# Toward Measuring the Effect of Robot Competency on Human Kinesthetic Feedback in Long-Term Task Learning

Shuangge Wang
Yale University
New Haven, CT, USA
shuangge.wang@yale.edu

Brian Scassellati
Yale University
New Haven, CT, USA
brian.scassellati@yale.edu

Tesca Fitzgerald
Yale University
New Haven, CT, USA
tesca.fitzgerald@yale.edu

## ABSTRACT

Robots face tremendous challenges when learning tasks where labeled data is not abundantly available. In these circumstances, human feedback can be adopted as a supervision substitute on the robot's performance. Existing learning approaches rely on a set of statistical assumptions about the human feedback that robots receive. Nominally, they assume that human feedback is Markovian, time consistent, and retrospective, while real human feedback does not necessarily match these assumptions. They also tend to assume an incentive-driven model to interpret the human feedback and their counterfactuals, enabling human action prediction and easy machine learning (ML) optimization but failing to capture other important intentions behind human feedback. Realistically, people routinely and systematically alter the way that they provide feedback based on the history of their interaction and the context in which the feedback is provided. This adaptability is particularly evident in long-term human-robot interactions where lifelong learning and personalization occur. In this work, we describe our preliminary work studying how a robot's competency influences how people correct its motion. We first survey existing work on learning from human feedback and highlight some of the assumptions that they impose on the human feedback. We then present our research question and hypotheses and, finally, describe a user study to evaluate them.

## KEYWORDS

Interactive Robot Learning, Learning from Corrections, Reinforcement Learning

**ACM Reference Format:**
Shuangge Wang, Brian Scassellati, and Tesca Fitzgerald. 2024. Toward Measuring the Effect of Robot Competency on Human Kinesthetic Feedback in Long-Term Task Learning. In *Proceedings of International Conference on Human-Robot Interaction (HRI).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

**Figure 1: A human providing kinesthetic feedback to a robot arm while it is performing pick-and-place tasks.**

## 1 INTRODUCTION

Thanks to the advancements in sensory technology, control hardware, and machine learning (ML) techniques, robots have become more and more capable in leveraging labeled data to learn tasks that were previously deemed challenging, such as reconfigurable manufacturing [50], general surgery [51], drone delivery [18], and autonomous driving [40].

In scenarios where labeled data is sparse, techniques like learning from human feedback are more widely adopted. Such techniques are particularly favored in settings where robots need to learn and adapt to specific users, like preference learning [56] and assistive robotics [7]. Under this paradigm, the robot receives feedback information from the human to learn a task model.

Traditional ML algorithms rely on several assumptions of human feedback. They assume that human feedback is Markovian, time consistent, and retrospective. Some also assume human feedback follows a Boltzmann distribution, i.e., that humans are exponentially more likely to choose feedback that results in higher utility. Natural human feedback, however, does not necessarily comply with these assumptions, especially in long-term HRI where people change their expectations of the robot over time. For instance, human feedback assesses not only the robot's most recent trial, but also a history of its performance [33]. People gradually adapt their internal model


*HRI, March 11–15, 2024, Boulder, CO*

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

of robot safety, so each feedback is not time consistent [14]. People also provide feedback as guidance to future robot actions which is not strictly retrospective [58].

One particular modality of feedback that has gained traction in recent years is corrections [21, 43, 44]. This form of feedback involves a human intervening in a robot's trajectory and modifying it with kinesthetic feedback (Fig. 1). Although correction feedback lacks a direct mapping to numerical rewards, its physical intuitiveness requires less mental math from the feedback providers compared to numerical evaluations, making it a good fit for studying natural human feedback. Moreover, as robotics technology advances, correction feedback might also be a more realistic interaction type where the human serves more as a supervisor who occasionally intervenes, rather than as a direct teacher who evaluates every robot trial.

In this work, we investigate how the robot's competency (i.e., the robot's success rate at a task) shapes how people provide feedback via corrections in a long-term task learning scenario. For the rest of this paper, we will survey prior work on learning from human feedback literature, and address some of the assumptions that they impose on the human feedback. We then discuss other prior works that challenge these assumptions. Finally, we propose our research question and design an experiment to test our hypotheses.

