Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)
Early Career Track

Irrational, but Adaptive and Goal Oriented: Humans Interacting with Autonomous Agents

Amos Azaria
Computer Science Department, Ariel University, Israel
amos.azaria@ariel.ac.il

Abstract

Autonomous agents that interact with humans are becoming more and more prominent. Currently, such agents usually take one of the following approaches for considering human behavior. Some methods assume either a fully cooperative or a zero-sum setting; these assumptions entail that the human's goals are either identical to that of the agent, or their opposite. In both cases, the agent is not required to explicitly model the human's goals and account for humans' adaptation nature. Other methods first compose a model of human behavior based on observing human actions, and then optimize the agent's actions based on this model. Such methods do not account for how the human will react to the agent's actions and thus, suffer an overestimation bias. Finally, other methods, such as model free reinforcement learning, merely learn which actions the agent should take at which states. While such methods can, theoretically, account for human adaptation nature, since they require extensive interaction with humans, they usually run in simulation. By not considering the human's goals, autonomous agents act selfishly, lack generalization, require vast amounts of data, and cannot account for human's strategic behavior. Therefore, we call for pursuing solution concepts for autonomous agents interacting with humans that consider the human's goals and adaptive nature.

1 Introduction

Autonomous agents interacting with humans are becoming ubiquitous. They are present in smart home environments, such as Alexa, Cortana and Google Assistant, on the internet, as a form of chatbots or assisting bots, and in the physical world, such as robotic vacuum cleaners and mopers. Clearly, the presence of such agents will grow significantly in the years to come, including new areas, in which autonomous agents are only beginning to enter, such as autonomous vehicles, drones and other autonomous robots. Autonomous agents also interact with humans in competitive game environments, such as Chess, Go, Dota, and Starcraft [Skinner and Walmsley, 2019; Silver *et al.*, 2016; Hsu *et al.*, 1995].

Autonomous agents attempting to proficiently interact with humans must model human behavior. Such agents cannot rely on game theory or platforms assuming that humans are perfectly rational for composing a human model, as people often deviate from what is thought to be rational behavior. People are affected by a variety of factors: a lack of knowledge of one's own preferences, the effects of the task complexity, framing effects, the interplay between emotion and cognition, the problem of self-control, the value of anticipation, future discounting, anchoring and many other effects [Tversky and Kahneman, 1981; Loewenstein, 2000; Ariely *et al.*, 2003; Camerer, 2003]. Therefore, algorithmic approaches that use a pure theoretically analytic objective often perform poorly with real humans [Peled *et al.*, 2011; Nay and Vorobeychik, 2016]. Nevertheless, the concept of using a utility function to explain human behavior is very common and has been widely used in economics and psychology for centuries. Prior to the 1700s it was common to use the expected monetary gain as the utility function, which people are assumed to maximize. The expected utility hypothesis, formulated by Daniel Bernoulli, states that people attempt to maximize the expected utility, rather than their expected monetary gain. People may assign some value to a total wealth of $\$100,000$, but only double that value to a total wealth of $\$1,000,000$. The Prospect theory and the cumulative prospect theory, which are a further refinement to the expected utility hypothesis, awarded Daniel Kahneman a Nobel prize [Tversky and Kahneman, 1992]. It is clearly a fundamental assumption that humans performing some task, are doing it in order to achieve some goal. Therefore, modeling human behavior must account for human goals.

