MEAN-PAYOFF ADVERSARIAL STACKELBERG GAMES

Shibashis Guha

Tata Institute of Fundamental Research, India

Based on:

1. The Adversarial Stackelberg Value in Quantitative Games (ICALP' 20)
Emmanuel Filiot, Raffaella Gentilini, Jean-François Raskin

2. Fragility and Robustness in Mean-payoff Adversarial Stackelberg Games (CONCUR' 21)
Mrudula Balachander, Shibashis Guha, Jean-François Raskin

FM Update, July 10 2021

Stackelberg Games

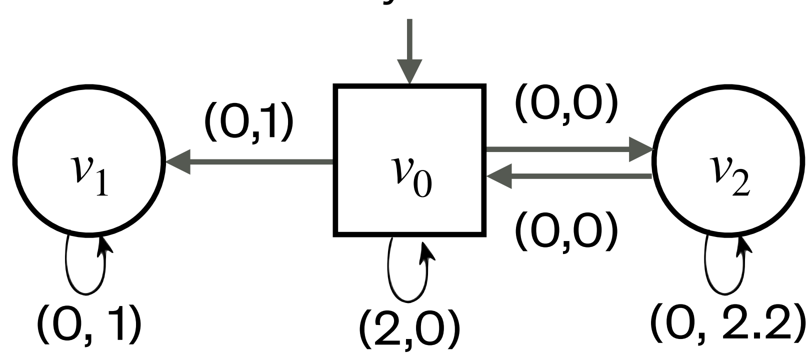
Two (types of) Players:

Leader   Follower

Both players are rational

Game:

Mean-Payoff Games



Bi-Matrix Games

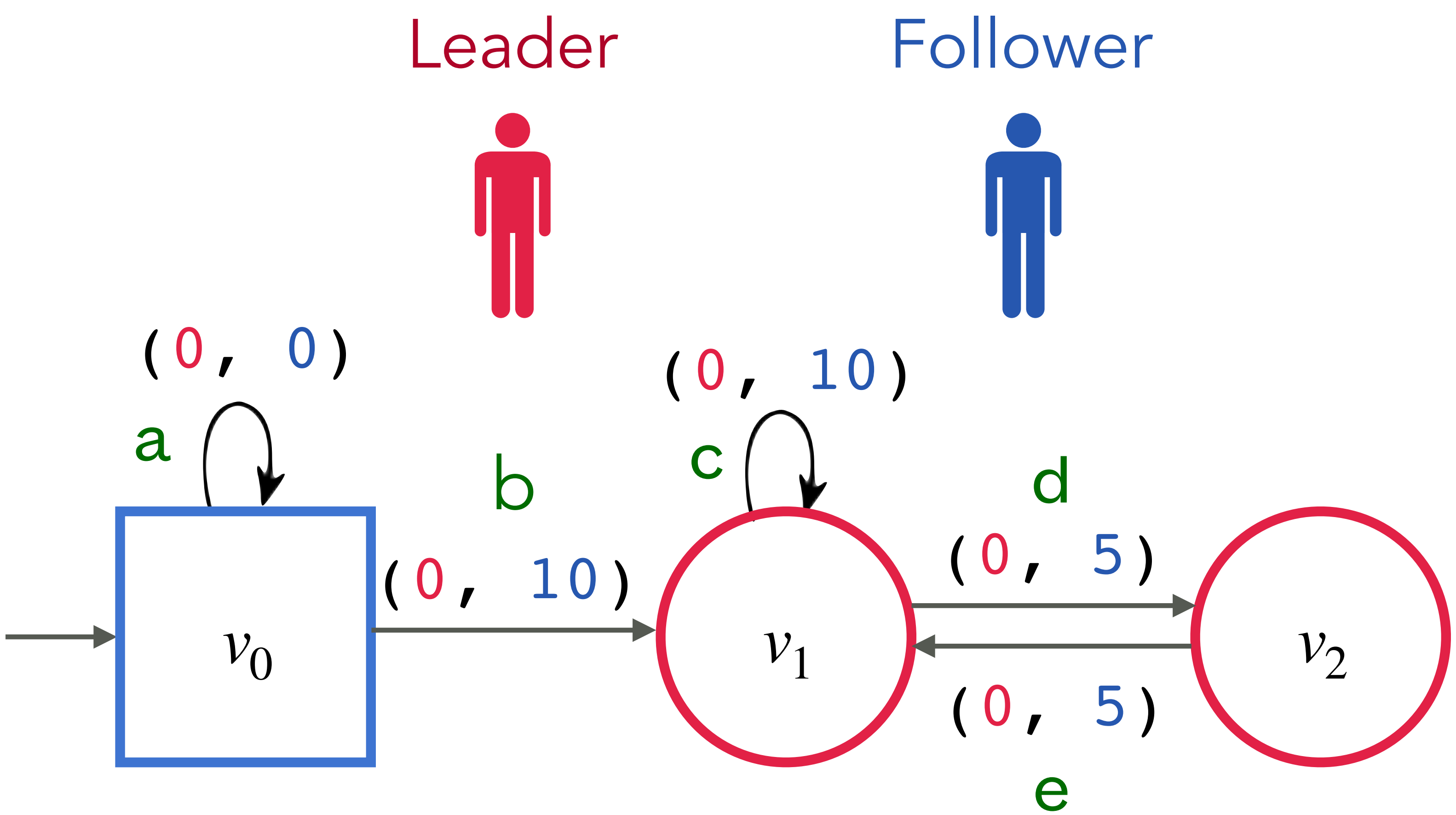
	I	II
I	(1,2)	(3,0)
II	(0,0)	(2,1)

Sequential Moves:

1. Leader announces her strategy

2. Follower announces his response to leader's strategy

Game Setting



Game played on a finite arena

Vertices partitioned into Leader (V_L) and Follower (V_F) vertices

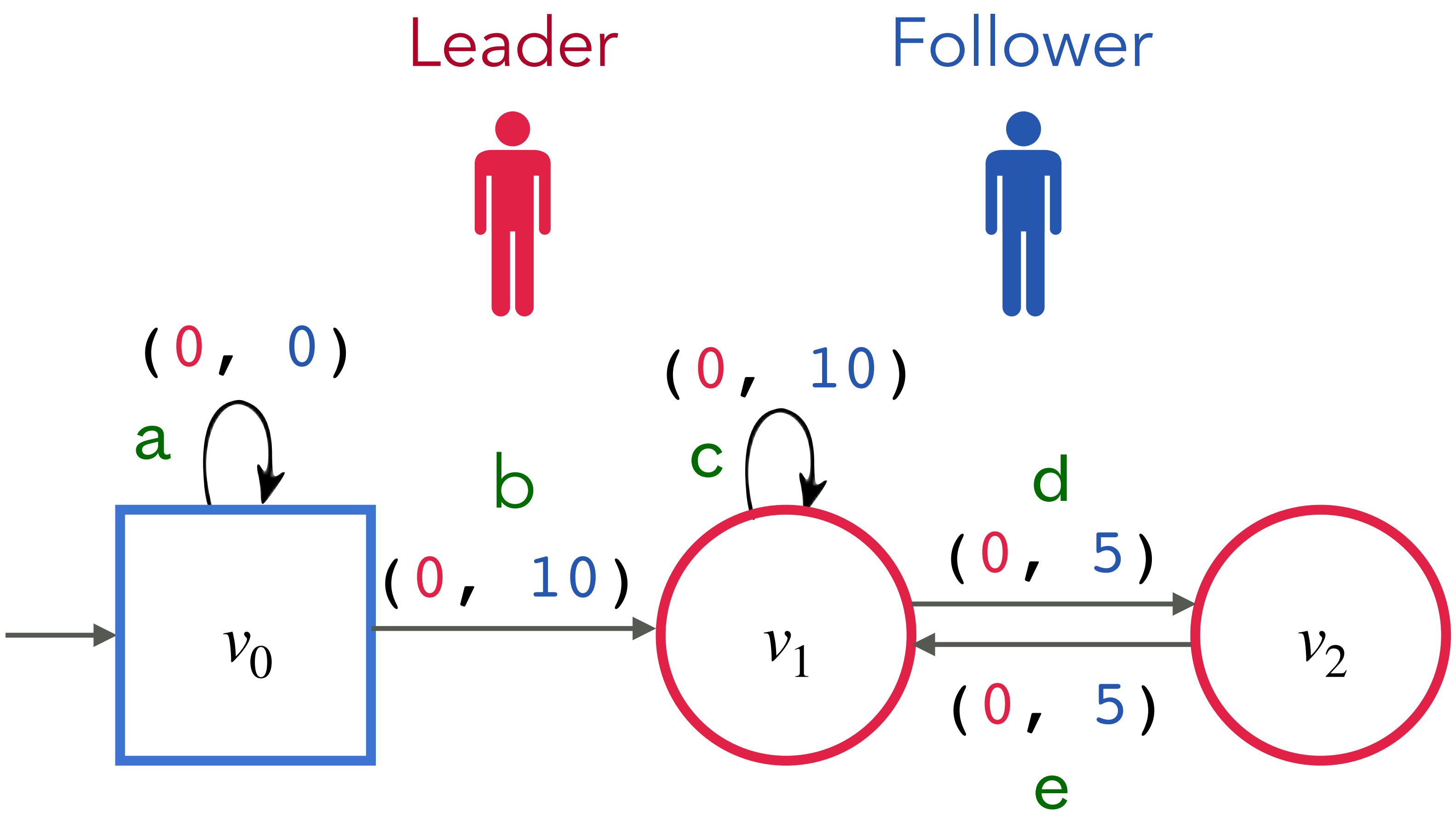
For infinitely many rounds producing an infinite path:

Player owning a vertex moves the token to the next vertex.

Payoffs along the path: $(c_1, d_1), (c_2, d_2), (c_3, d_3)\ldots$

Quantitative: mean-payoff a.k.a long-run average objective

Motivation



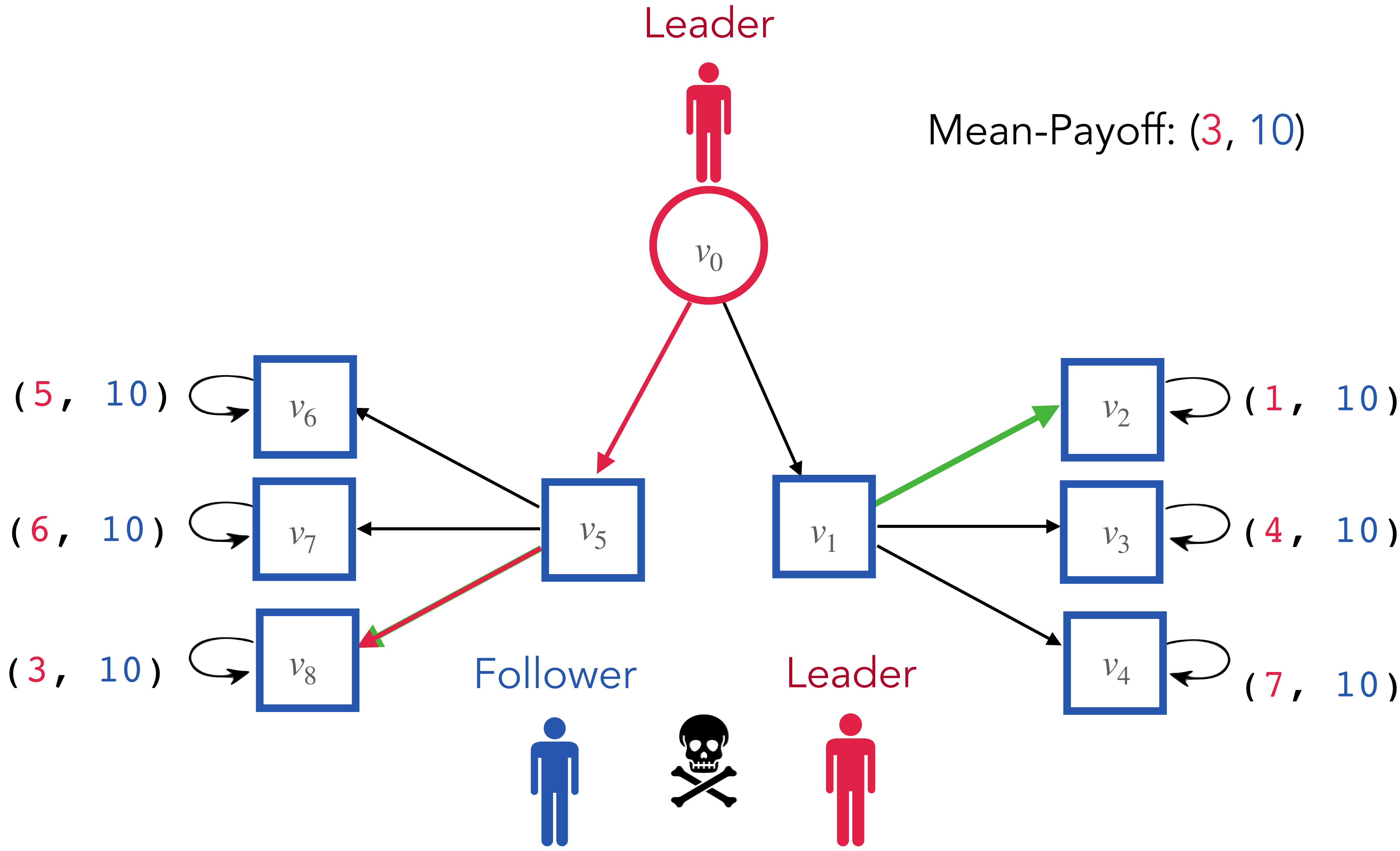
Contribution to rational synthesis framework

Nonzero-sum game where both the program and its environment are rational agents, with their own goals.

Have been studied for qualitative omega-regular objectives.

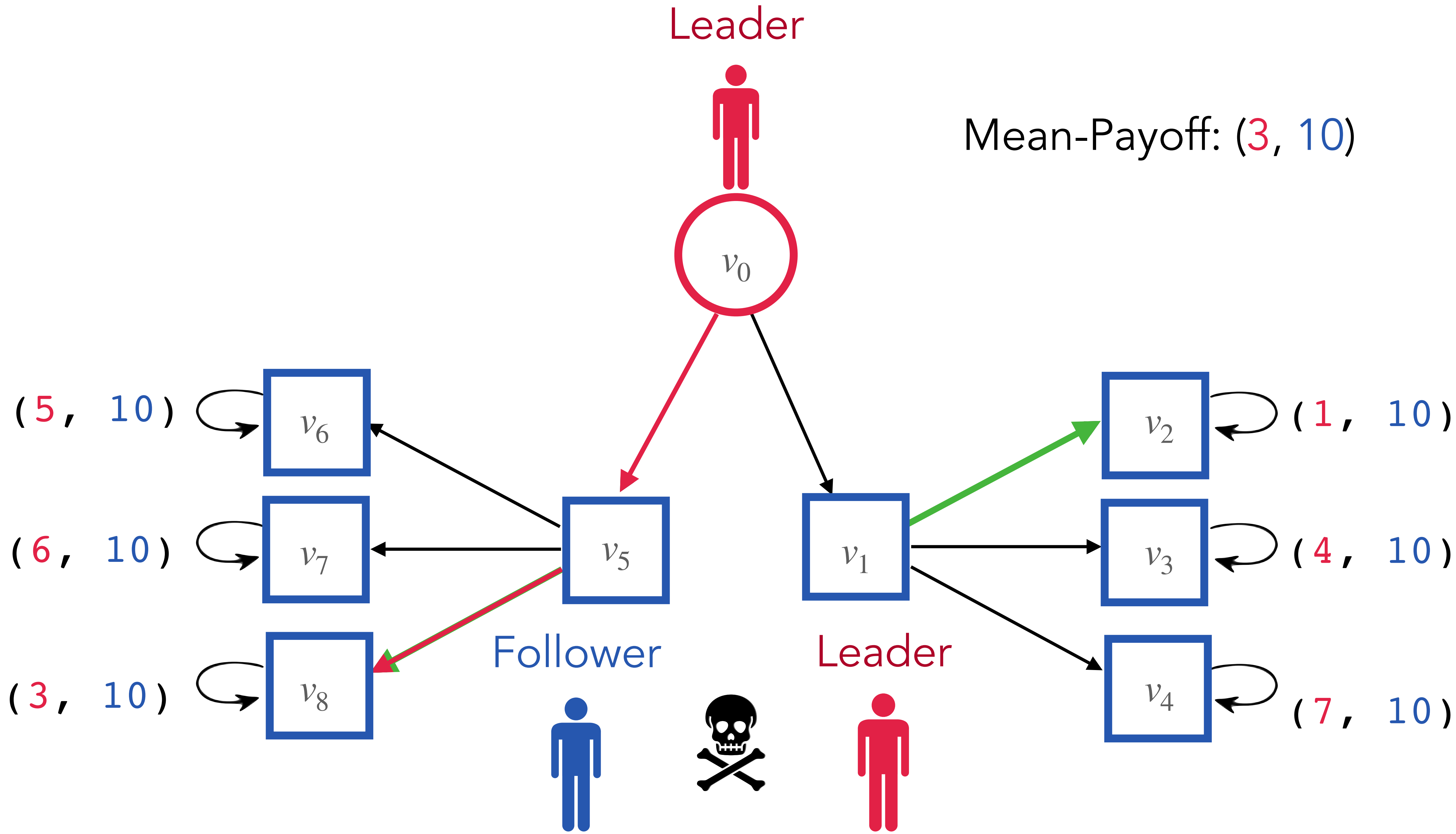
Fisman, Kupferman, Lustig' 10 and Kupferman, Perelli, Vardi' 16