# 2 RELATED WORK

## 2.1 Feedback Modalities

Research in robotics has progressed thus far in learning from human feedback, including learning from rewards [16, 34, 58], demonstrations [1], corrections [9], preferences [9], implicit feedback [17], and natural language instructions [38, 57]. Correction feedback, in particular, involves a robot attempting to complete a task while under supervision from a human teacher. The teacher can intervene and modify the robot's motion, producing a *corrected* trajectory that is assumed to be more optimal with respect to the hidden task objectives [3, 30]. While prior work has proposed methods for learning from corrections, they do not model the teacher's decision to intervene in the first place.

## 2.2 Common Assumptions for Interpreting Feedback

*Consistently & Noisily Optimal.* Prior work in human feedback modeling typically assumes that humans provide feedback to maximize some reward. The Boltzmann rational model has been used to represent the likelihood of a person preferring one robot action over another based on their relative rewards. This model is then used to construct a probabilistic interpretation of that feedback and their counterfactuals [30, 45, 46, 62], which is widely adopted in fields like psychology [4, 24, 25, 48], economics [12, 46, 49, 53], and artificial intelligence [11, 23, 30, 37]. Importantly, this model relies on a parameter representing the optimality of the human as they provide feedback. While this parameter can be tuned for a specific human teacher [22], it is assumed to be a static parameter. Realistically, however, the optimality of a person's feedback may change over time based on their workload or the requirements of the task.

*Markovian & Time-Consistent.* There are several assumptions about the human feedback provided in reinforcement learning (RL). First, they assume that the feedback is Markovian in that each feedback only evaluates the most recent robot behavior [31, 60]. Second, they assume the feedback are time consistent: that repeating the same behavior should result in the robot receiving the same feedback. As a result, the ordering of the action-feedback pairs does not affect learning performance [59]. Finally, feedback is retrospective in that it depends only on prior trials, not projected future performances [6, 54]. In some cases, other constraints on feedback might apply, such as the Bellman optimality equation [5] or the triangle inequality [52].

## 2.3 Evidence Against These Assumptions

Prior works have demonstrated that people's feedback does not necessarily satisfy these RL-based assumptions. Experiments have shown that people provide more positive feedback to a struggling robot when it succeeds after a series of failures than a robot that consistently performs well [33]. Studies have shown that the teacher's prior rewards influences the scaling of their future rewards [42, 55]. These findings all refute the Markovian assumption above.

Some works have found that human's adaptive mental model of the robot influences them to provide feedback that is not time consistent. For instance, some works have shown that human's mental model of the robot's capabilities can influence whether humans provide strict or lenient evaluation [29]. Others found that the robot's performance can affect the human's mental model of both the robot and their own teaching capability [28]. Other works have demonstrated that the predictability of robot motion can influence the human teacher's expectations and confidence in the robot [19].

The retrospective assumption is also challenged since researchers found some people use feedback as reward signals about past actions [35, 36] while others use feedback as future directed rewards to guide subsequent actions [58].

Also, an incentive-driven definition of feedback has been challenging to generalize since prior works have shown that people choose teaching style and use feedback very differently [15, 32, 41]. Some works showed that the human's perceived role in robot learning is not perfectly aligned with reward maximization [41]. Some are heavily influenced by the *Pratfall Effect*, providing positive feedback for the robot's attractiveness when making a mistake [47]. The timing of feedback has been shown to not be a sole product of human agency but also robot pauses that invite feedback [2].

**Overall**, prior work has shown that human feedback is not an objective measure of the robot's performance. Rather, these studies indicate that feedback may be biased by the human's expectation of the robot's performance and learning ability.

# 3 PROPOSED STUDY

People exhibit biases in their feedback that cannot be solely attributed to the task objectives they are trying to teach. Based on the previous section, we aim to study how a human teacher's expectation of the robot's performance influences their feedback. Particularly, we focus on bias that may be caused by the robot's *competency*, which we define as the robot's success rate for completing a task.

