Beneficial AI: Memo 1
Decision-Making for a Group of Individuals

Arnaud Fickinger
Department of EECS
University of California, Berkeley
arnaud.fickinger@berkeley.edu

1 Introduction

Any AI agent deployed in the real world will likely be confronted to a group of other agents, either human or artificial. We can think of multiple humans with different preferences sharing a robot, or a human-robot team trading off individual preferences for team performance. Yet most efforts in AI research are concentrated in the optimization of a single objective. Even the human-robot interaction (HRI) literature mainly consider interactions with a single human. In this memo we build on an interdisciplinary view of decision-making for an individual and for a group of individuals to understand what is needed for an AI agent to evolve among a group of individuals.

2 Formalizing Rational Decision-Making for an Individual

Every attempts to rationalize decision-making for an individual in a positive way, as opposed to a normative way, assume the existence of a preference ordering $\succsim$ between objects [1]. In Economics, in order to account for the uncertainties underlying real world decision-making, preferences are not made on the objects them-self but on the space $\mathcal{P}$ of lotteries $(p_1, A_1; ...; p_n, A_n)$ over a set of objects following a categorical distribution. To model sequential decision-making, the outcome of any lottery can be another lottery, creating a tree. We can relate it to the usual model of sequential decision-making in Computer Science where we choose actions that can lead to different states following a categorical distribution, assuming a finite number of states [2].

Both disciplines build on the Von Neumann-Morgenstern utility theorem to rationalize decision-making and make it computationally tractable. The theorem states that if the preference ordering follow the utility axioms (completeness, transitivity, monotonicity, independance), then there exists a function $u : \mathcal{P} \to \mathbb{R}$ such that:

$$L \succsim M \Leftrightarrow E(u(L)) \geq E(u(M)) \tag{1}$$

where $E(u(p_1, A_1; ...; p_n, A_n) = p_1 u(A_1) + ... + p_n u(A_n)$ and a decision-maker is defined as rational if he chooses action maximizing his expected utility. The existence of such a utility function enable to separate the true preferences from the stochasticity of the world since we just have to define the utility for the atomic objects. This assumption is implicit in Reinforcement Learning (RL) where the reward function is defined on the space of states (or states-action) and clearly separated from the transition probabilities of the Markov Decision Process (MDP).

In Economics these axioms have been often challenged which has led to non-Expected Utility (EU) theories. In practice, simple experiments show that humans often violate the utility axioms [3]. In Computer Science, almost every decision-making algorithms are built on expectation maximization, yet other models of (bounded) rationality are investigated in

HRI to model human decision-making. A widely used model among the HRI community is built on the Luce's choice axiom [4] and corresponds to a softmax policy [5]:

$$P(a|s) = \frac{\alpha e^{Q(a,s)}}{\sum_{a'} \alpha e^{Q(s,a')}} \tag{2}$$

Where $P(a|s)$ is the probability to choose action a while in state s and $Q(s,a)$ can be interpreted of the utility of the state-action pair. The rationality parameter α measure how far the human is from a perfectly rational decision-maker ($\alpha = \infty$). We can relate α to Afriat's critical cost efficiency index that measure the consistency with utility maximization [6].

3 Formalizing Rational Decision-Making for a Group of Individuals

A first attempt to rationalize the decision-making of a group of individuals is to built on the individualistic formalism discussed above. Social choice theory assume that individual preference orderings and utilities can be aggregated by what Arrow calls social welfare functions [7]. This first solution has the advantage to be easily implementable in the form of a single objective optimization problem. Yet social choice theory has been often challenged by philosophers and economists.

A first line of arguments challenge the formalism of preference ordering aggregation. Condorcet shows that the aggregation of transitive individual preferences can lead to cyclic collective preferences in the context of majority voting [8]. Arrow shows that no rank-order electoral system built on the aggregation on individual preferences can meet at the same time the following criteria: unrestricted domain, non-dictatorship, Pareto efficiency and independence of irrelevant alternatives [7].

In the case of utility aggregation, a more obvious argument is the subjectivity of the utility values and the difficulty to compare the mental states between individuals [9]. Two individuals that express the same preferences and have the same behavior can have two different mental states and react differently to the outcome of the solution to an optimization problem aggregating their preferences. A first solution would be to normalize the utilities. Furthermore the formal arguments mentioned for preference ordering aggregation can be generalized to utility aggregation [10].

4 Implementation in AI systems

An appealing paradigm to solve sequential decision-making problems is Reinforcement Learning (RL). By interacting with the world, the robot will explicitly (Model-Based RL) or implicitly (Model-Free RL) deduce the parameter of the underlying MDP and find an optimal policy. Yet we still need to be able to give a feedback signal to the robot which require tremendous effort in reward engineering. Inverse Reinforcement Learning (IRL) methods propose to alleviate the difficulty of specifying a reward by inferring it from human's behavior [11].

We notice that IRL is in agreement with the positive view of rationality discussed above. Following on from social choice theory, the agent would infer a reward for each individual and would be faced to the same challenge when trying to aggregate all the rewards. We can imagine a multi-objective optimization recovering the Pareto front, but it is computationally intensive and just postpone the aggregation problem since the agent will still need to choose a point on the Pareto front.

5 Critique of this view of rationality

A critique we can address to the utilitarian view of rationality is that it is fundamentally individualistic, whereas complex decision-making are influenced by others. Of course we can argue that the others can be part of the utility function in the form of envy, altruism or pride, but this doesn't solve the problem. The reason why social choice theory have a hard time to go from an individualistic view of rationality to a collective view of rationality is because individual utilities does not have much sense in a society, especially for sequential decision-making. In my opinion the right way to formalize rational decision-making is to start from the group and deduce individual utilities, which is the inverse of social choice theory. Mathematically this amounts to formalize rational decision-making for a group with games $(u_1(a_1,...,a_n),...,u_n(a_1,...,a_n))$ instead of multi-objective optimization $(u_1(a_1),...,u_n(a_n))$.

References

[1] Ariel Rubinstein. *Lecture Notes in Microeconomic Theory: The Economic Agent - Second Edition*. Princeton University Press, rev - revised, 2 edition, 2012.

[2] Stuart Russell and Peter Norvig. Artificial intelligence: a modern approach. 2002.

[3] M. Allais. *Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine*. 1953.

[4] R Duncan Luce. *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2012.

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

[6] Sidney N Afriat. Efficiency estimation of production functions. *International economic review*, pages 568–598, 1972.

[7] Kenneth J Arrow. *Social choice and individual values*, volume 12. Yale university press, 2012.

[8] Nicolas de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Cambridge Library Collection - Mathematics. Cambridge University Press, 2014.

[9] John C Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of political economy*, 63(4):309–321, 1955.

[10] Amartya Sen. *Collective choice and social welfare*. Harvard University Press, 2018.

[11] Andrew Y Ng, Stuart J Russell, et al. Algorithms for inverse reinforcement learning. In *Icml*, volume 1, pages 663–670, 2000.