Dynamic Games and Applications (2020) 10:478–492
https://doi.org/10.1007/s13235-019-00328-w

Optimality, Equilibrium, and Curb Sets in Decision Problems Without Commitment

P. Jean-Jacques Herings[1] · **Andrey Meshalkin**[1] · **Arkadi Predtetchinski**[1]

Published online: 17 September 2019

Abstract
The paper considers a class of decision problems with an infinite time horizon that contains Markov decision problems as an important special case. Our interest concerns the case where the decision maker cannot commit himself to his future action choices. We model the decision maker as consisting of multiple selves, where each history of the decision problem corresponds to one self. Each self is assumed to have the same utility function as the decision maker. Our results are twofold: Firstly, we demonstrate that the set of subgame optimal policies coincides with the set of subgame perfect equilibria of the decision problem. Furthermore, the set of subgame optimal policies is contained in the set of optimal policies and the set of optimal policies is contained in the set of Nash equilibria. Secondly, we show that the set of pure subgame optimal policies is the unique minimal curb set of the decision problem. The concept of a subgame optimal policy is therefore robust to the absence of commitment technologies.

Keywords Game theory · Decision problem · Multiple selves · Subgame perfect equilibrium · Curb sets

JEL Classification C61 · C62 · C73

1 Introduction

In this paper, we study a class of decision problems with an infinite time horizon that contains discounted Markov decision problems with a finite set of states and actions as an important subclass. In every time period, nature selects a state. We take the perspective of a decision maker who is informed about the state and has to take an action out of a set of actions, thereby

✉ P. Jean-Jacques Herings
P.Herings@maastrichtuniversity.nl

Andrey Meshalkin
meshalkin9@gmail.com

Arkadi Predtetchinski
A.Predtetchinski@maastrichtuniversity.nl

[1] Department of Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

generating a, potentially probabilistic, transition to a new state. This process is repeated indefinitely. Contrary to Markov decision problems, we allow for history-dependent sets of available actions and history-dependent state transitions. Non-Markovian decision problems have also been studied by Schäl [25], who addresses the existence of optimal policies in such a setting under assumptions on the payoff functions that are somewhat different from ours. Our main interest is in the case where the decision maker cannot commit himself to his future actions. He is therefore modeled as consisting of multiple selves that have utility functions identical to the one of the decision makers. Our emphasis is on the characterization of policies that are consistent with Nash equilibrium, subgame perfect equilibrium, and sets of policies that are closed under rational behavior.

To obtain a benchmark, we start our analysis by considering a decision maker that can commit himself to his future action choices. A policy of such a decision maker specifies a profile of history-contingent action choices that are all feasible at the corresponding history. A policy is optimal at a history if it maximizes the utility of the decision maker conditional on reaching that history. A policy is subgame optimal if it maximizes the utility of the decision maker at every history. We show the existence of subgame optimal policies and show that the set of subgame optimal policies has a product structure. Moreover, we show that a policy is subgame optimal if and only if it is 1-day optimal at each history. These results are completely in line with those derived for Markov decision problems; see, for instance, the excellent overview by Puterman [21].

We continue our analysis by assuming that the decision maker cannot commit himself to his future choices. He is fully aware of the actions and the payoff consequences of his future actions, but cannot commit himself to any future choice. There are many examples where the inability to commit to one's future actions has drastic consequences. A famous example is given by Kydland and Prescott [12], where lack of commitment leads to socially suboptimal decision making.

The decision maker can only rely on the fact that his future self will make an optimal choice. To study this case, we represent a decision problem as a stochastic game with an infinite number of players. Each history of the decision problem is represented by one player, who corresponds to one particular self of the decision maker. The utility functions of the players are all assumed to be identical to each other and equal to the one of the decision makers.[1] A strategy profile as chosen by the players is in a one–one correspondence with a policy in the decision problem.

The standard way to solve a game is the concept of Nash equilibrium as proposed by Nash [14]. At a Nash equilibrium of the decision problem, there is no self of the decision maker who can benefit from taking another action, given that all other selves stick to their actions. This approach to multiple-selves problems corresponds to the one suggested by Peleg and Yaari [18] in the context of consumption choice with time-inconsistent preferences. We show that an optimal policy at the initial history is a Nash equilibrium, but not vice versa, so a Nash equilibrium may fail to be an optimal policy at the initial history. At a Nash equilibrium of the decision problem, the different selves may fail to coordinate in a satisfactory way, leading to suboptimal behavior.

A well-known problem with the concept of Nash equilibrium is that it does not require conditionally optimal behavior by players in subgames that are reached with probability zero. We continue the analysis by considering the concept of subgame perfect equilibrium

[1] Thus, there is no issue of time-inconsistent preferences as introduced in Strotz [24] and Pollak [19]. It is well know that when preferences exhibit dynamic inconsistencies, subgame perfect policies need not be optimal; see, for example, O'Donoghue and Rabin [15] or Cingiz et al. [5]. For an extensive overview of the literature on time-inconsistent preferences, we refer the reader to Frederick et al. [7].

as introduced in Selten [22]. We identify the subgames of a decision problem as decision problems conditional on reaching a particular history. A subgame perfect equilibrium of a decision problem is a strategy profile that is a Nash equilibrium of every decision problem that corresponds to a subgame. This approach to multiple-selves problems corresponds to the one suggested by Goldman [8] in the context of consumption choice with time-inconsistent preferences. We show that the set of subgame perfect equilibria of a decision problem is equal to the set of subgame optimal policies.

A Nash equilibrium requires only that deviations are not profitable. So even the concept of subgame perfect equilibrium, requiring that the strategy profile is a Nash equilibrium in every subgame, does not require that unilateral deviations actually involve a loss, which is required by the more demanding notion of strict equilibrium. Although a strict equilibrium, and a fortiori a strict subgame perfect equilibrium, is more convincing as a stable strategy profile, it is not guaranteed to exist in decision problems. We therefore turn to a set-valued version of strict equilibrium as proposed by Basu and Weibull [3] for games in normal form with a finite number of players.