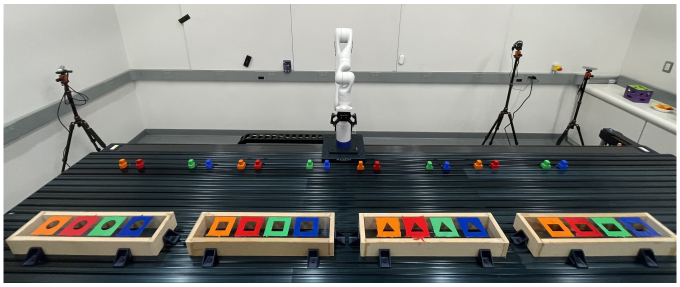

**Figure 2: Experimental setup: the robot will attempt to place blocks into the target hole, succeeding or failing based on a pre-set competency condition. As the robot makes each attempt, the participant may choose to intervene and correct the robot's motion as they deem necessary.**

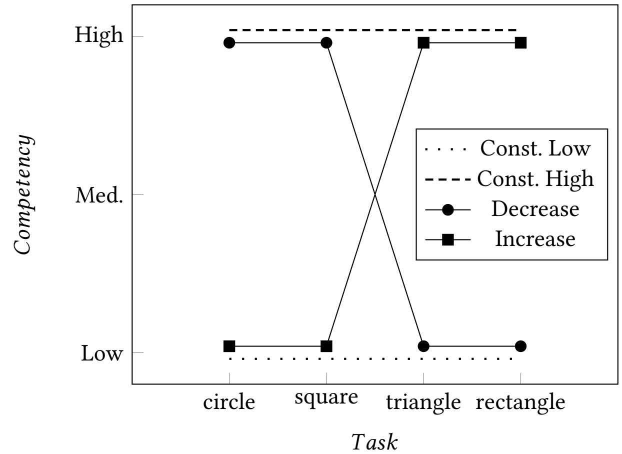


**Figure 3: Four Competency Conditions: *Const. Low:* The robot exhibits consistently low competency throughout all tasks. *Const. High:* The robot exhibits consistently high competency throughout all tasks. *Decrease:* The robot exhibits high competency in the first half of the tasks and low competency in the second half. *Increase:* The robot exhibits low competency in the first half of the tasks and high competency in the second half.**

Competency has been shown to be the most decisive factor that influences human trust towards the robot [26, 27].

Our overall research question is: **How does the robot's past and current competency at the task affect when and how people decide to intervene and correct the robot's motion?**

In this section, we propose a user study to address this question. Using a combination of quantitative measures, qualitative measures (e.g., usability, workload, and trust), and data analysis, we aim to study how people in each condition provide feedback differently to a robot arm via physical interaction.

## 3.1 Experimental Setup

We envision a series of pick-and-place tasks (Fig. 2) for the robot to perform under the supervision of the human participant. The robot's objective is to place each block into the hole with the same shape and color as the block. Participants will be welcomed to interrupt the robot's motion and provide a correction (i.e., kinesthetic feedback) whenever and however they see fit to guide the robot toward successfully completing the task (Fig. 1).

## 3.2 Conditions

Participants will be assigned to four robot competency conditions (visualized in Fig. 3) representing a diverse range of agent performance in a vanilla RL setting:

- **Consistently Low**: The robot exhibits consistently low competency throughout all tasks. This represents an agent that seemingly does not learn from feedback.
- **Consistently High**: The robot exhibits consistently high competency throughout all tasks. This represents an agent that requires little supervision.
- **Decreasing**: The robot exhibits high competency in the first half of the tasks and low competency in the second half. This condition resembles that of catastrophic forgetting.
- **Increasing**: The robot exhibits low competency in the first half of the tasks and high competency in the second half. This condition represents an agent that learns over time.

## 3.3 Hypotheses

We intend to analyze three ways (not necessarily independent) through which competency can influence people's feedback. Firstly, we will measure how the kinesthetic characteristics of each correction (i.e., torque, velocity, and displacement) are influenced by the robot's competency.

- **RQ-1**: Does the robot's prior competency have an effect on the effort, speed, and deviation from the nominal trajectory? If so, how long does this effect last?
- **H1**: In the high-competency condition, people will provide feedback with greater effort, velocity, and displacement from the nominal trajectory (compared to the low-competency condition).

