Thou Shalt Love Thy Neighbor as Thyself When Thou Playest: Altruism in Game Theory

Jörg Rothe
Institut für Informatik
Heinrich-Heine-Universität Düsseldorf
40225 Düsseldorf, Germany
rothe@hhu.de

Abstract

Game theory is typically used to model the interaction among (software) agents in multiagent systems and, therefore, is a key topic at leading AI conferences. Game-theoretic models, however, are often based on the assumption that agents are perfectly rational and narrowly selfish and are interested only in maximizing their own gains, no matter what the costs to the other agents are. This summary paper presents various ways of introducing certain notions of altruism into existing game-theoretic models in both noncooperative and cooperative games, in the hope that simulating altruistic behavior in AI systems will make AI better suit real-world applications—and thus may make the real world a better place.

Introduction

The breathtakingly rapid development of artificial intelligence (AI) is largely based on mimicking—by means of tools, methods, and insights from computer science, mathematics, and other fields of science—human intelligence and human properties, attributes, and behavior as individuals and in society. For example, the explosive recent successes of machine learning and deep learning, a true success story within AI, have been possible only by mimicking the human brain through neural networks[1] and by machines acquiring the human ability to learn from previous experiences.

Other subareas of AI are concerned with using game theory, social choice theory, voting, and economic paradigms so as to model interaction, coordination, collaboration, and collective decision-making among the agents in a multiagent system. For instance, social choice mechanisms and voting have been widely and successfully used in multiagent systems and AI for preference aggregation and collective decision-making in societies of (software) agents.

The rise of *computational social choice* (COMSOC, for short; see, e.g., the comprehensive textbooks edited by Brandt et al. (2016) and Rothe (2015) as well as the surveys written by Chevaleyre et al. (2007), Bredereck et al. (2014), Hemaspaandra (2018), and Rothe (2019a)) is another well-known success story within AI, as witnessed by the fact that numerous important COMSOC results have been presented at leading AI conferences in recent years, including AAAI, AAMAS, ECAI, and IJCAI. Interaction among agents in a multiagent system is typically modeled via game-theoretic means in AI. From the early beginnings of (noncooperative) game theory due to von Neumann and Morgenstern (1944), a player (or agent) in a game has been viewed as a *homo economicus*: Such players are perfectly rational, narrowly selfish, and interested only in maximizing their own gains, no matter what the costs to the other players are. In spirit, this assumption is somewhat related to Darwin's thesis of *"survival of the fittest,"* where "survival" essentially is measured by the ability of reproduction, propelling the biological evolution of the human species. However, even in terms of biology and evolution, there are reasonable doubts if selfishness alone (in the sense that more aggressive behavior yields more offspring) is really the key to success.

Recently, Hare and Woods (2020) countered Darwin's thesis with their *"survival of the friendliest."* Specifically, one of their many arguments is that of the two species making up the genus Pan among the great apes, bonobos and chimpanzees, the bonobos benefit from their much friendlier behavior: The most successful male bonobo has more progenies than the most successful male chimpanzee, i.e., has a higher reproduction rate. Hare and Woods (2020) also argue that the evolutionary supremacy of the human species is mainly due to their *friendly* behavior, which made it possible for them to form larger social groups and even societies.

Now, if we agree that AI is best off when mimicking natural life and simulating real-world human behavior, the *homo economicus* from the early days of game theory is obsolete and better models are needed. Indeed, relentlessly aiming at one's own advantage and maximizing one's own utility regardless of the consequences for others in fact not only diminishes the individual gains of the agents, but it may also harm the society of agents in a multiagent system as a whole.

This paper presents various ways of introducing altruism into existing game-theoretic models, focusing both on noncooperative and cooperative game theory. This adds to previous approaches of taking ethics, psychology, emotions, and behavioral dynamics into consideration in collective decision-making (Regenwetter et al. 2006; Popova, Regenwetter, and Mattei 2013; Rothe 2019b). The purpose of this paper is to make a case for better integrating altruism into AI systems and AI research.

[1] An interpretation that some researchers disagree with, though.

Altruism in Noncooperative Games

Game theory more or less started with the book by von Neumann and Morgenstern (1944) who explored *noncooperative* games in which all players are on their own, competing with each other to win the game and to maximize their own profit. The simplest noncooperative games are called *strategic games in normal form* (Borel 1921; von Neumann 1928) where each player can choose one out of a finite set of possible strategies, and a gain vector gives the gains of all players depending on which strategy was chosen by each of them. For more background on noncooperative game theory, the reader is referred, e.g., to the book by Nisan et al. (2007) and the book chapter by Faliszewski, Rothe, and Rothe (2015).

Altruism in games has been considered mainly for noncooperative games to date. We give a short overview in the remainder of this section.

Social Context Games

Ashlagi, Krysta, and Tennenholtz (2008) introduced *social context games* by embedding a strategic game in a social context that is modeled by a graph of neighborhood among players and aggregation functions. It has the same players and strategies as the underlying strategic game. However, while social context games may not capture an actual notion of altruism, the players' utilities in such a game do not depend only on their payoffs in the underlying strategic game but also on the neighborhood graph and the aggregation functions that express a social context. The goal is to explore the influence a social context may have on the properties of the underlying strategic game.