Basu and Weibull [3] define a set of strategy profiles to be closed under rational behavior (curb) if it contains all its best responses. A minimal curb set is a curb set that does not contain any other curb set as a proper subset. Pruzhansky [20], using a slight variation on this notion, establishes two results for extensive games with perfect information and a finite horizon. Firstly, he shows that any such game possesses only one minimal curb set; and secondly, that the minimal curb set includes all subgame perfect equilibria of the game. Myerson and Weibull [13] define tenable strategy blocks, leading to a refinement of curb sets.

We define a minimal curb set for a decision problem by requiring that every self of the decision maker has a best response in the curb set conditional on his history being reached. A curb set therefore captures the situation where every self of the decision maker chooses an action that is a best response to some belief over action choices that are best responses for the future selves conditional on the history of the current self being reached.

Voorneveld et al. [26] point toward an important advantage of minimal curb sets. Contrary to point-valued concepts as studied in the equilibrium selection literature, a minimal curb set satisfies the axiom of consistency, a notion that has been introduced by Peleg and Tijs [17] and Peleg et al. [16]. Consistency requires that if a set of players plays the game according to a particular solution, then the remaining players in the reduced game should not have an incentive to deviate from it. Using consistency, Voorneveld et al. [26] provide an axiomatization of minimal curb sets.

Another advantage of minimal curb sets is that they are robust in a dynamic sense. Hurkens [10] studies a stochastic version of fictitious play in the spirit of Young [27] and shows that such a dynamic process of strategy adjustment will eventually settle down in a minimal curb set. Similarly, Young [28] presents a fictitious play process with independent beliefs such that the stochastically stable states of the process correspond to the minimal curb sets minimizing the stochastic potential; see also Durieu et al. [6] for the analysis of a more general class of fictitious play processes. Balkenborg et al. [2] show how generalized best reply dynamics settle down within a minimal curb set based on the refined best reply correspondence. Further results on the connection between learning dynamics and minimal curb sets can be found in Kah and Walzl [11].

A curb set is said to be tight if it is exactly equal to its set of best responses. Since a strict equilibrium corresponds to a singleton tight curb set, a tight curb set is indeed the appropriate set-valued generalization of a strict equilibrium. We show that a minimal curb set of a decision problem is tight. We also prove that a minimal curb set always exists, is unique, and coincides with the set of pure subgame optimal policies.

The two main findings of the paper could be summarized as follows: Firstly, the set of subgame optimal policies coincides with the set of subgame perfect equilibria. Furthermore, the set of subgame optimal policies is contained in the set of optimal policies and the set of optimal policies is contained in the set of Nash equilibria. Examples show that both inclusions could be strict. Secondly, the minimal curb set is unique and equal to the set of pure subgame optimal policies.

The rest of the paper is organized as follows: In Sect. 2, we define a class of decision problems that contains Markov decision problems as a special case. Section 3 provides a benchmark for our analysis. There we take the point of view of a decision maker who exercises full control over all decisions to be taken. We analyze optimal and subgame optimal policies. In particular, we give a characterization of subgame optimal policies in terms of 1-day optimal policies. In Sects. 4 and 5, we take the perspective that the decision maker is unable to commit himself to his future actions. Accordingly, we adopt a multiple self model, whereby each history of the decision problem is controlled by a distinct self. Section 4 analyzes Nash equilibrium and subgame perfect equilibrium of the decision problem and establishes the first main result: The set of subgame perfect equilibria coincides with the set of subgame optimal policies. Section 5 introduces the notion of a minimal curb set of a decision problem and establishes our second main result: The minimal curb set is unique and equal to the set of pure subgame optimal policies. Section 6 concludes.

2 Decision Problems

A *decision problem* is described by the tuple $D = (S, A, H, \pi, f)$. Moves are made by nature and the decision maker in an alternating fashion, where the decision maker chooses *actions* from the set A and nature picks *states* in the set S. We let $\mathbb{N} = \{0, 1, \ldots\}$ denote the set of natural numbers. The set

$$H \subseteq \{s_0\} \times \bigcup_{t \in \mathbb{N}} (A \times S)^t$$

is the set of *histories*, where s_0 is a distinguished element of S called the *initial state*. Each element $h \in H$ is a finite sequence of the form $h = (s_0, a_0, \ldots, s_{\ell-1}, a_{\ell-1}, s_\ell)$ where ℓ is a natural number, $s_0, \ldots, s_\ell$ are elements of S, and $a_0, \ldots, a_{\ell-1}$ are elements of A. We refer to s_ℓ as the *current state*. Given a history $h = (s_0, a_0, \ldots, s_{\ell-1}, a_{\ell-1}, s_\ell)$ in H, we denote its length ℓ by $\ell(h)$.

Consider histories h and h' in H, where $h' = (s_0, a_0, \ldots, s_{\ell-1}, a_{\ell-1}, s_\ell)$. The history h is said to be a *subhistory* of h' if $h = (s_0, a_0, \ldots, s_{k-1}, a_{k-1}, s_k)$ for some $k \leq \ell$. It is said to be a proper subhistory of h' if $k < \ell$. We write $h \leq h'$ to denote that h is a subhistory of h' and $h < h'$ to denote that h is a proper subhistory of h'. The unique subhistory of a history $h \in H$ of length $k \leq \ell(h)$ is denoted by h^k.

The set of actions available at a history $h \in H$ is denoted by A_h, so

$$A_h = \{a \in A \mid \exists s \in S \text{ such that } (h, a, s) \in H\}.$$

It is convenient to define the set G of *nature histories*, i.e., histories after which nature selects the next state,

$$G = \{(h, a) \in H \times A \mid a \in A_h\}.$$

The notions of subhistories and length are extended to nature histories in a straightforward way.