Secondly, we investigate how human trust is afforded differently to the robot with different competency. Trust is a measure of the human teacher's confidence in the robot [28], and it is particularly relevant in the setting of correction feedback since humans provide feedback before the robot finishes a task, so whether the human chooses to intervene is an indication of the human's prediction of whether that robot will succeed. We believe that at what phase of the trajectory and how often a human intervenes can be leveraged to evaluate the human's trust on the robot.

- **RQ-2**: Does the robot's prior competency have an effect on people's trust of the robot?
- **H2a**: In the high-competency condition, people will intervene less frequently (i.e., predict that robot are less likely to fail) compared to the low-competency condition.

- **H2b**: In the high-competency condition, people will report higher measures of trust towards the robot compared to the low-competency condition.

Thirdly, we examine how competency shapes human effort in providing feedback. Many ML algorithms are extremely data-hungry. For instance, in preference learning, besides the individual demonstrations, an additional factorial-amount of pairwise comparison are needed, which is a significant labeling effort [13, 39]. Query-based robot learning also assumes that humans are readily available in their bandwidth [8, 20, 61], but there are inherent cost associated with providing feedback, whether being time, physical effort, cognitive load, or monetary loss. Existing work assumes human effort minimization as a regularizer while optimizing a main objective [10]. We believe the magnitude on which this regularization is enforced (i.e., human effort) is dependent on different robot competency. To test this theory, we define a normalized workspace area for each human participants, and we measure human effort as how often they step out of this workspace to provide feedback to the robot.

- **RQ-3**: Does the robot's prior competency have an effect on people's effort while providing feedback?
- **H3**: Given the same current robot competency, people will step out of their workspace more frequently and by a larger magnitude when the robot has previously exhibited a low competency at the task than when the robot has previously exhibited high competency.

## 4 CONCLUSION

Natural human feedback does not always comply with the assumptions built into many ML algorithms: particularly, the assumption that feedback is Markovian and time consistent. In order to develop lifelong, interactive learners, it is paramount that we understand and model how people continuously adapt their feedback based on their changing expectations of the robot's behavior. By acknowledging the dynamic nature of human-robot interaction and the need for online learning mechanisms that account for how people adapt their feedback based on the robot's competency, we can build models that inform how RL algorithm should interpret and learn from natural human feedback.

In this paper, we surveyed how prior work makes assumptions about how people provide feedback to robots. We proposed several research questions and hypotheses to guide future work on this topic and, finally, proposed a user study to evaluate how a robot's competency influences how people correct its motion. We expect that this work will reveal insights on how ML algorithms can more-effectively interpret and leverage natural human feedback.

## REFERENCES


[1] Christopher G Atkeson and Stefan Schaal. 1997. Robot learning from demonstration. In *ICML*, Vol. 97. 12–20.

[2] Agnes Axelsson and Gabriel Skantze. 2022. Multimodal User Feedback During Adaptive Robot-Human Presentations. *Frontiers in Computer Science* 3 (2022). https://doi.org/10.3389/fcomp.2021.741148

[3] Andrea Bajcsy, Dylan P Losey, Marcia K O'malley, and Anca D Dragan. 2017. Learning robot objectives from physical human interaction. In *Conference on Robot Learning*. PMLR, 217–226.

[4] Chris L Baker, Rebecca Saxe, and Joshua B Tenenbaum. 2009. Action understanding as inverse planning. *Cognition* 113, 3 (2009), 329–349.

[5] Richard Bellman. 1966. Dynamic programming. *Science* 153, 3731 (1966), 34–37.

[6] Dimitri Bertsekas. 2012. *Dynamic programming and optimal control: Volume I*. Vol. 4. Athena scientific.

[7] Tapomayukh Bhattacharjee, Ethan K Gordon, Rosario Scalise, Maria E Cabrera, Anat Caspi, Maya Cakmak, and Siddhartha S Srinivasa. 2020. Is more autonomy always better? exploring preferences of users with mobility impairments in robot-assisted feeding. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*. 181–190.