Cooperative vs Adversarial



In the adversarial setting, Follower chooses Best-Response which `minimises` payoff of Leader

Cooperative vs Adversarial



In the *adversarial setting*, the program (Leader) assumes less hypothesis on the behaviour of the user (Follower).

Satisfies specification for all rational behaviour of the user (Follower).

Strategies

Leader strategies: Given a finite path ending in a Leader vertex, the choice of the successor vertex.

$$\sigma_0 : V^* V_L \longrightarrow V$$

Similarly, Follower strategies: …

$$\sigma_1 : V^* V_F \longrightarrow V$$

Choosing a strategy for Leader and a strategy for Follower leads to a unique infinite path in the game graph, called the *outcome* of the two strategies.

Memoryless strategies:

$$\sigma_0 : V_L \longrightarrow V$$

$$\sigma_1 : V_F \longrightarrow V$$

Adversarial Stackelberg Value ($\mathbf{ASV}$)

(Filiot, Gentilini and Raskin - ICALP 2020)

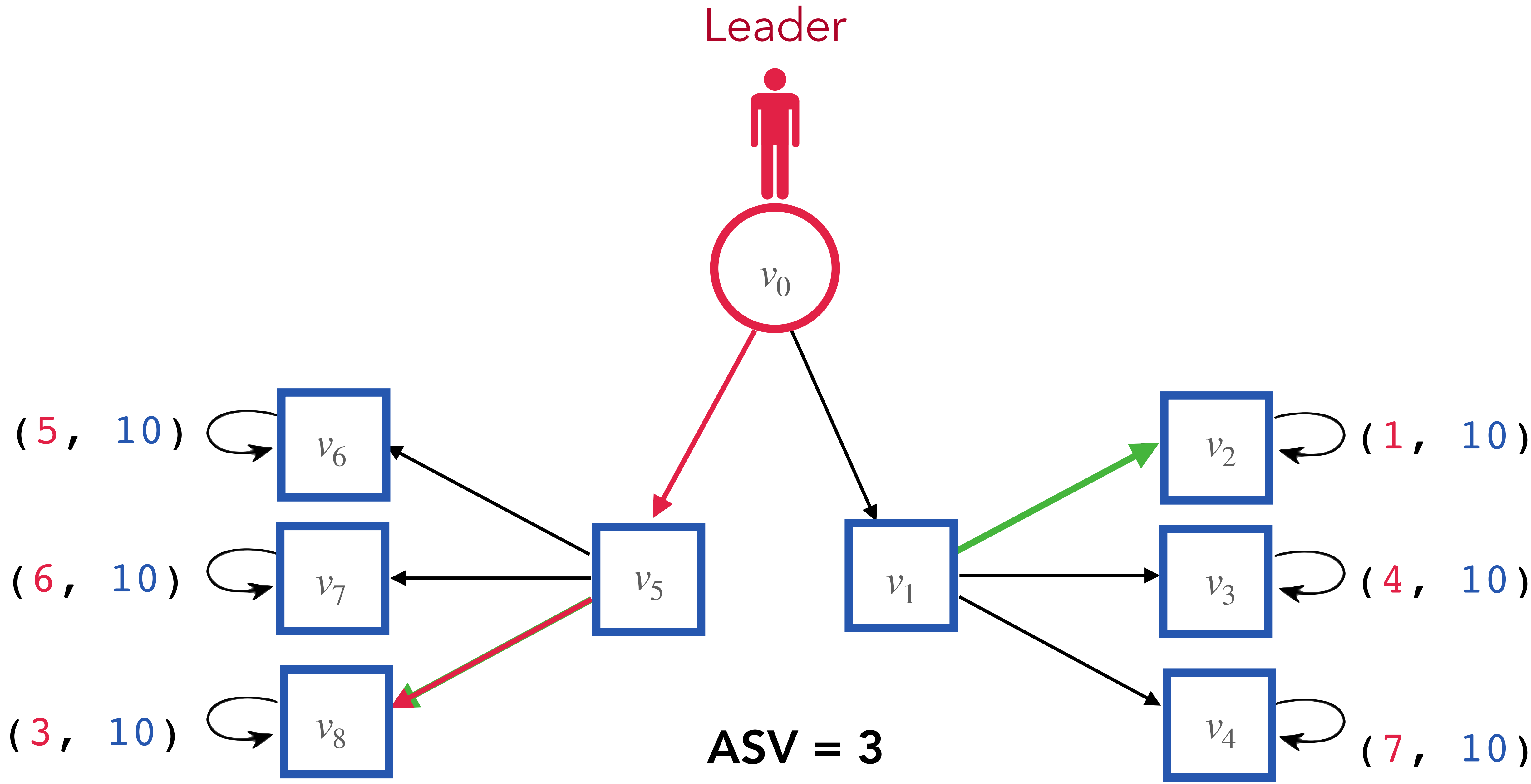
$\mathbf{ASV}$ is the largest mean-payoff value Leader can obtain when Follower plays an **adversarial** best response.

$$\mathbf{ASV}(\sigma_0)(v) = \inf_{\sigma_1 \in \mathbf{BR}(\sigma_0)} \text{Mean-Payoff}_L\,[\text{Outcome}(\sigma_0, \sigma_1)]$$