The set of states that may be reached at $g \in G$ is denoted by S_g, so

$$S_g = \{s \in S \mid (g, s) \in H\}.$$

The set of histories H is assumed to have the following properties:

1. The history (s_0) is an element of H.
2. For every $h \in H$, the set A_h is non-empty and finite.
3. For every $g \in G$, the set S_g is non-empty and finite.
4. For every $h \in H$, each subhistory of h is an element of H.

The function π is a *law of transition* that assigns to each nature history $g \in G$ a probability distribution on the set S_g and thereby specifies the *transition probabilities*. We let $\pi(s \mid g) \geq 0$ denote the probability that the system jumps from nature history $g \in G$ to state $s \in S_g$. Obviously, it holds that $\sum_{s \in S_g} \pi(s \mid g) = 1$.

Consider an infinite sequence $p = (s_0, a_0, s_1, a_1, \ldots)$. The sequence p is said to be a *play* if all the prefixes of p, that is the finite sequences (s_0), (s_0, a_0, s_1), $(s_0, a_0, s_1, a_1, s_2)$, ..., are elements of H. We let P be the set of plays. We endow P with the topology generated by the basis of cylinder sets, i.e., sets of the form $\{p \in P : h \text{ is a prefix of } p\}$ for some history $h \in H$. The *payoff function* $f : P \to \mathbb{R}$ assigns payoffs to plays. Throughout this paper, we assume that the function f is continuous.

An important subclass of decision problems is discounted Markov decision problems. The decision problem is said to be a *discounted Markov* decision problem if [1] the set of available actions at a history depends only on the current state, [2] the transitions probabilities depend only on the current state and the latest action, and [3] there is a function $u : S \times A \to \mathbb{R}$, called the *instantaneous payoff function*, and a number $\delta \in (0, 1)$, called the *discount factor*, such that for any play $p = (s_0, a_0, s_1, a_1, \ldots)$ we have

$$f(p) = \sum_{k=0}^{\infty} \delta^k u(s_k, a_k).$$

Given a decision problem D as above and a history $h = (s_0, a_0, \ldots, s_{\ell-1}, a_{\ell-1}, s_\ell)$ in H, we introduce a *conditional decision problem* D_h to be the problem that the decision maker faces once the history h has occurred. The idea of such conditional decision problems is similar to the idea of a subgame in game theory. In decision problem D_h the initial state is s_ℓ. The set of histories associated with D_h, denoted H_h, is the set of sequences $h' = (s'_0, a'_0, \ldots, s'_{k-1}, a'_{k-1}, s'_k)$ such that $s'_0 = s_\ell$ and the sequence $(h, h') = (s_0, a_0, \ldots, s_\ell, a'_0, \ldots, s'_{k-1}, a'_{k-1}, s'_k)$ is an element of H. We let P_h denote the set of plays of D_h. Transition probabilities in D_h are defined in the obvious way and the payoff function is given by $f_h(p) = f(h, p)$ for each play $p = (s'_0, a'_0, s'_1, a'_1, \ldots)$ in P_h, where (h, p) denotes the infinite sequence $(s_0, a_0, \ldots, s_\ell, a'_0, s'_1, a'_1, \ldots)$. In particular, it holds that $D_{s_0} = D$.

3 Single Decision Maker and Subgame Optimal Policies

This section provides a benchmark for our study. Here we take the point of view of a single decision maker who exercises full control over all the decisions throughout the entire duration of the decision problem. An appropriate solution concept is that of a subgame optimal policy: a policy that is optimal when evaluated after any finite history. We discuss a characterization of subgame optimal policies that is analogous to the dynamic programming principle. We also contrast subgame optimal policies with optimal policies, the latter being the policies that are only required to be optimal at the initial history.

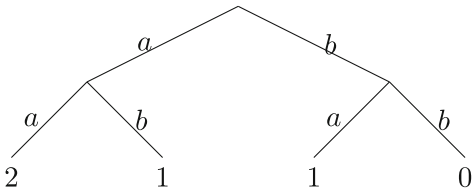

Fig. 1 The decision problem of Example 3.1

A *policy* is a function σ assigning to each history $h \in H$ a probability distribution $\sigma(h)$ on the set A_h. The set of policies is denoted by Σ. A policy is said to be pure if for each $h \in H$ the distribution $\sigma(h)$ assigns probability 1 to some particular action in A_h. For each history $h \in H$, a policy σ of a decision problem D induces a policy σ_h in the decision problem D_h by letting $\sigma_h(h') = \sigma(h, h')$ for each history $h' \in H_h$.

We let $U(\sigma)$ denote the expected payoff of a policy σ. Formally, $U(\sigma)$ is the expected value of the payoff function f with respect to the probability measure on P generated by the policy σ and the law of transition π. Similarly, we let $U_h(\sigma)$ denote the expected payoff of the policy σ conditional on the history h being reached. In particular, it holds that $U_{s_0}(\sigma) = U(\sigma)$. Note that $U_h(\sigma)$ is equal to the expected payoff of σ_h in the decision problem D_h.

For every $h \in H$, we let v_h denote the *value* of the decision problem D_h, that is the highest expected payoff that the decision maker can achieve, once history h has occurred,

$$v_h = \sup_{\sigma \in \Sigma} U_h(\sigma).$$

We write $v = v_{s_0}$ to denote the value of D. A policy $\sigma \in \Sigma$ is *optimal at the initial history*, or simply *optimal*, if $U(\sigma) = v$. It is *optimal at history* $h \in H$ if $U_h(\sigma) = v_h$. A policy $\sigma \in \Sigma$ is *subgame optimal* in D if it is optimal at every $h \in H$.

A subgame optimal policy is clearly optimal at the initial history, but the reverse may not be true. If σ is optimal at the initial history and if history h is reached with positive probability under σ, then σ is also optimal at h. In general, however, a policy that is optimal at the initial history may not be optimal at some other histories, and hence, it may fail to be subgame optimal.

