Self-Interested Recommendations in Multi-Agent Social Dilemmas

Francisco Carvalho[1]
francisco.de.carvalho@tecnico.ulisboa.pt

Instituto Superior Técnico
Av. Rovisco Pais 1, 1049-001 Lisboa

Abstract. Recommendation systems (RS) have become intertwined with collective human affairs: from social interaction, to governance and many sectors of economic activity. The abundance of information with which we find ourselves and the need to parse it, means RS will only become more pervasive. However, current advanced RS algorithms are rational learning agents [13] that choose their actions (recommendations) based on what would maximize metrics determined by their operators. These metrics often conflate user satisfaction with retention or ad-click rates, while satisfying business incentives. This results in an often ignored misalignment between the user's best interest and the actual incentives of the mediator (RS). To better understand the impact that mediators with their own agendas have in human coordination, we propose studying the impact of such self-interested recommendation systems on a population of agents organized in scale-free networks facing social dilemmas. Self-interested policies will be produced by reinforcement learning and then applied to mediate these environments in simulations using methods from evolutionary game theory. Relevant metrics include the distribution of utility in the population, its cooperation rate, and the network structure over time. We hoped that the methods developed might contribute in generating insights that inform design and legislation.

Keywords: Recommendation systems · Reinforcement learning · Complex Networks · Social dilemmas · Evolutionary game theory

Table of Contents

1 Introduction

Preparatory work for this proposal included surveying the landscape of work being done under the scope of *AI alignment*. This prompted an extensive review of the epistemics of the field of *AI risk*, one of the most vocal proponents of work in alignment. Out of this, a conclusion arose that recommendation systems (RS) represent one of the most urgent, concrete - and understudied - objects of scrutiny and intervention in AI risk and alignment today.

Here we will motivate the study of the impacts of self-interested recommendation systems on the performance of populations of agents facing social dilemmas. Observations will be produced by a modelling and simulation approach of a game played between users and RS. The game will alternate iterated social dilemmas between agents on a scale-free network, with a recommendation game between each user and a self-interested RS mediator. Self-interested policies for the RS mediators will be learned with reinforcement learning and insights will be extracted by analyzing how cooperation rates, aggregate utility, and network structure evolve over time.

We'll begin by looking at the high-level relevance of RS, making sense of the ethical challenges they face. We'll examine the inherent misalignment between stakeholders in current RS, and will lay out the range of social effects RS can have. Finally, we'll relate current RS to the alignment problem, mention a few current proposals of paths to alignment, and how to go about modelling RS as mediators of human coordination.

1.1 Recommendation systems and why they matter

Thanks to the abundance of information with which we find ourselves and the need to parse it, RS have become intertwined with collective human affairs from social interaction, to governance, to most sectors of economic activity. Our civics rely on public discourse which happens overwhelmingly online, on monolithic platforms like Facebook, Twitter, and Youtube, which are mediated by RS. Additionaly, they are some of the largest scale, most sophisticated and influential instances of machine learning being deployed in the world, which makes them prime case studies for AI alignment.

As of January 2019, according to The Global State of Digital Report 2019, there are 3.484 billion active social media users. Of those social media users 2.320 billion people are Facebook users, and 1.9 billion people are YouTube users ("Most popular social networks worldwide as of April 2019, ranked by numbers of active users " 2019). YouTube's chief product officer revealed[1] that 70% of views on YouTube result from recommendations of the YouTube algorithm.

Despite their clear societal importance, most of the work done in RS has been in order to improve their capabilities and efficiency [37] [1] [53], usually motivated by business incentives. However, research about the wider impact on users and

[1] <https://www.cnet.com/news/youtube-ces-2018-neal-mohan/>

on society has been scattered in part due to the recency of this technology and to proprietary algorithms being often treated as business secrets. [48]

In their prescient "Analysis of the Interaction Between Intelligent Software Agents and Human Users", Burr et al. (2018) consider the class of interactions between humans and "Intelligent Software Agents (ISA)". They adapt Russell and Norvig's (2010) [55] definition of learning agents, and define an ISA as *any program that can be described as having a model of its environment, which it uses to take actions that enable the ISA to achieve its goals, while also acquiring further information that it can use to update the parameters of its model.*

In the context of RS, environments include human users and their rewards depend on human actions, (clicks, purchases, etc) therefore what rewards the ISA obtains is - importantly - conditional on its ability to influence the behaviour of a human user. So the RS is effectively a mediator of user access to content - and in cases where content is user-generated, a mediator of user interactions - or it controls some other aspect of the user experience, without the need to remain neutral about outcomes of user choices, by design.

More specifically, Milano et al. (2020) [48] describe RS as functions that take information about a user's preferences (e.g. about movies) as an input, and output a prediction about the rating that a user would give of the items under evaluation (e.g., new movies available). Inevitably, they shape user experience and social interaction [68]. The task of a RS — the recommendation problem — is often summarised as that of finding good items [37]. To make it an operational definition, one must specify a) what the space of options is, b) what counts as a good recommendation; and, importantly c) how the RS's performance can be evaluated. [48]

1.2 Challenges of RS.

Milano et al. (2020) [48] survey the ethical challenges of RS in the first systematic literature review of its kind. They point out the parameters of an operational definition of an RS depend on the "level of abstraction" (LoA) at which the recommendation problem is being considered. At their simplest RS can be seen as catalogue-based, making recommendations that rely on predictions of immediate user feedback. RS as decision support involve appreciating a user's goals and decision-making and evaluation of system performance requires more complex metrics [53]. Finally, RS can be thought of as multi-stakeholder environments where multiple parties (users, providers, system operators) can derive different utilities from recommendations. (See Fig 8)

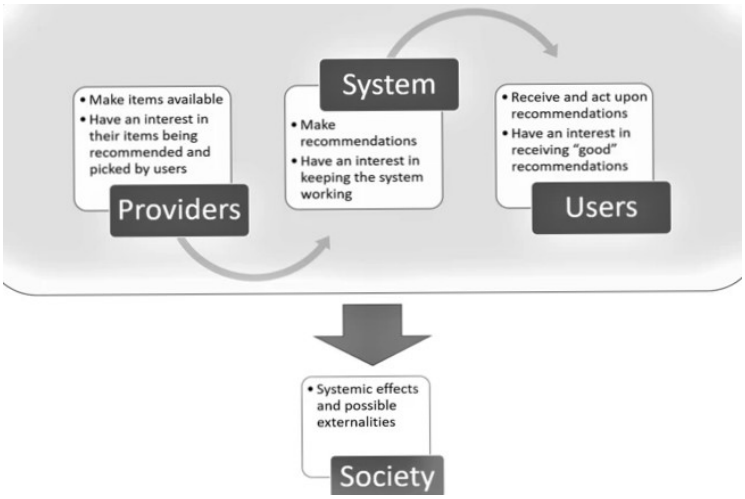


Fig. 1. RS as a multi-stakeholder environment. [48]

This third lens of a multi-stakeholder environment will be especially useful in our approach because it enables conceptualization of the impact that RS have at different levels, on individuals, and society more broadly, making it possible to explicitly trade-off between these possibly competing interests.

Burr et al. [14] remind us that an RS is usually deployed for the purpose of maximising a given utility function that reflects the goals of one of the stakeholders, usually its developers or operators, For example, to increase the revenue of a company or spread a political message. The RS chooses from a set of actions whose outcomes partially depend on the choices made by other stakeholders, namely, human users. This dependence might give rise to a tension between the RS and the user's incentives, which might end up effectively competing with each other instead of consistently having the RS attempt to advance its users' interests.

In their survey, Milano et al. (2020) [48] establish a taxonomy of the ethical issues facing RS and draw two axes we can use to analyze them. (See Fig 2)

1. Whether a RS negatively impacts the *utility* of some of its stakeholders or, instead, constitutes a *rights violation*, which is not necessarily measured in terms of utility.
2. Whether the negative impact constitutes an *immediate harm* or it exposes the relevant party to future *risk of harm or rights violation.*

	Immediate harm	Exposure to risk
Utility	Inappropriate content (4.1)	Opacity (4.4)
		Questionable content (4.1)
Rights	Unfair recommendations (4.5)	Privacy (4.2)
	Encroachment on individual autonomy and identity (4.3)	Social effects (4.6)

Fig. 2. Taxonomy of ethical issues in RS. [48]

Zoetekouw (2019) [72] surveys the negative consequences of the widespread use of RS and identifies 5 classes of challenges for RS:

1. *content diversity* (filter-bubbles, popularity feedback. Example: 2016 US election);
2. *over-personalization* (individual manipulation of opinion. Example: rabbit-hole/alt-right pipeline)
3. *data problems* (less popular items generate less information and get recommended less. Example: Contrarian opinions get drowned);
4. *adversarial exploitation of metrics.* (Example: Search engine optimization);
5. *external influence* (interested parties (like governments) can exert influence on the content in the platform. Example: Egyptian revolution 2011).

