Leveraging Interpretable Human Models to Personalize AI interventions for Behavior Change

Doctoral Consortium

Eura Nofshin
Harvard University
Cambridge, USA
eurashin@g.harvard.edu

ABSTRACT

Many important areas of behavior change, such as wellness or education, are frictionful; they require individuals to expend effort over a long period of time with little immediate gratification. Because of this, humans often act sub-optimally with respect to their stated long-term goal. Here, an artificial intelligence (AI) agent can provide personalized behavioral interventions to correct human policies. The AI must personalize rapidly (before the individual has a chance to disengage) and interpretably, to aid our scientific understanding of the behavioral interventions. This work focuses on crafting small, interpretable models of the human that capture the mechanism behind the human agent's sub-optimal policies. These human models provide the AI with enough inductive bias to quickly learn intervention policies for each individual it encounters.

KEYWORDS

Reinforcement learning; Personalization; Agent-based modeling of humans; Bounded rationality

ACM Reference Format:
Eura Nofshin. 2024. Leveraging Interpretable Human Models to Personalize AI interventions for Behavior Change: Doctoral Consortium. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

In many AI+human applications for behavior change, AI agents assist the human in performing *frictionful* tasks, where making progress toward the human's goal requires sustained effort over time with little immediate gratification. Examples include physical therapy programs, adherence to scheduled medication, or passing an online course. Two key challenges for AI agents in these settings are (1) rapid personalization and (2) learning interpretable intervention policies. In frictionful tasks, since effort exerted by the human does not reap immediate benefits, the AI agent must learn a personalized intervention policy from a small number of interactions for each human, or risk disengagement. These policies must also be interpretable, so that behavioral experts can discover which interventions work for which individuals, and why.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

Current RL approaches have two major drawbacks when used to solve for the AI agent's intervention policy. First, most planning methods are too data-intensive for our online setting. For example, online algorithms in robotics require thousands of interactions to learn reasonable policies (e.g. in Tebbe et al. [25], Thabet et al. [26], Yang et al. [28]), but in frictionful tasks, we are limited to *tens to hundreds* of interactions per person [27]. Second, existing planning methods solve for the AI agent's optimal policy by modeling the human as a black-box transition or value function. Unfortunately, in learning black-box representations of the human, we lose the ability to interpretably attribute human behavior to the model learned by the AI. In this work, I target interpretable and effective planning by the AI agent, through the use of carefully crafted human models. To create and work with these human models, this body of work bridges across machine learning, behavioral science, and human-computer interaction (HCI).

Behavioral science provides us with formal theories and models of human decision-making in frictionful tasks. However, there is a gap in how to instantiate high-level constructs from behavioral science (such as temporal discounting in humans) into computational models (which describe the scale and functional forms of how temporal discounting changes over time) [9]. Machine learning offers paradigms that can elegantly encode the behavioral assumptions needed to form computational models. For example, temporal discounting from behavior science [20] can be connected to the discount factor, γ, which is part of a Markov Decision Process (MDP). Such explicit computational models are powerful because they (1) provide the link between behavioral assumptions and the observed data; and (2) can be incorporated into the AI agent's planning. But, models that show promise in theory and simulation must be tested with real end-users, and user studies guided by HCI design principles can evaluate effectiveness.

To fill in these gaps, I aim to address the core questions below:

Q1. What is the model? Reducing complex behavioral models to simpler ones that can be used for AI planning
Q2. How to learn the human model? Updating the human model to individual-level data observed online.
Q3. How to use the human model for intervention? Learning and testing intervention policies that work with real users.

2 THE BEHAVIOR MODEL RL (BMRL) FRAMEWORK FOR AI INTERVENTIONS

I define a formal framework, called BMRL, in which an AI agent learns to intervene on a human. Like previous work where RL has been used to plan interventions, the AI's reward and transitions depend on the human's reactions. For example, the AI will maximize a reward related to whether the human performs the goal-oriented behavior of interest (e.g. the number of steps in a physical activity application [11] or the quality of brushing in an oral health application [27]).