Well-known examples of social context games are *ranking games* and *coalitional congestion games*, as explained below. Ashlagi, Krysta, and Tennenholtz (2008) focus on *resource selection games* (a famous subclass of *congestion games*[2]) where there are players and resources, each player selects a resource, and each player's costs are a nondecreasing function of the number of players who have chosen this player's selected resource. They ask under which conditions there exist pure Nash equilibria (see Footnote 2 for an informal definition) when resource selection games are embedded into one of the following social contexts:

- *Rank competition:* Players are partitioned into cliques and compete on their relative payoff within each clique. This generalizes ranking games (Brandt et al. 2009), which are based on graphs forming only single cliques.
- *Best-member collaboration:* Given a social network, players care about the highest payoff for themselves or one of their neighbors. This is akin to congestion games in which players may choose several resources and care about the one with the best performance.
- *Min-max collaboration:* Given a social network, players seek to maximize the worst payoff for themselves or one of their friends. This requirement is similar in spirit to min-max fairness.
- *Surplus collaboration:* Given a social network, players seek to maximize the average payoff for themselves and their friends. This requirement is similar in spirit to coalitional congestion games (Hayrapetyan, Tardos, and Wexler 2006; Kuniavsky and Smorodinsky 2011).

Hoefer et al. (2011) also consider players being embedded in a social network and assume that certain constraints specify which sets of coalitions may jointly deviate from their actual strategies in the game. When doing so, however, they are *considerate* not to hurt others: They ignore potentially profitable group deviations whenever those cause a decrease of the gains of their neighbors in the network. Exploring the properties of *considerate equilibria* in resource selection games, they show that there exists a state that is stable against selfish and considerate behavior at the same time.

Anagnostopoulos et al. (2013) study how the kind of altruistic behavior of players in social context games can make them in fact more inefficient, in the sense that the *price of anarchy* (relating the worst-case cost of a Nash equilibrium to the one of an optimal outcome) can thus be increased.

Bilò et al. (2013) apply the model of social context games by Ashlagi, Krysta, and Tennenholtz (2008) to *linear congestion games* and *Shapley cost sharing games* with the aggregation functions min, max, and sum (or average). They characterize the graph topologies modeling the social contexts in these cases such that the existence of pure Nash equilibria is guaranteed. They also establish (asymptotically) optimal bounds on the price of anarchy in many of these cases and they extend their results to *multicast games*, an important subclass of Shapley cost sharing games.

Congestion Games

Hoefer and Skopalik (2013) consider *atomic congestion games*: In the standard case, as sketched above, there are myopic selfish players and a set of resources which each have a nondecreasing delay function; each player chooses a strategy by selecting or allocating a subset of resources (e.g., a path in a network) and experiences a delay corresponding to the total delay on all selected resources, which depends on the number of players that have allocated any of these resources. The goal of each player is to minimize the experienced delay. A stable state in this game is expressed by a pure Nash equilibrium (see Footnote 2), which here means that every player allocates exactly one subset of resources, and no player can decrease the experienced delay by unilaterally deviating from the chosen strategy.

Now, altruism is introduced by Hoefer and Skopalik (2013) into such games as follows. They assume that the players are partly selfish and partly altruistic, which is formalized by an *altruism level* $\beta_i \in [0,1]$ for each player i, where $\beta_i = 0$ means i is purely selfish and $\beta_i = 1$ means i is purely altruistic. These players' incentive then is to optimize a linear combination of personal cost (their individual experienced delay) and social cost (the total cost—i.e., experienced delay—of all players).

[2]A fundamental property of congestion games is that they always have a *Nash equilibrium in pure strategies*, i.e., there always exists a profile of pure strategies such that no player has an incentive to deviate from her strategy in the profile, provided the other players also stay with their strategies in the profile.

Hoefer and Skopalik (2013) then study under which conditions there exist pure Nash equilibria in various types of such games. For example, in *symmetric singleton games* where only one resource is chosen by each player and all players have the same strategy space available, pure Nash equilibria might not exist, even when players are purely altruistic or purely selfish. However, the existence problem can be solved in polynomial time, and a Nash equilibrium with best or worst social cost, if some exists, can be determined efficiently for any population of players with a constant number of altruism levels. For *asymmetric singleton games* with different strategy spaces for the players, on the other hand, the existence problem for pure Nash equilibria is NP-hard. Perhaps a bit surprisingly, Hoefer and Skopalik (2013) show that there is a potential function if all delay functions are affine. Consequently, pure Nash equilibria exist and better-response dynamics converge in this case.

To answer the question of how many altruists are required to stabilize a social optimum, Hoefer and Skopalik (2013) also show that *optimal stability thresholds* (the minimum number of altruists such that there exists an optimal Nash equilibrium) and *optimal anarchy thresholds* (the minimum number of altruists such that every Nash equilibrium is optimal) can be computed in polynomial time.

Chen et al. (2014) also study this model of altruism in congestion games but extend it to nonatomic such games. While in *atomic congestion games*, players have a non-negligible size (which means that using or not using a resource by even one player will affect the cost—or delay—perceptibly), players in *nonatomic congestion games* are assumed to be infinitesimally small and there is a continuum of them (which means that using a resource by a single player will have no measureable effect on the cost; only the accumulation of players using it will affect its cost).