Comparing these to Milano et al.'s taxonomy, the first three roughly match instances of "inappropriate content", "encroachment on individual autonomy", and "unfair recommendations" respectively. The latter two are caused by third parties using the RS as an environment in which to exploit others, which isn't really contemplated, but fits well in the framing of RS as multi-stakeholder environments.

Given our focus on cooperation in populations of agents. Our proposal is primarily concerned with using modelling and simulation to study emergent social effects resulting from mechanics and aspects of RS that might cause "encroachment on individual autonomy". In the taxonomy such effects are classified as rights violations, but we will also try to quantify these by measuring a notion of aggregate utility for users as well as the population's cooperation rate. In contrast, in this work we are not concerned with inappropriate or questionable content, privacy, opacity/transparency, or unfair recommendations.

1.3 Misalignment of users and RS

Encroachment on individual autonomy and identity In the context of our chosen ethical issue, RS can encroach on individual users' autonomy in several ways. Including but not limited to "nudging" them in a particular direction, provoking "addiction" to some stimuli, or filtering the range of information they

see [14] [68] [63]. Instances of this aren't necessarily prejudicial, we might appreciate being nudged towards healthier habits, or that irrelevant information be hidden from me, but they can also be manipulative and coercive by, for example, exploiting predictable human biases. Either way, we should be aware of the role of the system as - beyond that of taking a pre-existing preference profile and tailoring its recommendations to it - that of contributing to the construction of the user identity dynamically [20].

Human biases can be exploited by content recommenders Human biases can be exploited by content recommenders. The RS of the most popular social media platforms are optimized for easy-to-measure metrics like click-through rate in ads, total time spent logged on, or number of daily active users [59], which are only weakly correlated with what users care about. Users often regret their use of social media. [61] One of the most powerful optimization processes in the world is being applied to increase these metrics, involving thousands of engineers, the most cutting-edge machine learning technology, and a significant fraction of global computing power. "The result is software that is extremely addictive[2], with a host of hard-to-measure side effects on users and society including harm to relationships[3], reduced cognitive capacity[4], and political radicalization[5]." [67]

We can see this as an instance of Goodhart's law, which states that "as soon as a measure becomes a target, it ceases to be a good measure". [21]

"As psychology keeps showing, the information we are exposed to radically biases our beliefs, preferences and habits[6], with both short-term[7] and long-term effects[8]" [67]

RS as traps An interesting take on the issue of personal autonomy resides in "Captivating algorithms: Describing Recommender Systems as Traps" [59], where Seaver applies his anthropological perspective to the "captology" of RS. In interviews to RS developers in the US a tendency is reported among these systems' makers to describe their purpose as 'hooking' people – enticing them into frequent or enduring usage.

> *Conflating retention and satisfaction has allowed developers to mediate the tension between users (whom they wanted to help) and business people who wanted to capture them.*

[2] <https://www.bbc.com/news/technology-44640959>
[3] <http://www.bierdoctor.com/papers/filterbubble_CHI2015_final.pdf>
[4] <http://www.journals.uchicago.edu/doi/abs/10.1086/691462#>
[5] <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>
[6] <https://www.youtube.com/watch?v=QDCcuCHOIyY>
[7] <https://www.youtube.com/watch?v=gQHvTow91FY>
[8] <https://www.youtube.com/watch?v=_RuyXyekx6g>

—Seaver (2019) [59]

The authors point out that due to the usefulness and scale of RS, the way should be to transform, not escape them. That is the work we intend to start here.

> *To be trapped at the level of infrastructure is also to be hosted. The question to ask of traps is not how to escape them but how to reconfigure them in the service of other aims.*
>
> —Seaver (2019) [59]

Taxonomy of human-RS interactions Burr et al. (2018) propose a taxonomy of RS/user interactions that may take place as an ISA acts to maximizes its expected utility, all cases where - as mentioned earlier - user autonomy is encroached upon. (See Fig 5) We would place emphasis on the distinction between first-order effects - Coercion, Deception, Persuasion - and second-order effects - change in human utility function and change in beliefs - which are harder to track and measure.
Our experiments will focus on assessing the second order effects of interactions that we expect to fall on the *Persuasion* category. *Coercion* stands for the undesired presence of advertisements and agreeing to share personal data. *Deception* is the presenting of deceptive content like clickbait, phishing, misleading ads, and fake news. *Persuasion* is subdivided into *Trading* and *Nudging*. *Trading* refers to instances where an ISA models user utility to satisfy it as far as it advances its own utility, a "win-win" situation but not optimal for the user. *Nudging* seeks to influence human behavior by exploiting its predictable biases and may guide them to actions that would otherwise be irrational [4] and may or may not increase user utility. For example, this could involve exploiting emotional associations, or a bias towards sensational content.

Second-order effects can be changes in human utility function and/or changes in beliefs. Changes to utility function may happen as a result of repeated exposure to a rewarding stimulus that generates behavioral addiction in human users. Changes in beliefs may result from repeated exposure to certain types of content that influence a user's belief system, like their assessment of what the mainstream consensus is on a given topic.

Although Burr et al. focus their examples on *click-through rate* as the basis for RS utility functions, any other measure of "success" could be used. We'll be considering user retention, as described in [59] to assess the impact of self-interested RS on utility and cooperation.

A slight critique of our model: When it comes to the second-order effects, the distinction between a change in utility function and a change in beliefs in our model isn't as explicit as it could be. The payoff that agents receive from playing a dilemma depends on the similarity between its and its partner's attributes. A change in attributes results in a change to the payoffs an agent can get from each of its neighbors, and so a change in utility function.

Beliefs aren't explicitly represented in our model, as users lack models of their environment beyond the information they have about their neighbors, which is their similarity and last action. On another hand, we could interpret the information they have access to as their sampling of the overall environment, which may become biased over time as a result of recommendations, even if it was representative in the beginning. However, our models only ever allow network rewiring as a result of recommendation, so in order to get a fair baseline for the impact of RS on beliefs, we might allow some form of local network rewiring.

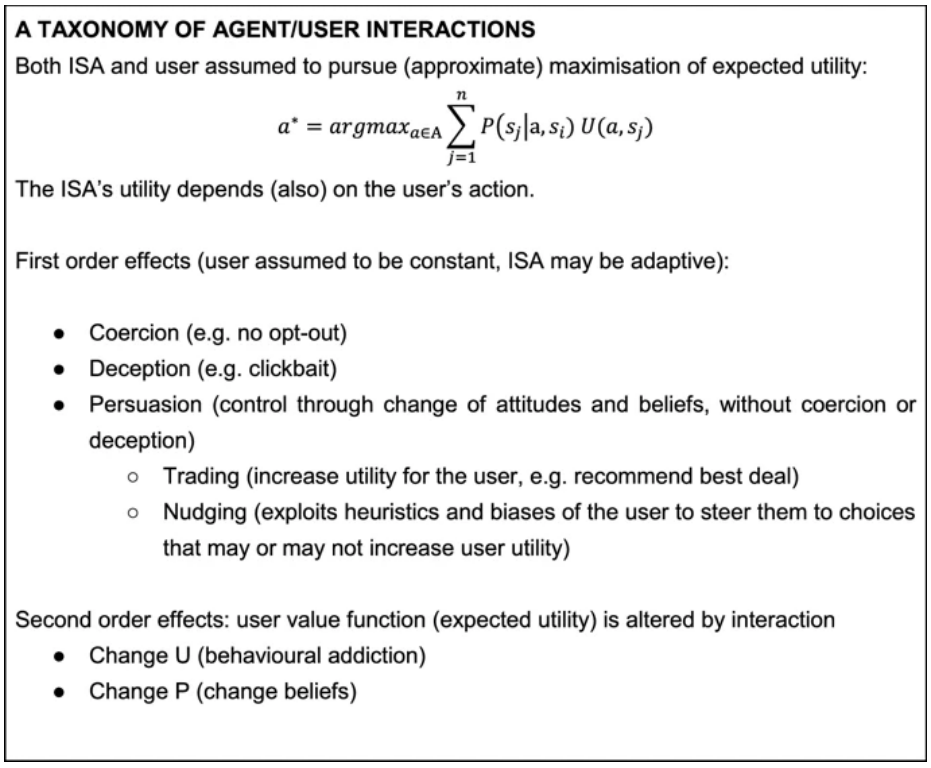


Fig. 3. Taxonomy of RS/human-user interactions.[14]

1.4 Social effects of recommendation systems

Social effects of RS are widely discussed, for example: the design of news recommender systems risks insulating users from exposure to different viewpoints, creating self-reinforcing biases and "filter bubbles" that are damaging to public debate, group deliberation, and democratic institutions more generally [10] [11] [25] [26] [52]. This feature of recommender systems can have negative effects on

social utility, but we can hardly tell how large the impacts of exploiting human bias at a simultaneously massive and personalized scale may be. A recent but poignant example is the spread of propaganda against vaccines, which has been linked to a decrease in herd immunity. [12]

1.5 Recommendation systems and the alignment problem

Burr et al. (2018) [14] use ideas from bounded rationality and surrounding fields to explore the impact of user interactions with RS on individual autonomy. As we've seen, the RS - or mediator - benefits from steering user behaviour towards outcomes that maximise its utility, outcomes which may not align with the ultimate best interest of its users. This creates incentives for the mediator to be designed to maximize metrics correlated with the business's bottom-line like ad-clicks or user retention, not necessarily user utility. This predicament relates to the value alignment problem. [9].

