adequate support to justify large scale human subject experiments aimed at investigating entropy rate as a component of a dynamic assessment of risk in human-robot collaboration.

VIII. ACKNOWLEDGEMENTS

We would like to acknowledge the Onassis Foundation as a sponsor.

References

[1] Pieter Abbeel and Andrew Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in Proceedings of the Twenty-first International Conference on Machine Learning, ACM Press, 2004.

[2] Baris Akgun, Maya Cakmak, Jae Wook Yoo, and Andrea Lockerd Thomaz, "Trajectories and keyframes for kinesthetic teaching: a human-robot interaction perspective", in HRI, pages 391–398, 2012.

[3] Brenna D. Argall, Sonia Chernova, Manuela Veloso, and Brett Browning, "A survey of robot learning from demonstration," *Robot. Auton. Syst*., 57(5):469–483, May 2009.

[4] Christopher G. Atkeson and Stefan Schaal, "Robot learning from demonstration," in ICML, pages 12–20, 1997.

[5] Bruce Blumberg, Marc Downie, Yuri Ivanov, Matt Berlin, Michael Patrick Johnson, and Bill Tomlinson, "Integrated learning for interactive synthetic charactersm," ACM Trans. Graph., 21(3):417–426, July 2002.

[6] Sonia Chernova and Manuela Veloso, "Teaching multi-robot coordination using demonstration of communication and state sharing," in Proc. AAMAS, Richland, SC, 2008.

[7] Finale Doshi and Nicholas Roy, "Efficient model learning for dialog management," in Proc. HRI, Washington, DC, March 2007.

[8] Y L. Ekroot and T.M. Cover., "The entropy of markov trajectories," Information Theory, IEEE Transactions on, 39(4):1418 –1421, Jul 1993.

[9] Frédéric Kaplan, Pierre-Yves Oudeyer, Enikö Kubinyi, and Adám Miklósi. Robotic clicker training. *Robotics and Autonomous Systems*, 38(3-4):197–206, 2002.

[10] M.A. Marks, M.J. Sabella, C.S. Burke, and S.J. Zaccaro. "The impact of cross-training on team effectiveness", *J Appl Psychol*, 87(1):3–13, 2002.

[11] Monica N. Nicolescu and Maja J. Mataric,"Natural methods for robot task learning: Instructive demonstrations, generalization and practice", In Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems, pages 241–248, 2003.

[12] Stefanos Nikolaidis and Julie Shah, "Human-robot cross-training: Computational formulation, modeling and evaluation of a human team training strategy," IEEE/ACM International Conference on Human-Robot Interaction, March 2013.

[13] Phasespace motion capture http://www.phasespace.com, 2012

[14] Stuart J. Russell and Peter Norvig. *Artificial Intelligence*: *A Modern Approach*. Pearson Education, 2003.

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

[16] Andrea L. Thomaz and Cynthia Breazeal, "Reinforcement learning with human teachers: evidence of feedback and guidance with implications for learning performance," In Proc. AAAI, pages 1000–1005, 2006.