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

Maastricht University

P. Jean-Jacques Herings,
Andrey Meshalkin,
Arkadi Predtetchinski

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

RM/16/021

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

P.O Box 616
NL- 6200 MD Maastricht
The Netherlands

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

P. Jean-Jacques Herings[*], Andrey Meshalkin[†], Arkadi Predtetchinski[‡]

May 10, 2016

Abstract

The paper considers a class of decision problems with 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. We introduce the notions of Nash equilibrium, subgame perfect equilibrium, and curb sets for decision problems. An optimal policy at the initial history is a Nash equilibrium but not vice versa. Both subgame perfect equilibria and curb sets are equivalent to subgame optimal policies. The concept of a subgame optimal policy is therefore robust to the absence of commitment technologies.

KEYWORDS: decision problem, multiple selves, subgame perfect equilibrium, curb sets.

JEL CODES: C61, C62, C73.

[*]P.J.J. Herings (P.Herings@maastrichtuniversity.nl). Department of Economics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands.

[†]A. Meshalkin (A.Rybakov@maastrichtuniversity.nl), Department of Economics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands.

[‡]A. Predtetchinski (A.Predtetchinski@maastrichtuniversity.nl), Department of Economics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands.

1 Introduction

In this paper, we study a class of decision problems with 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 that is informed about the state and has to take an action out of a set of actions, thereby obtaining an instantaneous payoff and 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.

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 maker. Our emphasis is on the characterization of policies that are consistent with Nash equilibrium and 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 characterize subgame optimal policies as a particular product of pure subgame optimal policies. These results are completely in line with those derived for Markov decision problems, see for instance the excellent overview by Puterman (1994).

We continue our analysis by assuming that the decision make 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 future actions has drastic consequences. Famous examples are given in the area of monetary policy by Kydland and Prescott (1977), where lack of commitment leads to socially suboptimal decision making, and in durable goods monopoly by Gul, Sonnenschein, and Wilson (1986), where the inability of the seller to commit to future prices exerts a negative externality on its current decisions and reduces its profits.

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 maker.[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 (1950). 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 (1973) 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 as introduced in Selten (1965). 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 (1980) 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 (1991) for games in normal form with a finite number of players.

Basu and Weibull (1991) 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 (2003) proposes a slight variation on this notion for extensive games with perfect information and a finite horizon.

[1] Even if the decision maker is modeled as consisting of multiple selves, there is therefore no issue of time-inconsistent preferences as introduced in Strotz (1956) and Pollak (1968). For an overview of that stream of the literature, we refer the reader to Frederick, Loewenstein, and O'Donoghue (2002).

Myerson and Weibull (2015) 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, Kets, and Norde (2005) point towards 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 (1996) and Peleg, Potters and Tijs (1996). 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, Kets, and Norde (2005) provide an axiomatization of minimal curb sets.

Another advantage of minimal curb sets is that they are robust in a dynamic sense. Hurkens (1995) studies a stochastic version of fictitious play in the spirit of Young (1993) and shows that such a dynamic process of strategy adjustment will eventually settle down in a minimal curb set. Similarly, Young (1998) 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, Solal, and Tercieux (2011) for the analysis of a more general class of fictitious play processes. Balkenborg, Hofbauer, and Kuzmics (2013) 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 (2015).

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 rest of the paper is organized as follows. In Section 2, we define a class of decision problems that contains Markov decision problems as a special case. Section 3 provides the basic definitions related to policies and gives a characterization of subgame optimal policies. We introduce multiple selves in Section 4, define and study the concepts of Nash equilibrium and subgame perfect equilibrium of a decision problem and show the set of subgame perfect equilibria to coincide with the set of subgame optimal policies. Section 5

introduces the notion of a minimal curb set of a decision problem and shows that it coincides with 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* in the non-empty, finite set of actions A and nature picks *states* in the non-empty, finite set of states S. The payoff to the decision maker depends on the entire sequence of states and actions thus produced. The class of decision problems that we study contains the class of discounted Markov decision problems as a subclass. The element s_0 is a distinguished element of S called the initial state.