BMRL incorporates sophisticated models of the human's decision-making, which are informed by behavoiral science, into the AI environment. Grounded in literature that treats humans as sequential decision-makers (e.g. [13, 19, 23, 24, 31]), we model the human as a Reinforcement Learning (RL) agent planning under a "maladapted" Markov Decision Process (MDP). In maladapted human MDPs, the optimal policy does not reach the human's stated goal (for example, the goal of an active lifestyle). One example of a maladapted MDP is having an extremely low discount rate, γ. This represents myopic decision-making, wherein an individual forgoes the long-term goal (being active) to avoid experiencing friction in the short-term (unpleasantness of exercising). Unlike prior work that is limited to inferring the source of suboptimality (e.g. [2, 5, 7, 10, 14, 16, 30]) or intervening on human rewards/states [4, 12, 15, 22, 29, 31], in our work, the AI agent *intervenes* on *any* of the human's maladapted MDP parameters to help them achieve their long-term goals.

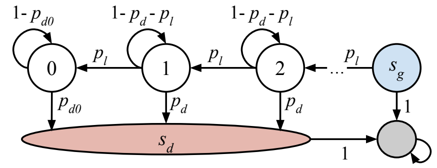
3 CHAINWORLD: A SIMPLE HUMAN MDP THAT IS BEHAVIORALLY GROUNDED

The defining aspect of BMRL is the definition of the human MDP. When used to inform the AI agent's policy, the human MDP can be simpler than expected; in [18], I introduce the concept of "AI equivalence" to identify a class of more complex human models for which AI policies learned in a simpler one can be lifted with provably no loss of performance. Simpler MDPs are preferred because they require less data to learn (**Q2**) and are easier to examine (can be more interpretable). Then, I introduce "chainworlds," a class of simple human MDPs (shown in fig. 1) (**Q1**). I prove that chainworlds produce equivalent AI optimal policies as if the AI had used a more behaviorally complex model, and produce results such as fig. 2 , which demonstrates chainworlds allow the AI to learn quickly (**Q3**).

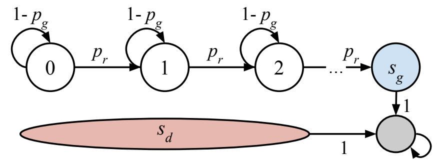
4 ONGOING AND FUTURE WORK

User studies. Related to **Q3**, I am developing a flashcard study app to test whether our theoretical and simulated results will hold on real human users. Flashcards are frictionful because the user must put in consistent effort to learn and retain the information to make progress toward their learning goal. Our measure of success will be whether or not the user meets their target number of study sessions for the week.

Relaxing behavioral assumptions of chainworld. Related to **Q1**, our chainworld made several simplifying assumptions regarding the human MDP, that if relaxed, would be interesting future work. For example, I avoided a POMDP formulation of the AI agent by assuming that there were no delayed effects of the AI's actions on the human MDP. However, habituation (reduced effectiveness of repeated interventions) is a well-studied phenomenon in the digital intervention space (e.g. [8]). Furthermore, I avoided the complexity of multi-agent RL by assuming that the human is *not learning*, and instead, is solving an (implicitly) known MDP. Finally, humans have



(a) When the human abstains from the behavior, they may lose progress or slip into disengagement s_d.



(b) When the human performs the behavior, they may transition toward the goal state s_g.

Figure 1: Graphical representation of the chainworld. Each state on the chain represents the progress toward the goal state, s_g.

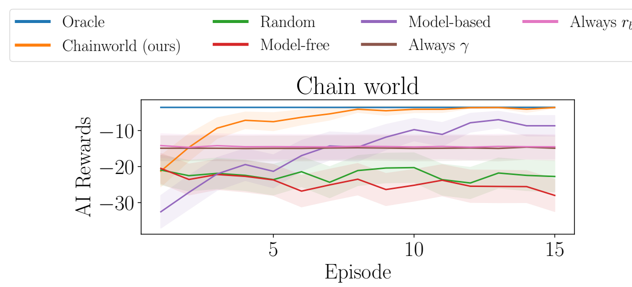


Figure 2: Using our chainworld model (orange), the AI reaches oracle-level performance (blue) quickest. Plot is AI rewards (y-axis) over multiple episodes (x-axis). Lines in upper-left corner mean the AI personalizes more quickly.

been observed to *hyperbolically* discount future rewards, as opposed to the exponential discounting assumed by the MDP formalism [21]. This formalism would have to change, as in Fedus et al. [6], in order to allow our human agent to perform other types of discounting.