Example 3.1 Consider the decision problem depicted in Fig. 1. It could be represented as a Markov decision problem with three states, s_0, s_1, and s_2. The transitions are deterministic and independent of actions: s_0 is the state in period 0. From state s_0 transition to s_1 occurs with probability 1, and from s_1 transition to s_2 occurs with probability 1. State s_2 is absorbing. In states s_0 and s_1, there are two actions, a and b. Action a gives an instantaneous reward of 1 and action b gives an instantaneous reward of 0. In state s_2, there is only one action and all rewards are equal to zero. The discount factor is taken to be 1. Obviously, playing a after each history in periods 0 and 1 is the only subgame optimal policy. The policy calling the decision maker to play a at the initial history (s_0), to play a at the history (s_0, a, s_1), and to play b at the history (s_0, b, s_1), is optimal but is not subgame optimal.

The set of policies Σ endowed with the product topology is compact by the Tychonoff product theorem. We show in "Appendix" that the payoff function U_h is continuous on Σ, so an optimal policy at history h always exists. We next consider a characterization of subgame optimal policies in terms of 1-day optimal actions, a result that is known in various guises in dynamic programming; see Blackwell [4], and in stochastic games, see Shapley [23].

The values defined above satisfy the following recursive relations:

$$v_h = \max_{a \in A_h} \sum_{s \in S_{h,a}} \pi(s|h,a) \cdot v_{h,a,s}. \tag{3.1}$$

For $h \in H$, we let O_h denote the set of actions $a \in A_h$ for which the maximum in (3.1) is attained. Elements of O_h are called *1-day optimal* actions at h.

Theorem 3.2 *A policy σ is subgame optimal if and only if for each $h \in H$ the distribution $\sigma(h)$ only assigns positive probability to the actions in O_h.*

Proof To prove the *only if* part of the theorem, consider a subgame optimal policy σ. For a history $h \in H$ we have

$$\begin{aligned} v_h &= U_h(\sigma) \\ &= \sum_{a \in A_h} \sum_{s \in S_{h,a}} \sigma(h)(a) \cdot \pi(s|h,a) \cdot U_{h,a,s}(\sigma) \\ &= \sum_{a \in A_h} \sigma(h)(a) \sum_{s \in S_{h,a}} \pi(s|h,a) \cdot v_{h,a,s}, \end{aligned}$$

implying that $\sigma(h)$ is supported by the set O_h.

To prove the *if* part of the theorem, consider a policy σ such that $\sigma(h)$ only assigns positive probability to the members of O_h. We show that σ is optimal at the initial history. We do so using a limit argument: We introduce a sequence $\sigma^0, \sigma^1, \ldots$ of policies converging to σ, where each member σ^t of the sequence is optimal at the initial history, and use the continuity of U.

For $t \in \mathbb{N}$, we define the policy σ^t as follows: Let the decision maker follow the strategy σ until period t. Now suppose that history $h = (s_0, \ldots, s_{t-1}, a_{t-1}, s_t)$ has been reached by period t. Then, at period t, the decision maker is required to switch to any strategy that is optimal at h.

In particular, σ^0 is optimal at the initial history. Unraveling the relation (3.1), we see that the strategy σ^t is optimal at the initial history for each $t \in \mathbb{N}$ since $U(\sigma^t) = v$. Moreover, σ^t coincides with σ on histories with length no longer than t. Hence, σ^t converges to σ as $t \to \infty$. Using the continuity of the function U, we obtain that $U(\sigma) = v$.

A similar argument shows that σ is optimal at each history h. □

It follows from the above theorem that a pure policy σ is subgame optimal if and only if $\sigma(h) \in O_h$ for every $h \in H$. The set of pure subgame optimal policies can therefore be written as a Cartesian product $\prod_{h \in H} O_h$, henceforth denoted by O. Similarly, the entire set of subgame optimal policies can be written as a Cartesian product $\prod_{h \in H} \Delta(O_h)$, where $\Delta(O_h)$ denotes the set of probability distributions on O_h. Henceforth, we denote $\prod_{h \in H} \Delta(O_h)$ by $\Delta(O)$.

If D is a discounted Markov decision problem with instantaneous payoff function u and discount factor δ, then the conditional decision problem D_h and its value v_h depend only on the current state s. We can then write v_s to denote the value v_h for any h with current state equal to s. The relation (3.1) takes the form

$$v_s = \max_{a \in A_s} \left\{ u(s,a) + \delta \sum_{s' \in S_{s,a}} \pi(s'|s,a) \cdot v_{s'} \right\}.$$

4 Multiple Selves and Subgame Perfect Equilibrium

In this section, we take the perspective that the decision maker cannot commit himself to his future actions. To do so, we model the decision maker as consisting of multiple selves. This leads us to a game-theoretic model in which each history of the decision problem is associated with a self. While all selves have the same payoff function, identical to the one of the decision makers, they make their decisions independently. This opens up the possibility of miscoordination. Indeed, we use Example 4.2 to demonstrate that Nash equilibria may fail to be optimal at the initial history. On the other hand, no such miscoordination occurs under the concept of subgame perfect equilibrium: Indeed we prove in Theorem 4.3 that a policy is subgame optimal if and only if it is a subgame perfect equilibrium.

Let D be a decision problem as in Sect. 2. At each history, the current self of the decision maker is free to take any action. Every possible history $h \in H$ therefore leads to a self of the decision maker, also referred to as a player. A pure strategy of player h is an element of A_h and a mixed strategy is an element of $\Delta(A_h)$, the set of probability distributions on A_h. A profile $\sigma = (\sigma(h))_{h \in H}$ of mixed strategies describes a strategy choice for each player. Notice that as a mathematical object, a strategy profile is equivalent to a policy. The utility function of every player $h \in H$ is the same and is identical to the one of the decision makers at the initial history, U.