The dilemma between the user's utility and the mediator's is not limited to social media or news; it is representative of a wider issue known as the 'value alignment problem. As Stuart Russell and co-authors put it:

> *For an autonomous system to be helpful to humans and to pose no unwarranted risks, it needs to align its values with those of the humans in its environment in such a way that its actions contribute to the maximization of value for the humans.*
>
> —Hadfield-Menell et al. (2016) [22]

They describe how a negative side effect of an intelligent agent's behaviour may result from a misspecified reward, and describe this as *a failure mode* of reward design (Hadfield-Menell et al. (2017) [23] where leaving out important aspects leads to poor behavior, in yet another example of Goodhart's law.

1.6 Work in AI alignment

As our problem can be seen in the light of AI alignment work, it benefits us to look at current work and proposals for full alignment to contextualize our work in the literature.

Jacob Steinhardt, co-author of Concrete Problems in AI Safety [2] presents AI Alignment Research Overview [62] as an updated version of their popular research agenda. He describes 4 categories of **technical work**: *Technical AI alignment* as in solving conceptual and engineering issues (See Figure 4), developing tools for *detecting failures*; getting empirical *methodological insight* for building AI systems and *system building* for doing it at scale.

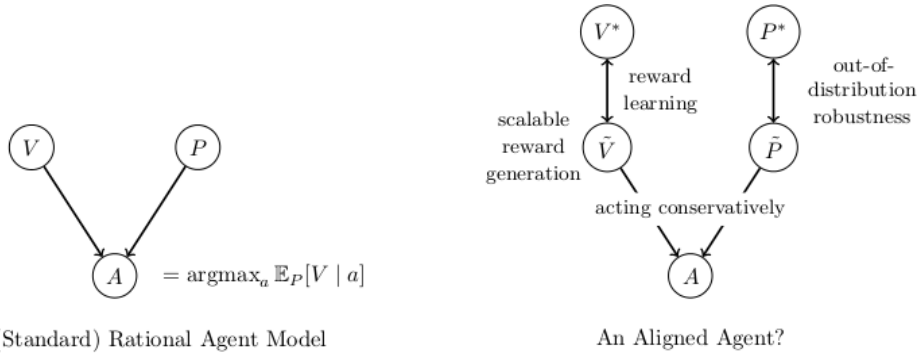


Fig. 4. Comparison between the standard model of a ration agent and a model of a potentially aligned agent according to Steinhardt (2019) [62].

"An overview of 11 proposals for building safe advanced AI" [30] describes eleven end-to-end alignment proposals where the goal is to build a powerful, beneficial AI system using current technological capabilities, and evaluates them on four axes: outer alignment - whether the optimal policy for the specified utility function would be aligned with humans; inner alignment - whether the actual model resulting from the training process would be aligned; training competitiveness; and performance competitiveness. Seven of the eleven proposals involve a recursive outer alignment technique (like scalable agent alignment via reward modelling [43] or debate [34]) combined with a technique for robustness (like relaxed adversarial training [31], or intermittent oversight by a competent supervisor). One of the non-recursive proposals for training a generally intelligent and aligned agent requires only vanilla reinforcement learning in a multi-agent setting where desirable features like cooperation and corrigibility are incentivized. The basic idea would be to mimic the evolutionary forces that led to humans' general cooperativeness. [41]

Hoang (2020) [28] lays out a roadmap that attempts to decompose the alignment problem into more tractable subproblems and lightly anthropomorphizes them with ABCDE names. In this light, the work we're proposing here can be described as one of analysis and search of ways to support "Bob" with systemic interventions.

Fig. 5. Hoang (2020) [28] segments alignment work into reliable data collection → world model inference → desirability score learning → the incentive problem → reinforcement learning. [28] Erin will be collecting data from the world, Dave will use these data to infer the likely states of the world, Charlie will compute the desirability of the likely states of the world, Bob will derive incentive-compatible rewards to motivate Alice to take the right decision, and Alice will optimize decision-making.

Clifton (2019) [17] presents a research agenda drawing on international relations, game theory, behavioral economics, machine learning, decision theory, and formal epistemology where they argue for the study of cooperation failure between current AI systems and between these and humans in the loop as one of their main research recommendations.

1.7 Objectives

The focus of this dissertation is the misalignment between the best interest of the users and the incentives of recommendation algorithms. Besides social networks, this effect should be equally relevant in the context of ecosystems of autonomous agents and AI services. [18]

Our objective is to assess the impacts of self-interested mediators on the aggregate utility and cooperation rate of multi-agent systems in solving underlying problems, namely social dilemmas. Basically answering the question "What happens to a population's ability to reach its goals when a recommendation mechanism has its own incentives?"

This will be done by:

1. Establishing an **abstract model** for the interaction between mediators and agents that could apply to many multi-agent domains.
2. Instantiating the above by **modelling and simulating** multi-agent social dilemmas where users are organized in a social network and receive recommendations from a mediator. Concretely for the proposed project, studying the iterated pairwise prisoner's dilemma for its simplicity.
3. Obtaining a self-interested mediator by training a reinforcement learning agent to provide recommendations while optimizing for their own utility function (user retention) rather than the users'.
4. Assessing the impact that **self-interested mediators** have on the distribution of utility in the systems when compared with **perfectly selfless** mediators, hypothetical anti-aligned mediators, and with the absence of mediators as baselines.
5. Evaluate the impact of **competition between mediators** on the relevant metrics for agents.

6. Examine a possible **intervention** on the system described in 5 if the results are still discouraging.

2 Background

2.1 Social dilemmas

Social dilemmas provide a platform to study the emergence of behavior and cooperation using simulated agents. [35] [46] They have featured heavily in behavioral economics, psychology and evolutionary biology [51] [45] [57]

We consider a classic iterated prisoner's dilemma (IPD) approach, where each iteration of the game can be characterized by a payoff matrix. (See Table 1) Agents play rounds of the dilemma and there is continuity between iterations as an agent's previous actions are used to inform the strategy and partner selection of the other agents. The resulting dilemma contains the tension between immediate individual gain and delayed but eventually greater benefit with an added risk of being exploited.

Table 1. Prisoner's dilemma. The game is modeled so that T ¿ R ¿ P ¿ S and 2R ¿ T + S. The motivation to defect comes from fear of an opponent defecting or acting greedily to gain the maximum reward when one anticipates the opponent might cooperate. The Nash equilibrium, the optimal game-theoretic strategy, is for both to defect, although payoffs are higher when both cooperate.

Table 2. Abstract prisoner's dilemma game payoff matrix.

Strats	**C**	**D**
C	R , R	S , T
D	T , S	S , T

Table 3. Example PD payoff matrix

Strats	**C**	**D**
C	3 , 3	0 , 4
D	4 , 0	1 , 1

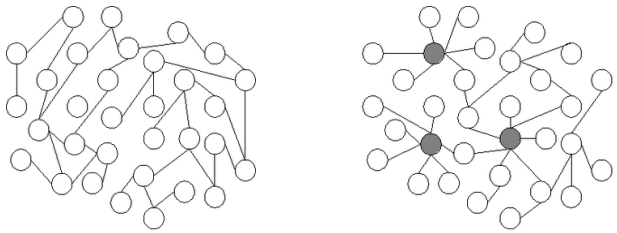
2.2 Scale-free networks

Many real world networks, including online social graphs, are associated with scale-free (See Figure 6), power-law degree distributions with exponent γ between 2 and 3. [6]

$$d(k) \sim k^{-\gamma} \tag{1}$$

Where d(k) is the degree distribution.