$$\mathbf{ASV}(v) = \sup_{\sigma_0} \mathbf{ASV}(\sigma_0)(v)$$

σ_0 : Leader Strategy σ_1 : Follower Strategy

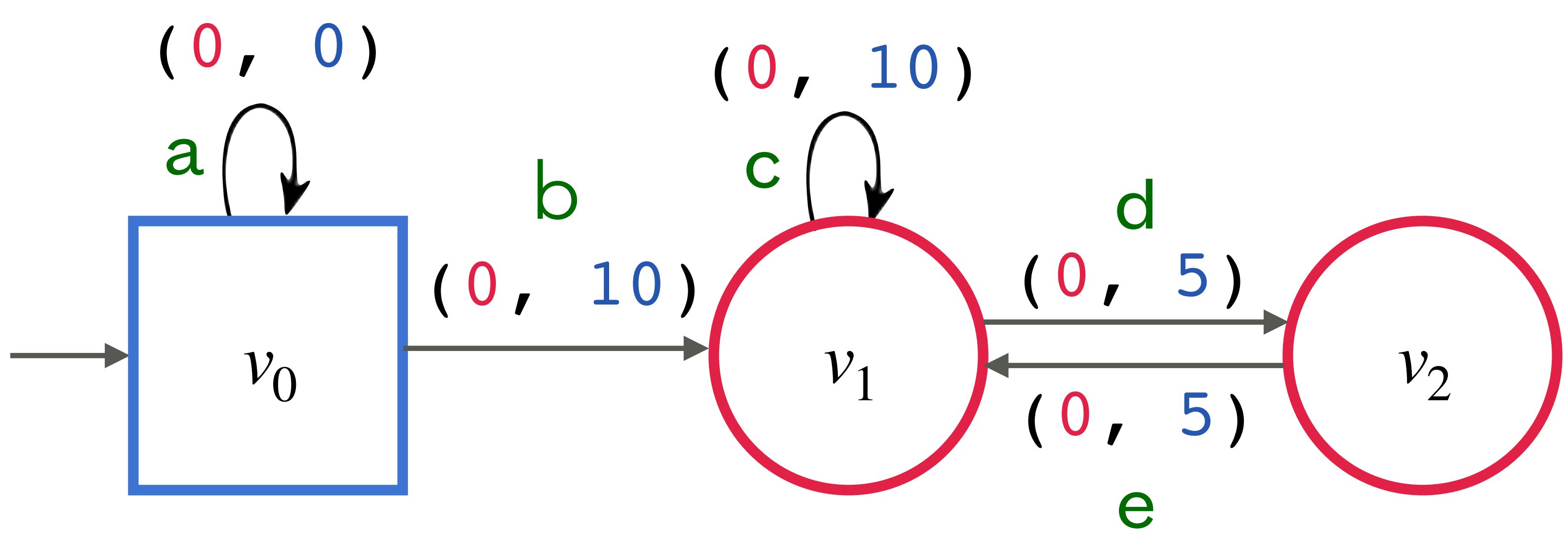
Adversarial Stackelberg Value (ASV)



$$\mathbf{ASV}(\sigma_0)(v) = \inf_{\sigma_1 \in \mathbf{BR}(\sigma_0)} \text{Mean-Payoff}\ [\text{Outcome}(\sigma_0, \sigma_1)]$$

$$\mathbf{ASV}(v) = \sup_{\sigma_0} \mathbf{ASV}(\sigma_0)(v)$$

Best Responses May Not Exist



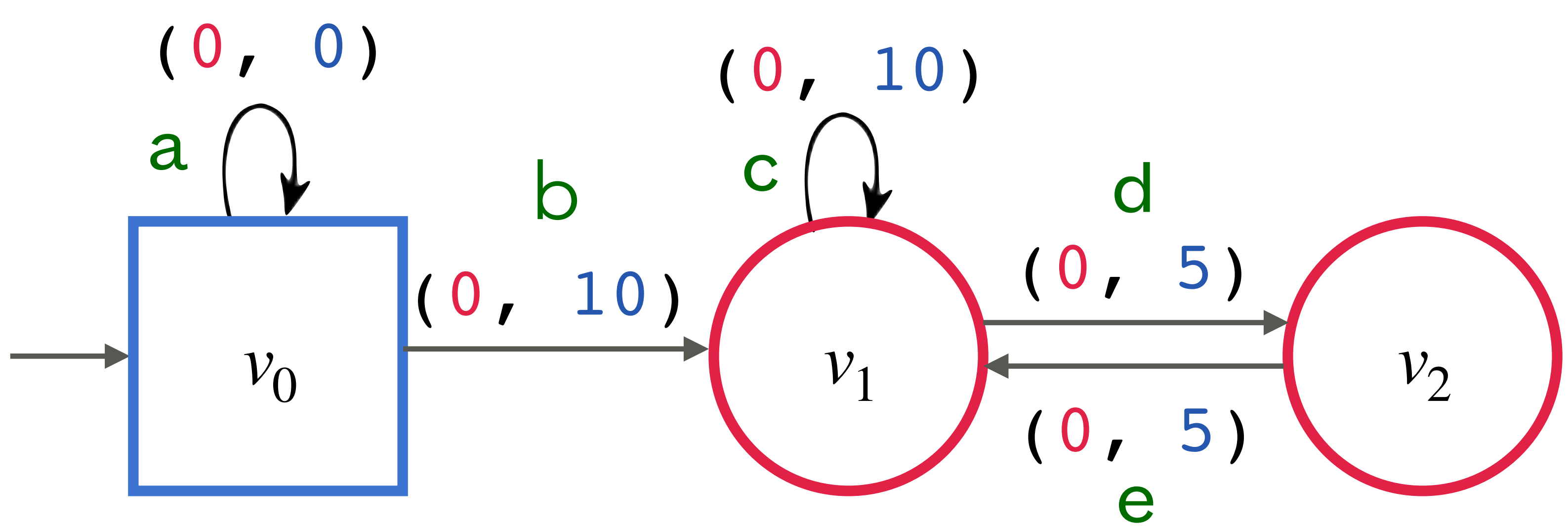
Leader strategy: If $a^k b$, then $(c^k de)^\omega$

Follower strategy: If $a^{1000} b$, then $(c^{1000} de)^\omega$

If $a^{100000} b$, then $(c^{100000} de)^\omega$

If $a^\infty b$, then the vertex v_1 is never reached.