We let $\mathbb{N} = \{0, 1, \dots\}$ denote the set of natural numbers with 0. Each element h of the set H is a finite sequence of the form $h = (s_0, a_0, \dots, s_{\ell-1}, a_{\ell-1}, s_\ell)$ where ℓ is a natural number, $s_0, \dots, s_\ell$ are elements of S and $a_0, \dots, a_{\ell-1}$ are elements of A. Elements of H are called *histories*.

Given a history $h = (s_0, a_0, \dots, s_{\ell-1}, a_{\ell-1}, s_\ell)$ in H, we denote its length ℓ by $\ell(h)$. Moreover, for $k = 0, \dots, \ell(h) - 1$, we denote the state in period k by $s^k(h)$, the action taken in period k by $a^k(h)$, and the current state $s^{\ell(h)}(h)$ by $s^*(h)$.

Consider histories h and h' in H, where $h' = (s_0, a_0, \dots, s_{\ell-1}, a_{\ell-1}, s_\ell)$. The history h is said to be a *subhistory* of h' if $h = (s_0, a_0, \dots, s_{k-1}, a_{k-1}, s_k)$ 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 h' of length k 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, as

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

We extend the notions of subhistories and length to nature histories in the 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. For every $h \in H$, $A_h \neq \emptyset$.

2. For every $g \in G$, $S_g \neq \emptyset$.

3. For every $h \in H$, each subhistory of h is an element of H.

We do not impose any restriction on the set of available actions at a history $h \in H$, apart from the existence of at least one such action. Similarly, we do not impose any restriction on the set of possible states at a nature history g, apart from the existence of at least one such state.

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 $p = (s_0, a_0, s_1, a_1, \dots)$. The sequence p is said to be a *play* if the finite sequence $(s_0, a_0, \dots, s_k, a_k)$ is an element of H for every $k \in \mathbb{N}$. We let P be the set of plays. We endow P with the topology generated by the basis of cylinder sets. The *payoff function* $f : P \to \mathbb{R}$ assigns payoffs to plays. Throughout this paper we assume that the function f is continuous.

A decision problem proceeds as follows. At stage 0 the state is given to be s_0. We let $h_0 = (s_0)$ be the first decision history encountered by the decision maker. The decision then chooses an action $a_0 \in A(h_0)$ and the transition to the next state s_1 occurs with probability $\pi(s_1|h_0, a_0)$. We let $h_1 = (s_0, a_1, s_1)$. The decision maker then chooses an action $a_1 \in A(h_1)$, and the transition to state s_2 occurs with probability $\pi(s_2|h_1, a_1)$. This process continues ad infinitum. Nature and the decision maker thus produce a play $p = (s_0, a_0, s_1, a_1, s_2, \dots)$. The decision maker receives the payoff $f(p)$.

An important subclass of decision problems are 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 [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 $p = (s_0, a_0, s_1, a_1, \dots)$ we have

$$f(p) = \sum_{t=0}^{\infty} \delta^t u(s_t, a_t).$$

Given a decision problem D as above and a history $h = (s_0, a_0, \dots, s_{\ell-1}, a_{t-1}, s_t)$ 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 the decision problem D_h the initial state is s_t. The set of histories in D_h, denoted H_h, is the set of sequences $h' = (s'_0, a'_0, \dots, s'_{\ell-1}, a'_{\ell-1}, s'_\ell)$ such that $s'_0 = s_t$ and the sequence $(h, h') = (s_0, a_0, \dots, s_t, a'_0, \dots, s'_{\ell-1}, a'_{\ell-1}, s'_\ell)$ 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 $p = (s'_0, a'_0, s'_1, a'_1, \dots)$ in P_h, where (h, p) denotes the infinite sequence $(s_0, a_0, \dots, s_t, a'_0, s'_1, a'_1, \dots)$. In particular, it holds that $D_{s_0} = D$.