These networks can be characterized by growth and linear preferential attachment and have been shown to foster cooperation in evolutionary settings of several social dilemmas. [19]



(a) Random network (b) Scale-free network

Fig. 6. Comparison between a random and a scale-free network. (From Wikipedia, no author listed) [48]

2.3 Reinforcement learning / Q-Learning

Q-Learning [69] and Deep Q-Learning [49] are popular algorithms in reinforcement learning that learn a policy by balancing exploration and exploitation and have been previously used in single and multi-agent settings. [3] Q-Learning makes use of a value-function Q that maps a state-action pair to a quantitative value: $Q : S$ x $A \rightarrow R$ function for a policy π. When dealing with high dimensional state-spaces, the Q-function is often characterized by a neural network. (DQN)

The policy of an agent is an ϵ-greedy policy and is defined by

$$\pi(s) = \begin{cases} argmax_{a \in A_i} Q_i(s,a) & \text{with probability} \quad 1-\epsilon \\ U(A_i) & \text{with probability} \quad \epsilon \end{cases} \tag{2}$$

where $U(A_i)$ denotes a uniform sampling from the action space. An agent stores its trajectories $(s, a, r_i, s')_t$ for each timestep t by interacting with the environment and updates its policy according to

$$Q_i(s,a) \leftarrow Q_i(s,a) + \alpha \left[r_i + \gamma * max_{a' \in A_i} Q_i(s',a') - Q_i(s,a) \right] \tag{3}$$

where s is the current state, a is the current action, r_i is the reward obtained by agent i and s' is the next state.

Finally, Q-Learning can be directly applied to multi-agent settings by having each agent i learn an independently optimal function Q_i. However, because agents are independently updating their policies as learning progresses, the environment appears non-stationary from the view of any one agent, violating Markov assumptions required for convergence of Q-learning.

3 Related Work

Our solution involves studying the effects of self-interested recommendation mechanisms in cooperation and utility in social dilemmas. To this end, we started

by surveying accounts of the societal impacts of recommendation systems. We proceed to reviewing a strain of research in game-theoretic RS, which is relevant when analyzing them as an environment with differently interested stakeholders. We assess instances of partner selection and recommendations in social dilemmas, as well as models of preference drift because they will be integral elements of our simulations. We examine some instances of RS modelling and then turn our attention to topics potentially more relevant to future work: literature on sequential social dilemmas, deep RL, and empirical game theory.

3.1 Impacts of recommendation systems

As mentioned in Section 1, some work has been put towards understanding the challenges and effects of human interaction with recommendation systems. [14], Zoetekouw et a. (2019) [72] and Milano et al. (2020) have similar goals of surveying the landscape of challenges of RS, ethical and technical. In Milano et al.'s taxonomy, the problems we're studying fall under "encroachment on individual autonomy" and "social effects", distinguishing ours from the comparatively common concerns of privacy, fairness, and inappropriate content. At Google, and in the context of A/B testing, Hohnhold et al. (2015) [29] devise methodology to determine the long-term effects that changes to a system can have on users. Burr (2018) [14] relates even more to our goals as it focuses specifically on surveying the impacts of interactions between users and intelligent software agents with potentially misaligned utility functions. In their terminology, we will focus on understanding the second-order effects (change in utility function and change in beliefs) that "nudging" and "trading" behaviors by the RS may have on users.

3.2 Misalignment and game-theoretic analysis of recommendation systems

Tennenholtz and Kurland (2019) [40] lay out a game theoretic research agenda for recommendation systems by interpreting them as multi-stakeholder environments. They note that strategic dynamics in content production and consumption may lead to the failure of classical principles of RS in maximizing social welfare, so they propose to revisit those principles. This agenda is posterior to a body of work that applies game-theoretic analysis and mechanism design to search engines, the optimality of current RS solutions, dueling and competing algorithms, and strategic users. Bahar et al. (2018) address a particular conflict in user and RS incentives that stem from the explore/exploit tradeoff. There has also been work on competing RS as seen in the search duel [36] [8] and in Babaioff et al. (2015) where they investigate the cost that strategic mediators in competition with one another would impose on the society of agents they mediate. Balduzzi et al. [5] construct a framework to organize gradient learners in a way that preserves local-to-global legibility, where one can draw qualitative conclusions about a collective from the nature of individuals. Halkidi et al. [24] address selfish users interacting with a recommender system in a privacy-protecting game where they try to balance the amount of information they share

about themselves with the quality of their recommendations. Ben-Porat and Tennenholtz (2018) [7] consider fair and stable treatment of content-providers by a recommendation system. These works relate to our own due to their focus on strategic behavior and conflict in RS.

3.3 Recommendations and partner selection in social dilemmas

We may understand the aggregate of human interaction as multi-agent social dilemmas and recommender systems as mediators of partner selection. Social relationships tend to form scale-free networks, which have been shown to be conducive to the emergence of cooperation under selective pressures. [19] Partner selection and adjustment of social ties have likewise been connected with the emergence of cooperation [58] [60] and reinforcement learning has been used to learn policies of partner selection for an iterated prisoner's dilemma. In this work we'll be using RL to train mediators instead of individuals in the task of partner selection, but the parallels are there. The role of recommendation mechanisms in spatial public goods games has been looked into before, although the recommendations were made by agents for agents instead of by a central self-interested mediator. [70] Finally, Santos et al. (2019) [56] have studied partner selection in collective risk dilemmas, a kind of game with uncertain, non-linear payoffs, using an outcome-based strategy based on empirical experiments with humans.

3.4 Preference drifts from social interaction

Zafari et al. (2018) [71] recognize preference drift as resulting from social interactions and validate their model of the process empirically. However, they do not focus on the recommender system as an actor able to induce preference drift strategically. Vasconcelos et al. (2019) present a model of competitive, complex contagion dynamics that expresses dynamical patterns of *Dominance*, *Polarisation*, and *Consensus* depending only on the relative complexity of the diffusing information. "These patterns are in many ways equivalent to the ones obtained in Evolutionary Game Theory", but require only simpler contagion dynamics.

3.5 Modelling recommendation systems

Most attempts at modelling RS have been made in the context of improving their classification performance, without a need to expressly model users apart from their preferences. Calero-Valdez et al. (2018) [15] makes the argument for multi-agent systems to model complex human/system interaction, or the "human-in-the-loop". Specifically, they remark upon its use in modelling the relationship between users and recommender system. Users may be modelled as rational, or semi-rational, or greedy agents; while RS may be modelled as the very types of algorithms that see deployment in the real world. (See the "Model" section) RecSim [33] and RecGym [54] are modeling frameworks designed to facilitate the development of RL-based RS.

3.6 Sequential social dilemmas and RL

Kumar et al. (2019) [39] used reinforcement-learning models based on Expectancy-Valence-Learning and Prospect-Valence-Learning to investigate human decisions in collective risk dilemmas.

A recent strain of literature has focused on using RL to solve sequential Markov social dilemmas, a Markov game where strategies are policies that satisfy the social dilemma inequalities instead of atomic actions. [44] [32] [27] [42] [38] [50] McKee et al. (2020) [47] pursue this line of research by introducing social diversity and intrinsic social preferences. Our hope would be to eventually extend this dissertation's work to RL agents mediated by self-interested recommenders in sequential social dilemmas. A related type of environment is the spatiotemporal Markov decision process defined in Chu et al. (2020) [16]

3.7 Evaluation via empirical game theory

Finally, Tuyls et al. (2018) [65] present a generalized method for empirical game theoretic analysis. They prove bounds and offer insight about how the meta game - the game resulting from a relevant set of meta strategies, or policies - reflects the underlying game. These techniques would be useful in understanding the strategic behavior of mediators towards users in our simulations and might be applicable when analysing competing mediators. Omidshafiei et al. (2020) develop a graph-theoretic toolkit that facilitates the obtention of high-level strategic insights about real-world games. The toolkit can also be used to automate generation of games, which could be instrumental for generating a comprehensive set of games in which we could analyze the impact of mediators on player performance.

4 Proposed Solution

4.1 The Model

Our model consists of a "core" set of features with extra modules that can be added and removed for simplicity.

Core game: Consider a population A of N users, henceforth called *agents*, organized in an undirected scale-free graph G generated with the Barabási-Albert algorithm [6] in order to approximate the topology of a social network. Each agent $a_i \in A$ is characterized by an intrinsic attribute (or preference) vector v_i and a strategy S_i. Attribute vectors are 5 dimensional by default. We denote the neighborhood of a_i as $a_i.neigh = \{a_j : (a_i, a_j) \in G\}$.

Consider also a recommendation system M, henceforth called a *mediator*, that has complete information about the attributes of users and strategies of users with uncertainty ϵ (estimating $v_i' = v_i + \epsilon * rand(-1, 1)$, we default to $\epsilon = 0$) and a utility function u_M. Mediators have a utility function by which they measure their success. In the case of the self-interested mediators, we will

use deep Q learning[49] to train one or more policies that maximize their utility functions.