Chen et al. (2014) propose that instead of using a convex combination of $(1 - \beta_i)$ times player i's personal cost (or payoff) and β_i times the social cost (or social welfare) as in atomic congestion games, for nonatomic congestion games the appropriate measure is to use the derivative of the social cost or social welfare. Their main modeling contribution is to provide a general definition that is applicable to other classes of games as well, such as *fair cost-sharing games* (the main difference is that if many players use a resource, instead of increasing its cost, this will actually decrease it by contributing towards its purchase: the fixed cost of a resource can be split among all individuals using it) and *valid utility games* (where, again, players select sets of resources but instead of seeking to minimize costs, they now seek to maximize the utility they each individually derive from the set of resources selected by all players jointly).

Strategic Games

Apt and Schäfer (2014) consider strategic games and introduce the notion of *selfishness level*, which are based on the "altruistic games" due to Ledyard (1995) (and, more recently, De Marco and Morgan (2007)) and measures the discrepancy in such games between the social welfare in a Nash equilibrium and in a social optimum. That is, they consider the smallest fraction of social welfare that each player must be offered so as to achieve that a social optimum can be realized in a pure Nash equilibrium. They show that the selfishness level is distinct from the *price of stability* and the *price of anarchy* (Nisan et al. 2007) and that it remains invariant under linear transformations of the gain functions.

The selfishness level can be finite (e.g., in *finite ordinal potential games*) or infinite (e.g., in *weakly acyclic games*). Apt and Schäfer (2014) establish explicit bounds on the selfishness level of *fair cost-sharing games* and *linear congestion games* and determine the selfishness levels of specific well-studied strategic games, such as the n-*player prisoner's dilemma*, the n-*player public goods game*, and the *traveler's dilemma game*. They show that other specific games like *Cournot competition* (which is an infinite ordinal potential game), *tragedy of the commons*, and *Bertrand competition* have an infinite selfishness level. Due to space restrictions, these interesting well-known games cannot be defined here in detail.

Social Contribution Games

Rahn and Schäfer (2013) introduce another class of games, which they call *social contribution games*. They are motivated by the fact that, as we have seen above, altruistic behavior (i.e., taking other players' preferences or utilities into account when making a decision) may actually render equilibria more inefficient (e.g., in congestion games) and thus may harm society as a whole (Anagnostopoulos et al. 2013). This is not the case for valid utility games, though, as Chen et al. (2014) have shown that the inefficiency of equilibria remains unaltered under altruistic behavior in these games. Therefore, a question naturally arises: What is it that causes or influences the inefficiency of equilibria in games with altruistic players?

In social contribution games, players' individual costs are set to the cost they cause for society just because of their presence, thus providing a useful abstraction of games with altruistic players when the robust price of anarchy is to be analyzed. Rahn and Schäfer (2013) show that social contribution games are what they call *altruism-independently smooth*, which means that the robust price of anarchy in these games remains unaltered under arbitrary altruistic extensions. In particular, they develop a general reduction technique by which the problem of establishing smoothness for an altruistic extension of an underlying game can be transferred into a corresponding social contribution game. This reduction can be used whenever the underlying game relates to a canonical social contribution game by satisfying a simple "social contribution boundedness" property, which means that the direct, personal cost of each player is bounded by this player's cost in the corresponding social contribution game.[3]

A variety of well-known classes of games, including congestion games and valid utility games, fulfill this condition and can thus be analyzed via this reduction technique, which establishes tight bounds on the robust price of anarchy of their altruistic extensions. Regarding the more distin-

[3] A slightly stronger condition is required to hold for the friendship model from social context games.

guished friendship setting from social context games, Rahn and Schäfer (2013) show that this model is amenable to their reduction technique whenever the underlying game satisfies three additional natural properties.

Altruism in Cooperative Games

In a cooperative game, players may work together by forming groups, so-called *coalitions*, and may take joint actions so as to realize their goals better than if they were on their own. In the cardinal setting, a utility function maps each possible coalition to some (usually nonnegative) real value; in the ordinal setting, players simply rank coalitions according to their preferences. If a *coalition structure* (i.e., a partition of the players into coalitions) has formed, the question arises how stable it is, i.e., whether some players may have an incentive to leave their coalition and to join another one. There are plenty of special types of cooperative games and of stability notions some of which we will encounter below. For more background on cooperative game theory, the reader is referred, e.g., to the books by Peleg and Sudhölter (2003) and Chalkiadakis, Elkind, and Wooldridge (2011) and to the book chapter by Elkind and Rothe (2015).

Social Distance Games

The world is small. Based on his psychological experiments, Milgram (1967) formulated the famous *"six degrees of separation"* hypothesis, which roughly speaking says that, on average, any two people in the world are connected by a path of length six. Motivated by this research, Brânzei and Larson (2011) introduced their *social distance games* as a family of coalitional games with nontransferable utility. Using a graph-theoretic approach with vertices representing players, they define the utility of players by measuring their (shortest-path) distance to the other members of their coalition. While this is not actually a notion of altruism, it does correspond to the principle of homophily (McPherson, Smith-Lovin, and Cook 2001), which says that people tend to form communities with similar others, so similarity breeds connection. Mathematically, the players' similarity with the other members of their coalition is expressed as the inverse social distance, indicating the players' centrality in them. Utility thus is a variant of closeness centrality in a network and has, as Brânzei and Larson (2011) show, a number of desirable properties reflecting the players' social nature. Therefore, social distance games are important to research on social and economic networks (Jackson 2008), which themselves are central to many AI applications.

