Equilibria in Multi-Objective Games: a Utility-Based Perspective

Roxana Rădulescu*
Artificial Intelligence Lab
Vrije Universiteit Brussel, Belgium
roxana.radulescu@vub.be

Patrick Mannion*
Department of Computer Science & Applied Physics
Galway-Mayo Institute of Technology, Ireland
patrick.mannion@gmit.ie

Diederik M. Roijers
Computational Intelligence
Vrije Universiteit Amsterdam, The Netherlands
d.m.roijers@vu.nl

Ann Nowé
Artificial Intelligence Lab
Vrije Universiteit Brussel, Belgium
ann.nowe@vub.be

ABSTRACT

In multi-objective multi-agent systems (MOMAS), agents explicitly consider the possible tradeoffs between conflicting objective functions. We argue that compromises between competing objectives in MOMAS should be analysed on the basis of the utility that these compromises have for the users of a system, where an agent's utility function maps their payoff vectors to scalar utility values. This utility-based approach naturally leads to two different optimisation criteria for agents in a MOMAS: expected scalarised returns (ESR) and scalarised expected returns (SER). In this paper, we explore the differences between these two criteria using the framework of multi-objective normal form games (MONFGs). We demonstrate that the choice of optimisation criterion (ESR or SER) can radically alter the set of equilibria in a MONFG when non-linear utility functions are used.

KEYWORDS

Multi-agent systems; game theory; solution concepts; Nash equilibrium; correlated equilibrium; multi-objective decision making

1 INTRODUCTION

Multi-agent systems (MAS) are ideally suited to model a wide range of real-world problems where autonomous actors participate in distributed decision making. Example application domains include urban and air traffic control [18, 38], autonomous vehicles [28, 30] and energy systems [20, 24, 34]. Although many such problems feature multiple conflicting objectives to optimise, most MAS research focuses on agents maximising their return w.r.t. a single objective. By contrast, in multi-objective multi-agent systems (MOMAS), agents explicitly consider the possible trade-offs between conflicting objective functions. Agents in a MOMAS receive vector-valued payoffs for their actions, where each component of a payoff vector represents the performance on a different objective. Following the utility-based approach [26], we assume that each agent has a utility function which maps vector-valued payoffs to scalar utility values. Compromises between competing objectives are then considered on the the basis of the utility that these trade-offs have for the users of a MOMAS.

The utility-based approach naturally leads to two different optimisation criteria for agents in a MOMAS: expected scalarised returns (ESR) and scalarised expected returns (SER). To date, the differences between the SER and ESR approaches have received little attention in multi-agent settings, despite having received some attention in single-agent settings (see e.g. [25, 26]). Consequently, the implications of choosing either ESR or SER as the optimisation criterion for a MOMAS are currently not well-understood. In this work, we use the framework of multi-objective normal form games (MONFGs) to explore the differences between ESR and SER in multi-agent settings.

In multi-agent systems, solution concepts such as Nash equilibria [21, 22] and correlated equilibria [2, 3] specify conditions under which each agent cannot increase its expected payoff by deviating unilaterally from an equilibrium strategy. Such solution concepts are well-studied in single objective settings, to capture stable multi-agent behaviour. However, in utility-based MOMAS the notion of an equilibrium must be redefined, as incentives to deviate from equilibrium strategies are now computed based on the relative utilities of vector-valued payoffs, rather than the relative values of scalar payoffs. Furthermore, the choice of optimisation criterion (ESR or SER) influences how equilibria are computed, as agents' incentives to deviate from an equilibrium strategy may be measured in terms of either differences in ESR or differences in SER.

The contributions of this work are:

(1) We provide the first comprehensive analysis of the differences between the ESR and SER optimisation criteria in multi-agent settings.
(2) We provide formal definitions of the criteria for Nash equilibria and correlated equilibria under ESR and SER.
(3) We prove that the ESR and SER criteria are equivalent in cases where linear utility functions are used.
(4) We demonstrate that the choice of optimisation criterion radically alters the set of equilibria in an MONFG.
(5) We propose two versions of correlated equilibria for MONFGs – single-signal and multi-signal – corresponding to different use-cases.
(6) We prove that in MONFGs under SER, Nash equilibria need not exist, whereas correlated equilibria can exist. These examples are supported by empirical results.

The next section of this paper introduces and discusses normal form games, relevant solution concepts and optimisation criteria for multi-objective decision making. Section 3 provides an overview of prior work on multi-objective games. Section 4 formally defines Nash and correlated equilibria in MONFGs under ESR and SER and discusses some important theoretical considerations arising from

*Both authors contributed equally to the paper

these definitions. Section 5 presents empirical results in support of the conclusions reached in Section 4. Finally, Section 6 concludes with a summary of our findings, a discussion of important open questions and promising directions for future work.

2 BACKGROUND

2.1 Normal-form Games and Equilibria

Normal-form (strategic) games (NFG) constitute a fundamental representation of interactions between players in game theory. Players are seen as rational decision-makers seeking to maximise their payoff. When multiple players are interacting, their strategies are interrelated, each decision depending on the choices of the others. For this reason, we usually try to determine interesting groups of outcomes, called solution concepts. Below we offer a formal definition for NFG and discuss two well-known solution concepts considered in this work: Nash equilibria and correlated equilibria.

Definition 2.1 (Normal-form game). An n-person finite normal-form game G is a tuple $(N, \mathcal{A}, \mathbf{p})$, with $n \geq 2$, where:

- $N = \{1, \ldots, n\}$ is a finite set of players.
- $\mathcal{A} = A_1 \times \cdots \times A_n$, where A_i is the finite action set of player i (i.e., the pure strategies of i). An *action (pure strategy) profile* is a vector $\mathbf{a} = (a_1, \ldots, a_n) \in \mathcal{A}$.
- $\mathbf{p} = (p_1, \ldots, p_n)$, where $p_i : \mathcal{A} \to \mathbb{R}$ is the real-valued payoff of player i, given an action profile.

Mixed-strategy profile. Let us denote by $P(X)$ the set of all probability distributions over X. We can then define the set of mixed strategies of player i as $\Pi_i = P(A_i)$. The set of *mixed-strategy profiles* is then the Cartesian product of all the individual mixed-strategy sets $\Pi = \Pi_1 \times \ldots \times \Pi_n$.

We define $\pi_{-i} = (\pi_1, \ldots, \pi_{i-1}, \pi_{i+1}, \ldots, \pi_n)$ to be a strategy profile without player's i strategy. We can thus write $\pi = (\pi_i, \pi_{-i})$.

A Nash equilibrium (NE) [22] can be defined based on a pure or mixed-strategy profile, such that each player has selected her best response to the other players' strategies. We offer a more formal definition below.