We start our discussion by applying the concept of Nash equilibrium to the decision problem D viewed as a game played by multiple selves. A strategy profile is a *Nash equilibrium* of D if no player can improve the payoff at the initial history by a unilateral deviation. More precisely, $\sigma \in \Sigma$ is a Nash equilibrium if for every $h \in H$ and for every $\eta(h) \in \Delta(A_h)$ it holds that $U(\sigma) \geq U(\sigma/\eta(h))$, where $\sigma/\eta(h)$ denotes the strategy profile obtained from σ after replacing $\sigma(h)$ by $\eta(h)$.

Lemma 4.1 *If a policy is optimal at the initial history, then it is a Nash equilibrium.*

Proof Let σ be optimal at the initial history. Then, σ maximizes the payoff function U over the entire set of policies. In particular, no unilateral deviation from σ can improve the payoff. □

The converse is not necessarily the case: A Nash equilibrium of D may fail to be an optimal policy at the initial history. Thus, under the concept of Nash equilibrium, multiple selves can severely fail to coordinate. The following example illustrates the point.

Example 4.2 We return to the game introduced in Example 3.1. Consider the policy σ given as follows: $\sigma(s_0) = b, \sigma(s_0, a, s_1) = b$, and $\sigma(s_0, b, s_1) = a$. Then, σ is a Nash equilibrium, but is not optimal at the initial history. The issue at hand is that Nash equilibrium fails to discipline the behavior of the current self at the history (s_0, a, s_1), because this history is reached with probability 0 under σ.

We now turn to the concept of subgame perfect equilibrium. We argue that in a subgame perfect equilibrium, full coordination obtains. More precisely, we show that the set of subgame perfect equilibrium strategy profiles coincides with the set of subgame optimal policies.

A strategy profile is a *subgame perfect equilibrium* of D if it induces a Nash equilibrium in each subgame. Thus, $\sigma \in \Sigma$ is a *subgame perfect equilibrium* if for every $h \in H$ the strategy profile σ_h is a Nash equilibrium of D_h. Equivalently, σ is a subgame perfect equilibrium of D if for every $h \in H$ and every $\eta(h) \in \Delta(A_h)$ it holds that $U_h(\sigma) \geq U_h(\sigma/\eta(h))$.

Theorem 4.3 *Let D be a decision problem. A policy is subgame optimal if and only if it is a subgame perfect equilibrium.*

Proof Let σ be a subgame optimal policy. Then, for every history $h \in H$, σ is optimal at history h. Hence, σ_h is a Nash equilibrium of D_h by Lemma 4.1. We conclude that σ is a subgame perfect equilibrium of D.

Conversely, let σ be a subgame perfect equilibrium of D. Let η be an arbitrary policy. For $t \in \mathbb{N}$, let σ^t be the strategy profile defined as follows: Let $\sigma^t(h)$ be equal to $\eta(h)$ for each history h with length smaller than t and equal to $\sigma(h)$ for a history h of length at least t. In particular, it holds that $\sigma^0 = \sigma$.

It is sufficient to show that for every $h \in H$, for every $t \in \mathbb{N}$,

$$U_h(\sigma^t) \geq U_h(\sigma^{t+1}). \tag{4.1}$$

Indeed, using the continuity of U_h and the fact that σ^t converges to η as t goes to infinity, one then concludes that $U_h(\sigma^0) \geq U_h(\eta)$ as desired.

We continue by proving (4.1). Take some $t \in \mathbb{N}$. If $\ell(h) \geq t+1$, then $\sigma_h^t = \sigma_h^{t+1} = \sigma_h$, so (4.1) holds with equality.

Consider a history h of length $\ell(h) = t$. Since σ_h is a Nash equilibrium in D_h, the player active at history h does not profit from a unilateral deviation from $\sigma(h)$ to $\eta(h)$. Notice that such a deviation induces the strategy profile σ_h^{t+1} in D_h. We conclude that (4.1) is satisfied.

Finally, we use induction to prove (4.1) for histories of length $0, \ldots, t$. We already know that (4.1) holds for all histories of length t. Suppose we have proven (4.1) for all histories of length $k+1 \in \{1, \ldots, t\}$. Consider a history h of length $\ell(h) = k$. It holds that

$$\begin{aligned} U_h(\sigma^t) &= \sum_{a \in A_h} \sum_{s \in S_{h,a}} \eta(h)(a)\pi(s|h,a) U_{(h,a,s)}(\sigma^t) \\ &\geq \sum_{a \in A_h} \sum_{s \in S_{h,a}} \eta(h)(a)\pi(s|h,a) U_{(h,a,s)}(\sigma^{t+1}) = U_h(\sigma^{t+1}), \end{aligned}$$

where the inequality follows by the induction hypothesis. This completes the induction step and the proof of the theorem. □

Theorem 4.3 is closely related to the principle of optimality of dynamic programming, also known in game theory as the one-stage-deviation principle. We are not aware of proofs of this principle at the level of generality of Theorem 4.3. The quite general treatment of Lemma 1 in Harris [9] does not allow for moves by nature. The extensive survey by Puterman [21] restricts attention to Markov decision problems.

The following corollary follows immediately from the preceding theorem.

Corollary 4.4 *Let D be a decision problem. A pure policy is subgame optimal if and only if it is a pure subgame perfect equilibrium.*

5 Multiple Selves and Curb Sets

Although we have derived an equivalence between subgame optimal policies and subgame perfect equilibria of a decision problem, one may still criticize the lack of stability of Nash equilibrium for the decision problems at the various histories. The problem is essentially that a Nash equilibrium only requires a deviation not to be profitable, rather than requiring that it actually involves a loss. The concept of strict equilibrium addresses this issue, but may fail to exist. For instance, in a decision problem where a player can choose between two distinct best responses, a strict equilibrium does not exist.

To address the instability of Nash equilibrium, Basu and Weibull [3] consider minimal sets of strategy profiles that are closed under rational behavior (curb) for normal-form games with a finite number of players. In this section, we define curb sets for decision problems. We show that the minimal curb set is unique and equal to the set of pure subgame perfect equilibria, so therefore equal to the set of pure subgame optimal policies by virtue of Corollary 4.4.