3 Policies

In this section, we define various notions related to policies and characterize optimal policies.

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 the 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 the 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 $U_{s_0}(\sigma) = U(\sigma)$. Equivalently, we let σ_h denote the policy in the decision problem D_h induced by σ and observe that $U_h(\sigma)$ is equal to the expected payoff of σ_h in the decision problem D_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 the 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* 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 $h \in H$. Equivalently, σ is subgame optimal if for every $h \in H$ the policy σ_h is optimal at the initial history of the conditional decision problem D_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 need not be optimal at all histories, and hence it may fail to be subgame optimal.

The set of policies Σ endowed with the product topology is compact by the Tychonoff product theorem. Since the payoff function U_h is continuous on Σ, an optimal policy at history h always exists. We next consider a characterization of subgame optimal policies in terms of one–day optimal actions, a result that is known in various guises in dynamic programming, see Blackwell (1965) and stochastic games, see Shapley (1953).

The values defined above satisfy the following recursive relations:

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

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 *one–day optimal* actions at h.

Theorem 3.1 *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

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

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. For $t \in \mathbb{N}$, let σ^t denote a policy such that [1] for every history h with length smaller than t, $\sigma^t(h) = \sigma(h)$, and [2] for every history h of length t, σ is optimal at h. Thus in particular σ^0 is optimal at the initial history. Unraveling the relation (3.1), we obtain $U(\sigma^t) = v$. Using the continuity of the function U, and the fact that σ^t converges to σ as $t \to \infty$, we obtain that $U(\sigma) = v$. A similar argument shows that σ is optimal at each history h. $\square$

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 to be 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 $s^*(h)$ at h. We can then write v_s to denote the value v_h for any h with $s^*(h) = s$. The relation (3.1) takes the form

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

For a Markov decision problem, a policy σ is optimal at the initial state s_0 if $U_{s_0}(\sigma) = v_{s_0}$. The policy σ is said to be *optimal* in the Markov decision problem D if it is optimal at each initial state, e.g. Puterman (1994). A classical result states that a discounted Markov decision problem D has a stationary optimal policy. A stationary optimal policy is also subgame optimal in the sense defined before. However, simple examples suffice to show that a history dependent optimal policy need not be subgame optimal. Subgame optimality is therefore a natural strengthening of optimality.

4 Multiple Selves

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 maker, they make their decisions independently. This potentially opens up a possibility for miscoordination. In this section we analyze the game using the concepts of both Nash and subgame perfect equilibrium.

Let D be a decision problem as in Section 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 maker 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 period by a unilateral deviation. More precisely, $\sigma \in \Sigma$ is a Nash equilibrium if for every $h \in H$ and for every

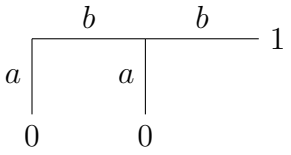

Figure 1: A decision problem with a Nash equilibrium that is not optimal at the initial history.

$\eta(h) \in \Delta(A_h)$ it holds that $U(\sigma) \geq U(\sigma/\eta(h))$, where $\sigma/\eta(h)$ denotes the strategy profile obtained from σ by replacing its coordinate $\sigma(h)$ by $\eta(h)$.

If a policy $\sigma^* \in \Sigma$ is optimal at the initial history, then it is a Nash equilibrium of D. For recall that by definition a policy that is optimal at the initial history maximizes the payoff function U over the entire set of policies. Thus in particular no unilateral deviation from σ^* can improve the payoff.

The following example shows that 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.