Definition 2.2 (Nash Equilibrium). A mixed strategy profile π^{NE} of a game G is a Nash equilibrium if for each player $i \in \{1, ..., N\}$ and for any alternative strategy $\pi_i \in \Pi_i$:

$$\mathbb{E}\, p_i(\pi_i^{NE}, \pi_{-i}^{NE}) \geq \mathbb{E}\, p_i(\pi_i, \pi_{-i}^{NE}) \tag{1}$$

Thus, under a Nash equilibrium, no player i can improve her payoff by unilaterally changing her strategy. The same definition applies for pure-strategy profiles. Nash [22] has proven that, allowing the use of mixed-strategies, any finite NFG has at least one Nash equilibrium.

A correlated equilibrium is a game theoretic solution concept proposed by Aumann [2] in order to capture correlation options available to the players when some form of communication can be established prior to the action selection phase (i.e, the players receive signals from an external device, according to a known distribution, allowing them to correlate their strategies). For the current work, we look at correlation signals taking the form of action recommendations.

Correlated strategy. A correlated strategy represents a probability vector σ on $\mathcal{A}$, that assigns probabilities for each possible action profile, i.e., $\sigma : \mathcal{A} \to [0, 1]$. The expected payoff of player i, given a correlated strategy σ is calculated as:

$$\mathbb{E}\, p_i(\sigma) = \sum_{\mathbf{a} \in \mathcal{A}} \sigma(\mathbf{a}) p_i(\mathbf{a})$$

Strategy modification. A strategy modification for player i is a function $\delta_i : A_i \to A_i$, such that given a recommendation a_i, player i will play action $\delta_i(a_i)$ instead. The expected payoff of player i, given a correlated strategy σ and a strategy modification δ_i is calculated as:

$$\mathbb{E}\, p_i(\delta(\sigma)) = \sum_{\mathbf{a} \in \mathcal{A}} \sigma(\mathbf{a}) p_i(\delta_i(a_i), a_{-i})$$

Definition 2.3 (Correlated equilibrium). A correlated strategy σ^{CE} of a game G is a correlated equilibrium if for each player $i \in \{1, ..., N\}$ and for any possible strategy modification δ_i:

$$\mathbb{E}\, p_i(\sigma^{CE}) \geq \mathbb{E}\, p_i(\delta_i(\sigma^{CE})) \tag{2}$$

Thus, a correlated equilibrium ensures that no player can gain additional payoff by deviating from the suggestions, given that the other players follow them as well. Although this definition strongly resembles the one of NE, there is one important aspect we emphasise here, namely the distinction between a mixed-strategy profile and a correlated strategy. Mixed-strategy profiles are composed of independent probability factors, while the action probabilities in correlated strategies are jointly defined.

Correlated equilibria can be computed via linear programming in polynomial time [23]. It has been also shown that no-regret algorithms converge to CE [9]. Furthermore, CE has the same existence guarantees in finite NFG [11] as NE, and any Nash equilibrium is an instance of a correlated equilibrium [3].

Example. Consider the game of Chicken with the payoffs described in Table 1. Each player has two actions: to continue driving towards the other player (D) or to swerve the car (S).

	S	D
S	6, 6	2, 7
D	7, 2	0, 0

Table 1: Payoff matrix for the game of Chicken.

There are three well-known Nash equilibria for this game with expected payoffs $(7, 2)$, $(2, 7)$ – pure strategy NE – and $(4\frac{2}{3}, 4\frac{2}{3})$ – mixed strategy NE where each player selects S and D with probabilities $\frac{2}{3}$ and $\frac{1}{3}$ respectively.

	S	D
S	0.5	0.25
D	0.25	0

Table 2: A possible correlated equilibrium for the game of Chicken.

A possible correlated equilibrium is represented in Table 2, by assigning 0.5 probability for the joint action (S, S), 0.25 for (D, S) and finally 0.25 for (S, D). The expected payoff for this CE is $(5\frac{1}{4}, 5\frac{1}{4})$, values higher than the ones obtained under any NE. Thus, the notion of correlated equilibrium not only extends Nash equilibrium,

but it also offers the potential for obtaining higher expected payoffs when players are able to receive a correlation signal (e.g., a recommended action).

2.2 Multi-Objective Normal-Form Games

Definition 2.4 (Multi-objective normal-form game). An n-person finite multi-objective normal-form game G is a tuple $(N, \mathcal{A}, \mathbf{p})$, with $n \geq 2$ and $d \geq 2$ objectives, where:

- $N = \{1, \ldots, n\}$ is a finite set of players.
- $\mathcal{A} = A_1 \times \cdots \times A_n$, where A_i is the finite action set of player i (i.e., the pure strategies of i). An *action (pure strategy) profile* is a vector $\mathbf{a} = (a_1, \ldots, a_n) \in \mathcal{A}$.
- $\mathbf{p} = (\mathbf{p_1}, \ldots, \mathbf{p_n})$, where $\mathbf{p_i} \colon \mathcal{A} \to \mathbb{R}^d$ is the vectorial payoff of player i, given an action profile.

In this work we adopt a utility-based perspective [26] and assume that each agent has a utility function that maps his vectorial payoff to a scalar utility value. A more detailed discussion of utility functions can be found in Section 2.4.

2.3 Multi-Objective Optimisation Criteria

When agents consider multiple conflicting objectives, they should balance these in such a way that the user utility derived from the outcome of a decision problem (such as a MONFG) is maximised. This is known as the utility-based approach [26]. Following this approach, we assume that there exists a utility function that maps a vector with a value for each objective to a scalar utility:

$$p_{u,i} = u_i(\mathbf{p}_i) \tag{3}$$

where $p_{u,i}$ is the utility that agent i derives from the vector $\mathbf{p}_i$. When deciding what to optimise in a multi-objective normal form game, we thus need to apply this function to the vector-valued outcomes of the decision problem in some way. There are two choices for how to do this [26, 27]. Computing the expected value of the payoffs of a joint strategy first and then applying the utility function, leads to the *scalarised expected returns (SER)* optimisation criterion, i.e.,

$$p_{u,i} = u(\mathbb{E}[\mathbf{p}_i \mid \pi]) \tag{4}$$

where π is the joint strategy for all the agents in a MONFG, and $\mathbf{p}_i$ is the payoff received by agent i. SER is employed in most of the multi-objective planning and reinforcement learning literature. Alternatively, the utility function can be applied before computing the expectation, leading to the *expected scalarised returns (ESR)* optimisation criterion [25], i.e.,

$$p_{u,i} = \mathbb{E}[u(\mathbf{p}_i) \mid \pi] \tag{5}$$

Which of these criteria should be considered best depends on how the games are used in practice. SER is the correct criterion if a game is played multiple times, and it is the average payoff over multiple plays that determines the user's utility. ESR is the correct formulation if the payoff of a single play is what is important to the user.

2.4 Utility Functions

From a single-objective game theoretic perspective the notions of utility and payoff functions are generally used interchangeably. When transitioning to the multi-objective domain, we usually denote by payoff function the vectorial return (containing a real-valued payoff for each objective) received by a player, given an action profile. The utility (scalarisation) function is then used to denote the mapping from this vectorial return to a scalar utility value for a player i: $u_i \colon \mathbb{R}^d \to \mathbb{R}$.