The game agents are embedded in consists of two phases: a recommendation phase, and a dilemma phase. The **recommendation phase** is where partner selection [3] happens. Each agent plays a recommendation game (See Table 5) with the mediator, who presents them with a recommendation that they may accept or reject. In this case, a recommendation is a potential partner (or set of partners, but we'll consider only sets of size 1) from anywhere in the network. We model agents as greedy in the recommendation game, (with an exploration parameter ϵ_i) so an agent compares the recommendation with its best alternative partner from its neighborhood and picks the one with the highest expected reward - in this case, by comparing its similarity.

Table 4. Recommendation game

Table 5. Abstract recommendation game payoff matrix. Usually $max_a \geq ok_a \geq alt$ and $max_{RS} \geq ok_{RS} \geq 0$.

Strats	**C**	**D**
C	max_a , ok_{RS}	ok_a , max_{RS}
D	alt , 0	alt , 0

Table 6. Example recommendation game payoff matrix.

Strats	**C**	**D**
C	3 , 1	2 , 2
D	1 , 0	1 , 0

Legend for Table 5:

- max_a: agent's expected utility for accepting an optimal recommendation;
- ok_rs: mediator's expected utility for providing an optimal recommendation;
- ok_a: agent's expected utility for accepting a sub-optimal recommendation;
- max_rs: mediator's expected utility for providing a selfish (suboptimal) recommendation;
- alt: agent's expected utility for refusing a recommendation and so playing with its best local alternative;

The **dilemma phase** consists of a set of prisoner's dilemmas where at least one agent plays one game with the partner they selected in the previous phase. For simplicity, a selected opponent cannot refuse to play.

Agent strategies and mediator utility functions are described in Section 4.2. Experiments. The following sub-sections contain the modules that we can use to extend the core game.

Evolutionary dynamics: (for experiments) To add evolutionary dynamics and enable the analysis of the emergence of cooperation we insert a new dynamic. After playing, agents may adopt a new strategy through the mechanism of pairwise comparison [64], whereupon an agent compares its fitness with a random neighbour's and imitates their strategy with probability

$$p = \frac{1}{1 + e^{-\beta(f_B - f_A)}} \tag{4}$$

where f represents fitness of agents A and B, and β parameterizes the dependency between probability and fitness difference.

Network rewiring: We might find that the evolutionary models reach equilibria quite soon, or that cooperation isn't able to emerge or dominate. Whatever the reason, we might want to extend the model by giving agents the reasonable ability to rewire their social ties after playing with a stranger with probability given by the pairwise comparison of the similarities of the new partner and an agents worst neighbor.

Extrinsic social preferences: A natural extension of network rewiring would be introducing an extrinsic component on the agent preference vector that is influenced by its neighbor's preferences. Allied with network rewiring, extrinsic social preferences would produce a "spreading" phenomenon.

$$v_i' = v_i + \frac{1}{|a_i.neigh|} \sum_{j \in a_i.neigh} v_j \tag{5}$$

Vasconcelos et al. (2019)[66] find that the spread of preferences by complex contagion produces effects very similar to opinion dynamics, which is one of the factors we're concerned might be negatively influenced by repeated exposure self-interested recommendation.

4.2 Experiments

The following experiments are proposed given our objectives of studying the impact of self-interested mediators on multi-agent systems facing social dilemmas. We start by describing the strategies agents have available and the mediator utility function we'll study.

Agent dilemma strategies: Agents will be randomly assigned one of 4 basic strategies at the start of the game: Always cooperate, Always defect; Discriminate - cooperate only with cooperators; Paradoxically discriminate - cooperate only with defectors. The last two strategies imply that the agent must be able to know an agent's last action. As a rule we assume that agents know their neighbor's last action, but make a random assumption about the last actions of agents from outside of their neighborhoods.

Mediator utility functions: The main focus of these experiments is to study self-interested mediators: our self-interested mediator maximizes "retention", or how many times its recommendations are accepted in a game with T rounds.

As baselines to compare the self-interested mediator to, we consider the absence of mediators, and an idealised "perfect" mediator, which makes the best possible recommendations for the agent based on agent attribute similarity and strategy. Policies more complex than "select most similar pair" - as is the case for our self-interested policy of maximizing retention - must be trained using the Q reinforcement learning algorithm for every variation of the environment we run. The learning process itself should be of interest and so we will be recording utility distribution, cooperation rate, and network parameters over time.
A fourth possible utility function would be based solely on user similarity, reflecting a "well-intentioned" mediator that isn't necessarily self-interested, but still optimizes for a metric that is correlated but not aligned with user utility. (because of strategies) A fifth possible utility function would be the "naive self-interested mediator", which would be implemented by offering only recommendations that are slightly better than the local alternatives but otherwise as bad as possible. It follows the naive reasoning of protecting its self-interest by immediately giving the best possible recommendation and then being made useless. These last two don't require RL training.

Initial experiment: Our initial experiment will consist on simulating the core game from section 3.2 under the 3 (or 4) different mediators: Agents with attribute-vectors play a 2-phased game with a recommendation phase and a dilemma phase, where they select their partner from neighbors and recommendations, and then proceed to play a match of iterated prisoner's dilemma. When played with a fixed mediator and no stochasticity (in the ϵs of observations and exploration), the behaviors should be the same in every phase.

Evolutionary dynamics: After simulating the core game for a few iterations, we will be able to observe differences in the utility distributions, but not much else. To investigate the impact of recommendation in the emergence of cooperation, we would need enable evolutionary dynamics so that agents may switch their strategy to the strategy of one of their neighbors in (exponential) proportion to the difference between their fitness scores. This lets us observe how the cooperation rate evolves under different mediator utility functions.

Network Rewiring: After introducing evolutionary dynamics, convergence might be simple or the effects of the self-interested mediator may not be as pronounced as one would expect from observing their real world impacts. The next proposed experience would be to momentarily forget evolutionary dynamics and introduce network rewiring. If we allow recommendations to create opportunities for rewiring, we can expect more strategic behavior to come from a self-interested mediator. Of course, in real life network rewiring can happen without RS albeit more slowly, so we may decide to introduce a similar mechanism that would allow agents to rewire to one of the neighbors of their last partner. Change in overall network structure would be measured in addition to the utility distribution.
Extrinsic social preferences: A natural extension to network rewiring, as

mentioned in 3.1, is adding a socially determined component to the attribute vector v_i of each agent. Besides network properties and utility distribution, this would allow us to measure the distribution of each of the vector components and analyze how attributes change over time.

Finally, a simulation would be run and duly analyzed with all modules: evolutionary dynamics, rewiring, and socially determined attributes.

Mediator competition: If results from self-interested recommendation systems are discouraging, we can analyze how they would fare in a competitive setting. For this, we would simply give an agent more than one recommendation and let it choose as usual. Competition is expected to moderate the selfishness of mediators, but is not a particularly realistic solution to the mediator dilemma since current recommendation systems hold effective monopolies over the markets they serve. (Facebook, Amazon, Google, Netflix, etc...) If this scenario remains grim, we could investigate a potential intervention in the market. A naive suggestion is introducing some notion of coalitions of users and conditional commitment between them for changing mediators, but this sounds like a public goods problem in itself.

5 Evaluation Methodology

We will be evaluating populations in systems exposed to self-interested mediators by comparing their utility, strategy distribution, network structure, and aggregate preferences, with those of systems exposed to idealized baseline mediators or to none at all.

The simplest metric we intend to use is the utility distribution (See Figure 7) and its total in a population after several iterations of a game, both after and during training of the mediators.

In experiments that feature evolutionary dynamics, we are interested in the distribution of strategies in evolutionarily stable states and in the cooperation rate η which is obtained from

$$\eta = \frac{1}{T} \sum_{i=0}^{T} \frac{C_i}{K_i} \tag{6}$$

Where T is the number of runs, C_i is the total number of cooperative acts in run i, and K_i is the total number of games played in run i;

For experiments that feature network rewiring, we are interested in the above but also in features of the network structure like average degree, the degree distribution, the clustering coefficient, and the average path length. It could also be interesting to plot utility as a function of certain node properties like degree or centrality.

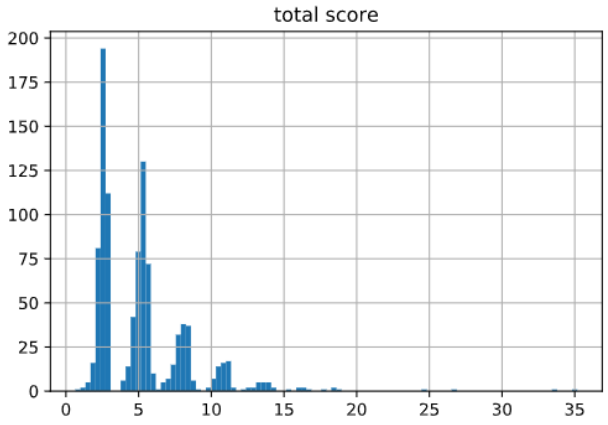


Fig. 7. Example of utility distribution after a simulation of the core environment using a naively self-interested mediator from our preliminary results. X-axis plots utility value, Y-axis frequency of agents with utility in a given interval.