Let D be a decision problem as in Sect. 2. Let $\mathcal{B}$ denote the collection of all non-empty product sets $X \subseteq \prod_{h\in H} A_h$, i.e., sets of the form $X = \prod_{h\in H} X_h$ for some $X_h \subseteq A_h$. Recall that, by Zorn's lemma, the product set X is non-empty precisely when X_h is non-empty for each $h \in H$. For every $X \in \mathcal{B}$, we define the subset $\Delta(X)$ of Σ as the set of strategy profiles σ such that for every $h \in H$ the support of $\sigma(h)$ is contained in X_h. Thus, $\Delta(X)$ is of the following form

$$\Delta(X) = \prod_{h\in H} \Delta(X_h).$$

We recall that the set of pure subgame optimal policies, denoted O, is a product set, with its factor O_h being the set of 1-day optimal actions at h, and that $\Delta(O)$ is the set of subgame optimal policies of D.

The set of *pure best responses* by player $h \in H$ at history h against a strategy profile $\sigma \in \Sigma$ is defined by

$$b_h(\sigma) = \arg\max_{a(h)\in A_h} U_h(\sigma/a(h)). \tag{5.1}$$

Note that the pure strategy profile σ is a subgame perfect equilibrium of D precisely when $\sigma(h) \in b_h(\sigma)$ for every $h \in H$.

We proceed to define the function $\mu : \mathcal{B} \to \mathcal{B}$, called the *curb operator* for D, as follows: For every $X \in \mathcal{B}$, let

$$\mu_h(X) = \bigcup_{\sigma\in\Delta(X)} b_h(\sigma),$$

and

$$\mu(X) = \prod_{h\in H} \mu_h(X) = \prod_{h\in H} \bigcup_{\sigma\in\Delta(X)} b_h(\sigma).$$

Thus, a pure policy η is an element of $\mu(X)$ if for every player $h \in H$ there exists a policy $\sigma \in \Delta(X)$ such that $\eta(h)$ is player h's best response to σ. Essential to this definition is the order of quantification. It reflects the fact that different players are allowed to hold different, and incompatible, beliefs about their future selves.

Definition 5.1 Let D be a decision problem. A set $X \in \mathcal{B}$ is *closed under rational behavior (curb)* if $\mu(X) \subseteq X$. A curb set is *minimal* if it does not contain any curb set as a proper subset.

The set of pure strategy profiles X is curb if in case all players believe that actions outside X are played with probability 0, then rational players will only play actions inside X. Since the curb criterion is met by the set $X = \prod_{h\in H} A_h$ of all pure strategy profiles, we are particularly interested in minimal curb sets.

For normal-form games with a finite number of players, Basu and Weibull [3] show that a minimal curb set always exists, though it may not be unique. The next result claims that also every decision problem has at least one minimal curb set.

Theorem 5.2 *Let D be a decision problem. Then, D has at least one minimal curb set.*

Proof Clearly, the set $\prod_{h\in H} A_h$ is a curb set. Let $\mathcal{C}$ be the collection of all curb sets. We define the partial order $\supseteq$ on $\mathcal{C}$ in the usual way, so for $X, X' \in \mathcal{C}$ it holds that $X \supseteq X'$ if and only if, for every $h \in H$, $X_h \supseteq X'_h$.

Let $\mathcal{D}$ be a subset of $\mathcal{C}$ that is totally ordered by $\supseteq$. We define $X' = \cap_{X\in\mathcal{D}} X$. Since $\mathcal{D}$ is totally ordered by $\supseteq$ and, for every $X \in \mathcal{D}$, for every $h \in H$, X_h is finite, it follows that X' is non-empty.

We show X' to be a curb set. For every $X \in \mathcal{D}$, since $X \supseteq X'$, it holds that $\mu(X) \supseteq \mu(X')$. It follows that

$$X' = \cap_{X\in\mathcal{D}} X \supseteq \cap_{X\in\mathcal{D}} \mu(X) \supseteq \mu(X'),$$

where the first inclusion follows from the fact that every X in $\mathcal{D}$ is a curb set. We have shown that X' is a curb set.

The set X' is an upper bound on $\mathcal{D}$; hence, by Zorn's lemma, it holds that $\mathcal{C}$ has at least one maximal element. A maximal element of $\mathcal{C}$ with respect to $\supseteq$ is a minimal curb set. □

We remark that the existence of a minimal curb set in a decision problem also follows from Theorem 5.5.

A strict subgame perfect equilibrium is a pure strategy profile σ such that $\mu(\{\sigma\}) = \{\sigma\}$, so is a singleton curb set. The set-valued generalization of a strict subgame perfect equilibrium is a curb set X such that $\mu(X) = X$.

Definition 5.3 Let D be a decision problem. A curb set X of D is *tight* if $\mu(X) = X$.

A tight curb set has the desirable property that none of its elements can be deleted if players hold beliefs in $\Delta(X)$. For normal-form games with a finite number of players, Basu and Weibull [3] show that a minimal curb set is always tight. The next result states that also for decision problems every minimal curb set is tight.

Theorem 5.4 *Let D be a decision problem. If X is a minimal curb set of D, then X is tight.*

Proof Let X be a minimal curb set of D. Since $\mu(X) \subseteq X$, it holds that $\mu(\mu(X)) \subseteq \mu(X)$, so $\mu(X)$ is a curb set. Since the curb set X is minimal and $\mu(X) \subseteq X$, it follows that $\mu(X) = X$, so the curb set X is tight. □

The next result shows that the set of pure subgame optimal policies is a minimal curb set. We have not yet ruled out the possibility that there are other minimal curb sets.