Linear combinations are a widely used canonical example of a scalarisation function:

$$u_i(\mathbf{p}_i) = \sum_{d \in D} w_d p_{i,d} \tag{6}$$

where D is the set of objectives, $\mathbf{w}$ is a weight vector[1], $w_d \in [0, 1]$ is the weight for objective d and $p_{i,d}$ is the payoff for objective d received by received by agent i. Non-linear, discontinuous utility functions may arise in the case where it is important for an agent to achieve a minimum payoff on one of the objectives; such a utility function may look like the following:

$$u_i(\mathbf{p}_i) = \begin{cases} p_{i,t_d} & \text{if } p_{i,d} \geq t_d \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $p_{i,d}$ represents the expected payoff for agent i on objective d, t_d is the required threshold value for d, and p_{i,t_d} is the utility to agent i of reaching the threshold value on d.

Utility functions may not always be known *a priori* and/or may not be easy to define depending on the setting. For example, in the *decision support scenario* [26] it may not be possible for users to specify utility functions directly; instead users may be asked to provide their preferences by scoring or ranking different possible outcomes. After the preference elicitation process is complete, users' responses may then be used to model their utility functions [42].

3 RELATED WORK

Since their introduction in Blackwell et al. [5], multi-objective (multicriteria) games have been discussed extensively in the literature. Below we present a non-exhaustive overview of this work, highlighting a few differences with the current considered perspective.

Most previous work in multi-objective games considers utility-function agnostic equilibria, i.e., the agents do not know their preferences. For this case, Shapley and Rigby [29] extend and characterise the set of mixed-strategy agnostic Nash equilibria for multicriteria two-person zero-sum games for linear utility functions: joint strategies that are undominated w.r.t. unilateral changes by either agent. They also note that if the preference functions differ, the scalarised game (implicitly assuming ESR) can possibly be no longer zero-sum. While the idea that utility functions could also be non-linear is discussed by Bergstresser and Yu [4], for analysis purposes they only consider linear utility functions and derive solution points from the resulting trade-off games. This is important because, as we will discuss in Section 4.2, there is no in-practice difference between ESR and SER in the linear case. The existence of Pareto[2] equilibria for two-person multi-objective games under linear utility functions

[1] A vector whose coordinates are all non-negative and sum up to 1.

[2] While the original paper refers to this type of equilibrium as "Pareto", we note that Pareto is a too loose domination concept when considering only linear utility functions, and would prefer "Convex" in this case. For consistency however, we keep the original term.

is proven by Borm et al. [6]. A further characterisation of Pareto equilibria can be found in [33].

Considering non-cooperative games, Wierzbicki [36] states that, in realistic scenarios, how to aggregate criteria might not be known, however some form of scalarisation function is necessary in order to compute possible solutions. This corresponds to explicitly taking the user utility into account, and we therefore fully agree with this approach. Conflict escalation and solution selection are discussed when considering linear or order-consistent scalarisation functions. Lozovanu et al. [15] formulate an algorithm for finding Pareto-Nash equilibria in multi-objective non-cooperative games (i.e. for every linear utility function for which the weights sum to one, compute the trade-off game, then find its NE). Finally, Lozan and Ungureanu [14] propose a method for computing Pareto-Nash equilibrium sets, also under linear utility functions. A third approach is to elicit preferences, i.e., information about the utility function, while determining equilibria [12]. As far as we know however, this also has only been done for linear utility functions.

Notice that, despite the fact that many works admit that it might not always be desirable for a player to share full information about her utility function or that utility functions could take any form (including non-linear), most analysis and theoretical contributions use linear utility functions only. Furthermore, the utility function is directly applied on the original game in order to derive and analyse the corresponding trade-off game, which corresponds to the expected scalarised return (ESR) case. However, due to the use of linear utility functions, there is no distinction to be made between the ESR and SER optimisation criteria, as we will show in Section 4.2. Interestingly enough, the field of multi-objective (single-agent) reinforcement learning typically focuses on the SER case [31, 32, 41], while in either field this vital choice is typically not made explicitly or explained in the individual papers. In this paper, we aim to make the choice between an ESR and SER perspective explicit, and show that this choice has profound consequences in multi-objective multi-agent systems.

4 COMPUTING EQUILIBRIA IN MONFGS

Now that we have covered the necessary background, we begin our exploration of the differences between the ESR and SER optimisation criteria in MOMAS. In Section 4.1 we formally define Nash and correlated equilibria in MONFGs under either ESR or SER. In Section 4.2 we discuss several important theoretical considerations arising from these definitions.

4.1 Definitions

As agents in MOMAS seek to optimise the utility of their vector-valued payoffs, rather than the value of scalar payoffs in single-objective settings, the standard solution concepts must be redefined based on the agents' utilities. Incentives to deviate from an equilibrium strategy may be defined based on utility, specifically the difference between the utility of an equilibrium action and the utilities of other possible actions. Here, we reformulate the conditions for Nash equilibria (Eqn. 1) and correlated equilibria (Eqn. 2) under the ESR optimisation criterion (Eqn. 5) and the SER optimisation criterion (Eqn. 4).

Definition 4.1 (Nash equilibrium in a MONFG under ESR). A mixed-strategy strategy profile π^{NE} is a Nash equilibrium in a MONFG under ESR if for all $i \in \{1, ..., N\}$ and all $\pi_i \in \Pi_i$:

$$\mathbb{E}\, u_i\left[\mathbf{p}_i(\pi_i^{NE}, \pi_{-i}^{NE})\right] \geq \mathbb{E}\, u_i\left[\mathbf{p}_i(\pi_i, \pi_{-i}^{NE})\right] \tag{8}$$

i.e. π^{NE} is a Nash equilibrium under ESR if no agent can increase the *expected utility of her payoffs* by deviating unilaterally from π^{NE}.

Definition 4.2 (Nash equilibrium in a MONFG under SER). A mixed-strategy strategy profile π^{NE} is a Nash equilibrium in a MONFG under SER if for all $i \in \{1, ..., N\}$ and all $\pi_i \in \Pi_i$:

$$u_i\left[\mathbb{E}\, \mathbf{p}_i(\pi_i^{NE}, \pi_{-i}^{NE})\right] \geq u_i\left[\mathbb{E}\, \mathbf{p}_i(\pi_i, \pi_{-i}^{NE})\right] \tag{9}$$

i.e. π^{NE} is a Nash equilibrium under SER if no agent can increase the *utility of her expected payoffs* by deviating unilaterally from π^{NE}.

Definition 4.3 (Correlated equilibrium in a MONFG under ESR). A probability vector σ^{CE} on $\mathcal{A}$ is a correlated equilibrium in a MONFG under ESR if for all players $i \in \{1, ..., N\}$ and for all strategy modifications δ_i:

$$\mathbb{E}\, u_i\left[\mathbf{p}_i(\sigma^{CE})\right] \geq \mathbb{E}\, u_i\left[\mathbf{p}_i(\delta_i(\sigma^{CE}))\right] \tag{10}$$

i.e. σ^{CE} is a correlated equilibrium under ESR if no agent can increase the *expected utility of her payoffs* by deviating unilaterally from the action recommendations in σ^{CE}.