Unfortunately, rather than accounting for human goals, a widely common assumption made by many works developing agents interacting with humans, is that an environment is either fully cooperative, and the agent's goal is identical to that of the human's or fully competitive, i.e., a zero-sum game. This assumption allows agent developers to ignore the human goals, and concentrate only on maximizing the agent's utility function. While zero-sum and fully cooperative games are simpler to analyze, it is nearly impossible to find *any* real-life interaction between a group of humans that adheres to one of these assumptions. For example, when two human players

play a zero-sum board, card or sport game, their goal is usually to enjoy the interaction. Albeit, the winner might enjoy the overall experience better. If two human players were to play a true zero-sum game, one would need to kill the other in order to gain maximal utility. Furthermore, even a life-or-death gunfight pistol duel cannot be seen as a zero-sum game, as each of the players may attempt to not show, escape, give-up or only injure the opponent—all actions leading to outcomes that are not directly opposite. Similarly, fully cooperative games are not present in real-life either. Consider a married couple; they surely have some shared goals, such as raising their children and living in a place they are happy to be in. However, each individual has goals she cares more about than others, such as success at her own career or social life. Even when considering a limited task given to a group of people, some might want to be more dominant and instruct the others, which may, in turn, want to come up with a solution themselves.

The same observation is true also for agents interacting with humans. While a chatbot at a website attempts to assist a user, its goal is usually to sell a product, which might not be directly aligned with the user's goal. Navigation apps may attempt to navigate users via specific locations in order to gain extra information. Home assistants may attempt to encourage a user to sign-up for a product or merely attempt to gain additional data by invoking responses that do not necessarily directly increase user satisfaction. Therefore, any agent interacting with a human must account for the fact that the human's utility function is likely to be different than its own. While there are some works on a group of complex games, which are non-zero sum nor fully cooperative, named social dilemmas, these games are merely a form of generalization of the iterated prisoner's dilemma, in which each agent may either decide to cooperate or defect [Sandholm and Crites, 1996; Wang *et al.*, 2016], and most work in this field considers autonomous agents only, rather than interaction with humans [Jaques *et al.*, 2019]. Similarly, there exists some work considering the theory of mind of agents interacting with each-other, and accounting for the agents' goals. These works as well do not consider interaction with humans [Freire *et al.*, 2019]. One notable exception, though in the context of negotiation, is the colored trails game, which was developed to allow humans and agents to interact with each-other [Grosz *et al.*, 2004].

A common approach for developing agents interacting with humans in a general game (which is neither zero-sum nor fully cooperative) is by encapsulating human behavior into a fixed model and ignoring the human's utility. This is usually performed by using machine learning techniques on a dataset, and possibly by also building upon psychological factors and human decision-making theory. The human behaviour model is then used by a planner to interact with humans [Gal and Pfeffer, 2007; Subrahmanian, 2000; Rosenfeld and Kraus, 2011; Azaria *et al.*, 2012]. Other approaches, such as model free reinforcement learning, treat the human as a part of the environment and merely learn which actions the agent should take at which situations, in order to maximize its own reward [Carroll *et al.*, 2019]. While such methods can, theoretically, account for human adaptation nature, since they require extensive interaction with humans, they usually run in simulation. Furthermore, an agent observing a sequence of actions performed by a human must ask why these actions were performed. Predicting future actions without accounting for the human's goals, is like predicting future words of an answer without accounting for the question asked.

There are several caveats with current approaches. The first is that current approaches always result in completely egoistic, selfish agents. One cannot expect taking others' goals into account if those are not considered at all, as in the black-box behavior approach. Environments modeled as a zero-sum game do not attempt to please the opponent, and those modeled as fully cooperative, assume humans have the exact same utility function as the agent, and thus are ignored. The second caveat is generalization; models ignoring the fact that humans attempt to pursue some value, cannot easily generalize to situations in which those humans attempt to pursue different values. Furthermore, even when the utility function is assumed to be fixed, ignoring it must require some compensation in terms of the dataset size. That is, modeling human behavior based on their actions alone must require a much larger dataset than a model that also considers human goals. Clearly, achieving human data is usually very costly. In addition, the dataset is unlikely to contain sufficient samples from situations similar to those encountered by the agent, since in many cases, the dataset is gathered before the agent is fully developed and the agent requires this dataset for its development. Composing a model based on limited resources or from a biased distribution without considering the human goals will result in an inaccurate model, which, in turn, will result in a suboptimal agent. Finally, an agent acting according to a model that does not account for user goals, cannot anticipate that the humans may adapt to its presence and change their behavior in order to achieve their original goal and might suffer from an overestimation bias. For example, if an autonomous vehicle stops for pedestrians in the middle of a highway road when it predicts that they attempt to cross, people may be tempted to cross a highway despite it being unsafe. Similarly, mischievous children, with no intention of crossing, may enjoy causing the vehicle to stop.