Example 4.1 Consider the decision problem depicted in Figure 1. Formally, this could be modeled as a decision problem D where the set of states S is a singleton, the set of actions is $A = \{a, b\}$, the set H consists of all sequences $(s_0, a_0, \dots, s_0)$ where $a_t \in A$ and s_0 is the only state, and $f(p)$ depends on the play $p = (s_0, a_0, s_0, a_0, \dots)$ only through its coordinates a_0 and a_1 as follows: $f(p) = 0$ if $a_0 = a$ or if $a_0 = b$ and $a_1 = a$, and $f(p) = 1$ otherwise.

Obviously, playing $a_0 = a_1 = b$ is an optimal policy at the initial history. However, the decision problem D has another Nash equilibrium: playing action a at both periods 0 and 1. Indeed, a unilateral deviation by player s_0, the player active at the root of the tree, to action b is not profitable, because player (s_0, b, s_0) is assumed to stick with action b. The unilateral deviation to action b by player (s_0, b, s_0) is not profitable, because player s_0 plays a, so that player (s_0, b, s_0) has no effect on the payoff. This second Nash equilibrium leads to a strictly lower payoff than an optimal policy at the initial history.

We presently 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 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 4.2 *Let D be a decision problem. A policy is subgame optimal if and only if it is a subgame perfect equilibrium of D.*

Proof: Suppose a policy σ is subgame optimal. Then for every history $h \in H$ the policy σ is optimal at history h. Hence σ_h is a Nash equilibrium of D_h. We conclude that σ is a subgame perfect equilibrium of D.

Conversely, suppose a policy σ is 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 be 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

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

where the inequality follows by the induction hypothesis. This completes the induction step and the proof of the theorem. $\square$

The following corollary follows immediately from the preceding theorem.

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

	L	R
T	4, 1	1, 2
M	1, 3	2, 1
B	6, 2	0, 3

Figure 2: A normal-form game with an unstable Nash equilibrium.

5 Sets Closed Under Rational Behavior

Although we have derived an equivalence between the subgame optimal policies and the 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 $h \in H$ can choose between two distinct best responses, a strict equilibrium does not exist.

The following example is taken from Basu and Weibull (1991) to illustrate that indifferences of a player about what action to take can make a Nash equilibrium unstable.

Example 5.1 Consider the normal-form game depicted in Figure 2. This game has a unique Nash equilibrium where player 1 randomizes between T and M with probability 2/3 and 1/3, respectively, and player 2 randomizes between L and R with probability 1/4 and 3/4, respectively. Under the Nash equilibrium beliefs, player 1 is indifferent between T and M and player 2 is indifferent between L and R. If player 2 believes that player 1 is going to choose M with probability above 1/3, a belief that is quite natural given that player 1 is indifferent between T and M, then player 2's unique best response is L. And if player 1 believes it to be quite likely that player 2 chooses L, then his unique best response is B. It therefore is not justified to exclude B as a reasonable choice for player 1.

To address the instability of Nash equilibrium, Basu and Weibull (1991) 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 a 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 4.3.

Let D be a decision problem as in Section 2. We say that a subset X of $\prod_{h\in H} A_h$ is a *product* set if $X = \prod_{h\in H} X_h$ for some $X_h \subseteq A_h$. Let $\mathcal{B}$ denote the collection of all non–empty products sets. 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 one–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 5.2 *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 implies that 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 (1991) 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 5.3 *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. $\square$

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 5.4 *Let D be a decision problem. A curb set X 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 (1991) 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 5.5 *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. $\square$

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 5.6 *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

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

Hence, the maximum in (5.1) equals v_h. 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 $X = O$. $\square$

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

Proof:

STEP 1: *Let X be a 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 5.6, 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 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. The result follows.

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

Take any $\sigma \in X$ and suppose that $\sigma \notin O$. By Corollary 4.3, 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. $\square$

6 Conclusions

The standard analysis of decision problems assumes perfect commitment of the decision maker. In this paper, we take the perspective that the decision maker cannot commit to his future action choices. 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.