Theorem 5.5 *Let D be a decision problem. Then, the set of pure subgame optimal policies O is a minimal curb set of D.*

Proof We first argue that for each $\sigma \in \Delta(O)$ it holds that $b_h(\sigma) = O_h$. To see this, take some $\sigma \in \Delta(O)$ and consider the maximization problem (5.1). Defining $a = a(h)$, we can write

$$U_h(\sigma/a(h)) = \sum_{s\in S_{h,a}} \pi(s|h,a) \cdot U_{h,a,s}(\sigma) = \sum_{s\in S_{h,a}} \pi(s|h,a) \cdot v_{h,a,s}.$$

Hence, the maximum in (5.1) equals v_h by (3.1). It is reached if and only if a is an element of O_h.

It follows that $\mu(O) = O$, so that O is a curb set. To see that O is a minimal curb set, let X be a curb set such that $X \subseteq O$. Take any $\sigma \in X$. Since $b_h(\sigma) = O_h$ for every $h \in H$, we have $O \subseteq \mu(X) \subseteq X$, and therefore $X = O$. □

A normal-form game can have several minimal curb sets. The next result shows that the minimal curb set of a decision problem is unique and therefore equal to O.

Theorem 5.6 *Let D be a decision problem. Then, O is the unique minimal curb set of D.*

Proof STEP 1: *Let X be a minimal curb set of D. Then, the function U is constant on X.*

Let $D' = (S, A, H', \pi', f')$ be the decision problem that is identical to D, except that the set of actions at a history $h \in H'$ is restricted to X_h, so H' consists of histories h such that $a_k(h) \in X_{h^k}$ for each $k \in \{0, \ldots, \ell(h) - 1\}$. The set $G' \subseteq G$ contains the nature histories corresponding to H' and π' is the restriction of π to G', and f' is the restriction of f to plays of D'. Let v' denote the value of D' and μ' its curb operator. Let O' be the set of pure subgame optimal policies of D'. By Theorem 5.5, O' is a minimal curb set of D'.

We prove Step 1 by showing that $U(\sigma) = v'$ for each $\sigma \in X$.

For $h \in H$, define X'_h to be equal to O'_h if $h \in H'$ and to be equal to X_h if $h \in H \setminus H'$. Let $X' = \prod_{h \in H} X'_h$. We argue that X' is a curb set of D. Since $X' \subseteq X$, we have that $\mu(X') \subseteq \mu(X) \subseteq X$. In particular, for $h \in H \setminus H'$ we have $\mu_h(X') \subseteq X_h = X'_h$. Now consider $h \in H'$. We argue that $\mu_h(X') \subseteq O'_h$. Take a policy $\sigma \in \Delta(X')$ in D and let σ' be the restriction of σ to histories in H'. Thus, σ' is a policy in D' and $\sigma' \in \Delta(O')$.

Consider an action $a \in b_h(\sigma)$, i.e., player h's best response to σ in D. Since $a \in \mu_h(X) \subseteq X_h$, a is a feasible action for player h in D'. It then follows that a is a best response of player h to σ' in D', so that $a \in b'_h(\sigma')$. This establishes that $\mu_h(X') \subseteq \mu'_h(O')$. Since $\mu'_h(O') \subseteq O'_h$, we obtain $\mu_h(X') \subseteq O'_h$. It holds that $\mu_h(X') \subseteq X'_h$ for all $h \in H$, as desired.

Since X' is a curb set of D, X is a minimal curb set of D, and $X' \subseteq X$, we conclude that $X' = X$. Thus, in particular, it holds that $O'_h = X_h$ for all $h \in H'$.

Now take a policy $\sigma \in X$ and let σ' be the restriction of σ to histories in H'. It is clear that the measure induced by σ from the initial state s_0 on the set P of plays of D is supported by P', the set of plays of D'. Consequently, $U(\sigma)$ equals the payoff of σ' in D'. But since $\sigma' \in O'$, the payoff on σ' in D' is exactly v'. We conclude that $U(\sigma) = v'$.

STEP 2: *Let X be a minimal curb set of D. Then, for every $h \in H$, the function U_h is constant on X.*

Take a history $h \in H$ and consider the decision problem D_h. The set $Y = \prod_{h' \in H_h} X_{(h,h')}$ is curb for the decision problem D_h. By Step 1, the payoff function U_h is constant on Y and the result follows.

STEP 3: *Let X be a minimal curb set of D. Then, it holds that $X \subseteq O$.*

Take any $\sigma \in X$ and suppose that $\sigma \notin O$. By Corollary 4.4, σ is not a subgame perfect equilibrium of D. Consequently, there exists $h \in H$ such that $\sigma(h) \notin b_h(\sigma)$. Hence, for $a(h) \in b_h(\sigma)$, we have $U_h(\sigma/a(h)) > U_h(\sigma)$. Since X is a curb set, it holds that $b_h(\sigma) \subseteq X_h$. Thus, $a(h)$ is an element of X_h, and hence, $\sigma/a(h)$ is an element of X. This is a contradiction to Step 2. The result of Step 3 follows.

Finally, since X is a curb set while O is a minimal curb set, we conclude that $X = O$. Thus, O is the only minimal curb set of D, as desired. □

Theorem 5.6 allows us to conclude that even when a decision maker is unable to commit to his future actions and even when one criticizes Nash equilibrium for its lack of stability and one considers the weaker solution concept of a curb set, one can safely restrict attention to subgame optimal policies in the analysis of decision problems.

6 Conclusions

The standard analysis of decision problems assumes perfect commitment of the decision maker. In this paper, we study a decision maker who is unable to commit to his future actions.

As a benchmark, we consider a single decision maker who exercises full control over all the decisions taken. Relevant to this benchmark are two solution concepts: optimality, and subgame optimality. Both implicitly assume that the decision maker is able to commit to his future action choices.

We next take the perspective that the decision maker cannot commit to his future actions. The decision maker is therefore modeled as consisting of multiple selves. The current self of the decision maker has to form beliefs regarding the behavior of his future selves. Relevant to this point of view are game-theoretic solution concepts that treat the multiple selves as players in a game. Accordingly, we consider Nash equilibrium, subgame perfect equilibria, and curb sets.