2 Research Directions

The primary goal of this call is to develop methods that allow agents to proficiently interact with humans, in a general complex environment, by considering not only human actions, but also human utility and adaptive nature.

2.1 Developing a Theoretical Framework

There is a need for developing a theoretical framework that will cover all situations and conditions that may be of interest, grounded in mathematical concepts. The theoretical framework should include several dimensions: (i) The human goals, or utility function may be observed by the agent, only by the human, or not available at all in the system (i.e., the human may not be fully aware of her own goals). (ii) The number of humans and agents in the system. (iii) The scope of interaction; a complex environment entails multiple actions

for each agent but each episode in which an agent interacts with the same human may be short, or long. Actions may be performed simultaneously or alternating. (iv) Available knowledge; a game may be of perfect knowledge, in which both players know all information, one side may have an information gain over the other, or both sides may have different information provided to them (either at the beginning of the game or throughout the game). (v) Communication mode between the agent and the humans; is this communication limited to actions, or is it broader? Perhaps enabling some form of speech. If the players differ in their available information, communication may include also sharing these pieces of information. Each condition should be associated with one or more testbed games that enable the evaluation of advancements in that condition.

2.2 Developing Solution Concepts

The main research direction should be toward the development of solution concepts. We present several options, but we expect that many other solution concepts will emerge. (i) The simplest solution concept is to design an agent that does not adapt at all to the human actions. This is basically a simple rule based agent, whose actions are obvious and predictable to any human who interacts with it. Such an agent will benefit from 'the power of stupidity'. This resembles trains traveling on a track; since they cannot suddenly stop, people have learned not to travel on train tracks, or do so with exceptional attention. If an agent can communicate that it does not care about the humans and always acts in a certain (and predictable) manner, the humans would be required to adapt to it and act in accordance. This shifts coordination and responsibility from the agent to the human. While such a solution does not account for the human's goals directly, it does assume that the human behaves strategically based on her own goals and thus, will adapt to the agent. However, while such a solution may be beneficial for the short-term and in specific settings, it is clearly far from the expectations we have from intelligent autonomous agents. We cannot risk a 'stupid' autonomous vehicle hitting an impaired child or participating in a deadly accident with another vehicle only because the human in that vehicle made a mistake.

(ii) Our next solution concept is simple as well, and was shown useful in several works. Since the human model is not accurate, one might consider adding a noise term to it or assume a stochastic human model such as quantile response. That is, the human has some distribution over the possible actions. This solution concept may reduce the over-estimation bias. The human's distribution over the possible actions may take into account a human utility function. This solution concept, however, does not account for humans' adaptive nature.

(iii) Another solution concept is the use of social optimization. That is, rather than optimizing toward the agent's goal only, the agent attempts to achieve a higher utility for itself by optimizing toward a linear combination of the agent's and the human's utility function. This solution concept still requires the use of a human model. However, since a human model is usually based on a limited data-set size, it is likely to be inaccurate. We expect that optimizing toward a linear combination will be beneficial for the agent, since the humans are likely to try and optimize their own utility function, so they are likely to deviate from the human model in a way that will indeed maximize their utility function. By optimizing toward a linear combination, the agent acts as if it already accounts for these deviations and is therefore more likely to adapt to them. Furthermore, reciprocation and cooperation may result in the human returning a favor. We define the parameter β, a value between 0 and 1, that quantifies the degree to which the agent considers its own outcome and the human's outcome. Namely, let $u(A)$ be the utility function of agent A, and $u(H)$ be the utility function of the human H; instead of optimizing toward $u(A)$, the agent optimizes toward $\beta u(A) + (1 - \beta)u(H)$. This parameter should account for the correlation between the outcome of the two players as well as for the accuracy of the human model. That is, a higher correlation between the outcomes allows the agent to be more altruistic, and a more accurate human model allows the agent to be more selfish. Our previous work has demonstrated the efficiency of this solution concept [Azaria *et al.*, 2016; Shapira and Azaria, 2022].