How much personalization is possible? Related to **Q2**, how precisely the AI can infer the human's MDP parameters depends on the data. When the data is *human demonstration data in a single environment*, there is inherent non-identifiability in the human MDP parameters, as we show in [1]. For example, a human with myopic discounting vs. a human that perceives low rewards on the goal state will both behave according to goal-avoidant policies. When there is demonstration data from *multiple environments*, it is possible to combat non-identifiability by aggregating information from demonstrations across environments [3], but this becomes a difficult search problem over which environment to show the user. When surveys are used to collect *self-reported* data, the trade-offs in information gained from self-report vs. direct observation of behavior has yet to be explored. Finally, across all data sources, our inference over the human model parameters must consider the noisiness and scale of data that is available, which I have explored in Shin et al. [17].

REFERENCES

[1] Lars L Ankile, Brian S Ham, Kevin Mao, Eura Shin, Siddharth Swaroop, Finale Doshi-Velez, and Weiwei Pan. 2023. Discovering User Types: Mapping User Traits by Task-Specific Behaviors in Reinforcement Learning. arXiv:2307.08169 [cs.AI]

[2] 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, California USA, 783–792.

[3] Thomas Kleine Buening and Christos Dimitrakakis. 2022. Environment Design for Inverse Reinforcement Learning. arXiv:2210.14972 [cs.AI]

[4] Kaiqi Chen, Jeffrey Fong, and Harold Soh. 2022. Mirror: Differentiable deep social projection for assistive human-robot communication. In *Robotics: Science and Systems*. Robotics: Science and Systems, New York USA.

[5] Owain Evans, Andreas Stuhlmüller, and Noah Goodman. 2016. Learning the preferences of ignorant, inconsistent agents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30. AAAI, Arizona USA.

[6] William Fedus, Carles Gelada, Yoshua Bengio, Marc G. Bellemare, and Hugo Larochelle. 2019. Hyperbolic Discounting and Learning over Multiple Horizons. arXiv:1902.06865 [stat.ML]

[7] Babatunde H Giwa and Chi-Guhn Lee. 2021. Estimation of Discount Factor in a Model-Based Inverse Reinforcement Learning Framework. https://hdl.handle.net/1807/125220

[8] Lisa Gotzian. 2023. Modeling the decreasing intervention effect in digital health: a computational model to predict the response for a walking intervention. https://doi.org/10.31219/osf.io/6v7d5

[9] Eric B Hekler, Susan Michie, Misha Pavel, Daniel E Rivera, Linda M Collins, Holly B Jimison, Claire Garnett, Skye Parral, and Donna Spruijt-Metz. 2016. Advancing models and theories for digital behavior change interventions. *American journal of preventive medicine* 51, 5 (2016), 825–832.

[10] Daniel Jarrett, Alihan Hüyük, and Mihaela Van Der Schaar. 2021. Inverse decision modeling: Learning interpretable representations of behavior. In *International Conference on Machine Learning*. PMLR, PMLR, Virtual, 4755–4771.

[11] Peng Liao, Kristjan Greenewald, Predrag Klasnja, and Susan Murphy. 2020. Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–22.

[12] Yonatan Mintz, Anil Aswani, Philip Kaminsky, Elena Flowers, and Yoshimi Fukuoka. 2023. Behavioral analytics for myopic agents. *European Journal of Operational Research* 310, 2 (2023), 793–811.

[13] Yael Niv. 2009. Reinforcement learning in the brain. *Journal of Mathematical Psychology* 53, 3 (2009), 139–154.

[14] Siddharth Reddy, Anca D. Dragan, and Sergey Levine. 2018. Where do you think you're going? inferring beliefs about dynamics from behavior. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Montréal, Canada) *(NIPS'18)*. Curran Associates Inc., Red Hook, NY, USA, 1461–1472.

[15] Siddharth Reddy, Sergey Levine, and Anca Dragan. 2021. Assisted perception: optimizing observations to communicate state. In *Conference on Robot Learning*. PMLR, PMLR, London UK, 748–764.

[16] Rohin Shah, Noah Gundotra, Pieter Abbeel, and Anca Dragan. 2019. On the feasibility of learning, rather than assuming, human biases for reward inference. In *International Conference on Machine Learning*. PMLR, PMLR, California, USA, 5670–5679.