Restricting attention to pure strategies, the following relationships between the various solution concepts emerge:

$$\begin{gathered}\text{subgame optimal policies}\\ =\\ \text{subgame perfect equilibria}\\ =\\ \text{the minimal curb set}\\ \subset\\ \text{policies optimal at the initial history}\\ \subset\\ \text{Nash equilibria.}\end{gathered}$$

We conclude with an open issue. Recall that each minimal curb set is necessarily tight. We have shown that a *minimal* curb set of a decision problem is unique. We do not know whether a *tight* curb set of a decision problem is unique, or equivalently, whether each tight curb set is necessarily minimal.

Appendix

In this appendix we prove that for every history $h \in H$, the function U_h is continuous on Σ, where Σ is endowed with the product topology.

Let Δ denote the set of Borel probability measures on P endowed with its usual weak* topology. Let $I : \Delta \to \mathbb{R}$ be given by $I(\mu) = \int f d\mu$. Let $J : \Sigma \to \Delta$ be the function that assigns to a policy $\sigma \in \Sigma$ the probability measure J_σ induced on the set of plays P by the policy σ together with the law of transition π, conditional on the history h being reached.

The function U_h is the composition of I and J. Thus, it is enough to argue that both I and J are continuous.

The function I is continuous because f is continuous.

We argue that the function $J : \Sigma \to \Delta$ is continuous. Fix $\sigma \in \Sigma$ and a sequence $(\sigma_n)_{n\in\mathbb{N}}$ of policies converging to σ. First notice that if $C \subseteq P$ is a cylinder set, then

$$\lim_{n\to\infty} J_{\sigma_n}(C) = J_\sigma(C).$$

Now any open subset of $W \subseteq P$ of plays can be written as a disjoint union of countably many cylinder sets, say as $W = C_0 \cup C_1 \cup \cdots$. It then follows that for each $m \in \mathbb{N}$,

$$\liminf_{n\to\infty} J_{\sigma_n}(W) \geq \liminf_{n\to\infty} J_{\sigma_n}(\cup_{i=1}^m C_i) = \sum_{i=0}^m \lim_{n\to\infty} J_{\sigma_n}(C_i) = \sum_{i=0}^m J_\sigma(C_i).$$

Taking the limit with respect to m, we obtain

$$\liminf_{n\to\infty} J_{\sigma_n}(W) \geq J_\sigma(W).$$

The fact that J is continuous now follows from Theorem 15.3 in Aliprantis and Border [1].

References

1. Aliprantis CD, Border K (2006) Infinite dimensional analysis. Springer, Berlin
2. Balkenborg D, Hofbauer J, Kuzmics C (2013) Refined best reply correspondence and dynamics. Theor Econ 8:165–192
3. Basu K, Weibull JW (1991) Strategy subsets closed under rational behavior. Econ Lett 36:141–146
4. Blackwell D (1965) Discounted dynamic programming. Ann Math Stat 36:226–235
5. Cingiz K, Flesch J, Herings PJJ, Predtetchinski A (2016) Doing it now, later, or never. Games Econ Behav 97:174–185
6. Durieu J, Solal P, Tercieux O (2011) Adaptive learning and p-best response sets. Int J Game Theory 40:735–747
7. Frederick S, Loewenstein G, O'Donoghue T (2002) Time discounting and time preference: a critical review. J Econ Lit 40:351–401
8. Goldman SM (1980) Consistent plans. Rev Econ Stud 47:533–537
9. Harris C (1985) Existence and characterization of perfect equilibrium in games of perfect information. Econometrica 53:613–628
10. Hurkens S (1995) Learning by forgetful players. Games Econ Behav 11:304–329
11. Kah C, Walzl M (2015) Stochastic stability in a learning dynamic with best response to noisy play. In: Working papers in economics and statistics, 2015–2015. University of Innsbruck, pp 1–29
12. Kydland FE, Prescott EC (1977) Rules rather than discretion: the inconsistency of optimal plans. J Polit Econ 85:473–491
13. Myerson RB, Weibull J (2015) Tenable strategy blocks and settled equilibria. Econometrica 83:943–976
14. Nash JF (1950) Equilibrium points in n-person games. Proc Natl Acad Sci 36:48–49
15. O'Donoghue T, Rabin M (1999) Doing it now or later. Am Econ Rev 89:103–124
16. Peleg B, Potters J, Tijs S (1996) Minimality of consistent solutions for strategic games in particular for potential games. Econ Theory 7:81–93
17. Peleg B, Tijs S (1996) The consistency principle for games in strategic form. Int J Game Theory 25:13–34
18. Peleg B, Yaari ME (1973) On the existence of a consistent course of action when tastes are changing. Rev Econ Stud 40:391–401
19. Pollak RA (1968) Consistent planning. Rev Econ Stud 35:201–208
20. Pruzhansky V (2003) On finding CURB sets in extensive games. Int J Game Theory 32:205–210
21. Puterman ML (1994) Markov decision processes. Discrete stochastic dynamic programming. Wiley, Hoboken
22. Selten R (1965) Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit, Teil I: Bestimmung des dynamischen Preisgleichgewichts. Zeitschrift für die gesamte Staatswissenschaft 121:301–324

23. Shapley LS (1953) Stochastic games. In: Proceedings of the national academy of sciences of the USA, vol 39, pp 1095–1100
24. Strotz RH (1956) Myopia and inconsistency in dynamic utility maximization. Rev Econ Stud 23:165–180
25. Schäl M (1975) On dynamic programming: compactness of the space of policies. Stoch Process Their Appl 3:345–364
26. Voorneveld M, Kets W, Norde H (2005) An axiomatization of minimal CURB sets. Int J Game Theory 33:479–490
27. Young HP (1993) The evolution of conventions. Econometrica 61:57–84
28. Young HP (1998) Individual strategy and social structure: an evolutionary theory of institutions. Princeton University Press, Princeton