We study a class of infinite horizon decision problems that contains the class of Markov decision problems as a special case. We formulate the concepts of optimality at a history

and subgame optimality for a policy as a benchmark. These concepts implicitly assume that the decision maker can commit to his future action choices. We argue that these concepts are robust with respect to the multiple selves model under fairly weak assumptions regarding the rationality of future selves. Both the concept of subgame perfect equilibrium and the concept of closed under rational behavior yields the set of subgame optimal policies as the unique prediction. Only a concept like Nash equilibrium that makes significantly weaker assumptions with respect to the rationality of future selves leads to a wider class of policies.

References

Balkenborg, D., J. Hofbauer, and C. Kuzmics (2013), "Refined Best Reply Correspondence and Dynamics," *Theoretical Economics, 8,* 165–192.

Basu, K., and J.W. Weibull (1991), "Strategy Subsets Closed under Rational Behavior," *Economics Letters, 36,* 141–146.

Blackwell, D. (1965), "Discounted Dynamic Programming," *The Annals of Mathematical Statistics, 36,* 226-235.

Durieu, J., P. Solal, and O. Tercieux (2011), "Adaptive Learning and p-Best Response Sets," *International Journal of Game Theory, 40,* 735–747.

Frederick, S., G. Loewenstein, and T. O'Donoghue (2002), "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature, 40,* 351–401.

Goldman, S.M. (1980), "Consistent Plans," *Review of Economic Studies, 47,* 533–537.

Gul, F., H. Sonnenschein, and R. Wilson (1986), "Foundations of Dynamic Monopoly and the Coase Conjecture," *Journal of Economic Theory, 39,* 155–190.

Hurkens, S. (1995), "Learning by Forgetful Players," *Games and Economic Behavior, 11,* 304–329.

Kah, C., and M. Walzl (2015), "Stochastic Stability in a Learning Dynamic with Best Response to Noisy Play," Working Papers in Economics and Statistics, 2015-15, University of Innsbruck, 1–29.

Kydland, F.E., and E.C. Prescott (1977), "Rules Rather than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy, 85,* 473–491.

Myerson, R.B., and J. Weibull (2015), "Tenable Strategy Blocks and Settled Equilibria," *Econometrica, 83,* 943–976.

Nash, J.F. (1950), "Equilibrium Points in n-Person Games," *Proceedings of the National Academy of Sciences, 36,* 48–49.

Peleg, B., J. Potters, and S. Tijs (1996), "Minimality of Consistent Solutions for Strategic Games, in Particular for Potential Games," *Economic Theory, 7,* 81–93.

Peleg, B., and S. Tijs (1996), "The Consistency Principle for Games in Strategic Form," *International Journal of Game Theory, 25,* 13–34.

Peleg, B., and M.E. Yaari (1973), "On the Existence of a Consistent Course of Action when Tastes Are Changing," *Review of Economic Studies, 40,* 391–401.

Pollak, R.A. (1968), "Consistent Planning," *Review of Economic Studies, 35,* 201–208.

Pruzhansky, V. (2003), "On Finding CURB Sets in Extensive Games," *International Journal of Game Theory, 32,* 205–210.

Puterman, M.L. (1994), *Markov Decision Processes, Discrete Stochastic Dynamic Programming,* John Wiley and Sons, Hoboken, New Jersey.

Selten, R. (1965), "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit, Teil I: Bestimmung des dynamischen Preisgleichgewichts," *Zeitschrift für die gesamte Staatswissenschaft, 121,* 301–324.

Shapley, L.S., (1953), "Stochastic Games," *Proc. Natl. Acad. Sci. USA 39,* 1095–1100.

Strotz, R.H. (1956), "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies, 23,* 165–180.

Voorneveld, M., W. Kets, and H. Norde (2005), "An Axiomatization of Minimal CURB Sets," *International Journal of Game Theory, 33,* 479–490.

Young, H.P. (1993), "The Evolution of Conventions," *Econometrica, 61,* 57–84.

Young, H.P. (1998), *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions,* Princeton University Press, Princeton, New Jersey.