Epsilon-Best Responses Always Exist

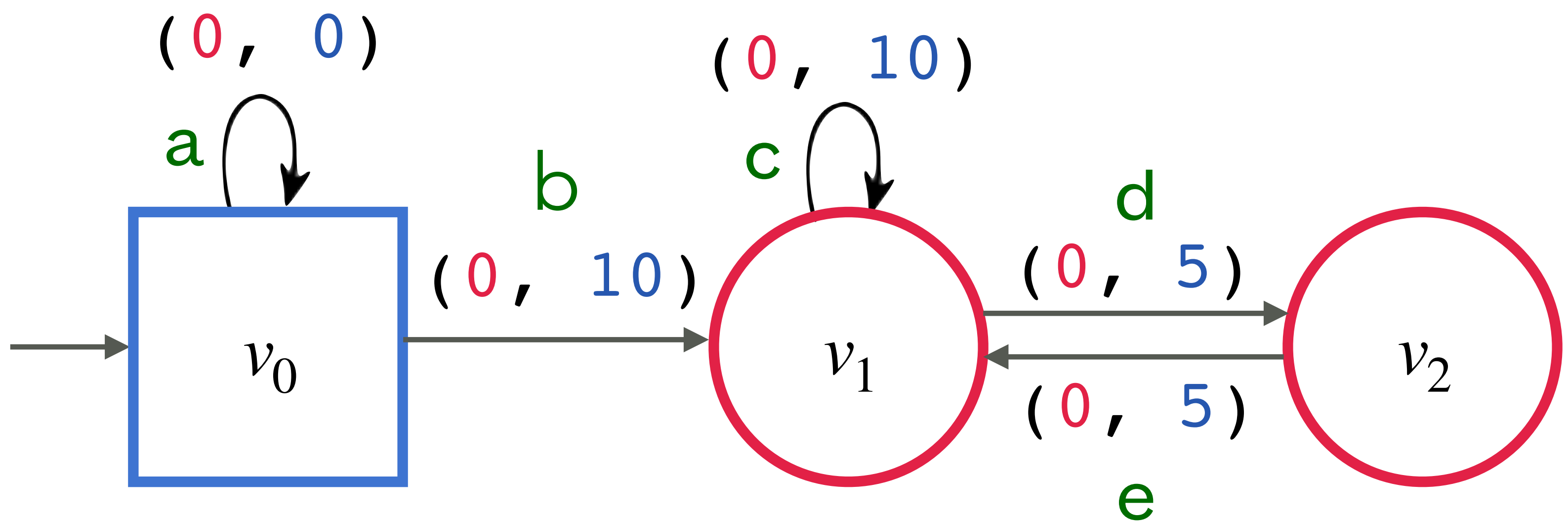


Leader strategy: If $a^k b$, then $(c^k de)^\omega$

Follower strategy: For $\epsilon = 0.1$, play $a^{1000}b$

For $\epsilon = 0.001$, play $a^{100000}b$

Best Responses May Not Exist



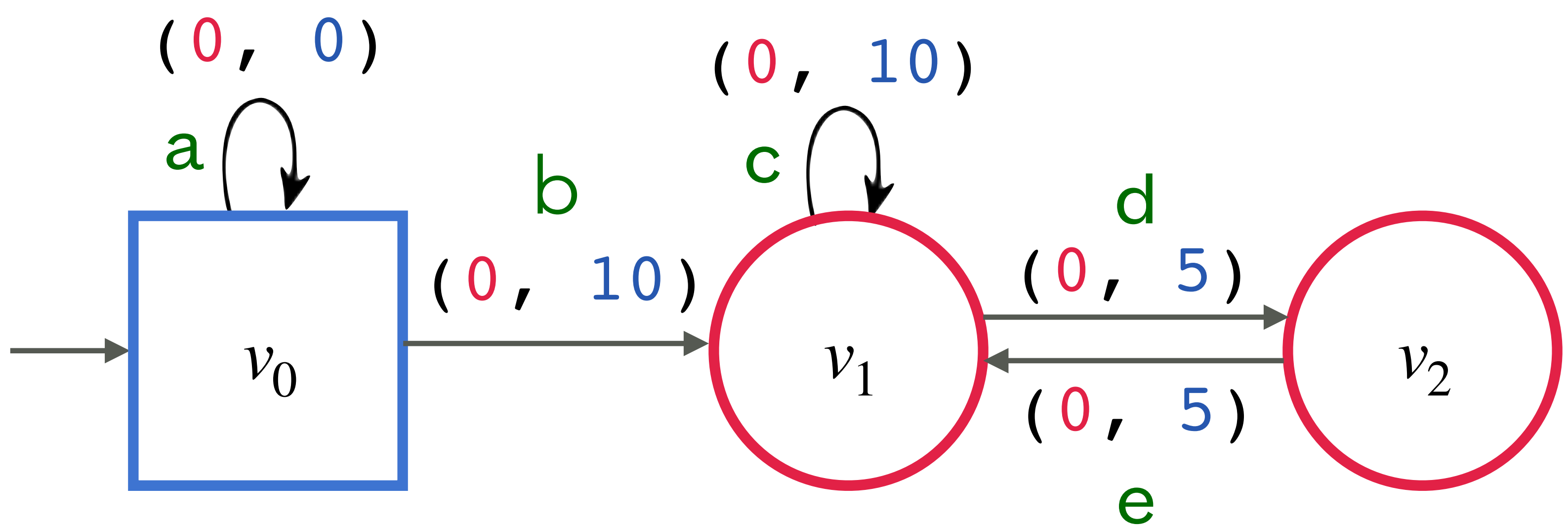
Leader strategy: If $a^k b$, then $(c^k de)^\omega$

This set may be empty

$$\mathbf{ASV}(\sigma_0)(v) = \inf_{\sigma_1 \in \mathbf{BR}(\sigma_0)} \text{Mean-Payoff}_L\,[\text{Outcome}(\sigma_0, \sigma_1)]$$

$$\mathbf{ASV}(v) = \sup_{\sigma_0} \mathbf{ASV}(\sigma_0)(v)$$

Epsilon-Best Responses Always Exist



Leader strategy: If $a^k b$, then $(c^k de)^\omega$

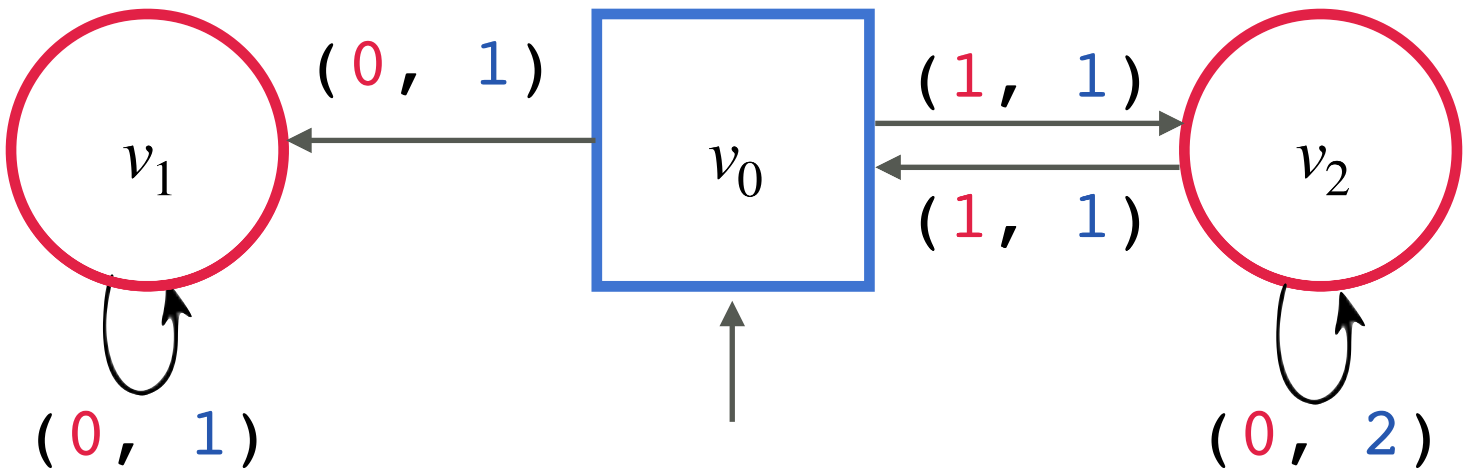
$$\mathbf{ASV}(\sigma_0)(v) = \sup_{\epsilon > 0} \inf_{\sigma_1 \in \mathbf{BR}^\epsilon(\sigma_0)} \text{Mean-Payoff}_L\,[\text{Outcome}(\sigma_0, \sigma_1)]$$

$$\mathbf{ASV}(v) = \sup_{\sigma_0} \mathbf{ASV}(\sigma_0)(v)$$

An ϵ-best response of Follower to a Leader strategy is one which is at most ϵ worse than every other response of Follower.

ASV May Not Be Achievable

$$\mathbf{ASV}(v_0) = 1$$



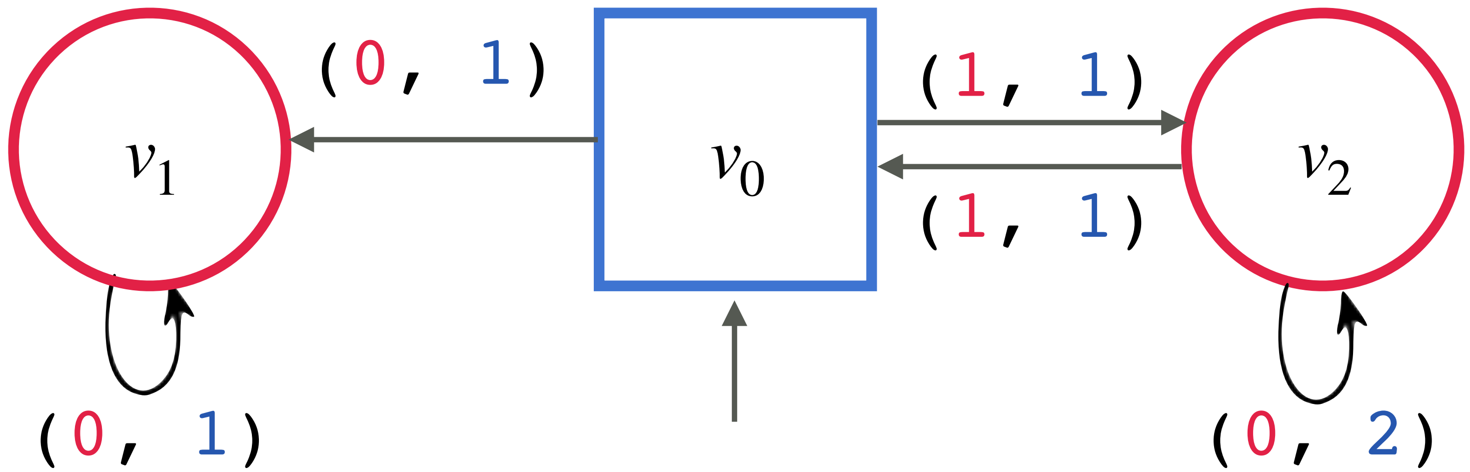
Follower must be given mean-payoff > 1
else he will play $v_0 \rightarrow v_1$

No strategy for Leader to get a mean-payoff of 1 for herself.

Threshold Problem:

Is $\mathbf{ASV}$ > c for some threshold c?