When applying the SER optimisation criterion for correlated equilibrium, there are two cases we can distinguish between, due to the two expectations that CE incorporates for every player i. First, we can define the expected payoff given a signal a_i^r due to the uncertainty about the other players' actions. Second, we can define the expected payoff given the correlated strategy (i.e., a certain probability distribution over the joint action space). Depending on where we place the utility function for taking the scalarised expectation, we distinguish between the *single-signal* and *multi-signal* cases.

Single-signal CE under SER. In the case of a single-signal correlated equilibrium, we assume that the signal is only given once, and that the expected payoffs over which the utility must be computed is conditioned on the signal. Even if the MONFG is played multiple times, the signal does not change. An example of a single persistent signal in a multi-agent decision problem can be a smart-grid in which the correlation signal corresponds to the price of electricity in a longer interval (e.g., one or more hours), and the actions of the agents are whether to perform a given task or not within a small interval (e.g., 10 min). In such cases, the utility of the other signals that might have been possible do not matter; they did not occur. Hence, the agent must maximise the utility of its expected vector-valued payoff given a single signal. Or, if the signal is not known at plan-time, for each signal separately.

Definition 4.4 (Single-signal CE in a MONFG under SER). A probability vector σ^{CE} on $\mathcal{A}$ is a single-signal correlated equilibrium in a MONFG under SER if for all players $i \in \{1, ..., N\}$, given a recommended action a_i^r, and for any alternative action a_i:

$$u_i\left[\frac{\sum_{a_{-i}\in\mathcal{A}_{-i}}\sigma^{CE}(a_{-i},a_i^r)\mathbf{p}_i(a_{-i},a_i^r)}{\sum_{a_{-i}\in\mathcal{A}_{-i}}\sigma^{CE}(a_{-i},a_i^r)}\right] \geq u_i\left[\frac{\sum_{a_{-i}\in\mathcal{A}_{-i}}\sigma^{CE}(a_{-i},a_i^r)\mathbf{p}_i(a_{-i},a_i)}{\sum_{a_{-i}\in\mathcal{A}_{-i}}\sigma^{CE}(a_{-i},a_i^r)}\right] \quad (11)$$

i.e. σ^{CE} is a single-signal correlated equilibrium under SER if no agent can increase the *utility of her expected payoffs* by deviating unilaterally from the given action recommendation in σ^{CE}.

Multi-signal CE under SER. The single-signal CE for MONFGs assumes that even if the MONFG is played multiple times, there will be one possible signal. Alternatively, the signal may change every time the game is played. I.e., the scalarisation is performed after marginalising over the entire correlated strategy probability distribution.

Definition 4.5 (Multi-signal CE in a MONFG under SER). A probability vector σ^{CE} on $\mathcal{A}$ is a multi-signal correlated equilibrium in a MONFG under SER if for all players $i \in \{1,...,N\}$ and for any strategy modification δ_i:

$$u_i\left[\mathbb{E}\,\mathbf{p}_i(\sigma^{CE})\right] \geq u_i\left[\mathbb{E}\,\mathbf{p}_i(\delta_i(\sigma^{CE}))\right] \quad (12)$$

i.e. σ^{CE} is a multi-signal correlated equilibrium under SER if no agent can increase the *utility of her expected payoffs* by deviating unilaterally from the given action recommendations in σ^{CE}.

Notice that while the ESR case is equivalent to solving the CE for the corresponding single-objective trade-off game, the SER case leads to a much more complicated situation. In a general case, when no restriction is imposed on the form of the utility function, we may end up having to solve a non-linear optimisation problem.

4.2 Theoretical Considerations

Theorem 4.6. *Every finite MONFG where each agent seeks to maximise the expected utility of its payoff vectors (ESR) has at least one Nash equilibrium.*

Proof. In the ESR case, any MONFG can be reduced to its corresponding single-objective trade-off game G', as players will apply the utility function on their payoff vectors after every interaction. We proceed with showing how one can construct G'.

Consider the following finite normal-form game $G' = (N, \mathcal{A}, f)$, where N and $\mathcal{A}$ are the same as in the original MONFG. According to Definition 2.1, the payoff function for G': $f = (f_1, \ldots, f_n)$.

We define each component $f_i : \mathcal{A} \to \mathbb{R}$ as the composition between player's i utility function $u_i : \mathbb{R}^d \to \mathbb{R}$ and her vectorial payoff function $\mathbf{p}_i : \mathcal{A} \to \mathbb{R}^d$:

$$f_i(a) = (u_i \circ \mathbf{p}_i)(a) = u_i(\mathbf{p}_i(a)), \forall a \in \mathcal{A}$$

Thus, in the ESR case, any MONFG is reduced to a corresponding single-objective trade-off finite NFG that can be constructed as shown above. According to the Nash equilibrium existence theorem [22], the resulting finite NFG G' has at least one Nash equilibrium. □

Theorem 4.7. *In finite MONFGs, when linear utility functions are used, the ESR and SER optimisation criteria are equivalent.*

Proof. Let π^{NE} be the NE strategy profile under the ESR optimisation criteria, according to Definition 4.1 and for each player i let u_i be a linear scalarisation function, according to Equation 6.

Due to the fact that u_i is a linear function, Jensen's inequality [13] allows us to rewrite each term of Equation 8 as follows:

$$\mathbb{E}\,u_i\left[\mathbf{p}_i(\pi_i^{NE} \cup \pi_{-i}^{NE})\right] = u_i\left[\mathbb{E}\,\mathbf{p}_i(\pi_i^{NE} \cup \pi_{-i}^{NE})\right] \quad (13)$$

$$\mathbb{E}\,u_i\left[\mathbf{p}_i(\pi_i \cup \pi_{-i}^{NE})\right] = u_i\left[\mathbb{E}\,\mathbf{p}_i(\pi_i \cup \pi_{-i}^{NE})\right] \quad (14)$$

Notice that by replacing the terms from Equation 8 according to Equations 13 and 14 we obtain the definition of the NE under SER (Equation 9). The same procedure can be applied for CE, to transition from Equation 10 to 12 and prove that, under a linear utility function, the ESR and SER criteria are also equivalent for CE. □

When considering a more general case, with u_i being a non-linear function, despite the fact that Jensen's inequality [13] would allow us to define inequality relations between the terms in Equations 13 and 14 (when constraining u_i to be convex or concave), we have no guarantee that the set of NE and CE remains the same under the two optimisation criteria ESR and SER. Thus, no clear conclusions can be drawn when generalising the form of the utility function. Furthermore, as we show below using a concrete example, in the general case, the ESR and SER criteria are not equivalent.

Theorem 4.8. *In finite MONFGs, where each agent seeks to maximise the utility of its expected payoff vectors (SER), Nash equilibria need not exist.*

Proof. Consider the following game. There are two agents that can each choose from three actions: *left*, *middle*, or *right*. The payoff vectors are identical for both agents, and are specified by the payoff matrix in Table 3.