Finding a coalition structure that maximizes utilitarian social welfare in social distance games is NP-hard, which is why Brânzei and Larson (2011) provide an algorithm that approximates the maximum utilitarian social welfare within a factor of two in such games. They also investigate the stability notion of the core[4] in social distance games and show that core stable coalition structures have small-world characteristics (i.e., most vertices can be reached from any other vertex by only a few steps through intermediate vertices). Relatedly, they analyze the notion of *stability gap* (Brânzei and Larson 2009) for social distance games, which is defined for games with a nonempty core and measures the loss of social welfare that comes from being in the core. The stability gap can be seen as an analogue of the notions of *price of stability* and *price of anarchy* (Nisan et al. 2007).

Social distance games are related to the general framework of hedonic games, which we consider next.

Hedonic Games

Hedonic games are cooperative games with nontransferable utility that were originally proposed by Drèze and Greenberg (1980) and then formally modeled by Banerjee, Konishi, and Sönmez (2001) and Bogomolnaia and Jackson (2002). In such coalition formation games, players have (ordinal) preferences over the coalitions they can be a member of. Hedonic games have been thoroughly studied in the past decade and have generated a rich body of results. For more background on hedonic games, the reader is referred to the book chapters by Aziz and Savani (2016) and Elkind and Rothe (2015) and to the survey by Woeginger (2013a).

Since every player in a hedonic game needs to rank (by a weak order) exponentially many (in the number of players) coalitions, it is crucial to find compact representations for these games. Among the most important representations of hedonic games are: the *individually rational encoding* by Ballester (2004), *hedonic coalition nets* due to Elkind and Wooldridge (2009), the *singleton encoding* (Cechlárová and Romero-Medina 2001), the *additive encoding* (Sung and Dimitrov 2007, 2010; Aziz, Brandt, and Seedig 2013; Woeginger 2013b), the *fractional hedonic games* by Aziz et al. (2019) (see also, e.g., the work of Bilò et al. (2014; 2015)), *boolean hedonic games* (Aziz et al. 2016; Peters 2016), the *friends-and-enemies encoding* (Dimitrov et al. 2006; Sung and Dimitrov 2007; Rey et al. 2016; Nguyen et al. 2016), the *anonymous encoding* (Ballester 2004; Darmann et al. 2018), and *FEN-hedonic games* (where FEN stands for "friends, enemies, and neutral players") with preferences encoded using the polarized responsive principle (Lang et al. 2015; Rothe, Schadrack, and Schend 2018; Kerkmann and Rothe 2019; Kerkmann et al. 2020).

Nguyen et al. (2016) introduced the notion of altruism in hedonic games, based on the *friends-oriented extension* of the players' preferences (Dimitrov et al. 2006; Sung and Dimitrov 2007): Every player partitions the other players into a set of friends and a set of enemies, and among any two coalitions containing a player i, the one with more friends of i's is preferred by i; if they both contain the same number of i's friends, the one with fewer of i's enemies is preferred by i; and if both coalitions contain the same number of i's friends and the same number of i's enemies, i is indifferent between them. Like social distance games, such hedonic games can be compactly represented by an undirected graph whose vertices are the players and whose edges express the friendship relations, which are assumed to be symmetric. This graph is called a *network of friends*.

Specifically (and informally stated), Nguyen et al. (2016) propose three degrees of altruism in hedonic games, where

[4] A coalition structure is in the *core* if there is no *blocking coalition*, i.e., no coalition B whose members would be better off leaving their coalition in the coalition structure and joining B instead.

a player i's utility for coalitions containing i is used as a measure to compare, or rank, any two coalitions A and B:

1. **Selfish-First (SF) Preferences:** Player i first looks at which of A or B is preferred friend-orientedly, and if and only if i is indifferent between A and B, i asks her friends in A and B for a vote by looking at these friends' average utilities (resulting from the friend-oriented extension) in each of the two coalitions.
2. **Equal-Treatment (EQ) Preferences:** Player i treats her own and her friends' utilities (again according to the friend-oriented extension) equally when taking the average in both coalitions.
3. **Altruistic-Treatment (AL) Preferences:** Player i first looks at her friends' average utilities (again resulting from the friend-oriented extension) in each of the two coalitions, and if and only if they are indifferent between A and B, i looks at her own friend-oriented utility for A and B to make the decision.

These three degrees of altruism define three types of preferences in hedonic games that better help an agent to make her decision as to which coalition to join, since one way or the other also her friends' opinions (provided they belong to those coalitions) are taken into account. It can be expected that working in a happier environment (the coalition she joins) will make her happier herself.