[8] Erdem Bıyık, Malayandi Palan, Nicholas C Landolfi, Dylan P Losey, and Dorsa Sadigh. 2019. Asking easy questions: A user-friendly approach to active reward learning. *arXiv preprint arXiv:1910.04365* (2019).

[9] Andreea Bobu, Andrea Bajcsy, Jaime F Fisac, Sampada Deglurkar, and Anca D Dragan. 2020. Quantifying hypothesis space misspecification in learning from human–robot demonstrations and physical corrections. *IEEE Transactions on Robotics* 36, 3 (2020), 835–854.

[10] Andreea Bobu, Andrea Bajcsy, Jaime F. Fisac, and Anca D. Dragan. 2018. Learning under Misspecified Objective Spaces. In *Proceedings of The 2nd Conference on Robot Learning (Proceedings of Machine Learning Research, Vol. 87)*, Aude Billard, Anca Dragan, Jan Peters, and Jun Morimoto (Eds.). PMLR, 796–805. https://proceedings.mlr.press/v87/bobu18a.html

[11] Andreea Bobu, Dexter RR Scobee, Jaime F Fisac, S Shankar Sastry, and Anca D Dragan. 2020. Less is more: Rethinking probabilistic models of human behavior. In *Proceedings of the 2020 acm/ieee international conference on human-robot interaction*. 429–437.

[12] Ralph Allan Bradley and Milton E Terry. 1952. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* 39, 3/4 (1952), 324–345.

[13] Daniel Brown, Wonjoon Goo, Prabhat Nagarajan, and Scott Niekum. 2019. Extrapolating beyond suboptimal demonstrations via inverse reinforcement learning from observations. In *International conference on machine learning*. PMLR, 783–792.

[14] Roland Buchner, Daniela Wurhofer, Astrid Weiss, and Manfred Tscheligi. 2013. Robots in time: How user experience in human-robot interaction changes over time. In *Social Robotics: 5th International Conference, ICSR 2013, Bristol, UK, October 27-29, 2013, Proceedings 5*. Springer, 138–147.

[15] Lawrence Chan, Andrew Critch, and Anca Dragan. 2021. Human irrationality: both bad and good for reward inference. *arXiv preprint arXiv:2111.06956* (2021).

[16] Sonia Chernova and Andrea L Thomaz. 2014. *Robot learning from human teachers*. Morgan & Claypool Publishers.

[17] Yuchen Cui, Qiping Zhang, Brad Knox, Alessandro Allievi, Peter Stone, and Scott Niekum. 2021. The empathic framework for task learning from implicit human feedback. In *Conference on Robot Learning*. PMLR, 604–626.

[18] Kevin Dorling, Jordan Heinrichs, Geoffrey G Messier, and Sebastian Magierowski. 2016. Vehicle routing problems for drone delivery. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 47, 1 (2016), 70–85.

[19] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. 2013. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 301–308.

[20] Tesca Fitzgerald, Pallavi Koppol, Patrick Callaghan, Russell Q. Wong, Reid Simmons, Oliver Kroemer, and Henny Admoni. 2022. INQUIRE: INteractive Querying for User-aware Informative REasoning. In *Proceedings of (CoRL) Conference on Robot Learning*.

[21] Tesca Fitzgerald, Elaine Short, Ashok Goel, and Andrea Thomaz. 2019. Human-guided Trajectory Adaptation for Tool Transfer. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* (Montreal QC, Canada) *(AAMAS '19)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1350–1358.

[22] Gaurav R Ghosal, Matthew Zurek, Daniel S Brown, and Anca D Dragan. 2023. The effect of modeling human rationality level on learning rewards from multiple feedback types. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 5983–5992.

[23] Cleotilde Gonzalez. 2023. Building Human-Like Artificial Agents: A General Cognitive Algorithm for Emulating Human Decision-Making in Dynamic Environments. *Perspectives on Psychological Science* (2023), 17456916231196766.

[24] Noah D Goodman, Chris L Baker, and Joshua B Tenenbaum. 2009. Cause and intent: Social reasoning in causal learning. In *Proceedings of the 31st annual conference of the cognitive science society*. Cognitive Science Society Amsterdam, 2759–2764.