Threshold problem and witness

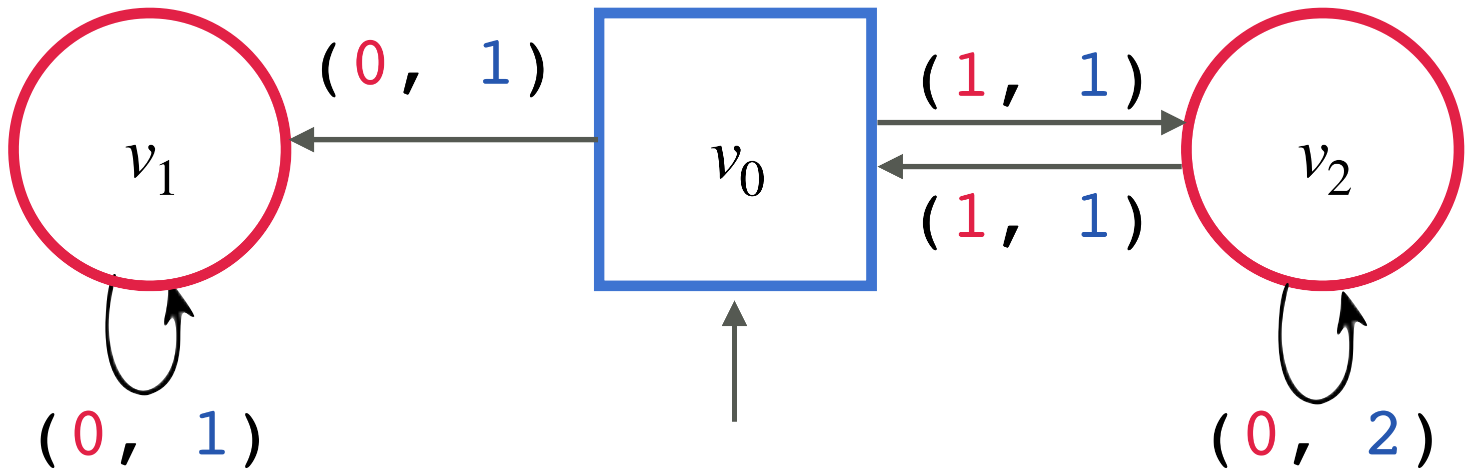


A path π is a witness for $\mathbf{ASV} > c$ if

(i) mean-Payoff of π is (c', d), where $c' > c$ and

(ii) …

Threshold problem and witness

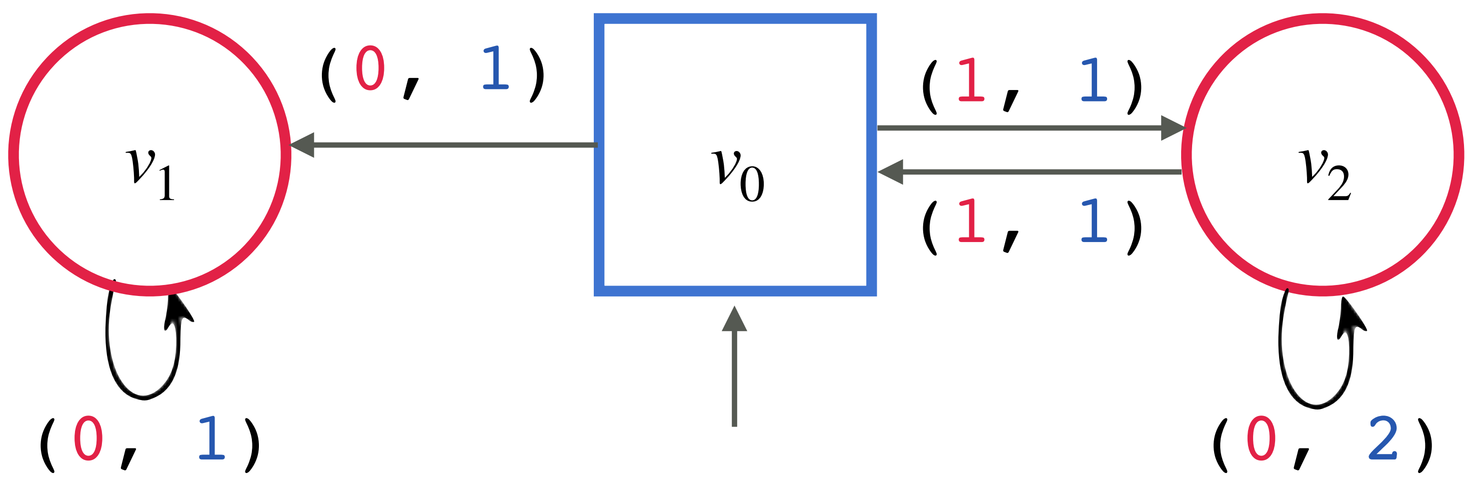


A path π is a witness for $\mathbf{ASV} > c$ if

(i) mean-Payoff of π is (c', d), where $c' > c$ and

(ii) and π does not cross a (c, d)-bad vertex.

Threshold problem and witness



A path π is a witness for $\mathbf{ASV} >$ c if

(i) mean-Payoff of π is (c′, d), where c′ > c and

(ii) and π does not cross a (c, d)-bad vertex.

$\Lambda(v) =$ { (c, d) $\in \mathbb{R}^2$ | From vertex v, Follower can ensure that Leader's payoff $\leqslant$ c and Follower's payoff $\geqslant$ d }

A vertex v is (c, d)-bad if (c, d) $\in \Lambda(v)$

Threshold problem and witness

A path π is a witness for $\mathbf{ASV}(v) > c$ if

(i) mean-Payoff of π is (c', d), where $c' > c$ and

(ii) and π does not cross a (c, d)-bad vertex.

$$\Lambda(v) = \left\{ (c, d) \in \mathbb{R}^2 \mid \text{From vertex } v\text{, Follower can ensure that Leader's payoff} \leqslant c \text{ and Follower's payoff} \geqslant d \right\}$$

A vertex v is (c, d)-bad if $(c, d) \in \Lambda(v)$

Theorem: ASV(v) > c if and only if there exists a *witness* for ASV(v) > c.

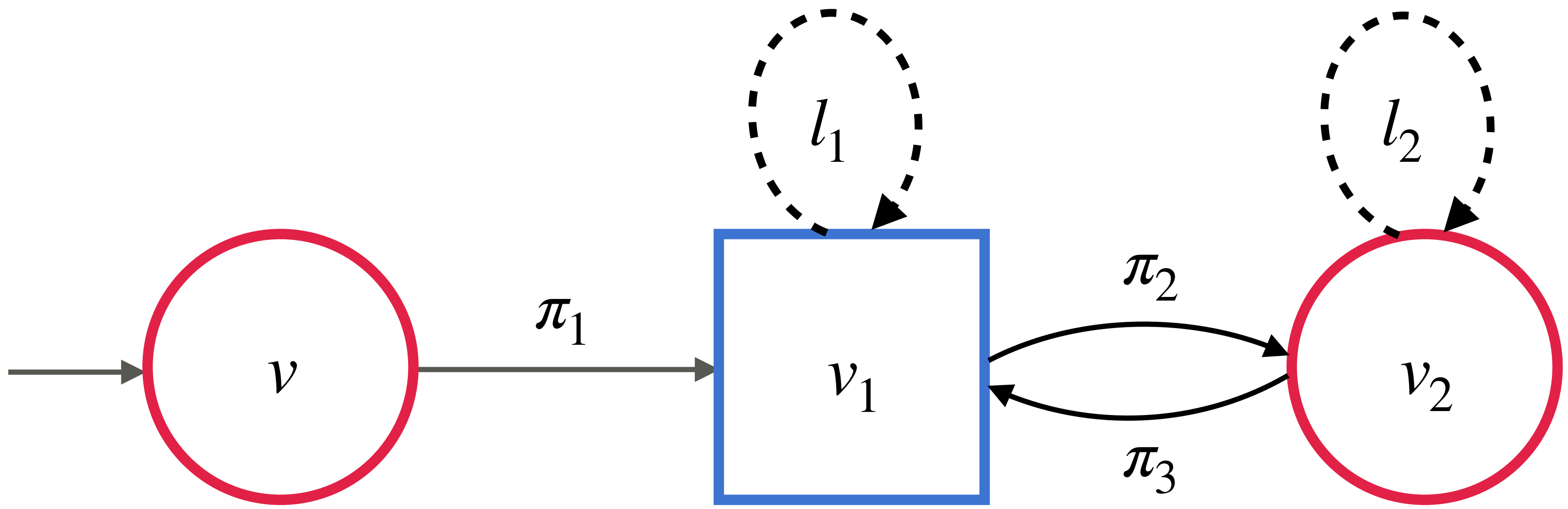
Threshold problem and witness

Theorem: ASV(v) > c if and only if there exists a *witness* for ASV(v) > c.