Nguyen et al. (2016) then studied various well-known properties and stability notions under these three degrees of altruism in hedonic games, which can be informally described as follows. A coalition structure Γ is said to be *individually rational* if no player prefers being alone to being in her current coalition in Γ; *individually stable* if no player prefers moving from her current to another coalition in Γ (without harming any player in that coalition); *contractually individually stable* if no player prefers moving from her current to another coalition in Γ (without harming any player in the new coalition by joining nor any player in the old coalition by leaving); *Nash stable* if no player prefers another coalition to her current coalition in Γ; *core stable* if no nonempty coalition B *blocks* Γ, i.e., there is no nonempty coalition B whose members all would prefer being in B to being in their current coalitions in Γ; *strictly core stable* if no coalition B *weakly blocks* Γ, i.e., there is no coalition B with at least one player who prefers being in B to being in her current coalition in Γ, while the other members of B do not prefer their current coalitions in Γ to B (i.e., they prefer B or are indifferent); *popular* if for every other coalition structure Δ, a weak majority of players prefer Γ to Δ; *strictly popular* if for every other coalition structure Δ, a strict majority of players prefer Γ to Δ; and *perfect* if no player prefers any other coalition structure to Γ.[5]

Nguyen et al. (2016) investigated the verification and the existence problem for the three degrees of altruism in hedonic games in terms of their computational complexity. *Verification* means that, given a hedonic game (by its network of friends) and a coalition structure Γ for it, we ask whether Γ satisfies any one of the above stability notions. In the *existence* problem, we are given a hedonic game (by its network of friends) and ask whether there exists some coalition structure satisfying any one of these stability notions. Nguyen et al. (2016) established the following results:

- For the four notions of stability regarding single-player deviations (individual rationality, individual stability, contractually individual stability, and Nash stability), they showed that both verification and existence can be solved in polynomial time (where the latter is trivial in each case, as there *always* exists a coalition structure satisfying the corresponding property).
- For stability regarding group deviations (core stability and strict core stability), only upper bounds are known: Verification is in coNP under all three degrees of altruism, whereas existence is trivial under selfish-first preferences and is in the second level of the polynomial hierarchy (Stockmeyer 1976) for the other two degrees of altruism.
- Popularity and strict popularity yield the same upper bounds except that the existence problem is only known to be in the second level of the polynomial hierarchy, yet for strict popularity under selfish-first preferences Nguyen et al. (2016) showed that the verification problem is even coNP-complete and the existence problem coNP-hard.
- For perfectness, finally, verification is polynomial-time solvable under selfish-first preferences and in coNP under the other two degrees of altruism, and the existence problem is polynomial-time solvable under selfish-first preferences, in coNP under equal-treatment preferences, and at least known not to be trivial under altruistic-treatment preferences.

Based on the work by Nguyen et al. (2016), Schlueter and Goldsmith (2020) introduced *super altruistic hedonic games* and studied them with respect to various stability notions. In their model, friends have a different impact on a player based on their distances in the underlying network of friends, just as in the social distance games by Brânzei and Larson (2011) that we discussed above. In addition to showing that both existence and verification for strict popularity are coNP-hard in super altruistic hedonic games as well, they also show that verification for core stability and strict core stability are coNP-hard in such games and, interestingly, that existence for strict core stability and perfectness are DP-hard in such games (just as Rey et al. (2016) show for these problems in enemy-oriented hedonic games), where DP is the class "Difference NP," introduced by Papadimitriou and Yannakakis (1984) as the class of problems that can be written as the difference of two NP problems.

Wiechers and Rothe (2020) consider an analogue of egalitarian social welfare[6] by replacing the average by the min-

[5]Perfectness is also called *wonderful stability*; see, e.g., (Woeginger 2013a; Rey et al. 2016; Schlueter and Goldsmith 2020).

[6]Remotely related is the work of Monaco, Moscardelli, and Velaj (2018; 2019) who study *"hedonic games with social context"* (which, in fact, are *nonhedonic* games based on additive separable utilities, an altruism factor, and a social network, and the players' utilities are the sum of their own utilities and the utilities of their friends in the network, the latter weighted by the altruism factor) as well as *modified fractional hedonic games* (which behave qual-

imum in the definitions of the same three degrees of altruism studied by Nguyen et al. (2016). The corresponding *minimum-based* variants of altruistic hedonic games are investigated in terms of the above notions of stability, and the related decision problems in terms of their computational complexity. Wiechers and Rothe (2020) show that some results for altruistic hedonic games (in the sense of Nguyen et al. (2016)) also hold for their minimum-based variants of such games: For strict popularity, verification is coNP-complete and existence coNP-hard under selfish-first preferences. In addition, they can strengthen some of the results of Nguyen et al. (2016) for the minimum-based counterparts: For strict popularity, verification is coNP-complete and existence coNP-hard under equal-treatment and altruistic-treatment preferences.

Coalition Formation Games

What makes a hedonic game hedonic is that the players' utilities depend on their own coalition alone. Kerkmann and Rothe (2020) drop this restriction and extend the definition of the above three degrees of altruism due to Nguyen et al. (2016) to coalition formation games in general. To motivate this, they look at the network of four friends that forms a path 1 —— 2 —— 3 —— 4: Player 1 is friends with player 2, 2 with 3, and 3 with 4. Consider the coalition structures $\Gamma = \{\{1,2,3\},\{4\}\}$ and $\Delta = \{\{1,2,4\},\{3\}\}$ in which players 3 and 4 are swapped. Under friend-oriented preferences, player 1 is indifferent between coalitions $\{1,2,3\}$ and $\{1,2,4\}$ because they both contain one of 1's friends and one of 1's enemies. Under the altruistic hedonic preferences of Nguyen et al. (2016), however, 1 cares for her friend 2 being friends with 3 but not with 4 and thus prefers $\{1,2,3\}$ to $\{1,2,4\}$ and Γ to Δ. Now suppose that 1 is falling out with 2 (over him calling her boring and spending the weekend with his buddy 3 and not with her—this is not a serious thing, she still is friends with 2, so the network of friends remains unchanged; still, she wants to be alone for a while). So Γ and Δ are changed to $\Gamma' = \{\{1\},\{2,3\},\{4\}\}$ and $\Delta' = \{\{1\},\{2,4\},\{3\}\}$. Should 1 still behave altruistically towards her friend 2? Kerkmann and Rothe (2020) guess so! However, under any *hedonic* preference relation, 1 *must* disregard $\{2,3\}$ and $\{2,4\}$ which do not contain 1, and so 1 *must* be indifferent between between Γ' to Δ', simply because 1's hedonism requires her to care for her own coalition alone. Indeed, this shows the tension between *altruism* (caring for one's friends) and *hedonism* (caring only for members of one's own group).