[25] Noah D Goodman and Andreas Stuhlmüller. 2013. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science* 5, 1 (2013), 173–184.

[26] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors* 53, 5 (2011), 517–527.

[27] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. 2011. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors* 53, 5 (2011), 517–527. https://doi.org/10.1177/0018720811417254 arXiv:https://doi.org/10.1177/0018720811417254 PMID: 22046724.

[28] Erin Hedlund, Michael Johnson, and Matthew Gombolay. 2021. The Effects of a Robot's Performance on Human Teachers for Learning from Demonstration Tasks. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 207–215.

[29] José Hernández-Orallo. 2017. Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement. *Artificial Intelligence Review* 48 (2017), 397–447.

[30] Hong Jun Jeon, Smitha Milli, and Anca Dragan. 2020. Reward-rational (implicit) choice: A unifying formalism for reward learning. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 4415–4426. https://proceedings.neurips.cc/paper_files/paper/2020/file/2f10c1578a0706e06b6d7db6f0b4a6af-Paper.pdf

[31] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. 1996. Reinforcement learning: A survey. *Journal of artificial intelligence research* 4 (1996), 237–285.

[32] Tasneem Kaochar, Raquel Torres Peralta, Clayton T Morrison, Ian R Fasel, Thomas J Walsh, and Paul R Cohen. 2011. Towards understanding how humans teach robots. In *User Modeling, Adaption and Personalization: 19th International Conference, UMAP 2011, Girona, Spain, July 11-15, 2011. Proceedings 19*. Springer, 347–352.

[33] Elizabeth S Kim, Dan Leyzberg, Katherine M Tsui, and Brian Scassellati. 2009. How people talk when teaching a robot. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. 23–30.

[34] W Bradley Knox and Peter Stone. 2009. Interactively shaping agents via human reinforcement: The TAMER framework. In *Proceedings of the fifth international conference on Knowledge capture*. 9–16.

[35] W. Bradley Knox and Peter Stone. 2009. Interactively shaping agents via human reinforcement: the TAMER framework. In *Proceedings of the Fifth International Conference on Knowledge Capture* (Redondo Beach, California, USA) *(K-CAP '09)*. Association for Computing Machinery, New York, NY, USA, 9–16. https://doi.org/10.1145/1597735.1597738

[36] Jens Kober, J. Andrew Bagnell, and Jan Peters. 2013. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research* 32, 11 (2013), 1238–1274. https://doi.org/10.1177/0278364913495721 arXiv:https://doi.org/10.1177/0278364913495721

[37] Cassidy Laidlaw and Anca Dragan. 2022. The boltzmann policy distribution: Accounting for systematic suboptimality in human models. *arXiv preprint arXiv:2204.10759* (2022).

[38] Stanislao Lauria, Guido Bugmann, Theocharis Kyriacou, Johan Bos, and Ewan Klein. 2001. Personal robot training via natural-language instructions. *IEEE Intelligent systems* 16, 3 (2001), 38–45.

[39] Kimin Lee, Laura Smith, and Pieter Abbeel. 2021. Pebble: Feedback-efficient interactive reinforcement learning via relabeling experience and unsupervised pre-training. *arXiv preprint arXiv:2106.05091* (2021).

[40] Jesse Levinson, Jake Askeland, Jan Becker, Jennifer Dolson, David Held, Soeren Kammel, J Zico Kolter, Dirk Langer, Oliver Pink, Vaughan Pratt, et al. 2011. Towards fully autonomous driving: Systems and algorithms. In *2011 IEEE intelligent vehicles symposium (IV)*. IEEE, 163–168.

[41] David Lindner and Mennatallah El-Assady. 2022. Humans are not Boltzmann Distributions: Challenges and Opportunities for Modelling Human Feedback and Interaction in Reinforcement Learning. arXiv:2206.13316 [cs.LG]

[42] Robert L Linn. 2008. *Measurement and assessment in teaching*. Pearson Education India.