The utility functions of the agents are given by $u_1([p^1, p^2]) = p^1 \cdot p^1 + p^2 \cdot p^2$ for agent 1, and $u_2([p^1, p^2]) = p^1 \cdot p^2$ for agent 2.[3] In this game, it is easy to see that agent 1 will always want

	L	M	R
L	(4, 0)	(3, 1)	(2, 2)
M	(3, 1)	(2, 2)	(1, 3)
R	(2, 2)	(1, 3)	(0, 4)

Table 3: The (Im)balancing act game.

to move towards an as imbalanced payoff vector as possible, i.e., concentrate as much of the value in one objective, while agent 2 will always want to move to a balanced solution, i.e, spread out the value across the objectives equally. Under SER, the expectation is taken before the utility function is applied. Therefore, a mixed strategy will lead to an expected payoff vector for both agents. If the expected payoff vector is balanced, i.e., $[2, 2]$, agent 1 will have an incentive to deterministically take action L or R, irrespective of its current strategy. If the payoff vector is imbalanced, e.g., $[2-x, 2+x]$, agent 2 will have an incentive to compensate for this imbalance, and play *left* more often to compensate if x is positive, and *right* more often if x is negative, and he is always able to do so. Hence, at least one of the agents will always have an incentive to deviate

[3] Please note that this is a monotonically increasing payoff function for positive-only payoffs. In the case of negative payoffs we can set the utility to 0 as soon as the payoff value for one of the objectives becomes negative.

from its strategy, and therefore there is no Nash equilibrium under SER. □

	L	M	R
L	$(16, 0)$	$(10, 3)$	$(8, 4)$
M	$(10, 3)$	$(8, 4)$	$(10, 3)$
R	$(8, 4)$	$(10, 3)$	$(16, 0)$

Table 4: The (Im)balancing act game under ESR with utility functions $u_1(\mathbf{p}) = {\mathbf{p}_1}^2 + {\mathbf{p}_2}^2$ and $u_2(\mathbf{p}) = \mathbf{p}_1 \cdot \mathbf{p}_2$ applied.

We also note that under ESR there is a mixed Nash equilibrium for the game in Table 3, i.e., agent 2 plays *middle* deterministically, and agent 1 plays *left* with a probability 0.5 and *right* with a probability 0.5, leading to an expected utility of $3^2 + 1^2 = 10$ for agent 1, and $3 \cdot 1 = 3$ for agent 2. This is not a Nash equilibrium under SER, as the expected payoff vector is $[2, 2]$ for this strategy, and agent 1 has an incentive to play either *left* or *right* deterministically, which would lead to an expected payoff vector of $[3, 1]$ or $[1, 3]$, yielding a higher utility for agent 1 if agent 2 does not adjust its strategy. Hence, the SER and ESR case are fundamentally different.

Theorem 4.9. *In finite MONFGs, where each agent seeks to maximise the utility of its expected payoff vectors given a signal (single-signal CE under SER), correlated equilibria can exist when Nash equilibria do not.*

Proof. Consider the action suggestions in Table 5 for the (Im)balancing act game.

	L	M	R
L	0	0.75	0
M	0	0	0
R	0	0.25	0

Table 5: A correlated equilibrium in the (Im)balancing act game under SER.

It may easily be shown that the action suggestions in Table 5 satisfy the conditions given in Eqn. 11 for a single-signal CE in a MONFG under SER:

- When L is suggested to the row player, the expected payoff vectors and SER for it to play L, M or R are:
  - L: $\mathbb{E}(\mathbf{p}) = (0.75 \cdot [3, 1])/0.75 = [3, 1]$, SER $= 3^2 + 1^2 = 10$
  - M: $\mathbb{E}(\mathbf{p}) = (0.75 \cdot [2, 2])/0.75 = [2, 2]$, SER $= 2^2 + 2^2 = 8$
  - R: $\mathbb{E}(\mathbf{p}) = (0.75 \cdot [1, 3])/0.75 = [1, 3]$, SER $= 1^2 + 3^2 = 10$
- When R is suggested to the row player, the expected payoff vectors and SER for it to play L, M or R are:
  - L: $\mathbb{E}(\mathbf{p}) = (0.25 \cdot [3, 1])/0.25 = [3, 1]$, SER $= 3^2 + 1^2 = 10$
  - M: $\mathbb{E}(\mathbf{p}) = (0.25 \cdot [2, 2])/0.25 = [2, 2]$, SER $= 2^2 + 2^2 = 8$
  - R: $\mathbb{E}(\mathbf{p}) = (0.25 \cdot [1, 3])/0.25 = [1, 3]$, SER $= 1^2 + 3^2 = 10$
- When M is suggested to the column player, the expected payoff vectors and SER for it to play L, M or R are:
  - L: $\mathbb{E}(\mathbf{p}) = (0.75 \cdot [4, 0] + 0.25 \cdot [2, 2])/(0.75 + 0.25) = [3.5, 0.5]$, SER $= 3.5 \cdot 0.5 = 1.75$
  - M: $\mathbb{E}(\mathbf{p}) = (0.75 \cdot [3, 1] + 0.25 \cdot [1, 3])/(0.75 + 0.25) = [2.5, 1.5]$, SER $= 2.5 \cdot 1.5 = 3.75$
  - R: $\mathbb{E}(\mathbf{p}) = (0.75 \cdot [2, 2] + 0.25 \cdot [0, 4])/(0.75 + 0.25) = [1.5, 2.5]$, SER $= 1.5 \cdot 2.5 = 3.75$

In all the cases above, neither of the agents may increase the utility of their expected payoff vectors given the recommendations, by deviating from the suggested actions in Table 5, assuming that the other agent follows the suggestions. Therefore CE may exist in MONFGs under SER when conditioning the expectation on a given signal, even in cases where Nash equlilibria do not exist. □

Theorem 4.10. *In finite MONFGs, where each agent seeks to maximise the utility of its expected payoff vectors over all the given signals (multi-signal CE under SER), correlated equilibria need not exist.*

Proof. In the case of a multi-signal CE, the agents are interested in their expected payoff vectors across all possible signals. In other words, to compute the expected payoff vectors, the signal must be marginalised out first. Therefore, the CE previously discussed for the single-signal case (Table 5) is not a CE for the multi-signal case, i.e., Player 1 will have an incentive to deterministically take action L or R, irrespective of the given signal. If the correlated strategy tries to incorporate this tendency, player 2 will have an incentive to deviate towards the options that offer her the most balanced outcome. Hence, similar to the proof for the non-existence of Nash-equilibria under SER, at least one of the agents will always have an incentive to deviate from the given recommendation, and therefore there is no multi-signal correlated equilibrium under SER. □

We thus conclude that an MONFG under ESR with *known* utility functions is equivalent to a single-objective NFG, and therefore all theory, including the existence of Nash equilibria and correlated equilibria, is implied. Under SER however, Nash equilibria and multi-signal correlated equilibria need not exist, and MONFGs are fundamentally more difficult than single-objective NFGs, even when the utility functions are known in advance.