[17] Eura Shin, Predrag Klasnja, Susan Murphy, and Finale Doshi-Velez. 2023. Online model selection by learning how compositional kernels evolve. *Transactions on Machine Learning Research* (2023). https://openreview.net/forum?id=23WZFQBUh5

[18] Eura Shin, Siddharth Swaroop, Weiwei Pan, Susan Murphy, and Doshi-Velez Finale. 2024. Reinforcement learning intervention on boundedly rational human agents in frictionful tasks. arXiv:2401.14923 [cs.AI]

[19] Hanan Shteingart and Yonatan Loewenstein. 2014. Reinforcement learning and human behavior. *Current opinion in neurobiology* 25 (2014), 93–98.

[20] Giles W Story, Ivo Vlaev, Ben Seymour, Ara Darzi, and Raymond J Dolan. 2014. Does temporal discounting explain unhealthy behavior? A systematic review and reinforcement learning perspective. *Frontiers in behavioral neuroscience* 8 (2014), 76.

[21] Giles W Story, Ivo Vlaev, Ben Seymour, Ara Darzi, and Raymond J Dolan. 2014. Does temporal discounting explain unhealthy behavior? A systematic review and reinforcement learning perspective. *Frontiers in behavioral neuroscience* 8 (2014), 76.

[22] Aaquib Tabrez, Shivendra Agrawal, and Bradley Hayes. 2019. Explanation-Based Reward Coaching to Improve Human Performance via Reinforcement Learning. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, Korea, 249–257. https://doi.org/10.1109/HRI.2019.8673104

[23] Véronique A Taylor, Isabelle Moseley, Shufang Sun, Ryan Smith, Alexandra Roy, Vera U Ludwig, and Judson A Brewer. 2021. Awareness drives changes in reward value which predict eating behavior change: Probing reinforcement learning using experience sampling from mobile mindfulness training for maladaptive eating. *Journal of behavioral addictions* 10, 3 (2021), 482–497.

[24] Véronique A Taylor, Isabelle Moseley, Shufang Sun, Ryan Smith, Alexandra Roy, Vera U Ludwig, and Judson A Brewer. 2021. Awareness drives changes in reward value which predict eating behavior change: Probing reinforcement learning using experience sampling from mobile mindfulness training for maladaptive eating. *Journal of behavioral addictions* 10, 3 (2021), 482–497.

[25] Jonas Tebbe, Lukas Krauch, Yapeng Gao, and Andreas Zell. 2021. Sample-efficient reinforcement learning in robotic table tennis. In *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, IEEE, China, 4171–4178.

[26] Mohammad Thabet, Massimiliano Patacchiola, and Angelo Cangelosi. 2019. Sample-efficient deep reinforcement learning with imaginary rollouts for human-robot interaction. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, IEEE, Macau, 5079–5085.

[27] Anna L Trella, Kelly W Zhang, Inbal Nahum-Shani, Vivek Shetty, Finale Doshi-Velez, and Susan A Murphy. 2022. Designing reinforcement learning algorithms for digital interventions: pre-implementation guidelines. *Algorithms* 15, 8 (2022), 255.

[28] Yuxiang Yang, Ken Caluwaerts, Atil Iscen, Tingnan Zhang, Jie Tan, and Vikas Sindhwani. 2020. Data efficient reinforcement learning for legged robots. In *Conference on Robot Learning*. PMLR, PMLR, Virtual, 1–10.

[29] Guanghui Yu and Chien-Ju Ho. 2022. Environment Design for Biased Decision Makers. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. International Joint Conferences on Artificial Intelligence Organization, Austria, 592–598.

[30] Tan Zhi-Xuan, Jordyn Mann, Tom Silver, Josh Tenenbaum, and Vikash Mansinghka. 2020. Online bayesian goal inference for boundedly rational planning agents. *Advances in neural information processing systems* 33 (2020), 19238–19250.

[31] Mo Zhou, Yonatan Mintz, Yoshimi Fukuoka, Ken Goldberg, Elena Flowers, Philip Kaminsky, Alejandro Castillejo, and Anil Aswani. 2018. Personalizing mobile fitness apps using reinforcement learning. In *CEUR workshop proceedings*, Vol. 2068. NIH Public Access, CEUR workshop proceedings, Japan.