In cases where the spreading of preferences takes place, it would further be of interest to plot how their variance changes over time.

Finally the evaluation of the scenario of competition between mediators would be performed through empirical game theory and replicator dynamics [65]

With all this, we hope to begin to understand the space of possibilities in social dilemmas where interactions between individuals are strategically mediated to the benefit of a third party.

6 Timeline

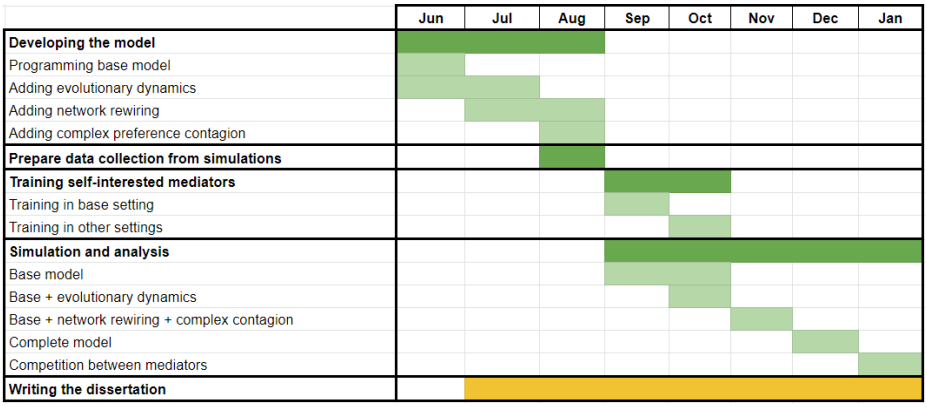


Fig. 8. Schedule of the work to be done to fulfill the objectives of this Master's thesis.

7 Conclusion

In this proposal we have expressed and contextualized our motivations with recourse to AI alignment and the societal impacts of recommendation systems, as well as laid out our theoretical foundations in game theory, complex network science, and reinforcement learning. We introduce a novel mechanic of recommendation in social dilemmas to produce an environment in which to study the societal impacts of recommendation systems, and scheduled a set of experiments.

In the future, it would make sense that the primary extension to these experiments would be to introduce the self-interested recommendation mechanism in other social dilemmas, such as the collective risk dilemma [56], a multi-player game like the public goods dilemma but with uncertainty, which is different from 2-player prisoner's dilemmas but equally significant to the study of the effects of RS on society. Afterwards, more generally, an extension to Markov games.

With this thesis we hope to create a precedent in the study of the impacts of recommendation systems by focusing on the emergent properties of a simple abstract system when exposed to mediators characterized by the incentives of their utility functions. With this, we hope to inform the establishing of design principles of RS, online platforms, and legislation.

References

[1] G. Adomavicius and A. Tuzhilin. "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions". In: *IEEE Transactions on Knowledge and Data Engineering* 17.6 (June 2005), pp. 734–749. ISSN: 1041-4347. DOI: 10.1109/TKDE.2005.99. URL: http://ieeexplore.ieee.org/document/1423975/ (visited on 06/07/2020).

[2] Dario Amodei et al. "Concrete Problems in AI Safety". In: *arXiv:1606.06565 [cs]* (July 2016). arXiv: 1606.06565. URL: http://arxiv.org/abs/1606.06565 (visited on 06/07/2020).

[3] Nicolas Anastassacos, Stephen Hailes, and Mirco Musolesi. "Partner Selection for the Emergence of Cooperation in Multi-Agent Systems Using Reinforcement Learning". In: *arXiv:1902.03185 [cs]* (Nov. 2019). arXiv: 1902.03185. URL: http://arxiv.org/abs/1902.03185 (visited on 05/29/2020).

[4] Dan Ariely. *Predictably irrational: the hidden forces that shape our decisions*. 1st ed. OCLC: ocn182521026. New York, NY: Harper, 2008. ISBN: 9780061353239.

[5] David Balduzzi et al. "Smooth markets: A basic mechanism for organizing gradient-based learners". In: *arXiv:2001.04678 [cs, stat]* (Jan. 2020). arXiv: 2001.04678. URL: http://arxiv.org/abs/2001.04678 (visited on 05/29/2020).

[6] Albert-László Barabási and Réka Albert. "Emergence of Scaling in Random Networks". en. In: *Science* 286.5439 (Oct. 1999), pp. 509–512. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.286.5439.509. URL: https://science.sciencemag.org/content/286/5439/509 (visited on 06/06/2020).

[7] Omer Ben-Porat and Moshe Tennenholtz. "A Game-Theoretic Approach to Recommendation Systems with Strategic Content Providers". In: *arXiv:1806.00955 [cs]* (Oct. 2018). arXiv: 1806.00955. URL: http://arxiv.org/abs/1806.00955 (visited on 06/03/2020).

[8] Omer Ben-Porat and Moshe Tennenholtz. "Competing Prediction Algorithms". In: *arXiv:1806.01703 [cs]* (June 2018). arXiv: 1806.01703 version: 1. URL: http://arxiv.org/abs/1806.01703 (visited on 05/29/2020).

[9] Nick Bostrom. *Superintelligence : Paths, Dangers, Strategies*. original-date: July 2014. July 2014. URL: https://nickbostrom.com/views/superintelligence.pdf.

[10] Engin Bozdag. "Bias in algorithmic filtering and personalization". en. In: *Ethics and Information Technology* 15.3 (Sept. 2013), pp. 209–227. ISSN: 1388-1957, 1572-8439. DOI: 10.1007/s10676-013-9321-6. URL: http://link.springer.com/10.1007/s10676-013-9321-6 (visited on 06/07/2020).

[11] Engin Bozdag and Jeroen van den Hoven. "Breaking the filter bubble: democracy and design". en. In: *Ethics and Information Technology* 17.4 (Dec. 2015), pp. 249–265. ISSN: 1388-1957, 1572-8439. DOI: 10.1007/

s10676-015-9380-y. URL: http://link.springer.com/10.1007/s10676-015-9380-y (visited on 06/07/2020).

[12] Talha Burki. "Vaccine misinformation and social media". en. In: *The Lancet Digital Health* 1.6 (Oct. 2019), e258–e259. ISSN: 25897500. DOI: 10.1016/S2589-7500(19)30136-0. URL: https://linkinghub.elsevier.com/retrieve/pii/S2589750019301360 (visited on 06/07/2020).

[13] Christopher Burr, Nello Cristianini, and James Ladyman. "An Analysis of the Interaction Between Intelligent Software Agents and Human Users". en. In: *Minds and Machines* 28.4 (Dec. 2018), pp. 735–774. ISSN: 1572-8641. DOI: 10.1007/s11023-018-9479-0. URL: https://doi.org/10.1007/s11023-018-9479-0 (visited on 06/07/2020).

[14] Christopher Burr, Nello Cristianini, and James Ladyman. "An Analysis of the Interaction Between Intelligent Software Agents and Human Users". en. In: *Minds and Machines* 28.4 (Dec. 2018), pp. 735–774. ISSN: 1572-8641. DOI: 10.1007/s11023-018-9479-0. URL: https://doi.org/10.1007/s11023-018-9479-0 (visited on 05/31/2020).

[15] André Calero Valdez and Martina Ziefle. "Human Factors in the Age of Algorithms. Understanding the Human-in-the-loop Using Agent-Based Modeling". en. In: *Social Computing and Social Media. Technologies and Analytics*. Ed. by Gabriele Meiselwitz. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, pp. 357–371. ISBN: 9783319914855. DOI: 10.1007/978-3-319-91485-5_27.

[16] Tianshu Chu, Sandeep Chinchali, and Sachin Katti. "Multi-agent Reinforcement Learning for Networked System Control". In: *arXiv:2004.01339 [cs, stat]* (Apr. 2020). arXiv: 2004.01339. URL: http://arxiv.org/abs/2004.01339 (visited on 06/03/2020).

[17] Jesse Clifton. *Cooperation, Conflict, and Transformative Artificial Intelligence: A Research Agenda*. Tech. rep. Dec. 2019.

[18] K.E. Drexler. *Reframing Superintelligence: Comprehensive AI Services as General Intelligence", Technical Report*. Tech. rep. Future of Humanity Institute, University of Oxford, Jan. 2019.