5 EXPERIMENTS

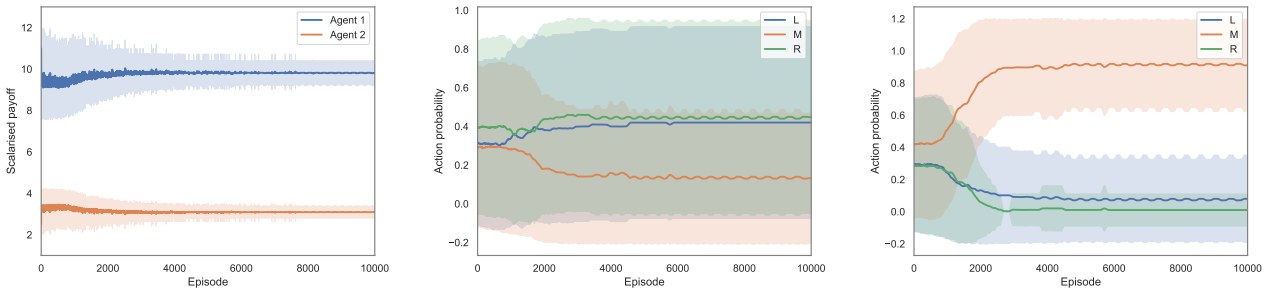
To demonstrate the effect of the SER optimisation criterion on equilibria in MONFGs, in the case of *single-signal correlated equilibrium*, we conducted two experiments using the (Im)balancing act game in Table 3. Both experiments were repeated 100 times and had a duration of 10,000 episodes, where the (Im)balancing act game was played once per episode. Agents implemented a simple algorithm[4] to learn estimates of the expected vectors for each action according to the following update rule (i.e. a "one-shot" vectorial version of Q-learning [35]):

$$\mathbf{Q}(s_i, a_i) \leftarrow \mathbf{Q}(s_i, a_i) + \alpha[\mathbf{p}_i(s_i, a_i) - \mathbf{Q}(s_i, a_i)] \tag{15}$$

where $\mathbf{Q}(s_i, a_i)$ is an estimate of the expected value vector for selecting action a_i when a private signal s_i is received, $\mathbf{p}_i(s_i, a_i)$ is the payoff vector received by agent i for selecting action a_i when observing s_i, and α is the learning rate.

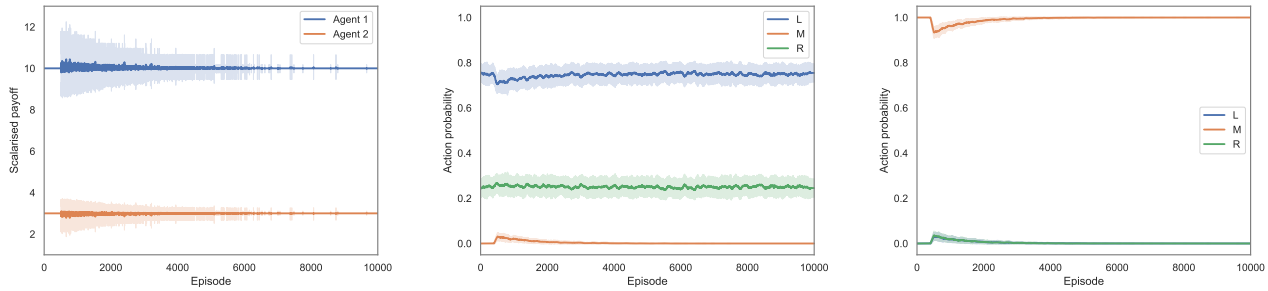
The private signals given to each agent allow us to test empirically whether agents will have an incentive to deviate from a single-signal correlate equilibrium in a MONFG under SER. In the first experiment, in each episode agents received unchanging private signals with probability 1 (i.e. equivalent to the case where no

[4]We note that specialised algorithms exist to learn mixed-strategy Nash equilibria (e.g. [10]) or correlated equilibria (e.g. [1]) in single-objective MAS. We leave the design and empirical evaluation of versions of these algorithms for learning or approximating equilibria in MOMAS under SER for future work.



(a) Scalarised payoffs obtained by each agent. (b) Action selection probabilities of Agent 1. (c) Action selection probabilities of Agent 2.

Figure 1: Experiment 1: The (Im)balancing act game under SER with no action recommendations.



(a) Scalarised payoffs obtained by each agent. (b) Action selection probabilities of Agent 1. (c) Action selection probabilities of Agent 2.

Figure 2: Experiment 2: The (Im)balancing act game under SER with action recommendations provided according to Table 5.

private signals are present). In the second experiment, the private signals received by each agent corresponded to the correlated action recommendations in Table 5 (i.e. in a given episode, (L,M) was recommended with probability 0.75, or else (R,M) was recommended with probability 0.25). For the first 500 episodes of the second experiment, both agents followed the action recommendations in their private signals deterministically, so that the correlated equilibrium behaviour could be learned. For the last 9,500 episodes of the second experiment, agents continues to receive action recommendations, but selected their actions autonomously.

Agents implemented the ϵ-greedy exploration strategy. As agents seek to optimise their action choices with respect to scalarised expected returns, they greedily selected the action with the highest SER, given the recommendation, with probability $1 - \epsilon$, or chose a random action with probability ϵ. Estimates of expected value vectors for each action were scalarised using the same utility functions as in Section 4.2.

All agents used a constant value of $\alpha = 0.05$ for the learning rate. For both agents in experiment 1, ϵ was initially set to 0.1 in the first episode, and decayed by a factor 0.999 in each subsequent episode. For both agents in experiment 2, ϵ was set to 0.0 in for the first 500 episodes where the agents deterministically followed the recommendations from their private signals, after which ϵ was set to 0.1 for episode 501 and decayed by a factor 0.999 in each subsequent episode. No attempt was made to conduct comprehensive parameter sweeps to optimise the values of α and ϵ which were used in either experiment.

The experimental results in terms of scalarised payoff are shown in Figs. 1a and 2a respectively. Both figures show the scalarised payoffs received by the agents in each episode, averaged over 100 trials. For each experiment we also plot the action selection probabilities for each of the two players. The probabilities are computed using a sliding window of size 100 over the history of taken actions and are also averaged over 100 trials. The shaded region around each plot shows one standard deviation from the mean. No smoothing was applied to any of the plots.

It is clear to see from the high standard deviations in Fig. 1a that agents do not reliably converge on any one joint strategy when no correlated action recommendations are provided. This conclusion is further strengthened when observing the action selection probabilities of player 1 (Fig. 1b) and player 2 (Fig. 1c).

Given our analysis in Theorem 4.8, this is to be expected, as agents will always have some incentive to deviate from a potential Nash equilibrium point in this game. As ϵ is decayed, the agents' behaviour becomes more deterministic, and the joint strategies learned in each run are always sub-optimal (i.e. not the best response to the other player's strategy) in terms of SER for one of the agents. Note as the strategies eventually become deterministic, this shows that no Nash equilibrium is reached by attempting to optimise action selections based on SER using pure strategies only

(a different action selection method such as softmax could be used to verify this for mixed strategies).