(iv) Alternating utility. This approach is somewhat similar to the social optimization approach, except that here, the agent sometimes acts with an attempt to optimize its own utility function and sometimes acts with an attempt to optimize the human's utility function. This approach is more suitable to situations in which the agent can communicate its current phase to the human (possibly by performing some gesture). The concept somewhat reassembles the idea of moral licensing, which states that if a person does a good deed, she is more likely to perform a bad deed, when compared to someone that has initially performed a bad deed and may feel guilt [Blanken *et al.*, 2015]. The portion of time that the autonomous agent acts in an attempt to optimize its own utility function is determined by some parameter, τ, which, like β, must depend on the extent of which the two utility functions differ. We believe that humans will understand this concept and will take advantage when the agent is being polite and allowing them to pursue their own utility function, and be willing to cooperate with the agent when it's the agent's 'turn' to optimize toward its own utility function. This approach may perform better if it is combined with the previous approach, as even if the agent is being altruistic, at the moment, it is not reasonable to completely ignore its own utility function (and encounter, for example, a huge loss to itself). Therefore, the agent may alternate between a high β value phase, close to 1, indicating an altruism phase, and a low β phase, close to 0, indicating a selfish phase.

(v) Our fifth proposed solution concept is the use of iterative modeling. The idea is to compose an initial human model, find the autonomous agent that performs best, given this model, and then update the model. This idea is somewhat similar to an expectation maximization process (EM). Updating the model can be done either by using a model that accounts also for the agent's actions, and then, updating it given the autonomous agent's behavior, or by simply running the autonomous agent against a new group of humans, and training the human model based on the new data-set. While this solution concept does not directly account for the human's utility function, it does assume that humans have strategic

behavior and adapt to the agent's actions. This approach may also encourage the agent to be more altruistic, if the human returns altruistic behavior toward the agent; however, such behavior is not guaranteed, and both players may converge to selfish behavior. Therefore, combining this solution with some form of the previous solutions may be beneficial.

(vi) Our sixth proposed solution concept is hierarchical training. This approach requires that the human model receives the utility function or reward as an input, and can thus account for different utility functions. This is in addition to the human model receiving the agent's actions as input. In this approach, the human model is composed of two levels: a high level, which accounts for the human's goals, and a low level, which receives the high level's output as input and determines which action to take at a specific situation. While the high level must be trained only on a data-set with real human data, the lower level can be trained also in simulation. This structure can be achieved either by using different modules to compose the human model, or by a neural network architecture that accounts for the two different levels. This approach allows the human model to express strategic behavior, and account for its utility function. After training the human model on a real dataset, the human model and the agent can be trained together, adjusting the parameters and weights of the lower level of the human model, or trained iteratively, until they reach an equilibrium.

(vii) Finally, our seventh's proposed solution concept suggests the composition of multiple human models, which may differ merely by the original weights of a neural network, by their training data, or by using a dropout layer at inference. Each of these models are sampled, and the agent may assume that the human will play that action that is worst for the agent. This approach is likely to reduce overestimation bias. The agent may be less conservative if the correlation between the human's and the agent's utility is high (as might be observed from past interaction), and it may not necessarily want to assume the worst case, but some median or mean case.