[19] Santos Fc and Pacheco Jm. *Scale-free Networks Provide a Unifying Framework for the Emergence of Cooperation*. en. Aug. 2005. DOI: 10.1103/PhysRevLett.95.098104. URL: https://pubmed.ncbi.nlm.nih.gov/16197256/ (visited on 05/29/2020).

[20] Luciano Floridi. "The Construction of Personal Identities Online". en. In: *Minds and Machines* 21.4 (Nov. 2011), pp. 477–479. ISSN: 0924-6495, 1572-8641. DOI: 10.1007/s11023-011-9254-y. URL: http://link.springer.com/10.1007/s11023-011-9254-y (visited on 06/07/2020).

[21] Charles Goodhart. *Problems of Monetary Management: The UK Experience*. original-date: 1981. URL: https://books.google.ch/books?id=OMe6UQxu1KcC&pg=PA111&redir_esc=y#v=onepage&q&f=false.

[22] Dylan Hadfield-Menell et al. "Cooperative Inverse Reinforcement Learning". In: *Advances in Neural Information Processing Systems 29*. Ed. by D. D. Lee et al. Curran Associates, Inc., 2016, pp. 3909–3917. URL: http://

papers.nips.cc/paper/6420-cooperative-inverse-reinforcement-learning.pdf (visited on 06/07/2020).

[23] Dylan Hadfield-Menell et al. "Inverse Reward Design". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 6765–6774. URL: http://papers.nips.cc/paper/7253-inverse-reward-design.pdf (visited on 06/07/2020).

[24] Maria Halkidi and Iordanis Koutsopoulos. "Recommender systems with selfish users". en. In: *Knowledge and Information Systems* (Mar. 2020). ISSN: 0219-3116. DOI: 10.1007/s10115-020-01460-5. URL: https://doi.org/10.1007/s10115-020-01460-5 (visited on 06/01/2020).

[25] Jaron Harambam, Natali Helberger, and Joris van Hoboken. "Democratizing algorithmic news recommenders: how to materialize voice in a technologically saturated media ecosystem". en. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2133 (Nov. 2018), p. 20180088. ISSN: 1364-503X, 1471-2962. DOI: 10.1098/rsta.2018.0088. URL: https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0088 (visited on 06/07/2020).

[26] Natali Helberger, Kari Karppinen, and Lucia D'Acunto. "Exposure diversity as a design principle for recommender systems". en. In: *Information, Communication & Society* 21.2 (Feb. 2018), pp. 191–207. ISSN: 1369-118X, 1468-4462. DOI: 10.1080/1369118X.2016.1271900. URL: https://www.tandfonline.com/doi/full/10.1080/1369118X.2016.1271900 (visited on 06/07/2020).

[27] Pablo Hernandez-Leal, Bilal Kartal, and Matthew E. Taylor. "A Survey and Critique of Multiagent Deep Reinforcement Learning". In: *Autonomous Agents and Multi-Agent Systems* 33.6 (Nov. 2019). arXiv: 1810.05587, pp. 750–797. ISSN: 1387-2532, 1573-7454. DOI: 10.1007/s10458-019-09421-1. URL: http://arxiv.org/abs/1810.05587 (visited on 05/29/2020).

[28] Lê Nguyên Hoang. "A Roadmap for Robust End-to-End Alignment". In: *arXiv:1809.01036 [cs]* (Feb. 2020). arXiv: 1809.01036. URL: http://arxiv.org/abs/1809.01036 (visited on 06/02/2020).

[29] Henning Hohnhold, Deirdre O'Brien, and Diane Tang. "Focusing on the Long-term: It's Good for Users and Business". In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '15. Sydney, NSW, Australia: Association for Computing Machinery, Aug. 2015, pp. 1849–1858. ISBN: 9781450336642. DOI: 10.1145/2783258.2788583. URL: https://doi.org/10.1145/2783258.2788583 (visited on 05/29/2020).

[30] Evan Hubinger. *An overview of 11 proposals for building safe advanced AI - AI Alignment Forum*. 2020. URL: https://www.alignmentforum.org/posts/fRsjBseRuvRhMPPE5/an-overview-of-11-proposals-for-building-safe-advanced-ai (visited on 06/07/2020).

[31] Evan Hubinger. *Relaxed adversarial training for inner alignment - AI Alignment Forum*. 2019. URL: https://www.alignmentforum.org/

posts/9Dy5YRaoCxH9zuJqa/relaxed-adversarial-training-for-inner-alignment (visited on 06/07/2020).

[32] Edward Hughes et al. "Inequity aversion improves cooperation in intertemporal social dilemmas". In: *arXiv:1803.08884 [cs, q-bio]* (Sept. 2018). arXiv: 1803.08884. URL: http://arxiv.org/abs/1803.08884 (visited on 06/06/2020).

[33] Eugene Ie et al. "RecSim: A Configurable Simulation Platform for Recommender Systems". In: *arXiv:1909.04847 [cs, stat]* (Sept. 2019). arXiv: 1909.04847. URL: http://arxiv.org/abs/1909.04847 (visited on 05/29/2020).

[34] Geoffrey Irving, Paul Christiano, and Dario Amodei. "AI safety via debate". In: *arXiv:1805.00899 [cs, stat]* (Oct. 2018). arXiv: 1805.00899. URL: http://arxiv.org/abs/1805.00899 (visited on 06/07/2020).

[35] Luis R. Izquierdo, Luis R. Izquierdo, and Segismundo Izquierdo. "Reinforcement learning dynamics in social dilemmas". en. In: (). URL: https://www.academia.edu/15281537/Reinforcement_learning_dynamics_in_social_dilemmas (visited on 06/06/2020).

[36] Peter Izsak et al. "The search duel: a response to a strong ranker". In: *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. SIGIR '14. Gold Coast, Queensland, Australia: Association for Computing Machinery, July 2014, pp. 919–922. ISBN: 9781450322577. DOI: 10.1145/2600428.2609474. URL: https://doi.org/10.1145/2600428.2609474 (visited on 05/29/2020).

[37] Dietmar Jannach and Gediminas Adomavicius. "Recommendations with a Purpose". en. In: *Proceedings of the 10th ACM Conference on Recommender Systems*. Boston Massachusetts USA: ACM, Sept. 2016, pp. 7–10. ISBN: 9781450340359. DOI: 10.1145/2959100.2959186. URL: https://dl.acm.org/doi/10.1145/2959100.2959186 (visited on 06/07/2020).

[38] Natasha Jaques et al. "Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning". In: *arXiv:1810.08647 [cs, stat]* (June 2019). arXiv: 1810.08647 version: 3. URL: http://arxiv.org/abs/1810.08647 (visited on 05/29/2020).

[39] Medha Kumar, Kapil Agrawal, and Varun Dutt. "Modeling Decisions in Collective Risk Social Dilemma Games for Climate Change Using Reinforcement Learning". In: *2019 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*. ISSN: 2379-1675. Apr. 2019, pp. 26–33. DOI: 10.1109/COGSIMA.2019.8724273.

[40] Moshe Tennenholtz Kurland Oren. *Rethinking Search Engines and Recommendation Systems: A Game Theoretic Perspective*. en. Dec. 2019. URL: https://cacm.acm.org/magazines/2019/12/241056-rethinking-search-engines-and-recommendation-systems/fulltext (visited on 05/29/2020).

[41] Joel Z. Leibo et al. "Autocurricula and the Emergence of Innovation from Social Interaction: A Manifesto for Multi-Agent Intelligence Research". In:

arXiv:1903.00742 [cs, q-bio] (Mar. 2019). arXiv: 1903.00742. URL: http://arxiv.org/abs/1903.00742 (visited on 05/29/2020).

[42] Joel Z. Leibo et al. "Multi-agent Reinforcement Learning in Sequential Social Dilemmas". In: *arXiv:1702.03037 [cs]* (Feb. 2017). arXiv: 1702.03037. URL: http://arxiv.org/abs/1702.03037 (visited on 05/29/2020).

[43] Jan Leike et al. "Scalable agent alignment via reward modeling: a research direction". In: *arXiv:1811.07871 [cs, stat]* (Nov. 2018). arXiv: 1811.07871. URL: http://arxiv.org/abs/1811.07871 (visited on 06/07/2020).

[44] Adam Lerer and Alexander Peysakhovich. "Maintaining cooperation in complex social dilemmas using deep reinforcement learning". In: *arXiv:1707.01068 [cs]* (Mar. 2018). arXiv: 1707.01068. URL: http://arxiv.org/abs/1707.01068 (visited on 05/29/2020).

[45] Mohtashemi M and Mui L. *Evolution of Indirect Reciprocity by Social Information: The Role of Trust and Reputation in Evolution of Altruism*. en. Aug. 2003. DOI: 10.1016/s0022-5193(03)00143-7. URL: https://pubmed.ncbi.nlm.nih.gov/12875829/ (visited on 06/06/2020).