In Fig. 2a, the effect of the single-signal correlated equilibrium may clearly be seen. As we would expect, for the first 500 episodes a consistent scalarised payoff is received by both agents while they learn the correlated equilibrium. From episode 501 both agents are free to select actions autonomously and to explore and learn the effects of deviating from the action suggestions. As ϵ is gradually decayed towards zero, the agents consistently converge back to the correlated equilibrium, evidenced by the low standard deviations around the means of the scalarised payoffs near episode 10,000. Furthermore, Fig. 2b and 2c show that the action selection probabilities for each player nicely converge to the probabilities of the correlated equilibrium in Table 5 (i.e., agent 1 will select L with 25% probability and R with 75% probability, while agent 2 ends up selecting M 100% of the time).

This provides empirical support for our claim in Theorem 4.9 that single-signal correlated equilibria can exist in MONFGs under SER, demonstrating that neither agent has an incentive to deviate unilaterally given an action recommendation, when learning in this MONFG under SER.

6 CONCLUSION AND FUTURE WORK

In this work, we explored the differences between two optimisation criteria for MOMAS: expected scalarised returns and scalarised expected returns. Using the framework of MONFGs, we constructed sets of conditions for the existence of Nash and correlated equilibria, two of the most commonly-used solution concepts in the single-objective MAS literature. Our analysis demonstrated that fundamental differences exist between the ESR and SER criteria in multi-agent settings.

While we have provided some theoretical results concerning the existence of equilibria in utility-based MONFGs, a number of deep and interesting open questions remain unanswered. Thus far, we have not found an example of a Nash equilibrium or multi-signal correlated equilibrium in a MONFG under SER with non-linear utility functions, although we provided examples in the proof of Theorems 4.8 and 4.10 where these type of equilibria do not exist. It is currently unclear whether or under what conditions Nash equilibria or multi-signal correlated equilibria could exist in this setting; therefore, further detailed theoretical analysis is required.

In the proof of Theorem 4.9 we provide an example where a single-signal correlated equilibrium does exist under SER, although it not known whether correlated equilibria always exist in this setting. The existence of correlated equilibria in single-objective NFGs has been proven by Hart and Schmeider based on linear duality [11], an argument which does not rely on the existence of Nash equilibria (or by extension, the use of a fixed point theorem as per Nash [22]) as Aumann's original proof did [2]. Extending the work of Hart and Schmeider for utility-based MONFGs under SER is a promising direction for future work.

As we saw in the example Chicken game in Table 1, correlated equilibria allow for better compromises to be achieved between conflicting payoff functions in single-objective NFGs, when compared with Nash equilibria. In utility-based MONFGs, we demonstrated that this property translates well, allowing compromises to be achieved between conflicting utility functions (and allowing a stable compromise solution to be reached in an instance where no stable compromise may be reached using Nash equilibria, when conditioning on the received signal).

The analysis in this paper has a number of important limitations which should be addressed in future work. Our worked examples considered MONFGs with two agents only, so the interaction between equilibria and optimisation criteria should be further explored in larger MOMAS. It would also be worthwhile to conduct larger and more rigorous empirical studies to further expand upon our findings, and test whether agents can actually converge on equilibrium points in a range of different MONFGs when learning or evolving strategies. By adopting the MONFG model, we considered stateless decision making problems only; our analysis should be extended to stateful MOMAS models such as multi-objective stochastic games (MOSGs) [16], or even multi-objective versions of partially observable stochastic games [37]. We note that a similar equilibrium concept to the correlated equilibrium exists for single-objective stochastic games; the cyclic equilibrium (or cyclic correlated equilibrium) [40]. Little is currently known about the existence of equilibria in utility-based multi-objective multi-agent sequential decision making settings. If the existence of Nash equilibria cannot be proven or demonstrated for MOSGs with non-linear utility functions under SER in the future, the cyclic equilibrium is one alternate solution concept which is worthy of exploration.

Another interesting line of future research concerns the interaction between MOMAS, optimisation criteria (ESR vs. SER) and reward shaping. Although reward shaping in MOMAS has received some attention to date (see e.g. [16, 17, 39]), it has been primarily from the ESR perspective, and using linear and hypervolume scalarisation functions only. Principled reward shaping techniques such as potential-based reward shaping and difference rewards come with convenient theoretical guarantees (e.g. preserving the relative value of policies and/or actions, and therefore Nash and Pareto relations between policies and/or actions in MAS/MOMAS [7, 8, 17, 19]); how well these techniques will work under SER with non-linear utility functions is currently unknown.

How to best model users' utility functions for MOMAS remains a significant open question. Recent work on preference elicitation strategies for multi-objective decision support settings [42] has delivered promising results in single agent settings with non-linear utility; this approach could feasibly be extended to generate utility functions for decision making in MOMAS.

Finally, as we mentioned in Section 2.3, users may prefer either the SER or ESR criterion depending on their needs (e.g. whether they care more about average performance over a number of policy executions, or just the performance of a policy single execution [25]). In larger MOMAS, it is possible that not all users would choose the same optimisation criterion, or that their preference for a specific optimisation criterion may change over time, potentially adding further complexity to the process of computing equilibria.

Acknowledgements

We thank Bart Bogaerts for his useful feedback and discussions on this work.

REFERENCES

[1] Jasmina Arifovic, Joshua F. Boitnott, and John Duffy. 2016. Learning correlated equilibria: An evolutionary approach. *Journal of Economic Behavior & Organization* (2016).

[2] Robert J Aumann. 1974. Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics* 1, 1 (1974), 67–96.

[3] Robert J Aumann. 1987. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica: Journal of the Econometric Society* (1987), 1–18.

[4] K Bergstresser and PL Yu. 1977. Domination structures and multicriteria problems in N-person games. *Theory and Decision* 8, 1 (1977), 5–48.

[5] David Blackwell et al. 1956. An analog of the minimax theorem for vector payoffs. *Pacific J. Math.* 6, 1 (1956), 1–8.

[6] PEM Borm, SH Tijs, and J van den Aarssen. 1990. Pareto equilibria in multiobjective games. *Methods of Operations Research* 60 (1990), 303–312.

[7] Mitchell Colby and Kagan Tumer. 2015. An evolutionary game theoretic analysis of difference evaluation functions. In *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*. ACM, 1391–1398.

[8] Sam Devlin and Daniel Kudenko. 2011. Theoretical Considerations of Potential-Based Reward Shaping for Multi-Agent Systems. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 225–232.

[9] Dean P Foster and Rakesh Vohra. 1999. Regret in the on-line decision problem. *Games and Economic Behavior* 29, 1-2 (1999), 7–35.