Another example of Kerkmann and Rothe (2020) shows that under altruistic treatment in the sense of Nguyen et al. (2016), a player can prefer a coalition structure, even though it makes all her friends worse off (just because they happen to be not in the same coalition). Motivated by these examples, Kerkmann and Rothe (2020) define the three degrees of altruism more generally in *coalition formation games* and show that they differ from the original notions due to Nguyen et al. (2016). They then study the common stability concepts (defined above) in their model, and the corresponding problems in terms of their computational complexity. Most of these complexity results remain the same as for the original notions of altruism, but two results can be strengthened for altruism in coalition formation games: Verification for core stability and strict popularity under (this new notion of) selfish-first preferences is even coNP-complete.

itatively different than the fractional hedonic games due to Aziz et al. (2019)) for which Nash (and, to some extent, core) stable outcomes are compared with *egalitarian* social welfare, where instead of taking the sum of the utilities we look at their minimum.

Conclusions

We have seen an overview of various notions of altruism and how they can be introduced into game-theoretic models, ranging from social context games over to social contribution games in the noncooperative setting and from social distance games over hedonic games to general coalition formation games in the cooperative setting. During the last two decades, this has been—and it continues to be—a vibrant research stream within algorithmic game theory and AI.

Some of the surveyed models consider altruism toward all players and other models with respect to one's friends only; some models allow for weighted combinations of self-regarding and others-regarding utility and other models specific ways of combining them. Future work should therefore propose and study a unifying framework that captures several of these models of altruism and allows to compare them.

In all these models, we have seen how certain well-studied stability concepts (e.g., Nash equilibria for noncooperative games and core stability or Nash stability for cooperative games) are affected by players being altruistic. Formalizing altruism, on the one hand, may make some of the existing game-theoretic models more complex, more subtle, and more difficult to analyze. On the other hand, altruism may exact a price by making stability less efficient in terms of an increased price of anarchy (and so societies of players may be harmed by some players behaving altruistically). However, altruism certainly does help to model the interaction among agents (e.g., in multiagent systems) more realistically and to explain better what we observe in real life. Since AI is at its best when it mimics real-world human behavior, there is some hope that, when modeling the interaction among agents, integrating altruism to a larger extent into AI systems and AI research will make them both better.

Acknowledgments

I am grateful to Martin Bullinger, Edith Hemaspaandra, Lane A. Hemaspaandra, and the anonymous AAAI'21 reviewers for helpful comments. This work was supported in part by DFG grants RO 1202/14-2 and RO 1202/21-1.

References

Anagnostopoulos, A.; Becchetti, L.; de Keijzer, B.; and Schäfer, G. 2013. Inefficiency of Games with Social Context. In *Proc. SAGT'13*, 219–230. Springer-Verlag *LNCS #8146*.

Apt, K.; and Schäfer, G. 2014. Selfishness Level of Strategic Games. *Journal of Artificial Intelligence Research* 49: 207–240.

Ashlagi, I.; Krysta, P.; and Tennenholtz, M. 2008. Social Context Games. In *Proc. WINE'08*, 675–683. Springer-Verlag *LNCS #5385*.

Aziz, H.; Brandl, F.; Brandt, F.; Harrenstein, P.; Olsen, M.; and Peters, D. 2019. Fractional Hedonic Games. *ACM Transactions on Economics and Computation* 7(2): Article 6, 6:1–6:29.

Aziz, H.; Brandt, F.; and Seedig, H. 2013. Computing Desirable Partitions in Additively Separable Hedonic Games. *Artificial Intelligence* 195: 316–334.

Aziz, H.; Harrenstein, P.; Lang, J.; and Wooldridge, M. 2016. Boolean Hedonic Games. In *Proc. KR'16*, 166–175. AAAI Press.

Aziz, H.; and Savani, R. 2016. Hedonic Games. In Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A., eds., *Handbook of Computational Social Choice*, 356–376. Cambridge University Press.

Ballester, C. 2004. NP-Completeness in Hedonic Games. *Games and Economic Behavior* 49(1): 1–30.

Banerjee, S.; Konishi, H.; and Sönmez, T. 2001. Core in a Simple Coalition Formation Game. *Social Choice and Welfare* 18(1): 135–153.

Bilò, V.; Celi, A.; Flammini, M.; and Gallotti, V. 2013. Social Context Congestion Games. *Theoretical Computer Science* 514: 21–35.