[46] Michael W. Macy and Andreas Flache. "Learning dynamics in social dilemmas". In: *Proceedings of the National Academy of Sciences of the United States of America* 99.Suppl 3 (May 2002), pp. 7229–7236. ISSN: 0027-8424. DOI: 10.1073/pnas.092080099. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC128590/ (visited on 06/06/2020).

[47] Kevin R. McKee et al. "Social diversity and social preferences in mixed-motive reinforcement learning". In: *arXiv:2002.02325 [cs]* (Feb. 2020). arXiv: 2002.02325. URL: http://arxiv.org/abs/2002.02325 (visited on 06/01/2020).

[48] Silvia Milano, Mariarosaria Taddeo, and Luciano Floridi. "Recommender systems and their ethical challenges". en. In: *AI & SOCIETY* (Feb. 2020). ISSN: 1435-5655. DOI: 10.1007/s00146-020-00950-y. URL: https://doi.org/10.1007/s00146-020-00950-y (visited on 05/29/2020).

[49] Volodymyr Mnih et al. "Human-level control through deep reinforcement learning". en. In: *Nature* 518.7540 (Feb. 2015), pp. 529–533. ISSN: 1476-4687. DOI: 10.1038/nature14236. URL: https://www.nature.com/articles/nature14236 (visited on 06/06/2020).

[50] Alexander Peysakhovich and Adam Lerer. "Prosocial learning agents solve generalized Stag Hunts better than selfish ones". In: *arXiv:1709.02865 [cs]* (Dec. 2017). arXiv: 1709.02865. URL: http://arxiv.org/abs/1709.02865 (visited on 06/06/2020).

[51] Axelrod R and Hamilton Wd. *The Evolution of Cooperation*. en. Mar. 1981. DOI: 10.1126/science.7466396. URL: https://pubmed.ncbi.nlm.nih.gov/7466396/ (visited on 06/06/2020).

[52] Urbano Reviglio. "Serendipity by Design? How to Turn from Diversity Exposure to Diversity Experience to Face Filter Bubbles in Social Media". In: *Internet Science*. Ed. by Ioannis Kompatsiaris et al. Vol. 10673. Cham: Springer International Publishing, 2017, pp. 281–300. ISBN: 9783319702834 9783319702841. DOI: 10.1007/978-3-319-70284-1_22. URL: http:

//link.springer.com/10.1007/978-3-319-70284-1_22 (visited on 06/07/2020).

[53] Francesco Ricci, Lior Rokach, and Bracha Shapira, eds. *Recommender Systems Handbook*. en. 2nd ed. Springer US, 2015. ISBN: 9781489976369. DOI: 10.1007/978-1-4899-7637-6. URL: https://www.springer.com/gb/book/9781489976369 (visited on 06/07/2020).

[54] David Rohde et al. "RecoGym: A Reinforcement Learning Environment for the problem of Product Recommendation in Online Advertising". In: *arXiv:1808.00720 [cs]* (Sept. 2018). arXiv: 1808.00720. URL: http://arxiv.org/abs/1808.00720 (visited on 05/29/2020).

[55] Stuart J. Russell and Peter Norvig. *Artificial intelligence: a modern approach*. 3rd ed. Prentice Hall series in artificial intelligence. Upper Saddle River, N.J: Prentice Hall/Pearson Education, 2010. ISBN: 9780137903955.

[56] Fernando P. Santos et al. "Outcome-based Partner Selection in Collective Risk Dilemmas". In: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS '19. Montreal QC, Canada: International Foundation for Autonomous Agents and Multiagent Systems, May 2019, pp. 1556–1564. ISBN: 9781450363099. (Visited on 05/29/2020).

[57] Francisco C. Santos, Marta D. Santos, and Jorge M. Pacheco. "Social diversity promotes the emergence of cooperation in public goods games". en. In: *Nature* 454.7201 (July 2008), pp. 213–216. ISSN: 1476-4687. DOI: 10.1038/nature06940. URL: https://www.nature.com/articles/nature06940 (visited on 06/06/2020).

[58] Francisco C Santos, Jorge M Pacheco, and Tom Lenaerts. "Cooperation Prevails When Individuals Adjust Their Social Ties". In: *PLoS Computational Biology* 2.10 (Oct. 2006). ISSN: 1553-734X. DOI: 10.1371/journal.pcbi.0020140. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1617133/ (visited on 05/29/2020).

[59] Nick Seaver. "Captivating algorithms: Recommender systems as traps". en. In: *Journal of Material Culture* 24.4 (Dec. 2019), pp. 421–436. ISSN: 1359-1835, 1460-3586. DOI: 10.1177/1359183518820366. URL: http://journals.sagepub.com/doi/10.1177/1359183518820366 (visited on 05/29/2020).

[60] Sven Van Segbroeck et al. "Selection pressure transforms the nature of social dilemmas in adaptive networks". In: *New Journal of Physics* 13.1 (Jan. 2011), p. 013007. ISSN: 1367-2630. DOI: 10.1088/1367-2630/13/1/013007. URL: https://iopscience.iop.org/article/10.1088/1367-2630/13/1/013007 (visited on 06/03/2020).

[61] Time Well Spent. *What's the difference between apps we cherish vs. regret?* Tech. rep. 2017. URL: http://www.timewellspent.io/app-ratings/.

[62] Jacob Steinhardt. *AI Alignment Research Overview (by Jacob Steinhardt) - LessWrong 2.0*. 2019. URL: https://www.lesswrong.com/posts/7GEviErBXcjJsbSeD/ai-alignment-research-overview-by-jacob-steinhardt (visited on 06/07/2020).

[63] Mariarosaria Taddeo and Luciano Floridi. "How AI can be a force for good". en. In: *Science* 361.6404 (Aug. 2018), pp. 751–752. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aat5991. URL: https://www.sciencemag.org/lookup/doi/10.1126/science.aat5991 (visited on 06/07/2020).

[64] Arne Traulsen, Martin A. Nowak, and Jorge M. Pacheco. "Stochastic Dynamics of Invasion and Fixation". In: *Physical Review E* 74.1 (July 2006). arXiv: q-bio/0609020, p. 011909. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.74.011909. URL: http://arxiv.org/abs/q-bio/0609020 (visited on 05/31/2020).

[65] Karl Tuyls et al. "A Generalised Method for Empirical Game Theoretic Analysis". In: *arXiv:1803.06376 [cs]* (Mar. 2018). arXiv: 1803.06376. URL: http://arxiv.org/abs/1803.06376 (visited on 05/29/2020).

[66] Vítor V. Vasconcelos, Simon A. Levin, and Flávio L. Pinheiro. "Consensus and Polarisation in Competing Complex Contagion Processes". In: *Journal of The Royal Society Interface* 16.155 (June 2019). arXiv: 1811.08525, p. 20190196. ISSN: 1742-5689, 1742-5662. DOI: 10.1098/rsif.2019.0196. URL: http://arxiv.org/abs/1811.08525 (visited on 05/29/2020).

[67] Ivan Vendrov and Jeremy Nixon. *Aligning Recommender Systems as Cause Area*. May 2019. URL: https://forum.effectivealtruism.org/posts/xzjQvqDYahigHcwgQ/aligning-recommender-systems-as-cause-area.

[68] Katja de Vries. "Identity, profiling algorithms and a world of ambient intelligence". en. In: *Ethics and Information Technology* 12.1 (Mar. 2010), pp. 71–85. ISSN: 1388-1957, 1572-8439. DOI: 10.1007/s10676-009-9215-9. URL: http://link.springer.com/10.1007/s10676-009-9215-9 (visited on 06/07/2020).

[69] Christopher J. C. H. Watkins and Peter Dayan. "Q-learning". en. In: *Machine Learning* 8.3-4 (May 1992), pp. 279–292. ISSN: 0885-6125, 1573-0565. DOI: 10.1007/BF00992698. URL: http://link.springer.com/10.1007/BF00992698 (visited on 06/06/2020).

[70] Zhihu Yang et al. "Role of recommendation in spatial public goods games". en. In: *Physica A: Statistical Mechanics and its Applications* 392.9 (May 2013), pp. 2038–2045. ISSN: 03784371. DOI: 10.1016/j.physa.2012.11.024. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378437112009892 (visited on 05/29/2020).

[71] F. Zafari, I. Moser, and T. Baarslag. "Modelling and Analysis of Temporal Preference Drifts Using A Component-Based Factorised Latent Approach". en. In: (Feb. 2018). URL: https://www.arxiv-vanity.com/papers/1802.09728/ (visited on 05/29/2020).

[72] K.F.A. Zoetekouw. *A critical analysis of the negative consequences caused by recommender systems used on social media platforms*. July 2019. URL: http://essay.utwente.nl/78500.