If **ASV**(v) > c, we can find a *regular* witness of the form

$$\pi_1 \cdot (l_1^{[\alpha.k]} \cdot \pi_2 \cdot l_2^{[\beta.k]} \cdot \pi_3)^\omega$$

$$u \cdot v^\omega$$



l_1 and l_2 are simple cycles,

π_1, π_2 and π_3 are finite acyclic plays

Threshold problem: NP membership

Theorem: ASV(v) > c if and only if there exists a *witness* for ASV(v) > c.

If $\mathbf{ASV}$(v) > c, we can find a *regular* witness of the form

$$\pi_1 \cdot (l_1^{[\alpha.k]} \cdot \pi_2 \cdot l_2^{[\beta.k]} \cdot \pi_3)^\omega$$

A path π is a witness for $\mathbf{ASV}$(v) > c if

(i) mean-Payoff of π is (c′, d), where c′ > c and

(ii) and π does not cross a (c, d)-bad vertex.

$$\Lambda(v) = \left\{ (c, d) \in \mathbb{R}^2 \mid \begin{array}{l}\text{From vertex } v\text{, Follower can ensure that Leader's} \\ \text{payoff} \leqslant c \text{ and Follower's payoff} \geqslant d\end{array} \right\}$$

Leads to an NP-membership for the threshold problem.

Adversarial Stackelberg Value for memoryless strategies ($\mathbf{ASV}_{\mathsf{ML}}$)

$$\mathbf{ASV}(\sigma_0)(v) = \inf_{\sigma_1 \in \mathbf{BR}(\sigma_0)} \text{Mean-Payoff}\ [\text{Outcome}(\sigma_0, \sigma_1)]$$

$$\mathbf{ASV}_{\mathsf{ML}}(v) = \sup_{\sigma_0 \in \Sigma_0^{\mathsf{ML}}} \mathbf{ASV}(\sigma_0)(v)$$

Deciding if $\mathbf{ASV}_{\mathsf{ML}}(v) > c$ is NP-complete.

Memoryless strategies:

$\sigma_0 : V_L \longrightarrow V$

$\sigma_1 : V_F \longrightarrow V$

Computing the ASV

Computing the $\mathbf{ASV}$: Using FO-Theory over Reals with Addition

Uses the notion of a witness

$$\mathbf{ASV}(v) = \sup\{c \,|\, \text{There is a -witness } \pi \text{ for } \mathbf{ASV}(v) > c\}$$

$$= \max_{S \in SCC} \sup\{c \,|\, \text{There is a -witness } \pi \text{ for } \mathbf{ASV}(v) > c \text{ and } \pi \text{ ends in } S\}.$$

Shows that there exist plays with mean-payoff (x, y) in the SCC S.

(Chatterjee et al. '10)

$$\rho(c) = \exists x, y : x > c \wedge \Phi_S(x, y) \wedge \neg \Psi_S(c, y)$$

Shows that the play does not cross a (c, y)- bad vertex in S.

(Brenguier, Raskin '15)

Computing the $\mathbf{ASV}$: Using FO-Theory over Reals with Addition

$$\mathbf{ASV}(v) = \sup\{c \mid \text{There is a -witness } \pi \text{ for } \mathbf{ASV}(v) > c\}$$

$$= \max_{S \in SCC} \sup\{c \mid \text{There is a -witness } \pi \text{ for } \mathbf{ASV}(v) > c \text{ and } \pi \text{ ends in } S\}.$$

Shows that there exist plays with mean-payoff (x, y).

$$\rho_S(c) = \exists x, y : x > c \wedge \Phi_S(x, y) \wedge \neg \Psi_S(c, y)$$

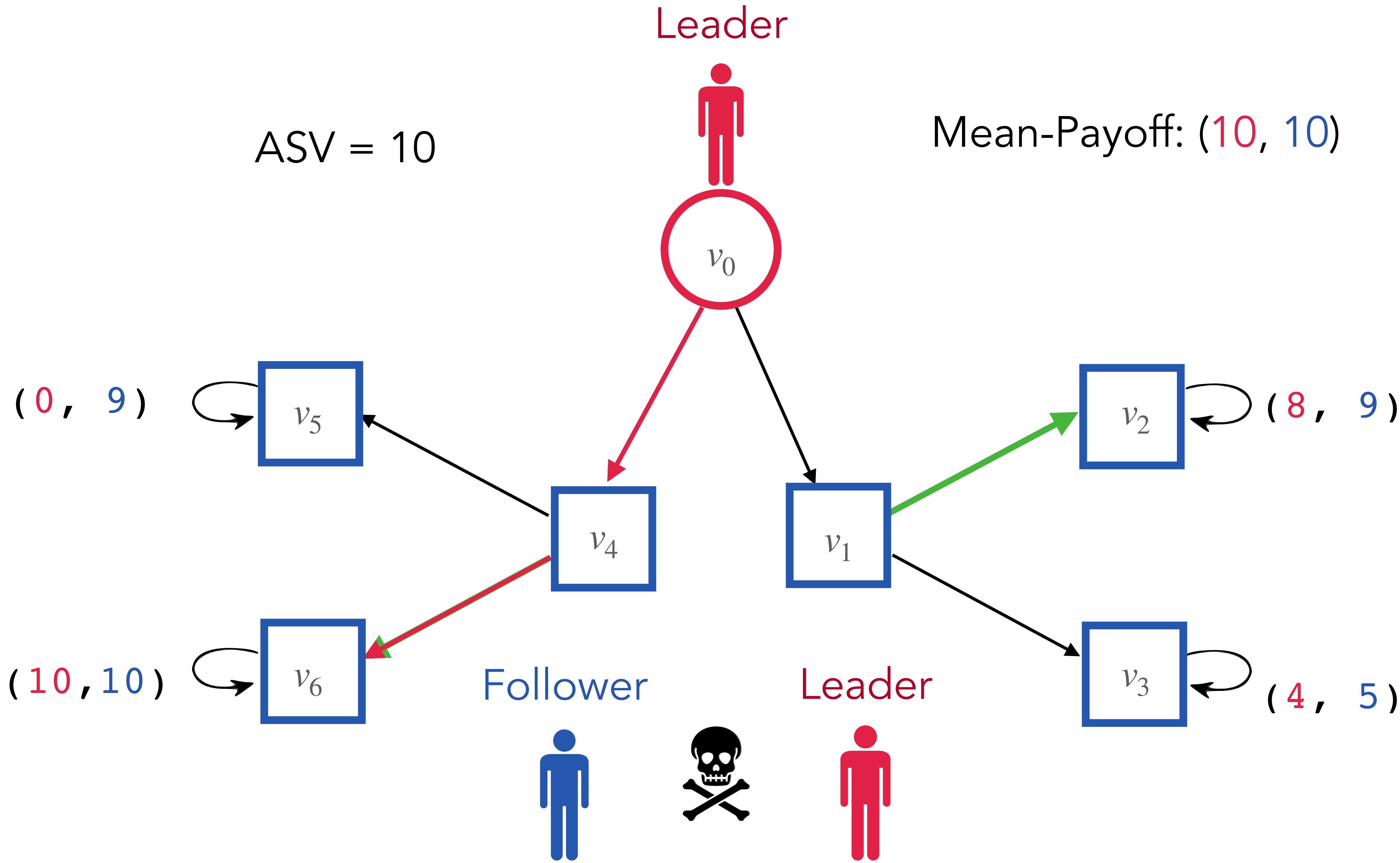
Shows that the play does not cross a (c, y)- bad vertex.

We can also express $\rho_S(c)$ as a set of linear programs.

In the linear program, we maximise **c**. (Gives an EXPTime algorithm.)