[10] Drew Fudenberg and David M. Kreps. 1993. Learning Mixed Equilibria. *Games and Economic Behavior* 5, 3 (1993), 320 – 367. https://doi.org/10.1006/game.1993.1021

[11] Sergiu Hart and David Schmeidler. 1989. Existence of correlated equilibria. *Mathematics of Operations Research* 14, 1 (1989), 18–25.

[12] Ayumi Igarashi and Diederik M Roijers. 2017. Multi-criteria coalition formation games. In *International Conference on Algorithmic DecisionTheory*. Springer, 197–213.

[13] Johan Ludwig William Valdemar Jensen et al. 1906. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta mathematica* 30 (1906), 175–193.

[14] Victoria Lozan and Valeriu Ungureanu. 2013. Computing the Pareto-Nash equilibrium set in finite multi-objective mixed-strategy games. (2013).

[15] Dmitrii Lozovanu, D Solomon, and A Zelikovsky. 2005. Multiobjective games and determining pareto-nash equilibria. *Buletinul Academiei de Ştiinţe a Republicii Moldova. Matematica* 3 (2005), 115–122.

[16] Patrick Mannion, Sam Devlin, Jim Duggan, and Enda Howley. 2018. Reward shaping for knowledge-based multi-objective multi-agent reinforcement learning. *The Knowledge Engineering Review* 33 (2018).

[17] Patrick Mannion, Sam Devlin, Karl Mason, Jim Duggan, and Enda Howley. 2017. Policy invariance under reward transformations for multi-objective reinforcement learning. *Neurocomputing* 263 (2017).

[18] Patrick Mannion, Jim Duggan, and Enda Howley. 2016. An Experimental Review of Reinforcement Learning Algorithms for Adaptive Traffic Signal Control. In *Autonomic Road Transport Support Systems*, Leo Thomas McCluskey, Apostolos Kotsialos, P. Jörg Müller, Franziska Klügl, Omer Rana, and René Schumann (Eds.). Springer International Publishing, 47–66.

[19] Patrick Mannion, Jim Duggan, and Enda Howley. 2017. A Theoretical and Empirical Analysis of Reward Transformations in Multi-Objective Stochastic Games. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

[20] Patrick Mannion, Karl Mason, Sam Devlin, Jim Duggan, and Enda Howley. 2016. Multi-Objective Dynamic Dispatch Optimisation using Multi-Agent Reinforcement Learning. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

[21] John Nash. 1950. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences* 36, 1 (1950), 48–49.

[22] John Nash. 1951. Non-Cooperative Games. *Annals of Mathematics* 54, 2 (1951), 286–295.

[23] Christos H Papadimitriou and Tim Roughgarden. 2008. Computing correlated equilibria in multi-player games. *Journal of the ACM (JACM)* 55, 3 (2008), 14.

[24] Mathieu Reymond, Christophe Patyn, Roxana Rădulescu, Geert Deconinck, and Ann Nowé. 2018. Reinforcement learning for demand response of domestic household appliances. In *Proceedings of the Adaptive and Learning Agents workshop at FAIM 2018*.

[25] Diederik M Roijers, Denis Steckelmacher, and Ann Nowé. 2018. Multi-objective Reinforcement Learning for the Expected Utility of the Return. In *Proceedings of the Adaptive and Learning Agents workshop at FAIM 2018*.

[26] Diederik M Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley. 2013. A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research* 48 (2013), 67–113.

[27] Diederik M. Roijers and Shimon Whiteson. 2017. Multi-Objective Decision Making. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 11, 1 (2017), 1–129.

[28] Roxana Rădulescu, Manon Legrand, Kyriakos Efthymiadis, Diederik M Roijers, and Ann Nowé. 2018. Deep Multi-Agent Reinforcement Learning in a Homogeneous Open Population. In *Proceedings of the 30th Benelux Conference on Artificial Intelligence (BNAIC 2018)*. 177–191.

[29] Lloyd S Shapley and Fred D Rigby. 1959. Equilibrium points in games with vector payoffs. *Naval Research Logistics Quarterly* 6, 1 (1959), 57–61.

[30] Victor Talpert, Ibrahim Sobh, Bangalore Ravi Kiran, Patrick Mannion, Senthil Yogamani, Ahmad El-Sallab, and Patrick Perez. 2019. Exploring applications of deep reinforcement learning for real-world autonomous driving systems. In *International Conference on Computer Vision Theory and Applications (VISAPP)*.

[31] Peter Vamplew, Richard Dazeley, Adam Berry, Rustam Issabekov, and Evan Dekker. 2011. Empirical evaluation methods for multiobjective reinforcement learning algorithms. *Machine learning* 84, 1-2 (2011), 51–80.

[32] Kristof Van Moffaert and Ann Nowé. 2014. Multi-objective reinforcement learning using sets of pareto dominating policies. *The Journal of Machine Learning Research* 15, 1 (2014), 3483–3512.

[33] Mark Voorneveld, Dries Vermeulen, and Peter Borm. 1999. Axiomatizations of Pareto equilibria in multicriteria games. *Games and economic behavior* 28, 1 (1999), 146–154.

[34] Erwin Walraven and Matthijs T. J. Spaan. 2016. Planning under Uncertainty for Aggregated Electric Vehicle Charging with Renewable Energy Supply. In *Proceedings of the European Conference on Artificial Intelligence*. 904–912.

[35] Christopher John Cornish Hellaby Watkins. 1989. *Learning from Delayed Rewards*. Ph.D. Dissertation. King's College, Cambridge, UK.

[36] Andrzej P Wierzbicki. 1995. Multiple criteria games – Theory and applications. *Journal of Systems Engineering and Electronics* 6, 2 (1995), 65–81.

[37] Auke J Wiggers, Frans A Oliehoek, and Diederik M Roijers. 2016. Structure in the value function of two-player zero-sum games of incomplete information. In *Proceedings of the Twenty-second European Conference on Artificial Intelligence*. IOS Press, 1628–1629.

[38] Logan Yliniemi, Adrian K Agogino, and Kagan Tumer. 2015. Simulation of the introduction of new technologies in air traffic management. *Connection Science* 27, 3 (2015), 269–287.

[39] Logan Yliniemi and Kagan Tumer. 2016. Multi-objective multiagent credit assignment in reinforcement learning and NSGA-II. *Soft Computing* 20, 10 (2016), 3869–3887.

[40] Martin Zinkevich, Amy Greenwald, and Michael L Littman. 2006. Cyclic equilibria in Markov games. In *Advances in Neural Information Processing Systems*. 1641–1648.

[41] Luisa M Zintgraf, Timon V Kanters, Diederik M Roijers, Frans A Oliehoek, and Philipp Beau. 2015. Quality assessment of MORL algorithms: A utility-based approach. In *Benelearn 2015: Proceedings of the Twenty-Fourth Belgian-Dutch Conference on Machine Learning*.

[42] Luisa M Zintgraf, Diederik M Roijers, Sjoerd Linders, Catholijn M Jonker, and Ann Nowé. 2018. Ordered Preference Elicitation Strategies for Supporting Multi-Objective Decision Making. In *Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1477–1485.