2.3 Obtaining Utility Functions

An agent attempting to consider the human's goals, might need to gain access to a human's utility function or compose such a function. This may be a problem when interacting with humans in a real-world situation, as the utility function is not always available. There are several methods for obtaining such a utility function. One option is by using inverse reinforcement learning, which by observing human actions, can elicit the underlying utility function [Ramachandran and Amir, 2007]. A somewhat similar approach is to compose a neural architecture that learns both the utility function and the prediction of future human actions simultaneously. However, these approaches may be problematic, as our general hypothesis is that one should account for human utility function when composing a human model and not only rely on observing human actions. These approaches result in the entire human model based solely on human actions. Another approach is to allow humans to communicate their preferences to the agent and tell it what their utility function is. Clearly, it is not a simple task to create an interface that allows a human to communicate her complete utility function. Furthermore, people may not be fully aware of their utility function, may not be willing to fully disclose it, or may act strategically and not reveal their true utility function, so that the agent's actions will be more beneficial to them. A more suitable approach would be to have humans answer survey questions, similarly to methods used for preference elicitation and in the field of ethics research. Humans may also be asked to explain the motive behind some actions, which are either randomly picked, or selected by the user herself. A utility function learned from the human's actions and annotations is likely to be more robust and accurate. In addition, a system may account for an active learning setting, in which the agent can ask, with respect to specific actions, why the user took them. The agent will need to learn for which actions it should obtain additional information from the user, and then use this information to improve its model of the user's utility function. Finally, an agent may use common-sense knowledge bases, or obtain such information by communicating with users (see for example [Arabshahi *et al.*, 2021]), in order to understand why people take specific actions. The agent may mine the web, and obtain common-sense and general information using natural language processing methods to understand why people take specific actions.

3 Conclusions

Game theory provides mathematical models, which attempt to explain human behavior in social environments, based on the expected utility earned by each individual. However, as stated by Ariel Rubinstein, game-theory has very limited usage in practice, and many scholars agree that game theory usually fails to provide useful solutions to real-life problems [Syll, 2018]. Therefore, in practice, autonomous agents must rely on concepts other than game theory in order to proficiently interact with humans, especially, when the interaction is complex and actions are long lasting.

Unfortunately, this has led to the extreme other case, in which agents interacting with humans do not account for human adaptive nature and being goal oriented. Autonomous agents act as if they suffer from the protagonist syndrome and the fundamental attribution error [Tetlock, 1985], not realizing that humans are not only as complex as they are, but much more. Some agent developers assume a zero-sum or fully cooperative game, others ignore the user utility by modeling human behavior based on human actions alone, and finally, some do not directly model human actions at all, but merely learn which actions are most beneficial for the agent. Suppose, for example, an autonomous vehicle approaches a pedestrian. The vehicle may attempt to predict future positions of the pedestrian to avoid collision, but will not consider her goals, and whether she is in a hurry and attempting to catch a bus. We argue that the community should pursue solution concepts that enable the composition of agents that account jointly for human behavior and utility, in a general game setting.

Acknowledgments

This research was supported in part by the Ministry of Science, Technology & Space, Israel.

References

[Arabshahi *et al.*, 2021] Forough Arabshahi, Jennifer Lee, Mikayla Gawarecki, Kathryn Mazaitis, Amos Azaria, and Tom Mitchell. Conversational neuro-symbolic commonsense reasoning. In *AAAI*, volume 35, pages 4902–4911, 2021.

[Ariely *et al.*, 2003] Dan Ariely, George Loewenstein, and Drazen Prelec. "coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, 118(1):73–106, 2003.

[Azaria *et al.*, 2012] Amos Azaria, Yonatan Aumann, and Sarit Kraus. Automated strategies for determining rewards for human work. In *AAAI*, 2012.

[Azaria *et al.*, 2016] Amos Azaria, Ya'akov Gal, Sarit Kraus, and Claudia V Goldman. Strategic advice provision in repeated human-agent interactions. *JAAMAS*, 30(1):4–29, 2016.

[Blanken *et al.*, 2015] Irene Blanken, Niels van de Ven, and Marcel Zeelenberg. A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, 41(4):540–558, 2015.

[Camerer, 2003] C. F. Camerer. *Behavioral Game Theory. Experiments in Strategic Interaction*, chapter 2. Princeton University Press, 2003.