Bilò, V.; Fanelli, A.; Flammini, M.; Monaco, G.; and Moscardelli, L. 2014. Nash Stability in Fractional Hedonic Games. In *Proc. WINE'14*, 486–491. Springer-Verlag *LNCS #8877*.

Bilò, V.; Fanelli, A.; Flammini, M.; Monaco, G.; and Moscardelli, L. 2015. On the Price of Stability of Fractional Hedonic Games. In *Proc. AAMAS'15*, 1239–1247. IFAAMAS.

Bogomolnaia, A.; and Jackson, M. 2002. The Stability of Hedonic Coalition Structures. *Games and Economic Behavior* 38(2): 201–230.

Borel, É. 1921. La théorie du jeu et les équations intégrales à noyau symétrique gauche. In *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, volume 173. 1304–1308.

Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A., eds. 2016. *Handbook of Computational Social Choice*. Cambridge University Press.

Brandt, F.; Fischer, F.; Harrenstein, P.; and Shoham, Y. 2009. Ranking Games. *Artificial Intelligence* 173(2): 221–239.

Brânzei, S.; and Larson, K. 2009. Coalitional Affinity Games and the Stability Gap. In *Proc. IJCAI'09*, 79–84. AAAI Press/IJCAI.

Brânzei, S.; and Larson, K. 2011. Social Distance Games. In *Proc. IJCAI'11*, 91–96. AAAI Press/IJCAI.

Bredereck, R.; Chen, J.; Faliszewski, P.; Guo, J.; Niedermeier, R.; and Woeginger, G. 2014. Parameterized Algorithmics for Computational Social Choice: Nine Research Challenges. *Tsinghua Science and Technology* 19(4): 358–373.

Cechlárová, K.; and Romero-Medina, A. 2001. Stability in Coalition Formation Games. *International Journal of Game Theory* 29(4): 487–494.

Chalkiadakis, G.; Elkind, E.; and Wooldridge, M. 2011. *Computational Aspects of Cooperative Game Theory*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool Publishers.

Chen, P.; B. de Keijzer; Kempe, D.; and Schäfer, G. 2014. Altruism and Its Impact on the Price of Anarchy. *ACM Transactions on Economics and Computation* 2(4): Article 17, 17:1–17:45.

Chevaleyre, Y.; Endriss, U.; Lang, J.; and Maudet, N. 2007. A Short Introduction to Computational Social Choice. In *Proc. SOFSEM'07*, 51–69. Springer-Verlag *LNCS #4362*.

Darmann, A.; Elkind, E.; Kurz, S.; Lang, J.; Schauer, J.; and Woeginger, G. 2018. Group Activity Selection Problem with Approval Preferences. *International Journal of Game Theory* 47(3): 767–796.

De Marco, G.; and Morgan, J. 2007. Slightly Altruistic Equilibria in Normal Form Games. Technical Report 185, Centre for Studies in Economics and Finance, University of Salerno, Fisciano, Italy.

Dimitrov, D.; Borm, P.; Hendrickx, R.; and Sung, S. 2006. Simple Priorities and Core Stability in Hedonic Games. *Social Choice and Welfare* 26(2): 421–433.

Drèze, J.; and Greenberg, J. 1980. Hedonic Coalitions: Optimality and Stability. *Econometrica* 48(4): 987–1003.

Elkind, E.; and Rothe, J. 2015. Cooperative Game Theory. In Rothe, J., ed., *Economics and Computation. An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*, 135–193. Springer.

Elkind, E.; and Wooldridge, M. 2009. Hedonic Coalition Nets. In *Proc. AAMAS'09*, 417–424. IFAAMAS.

Faliszewski, P.; Rothe, I.; and Rothe, J. 2015. Noncooperative Game Theory. In Rothe, J., ed., *Economics and Computation. An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*, 41–134. Springer.

Hare, B.; and Woods, V. 2020. *Survival of the Friendliest: Understanding Our Origins and Rediscovering Our Common Humanity*. New York: Penguin Random House.

Hayrapetyan, A.; Tardos, E.; and Wexler, T. 2006. The Effect of Collusion in Congestion Games. In *Proc. STOC'06*, 89–98. ACM Press.

Hemaspaandra, L. 2018. Computational Social Choice and Computational Complexity: BFFs? In *Proc. AAAI'18*, 7971–7977. AAAI Press.

Hoefer, M.; Penn, M.; Polukarova, M.; Skopalik, A.; and Vöcking, B. 2011. Considerate Equilibrium. In *Proc. IJCAI'11*, 234–239. AAAI Press/IJCAI.

Hoefer, M.; and Skopalik, A. 2013. Altruism in Atomic Congestion Games. *ACM Transactions on Economics and Computation* 1(4): Article 21, 21:1–21:21.

Jackson, M. 2008. *Social and Economic Networks*. Princeton University Press.

Kerkmann, A.; Lang, J.; Rey, A.; Rothe, J.; Schadrack, H.; and Schend, L. 2020. Hedonic Games with Ordinal Preferences and Thresholds. *Journal of Artificial Intelligence Research* 67: 705–756.

Kerkmann, A.; and Rothe, J. 2019. Stability in FEN-Hedonic Games for Single-Player Deviations. In *Proc. AAMAS'19*, 891–899. IFAAMAS.