[43] Dylan P. Losey, Andrea Bajcsy, Marcia K. O'Malley, and Anca D. Dragan. 2022. Physical interaction as communication: Learning robot objectives online from human corrections. *The International Journal of Robotics Research* 41, 1 (2022), 20–44. https://doi.org/10.1177/02783649211050958 arXiv:https://doi.org/10.1177/02783649211050958

[44] Dylan P. Losey and Marcia K. O'Malley. 2018. Including Uncertainty when Learning from Human Corrections. In *Proceedings of The 2nd Conference on Robot Learning (Proceedings of Machine Learning Research, Vol. 87)*, Aude Billard, Anca Dragan, Jan Peters, and Jun Morimoto (Eds.). PMLR, 123–132. https://proceedings.mlr.press/v87/losey18a.html

[45] R Duncan Luce. 1977. The choice axiom after twenty years. *Journal of mathematical psychology* 15, 3 (1977), 215–233.

[46] R Duncan Luce. 2005. *Individual choice behavior: A theoretical analysis*. Courier Corporation.

[47] Nicole Mirnig, Gerald Stollnberger, Markus Miksch, Susanne Stadler, Manuel Giuliani, and Manfred Tscheligi. 2017. To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI* 4 (2017), 21.

[48] Takayuki Osogami and Makoto Otsuka. 2014. Restricted Boltzmann machines modeling human choice. *Advances in Neural Information Processing Systems* 27 (2014).

[49] Lorenzo Pareschi and Giuseppe Toscani. 2014. Wealth distribution and collective knowledge: a Boltzmann approach. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 372, 2028 (2014), 20130396.

[50] Mikkel Rath Pedersen, Lazaros Nalpantidis, Rasmus Skovgaard Andersen, Casper Schou, Simon Bøgh, Volker Krüger, and Ole Madsen. 2016. Robot skills for manufacturing: From concept to industrial deployment. *Robotics and Computer-Integrated Manufacturing* 37 (2016), 282–291.

[51] Brian S Peters, Priscila R Armijo, Crystal Krause, Songita A Choudhury, and Dmitry Oleynikov. 2018. Review of emerging surgical robotic technology. *Surgical endoscopy* 32 (2018), 1636–1655.

[52] Silviu Pitis, Harris Chan, Kiarash Jamali, and Jimmy Ba. 2020. An inductive bias for distances: Neural nets that respect the triangle inequality. *arXiv preprint arXiv:2002.05825* (2020).

[53] Robin L Plackett. 1975. The analysis of permutations. *Journal of the Royal Statistical Society Series C: Applied Statistics* 24, 2 (1975), 193–202.

[54] Martin L Puterman. 2014. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

[55] Divya Ramesh, Anthony Z Liu, Andres J Echeverria, Jean Y Song, Nicholas R Waytowich, and Walter S Lasecki. 2020. Yesterday's Reward is Today's Punishment: Contrast Effects in Human Feedback to Reinforcement Learning Agents. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. 1090–1097.

[56] Dorsa Sadigh, Anca D Dragan, Shankar Sastry, and Sanjit A Seshia. 2017. *Active preference-based learning of reward functions*.

[57] Simon Stepputtis, Joseph Campbell, Mariano Phielipp, Stefan Lee, Chitta Baral, and Heni Ben Amor. 2020. Language-conditioned imitation learning for robot manipulation tasks. *Advances in Neural Information Processing Systems* 33 (2020), 13139–13150.

[58] Andrea Lockerd Thomaz, Cynthia Breazeal, et al. 2006. Reinforcement learning with human teachers: Evidence of feedback and guidance with implications for learning performance. In *Aaai*, Vol. 6. Boston, MA, 1000–1005.

[59] Christopher JCH Watkins and Peter Dayan. 1992. Q-learning. *Machine learning* 8 (1992), 279–292.

[60] Marco A Wiering and Martijn Van Otterlo. 2012. Reinforcement learning. *Adaptation, learning, and optimization* 12, 3 (2012), 729.

[61] Christian Wirth, Riad Akrour, Gerhard Neumann, Johannes Fürnkranz, et al. 2017. A survey of preference-based reinforcement learning methods. *Journal of Machine Learning Research* 18, 136 (2017), 1–46.

[62] Brian D Ziebart, J Andrew Bagnell, and Anind K Dey. 2010. Modeling interaction via the principle of maximum causal entropy. (2010).


Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009