[Carroll *et al.*, 2019] Micah Carroll, Rohin Shah, Mark K Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, and Anca Dragan. On the utility of learning about humans for human-ai coordination. *Advances in Neural Information Processing Systems*, 32:5174–5185, 2019.

[Freire *et al.*, 2019] Ismael T Freire, Xerxes D Arsiwalla, Jordi-Ysard Puigbò, and Paul Verschure. Modeling theory of mind in multi-agent games using adaptive feedback control. *arXiv preprint arXiv:1905.13225*, 2019.

[Gal and Pfeffer, 2007] Ya'akov Gal and Avi Pfeffer. Modeling reciprocal behavior in human bilateral negotiation. In *NCAI*, volume 22, page 815. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2007.

[Grosz *et al.*, 2004] Barbara Grosz, Sarit Kraus, Shavit Talman, Boaz Stossel, and Moti Havlin. The influence of social dependencies on decision-making: Initial investigations with a new game. 2004.

[Hsu *et al.*, 1995] Feng-hsiung Hsu, Murray S Campbell, and A Joseph Hoane Jr. Deep blue system overview. In *Supercomputing*, pages 240–244, 1995.

[Jaques *et al.*, 2019] Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Caglar Gulcehre, Pedro Ortega, DJ Strouse, Joel Z Leibo, and Nando De Freitas. Social influence as intrinsic motivation for multi-agent deep reinforcement learning. In *ICML*, pages 3040–3049. PMLR, 2019.

[Loewenstein, 2000] George Loewenstein. Willpower: A decision-theorist's perspective. *Law and Philosophy*, 19:51–76, 2000.

[Nay and Vorobeychik, 2016] John J Nay and Yevgeniy Vorobeychik. Predicting human cooperation. *PloS one*, 11(5):e0155656, 2016.

[Peled *et al.*, 2011] Noam Peled, Ya'akov Kobi Gal, and Sarit Kraus. A study of computational and human strategies in revelation games. In *AAMAS*, pages 345–352, 2011.

[Ramachandran and Amir, 2007] Deepak Ramachandran and Eyal Amir. Bayesian inverse reinforcement learning. In *IJCAI*, volume 7, pages 2586–2591, 2007.

[Rosenfeld and Kraus, 2011] Avi Rosenfeld and Sarit Kraus. Using aspiration adaptation theory to improve learning. In *AAMAS*, pages 423–430, 2011.

[Sandholm and Crites, 1996] Tuomas W Sandholm and Robert H Crites. Multiagent reinforcement learning in the iterated prisoner's dilemma. *Biosystems*, 37(1-2):147–166, 1996.

[Shapira and Azaria, 2022] Ido Shapira and Amos Azaria. Reinforcement learning agents for interacting with humans. In *CogSci*, 2022.

[Silver *et al.*, 2016] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

[Skinner and Walmsley, 2019] Geoff Skinner and Toby Walmsley. Artificial intelligence and deep learning in video games a brief review. In *ICCCS*, pages 404–408. IEEE, 2019.

[Subrahmanian, 2000] Ventatramanan S Subrahmanian. *Heterogeneous agent systems*. MIT press, 2000.

[Syll, 2018] Lars Pålsson Syll. Why game theory never will be anything but a footnote in the history of social science. *Real-World Economics Review*, 83:45–64, 2018.

[Tetlock, 1985] Philip E Tetlock. Accountability: A social check on the fundamental attribution error. *Social psychology quarterly*, pages 227–236, 1985.

[Tversky and Kahneman, 1981] Amos Tversky and Danliel Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981.

[Tversky and Kahneman, 1992] Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Risk and uncertainty*, 5(4):297–323, 1992.

[Wang *et al.*, 2016] Zhijian Wang, Yanran Zhou, Jaimie W Lien, Jie Zheng, and Bin Xu. Extortion can outperform generosity in the iterated prisoner's dilemma. *Nature communications*, 7(1):1–7, 2016.