Kerkmann, A.; and Rothe, J. 2020. Altruism in Coalition Formation Games. In *Proc. IJCAI'20*, 347–353. ijcai.org.

Kuniavsky, S.; and Smorodinsky, R. 2011. Equilibrium and Potential in Coalitional Congestion Games. Technical Report arXiv:1111.3933v1 [cs.GT], ACM Computing Research Repository (CoRR).

Lang, J.; Rey, A.; Rothe, J.; Schadrack, H.; and Schend, L. 2015. Representing and Solving Hedonic Games with Ordinal Preferences and Thresholds. In *Proc. AAMAS'15*, 1229–1237. IFAAMAS.

Ledyard, J. 1995. Public Goods: A Survey of Experimental Research. In Kagel, J.; and Roth, A., eds., *The Handbook of Experimental Economics*, 111–194. Princeton University Press.

McPherson, M.; Smith-Lovin, L.; and Cook, J. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27: 415–444.

Milgram, S. 1967. The Small World Problem. *Psychology Today* 2(1): 60–67.

Monaco, G.; Moscardelli, L.; and Velaj, Y. 2018. Hedonic Games with Social Context. In *Proc. ICTCS'18*, volume 2234, 24–35. CEUR-WS.org.

Monaco, G.; Moscardelli, L.; and Velaj, Y. 2019. On the Performance of Stable Outcomes in Modified Fractional Hedonic Games with Egalitarian Social Welfare. In *Proc. AAMAS'19*, 873–881. IFAAMAS.

Nguyen, N.; Rey, A.; Rey, L.; Rothe, J.; and Schend, L. 2016. Altruistic Hedonic Games. In *Proc. AAMAS'16*, 251–259. IFAAMAS.

Nisan, N.; Roughgarden, T.; Tardos, É.; and Vazirani, V., eds. 2007. *Algorithmic Game Theory*. Cambridge University Press.

Papadimitriou, C.; and Yannakakis, M. 1984. The Complexity of Facets (and some Facets of Complexity). *Journal of Computer and System Sciences* 28(2): 244–259.

Peleg, B.; and Sudhölter, P. 2003. *Introduction to the Theory of Cooperative Games*. Kluwer Academic Publishers.

Peters, D. 2016. Complexity of Hedonic Games with Dichotomous Preferences. In *Proc. AAAI'16*, 579–585. AAAI Press.

Popova, A.; Regenwetter, M.; and Mattei, N. 2013. A Behavioral Perspective on Social Choice. *Annals of Mathematics and Artificial Intelligence* 68(1–3): 5–30.

Rahn, M.; and Schäfer, G. 2013. Bounding the Inefficiency of Altruism through Social Contribution Games. In *Proc. WINE'13*, 391–404. Springer-Verlag *LNCS #8289*.

Regenwetter, M.; Grofman, B.; Marley, A.; and Tsetlin, I. 2006. *Behavioral Social Choice: Probabilistic Models, Statistical Inference, and Applications*. Cambridge University Press.

Rey, A.; Rothe, J.; Schadrack, H.; and Schend, L. 2016. Toward the Complexity of the Existence of Wonderfully Stable Partitions and Strictly Core Stable Coalition Structures in Enemy-Oriented Hedonic Games. *Annals of Mathematics and Artificial Intelligence* 77(3): 317–333.

Rothe, J., ed. 2015. *Economics and Computation. An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*. Springer.

Rothe, J. 2019a. Borda Count in Collective Decision Making: A Summary of Recent Results. In *Proc. AAAI'19*, 9830–9836. AAAI Press.

Rothe, J. 2019b. How Can We Model Emotional and Behavioral Dynamics in Collective Decision Making? In Laslier, J.; Moulin, H.; Sanver, R.; and Zwicker, W., eds., *The Future of Economic Design*, Studies in Economic Design, 245–251. Springer.

Rothe, J.; Schadrack, H.; and Schend, L. 2018. Borda-Induced Hedonic Games with Friends, Enemies, and Neutral Players. *Mathematical Social Sciences* 96: 21–36.

Schlueter, J.; and Goldsmith, J. 2020. Super Altruistic Hedonic Games. In *Proc. FLAIRS'20*, 160–165. AAAI Press.

Stockmeyer, L. 1976. The Polynomial-Time Hierarchy. *Theoretical Computer Science* 3(1): 1–22.

Sung, S.; and Dimitrov, D. 2007. On Core Membership Testing for Hedonic Coalition Formation Games. *Operations Research Letters* 35(2): 155–158.

Sung, S.; and Dimitrov, D. 2010. Computational Complexity in Additive Hedonic Games. *European Journal of Operational Research* 203(3): 635–639.

von Neumann, J. 1928. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen* 100(1): 295–320.

von Neumann, J.; and Morgenstern, O. 1944. *Theory of Games and Economic Behavior*. Princeton University Press.

Wiechers, A.; and Rothe, J. 2020. Stability in Minimization-Based Altruistic Hedonic Games. In *Proc. STAIRS'20*, volume 2655, paper 3. CEUR-WS.org.

Woeginger, G. 2013a. Core Stability in Hedonic Coalition Formation. In *Proc. SOFSEM'13*, 33–50. Springer-Verlag *LNCS #7741*.

Woeginger, G. 2013b. A Hardness Result for Core Stability in Additive Hedonic Games. *Mathematical Social Sciences* 65(2): 101–104.