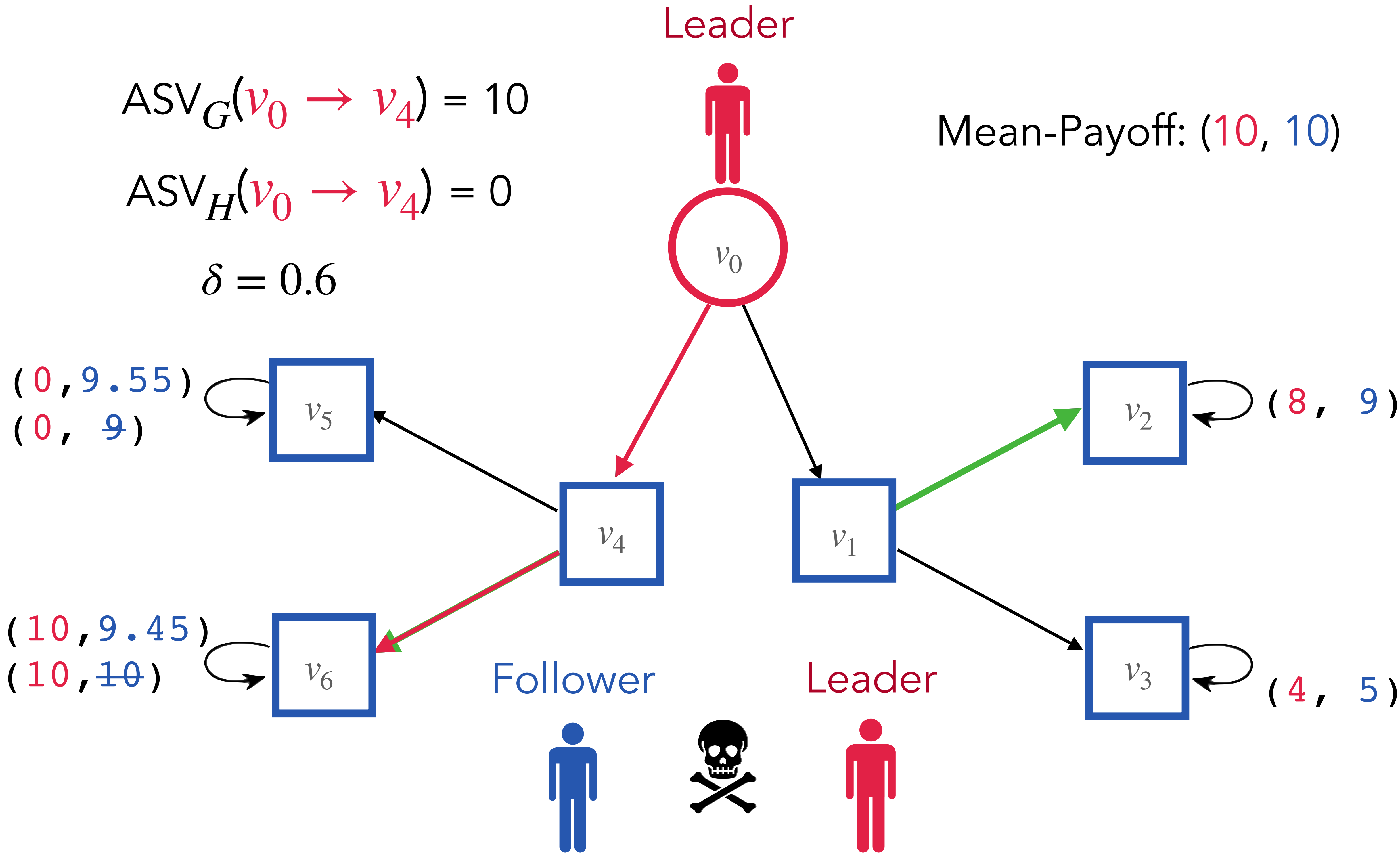
Fragility of ASV

Robustness: ASV under perturbation of weights



In the adversarial setting, Follower chooses Best-Response which `minimises` payoff of Leader

Robustness: ASV under perturbation of weights



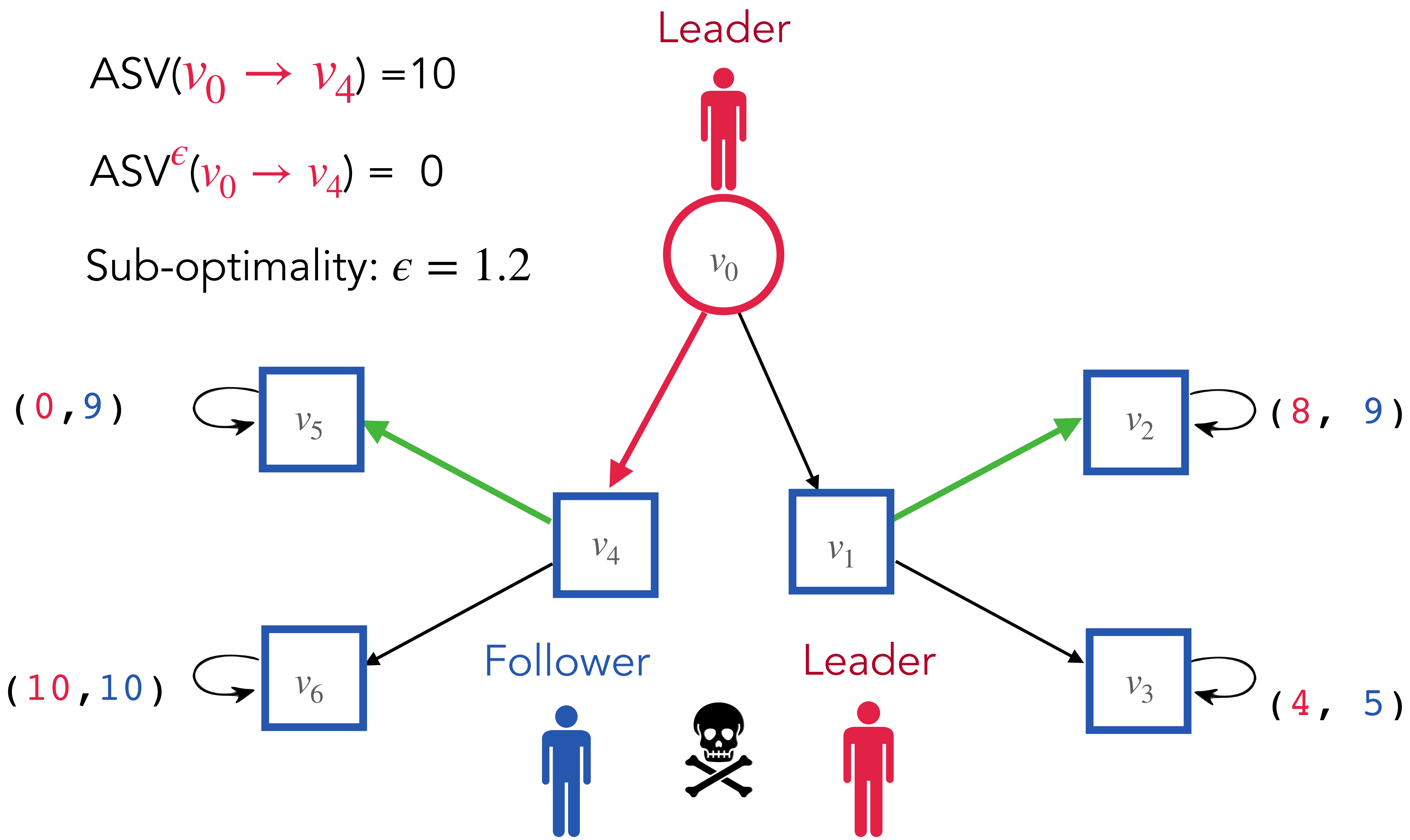
In the adversarial setting, Follower chooses Best-Response which `minimises` payoff of Leader

Robustness against sub-optimal responses

$\mathrm{ASV}(v_0 \rightarrow v_4) = 10$

$\mathrm{ASV}^{\epsilon}(v_0 \rightarrow v_4) = 0$

Sub-optimality: $\epsilon = 1.2$

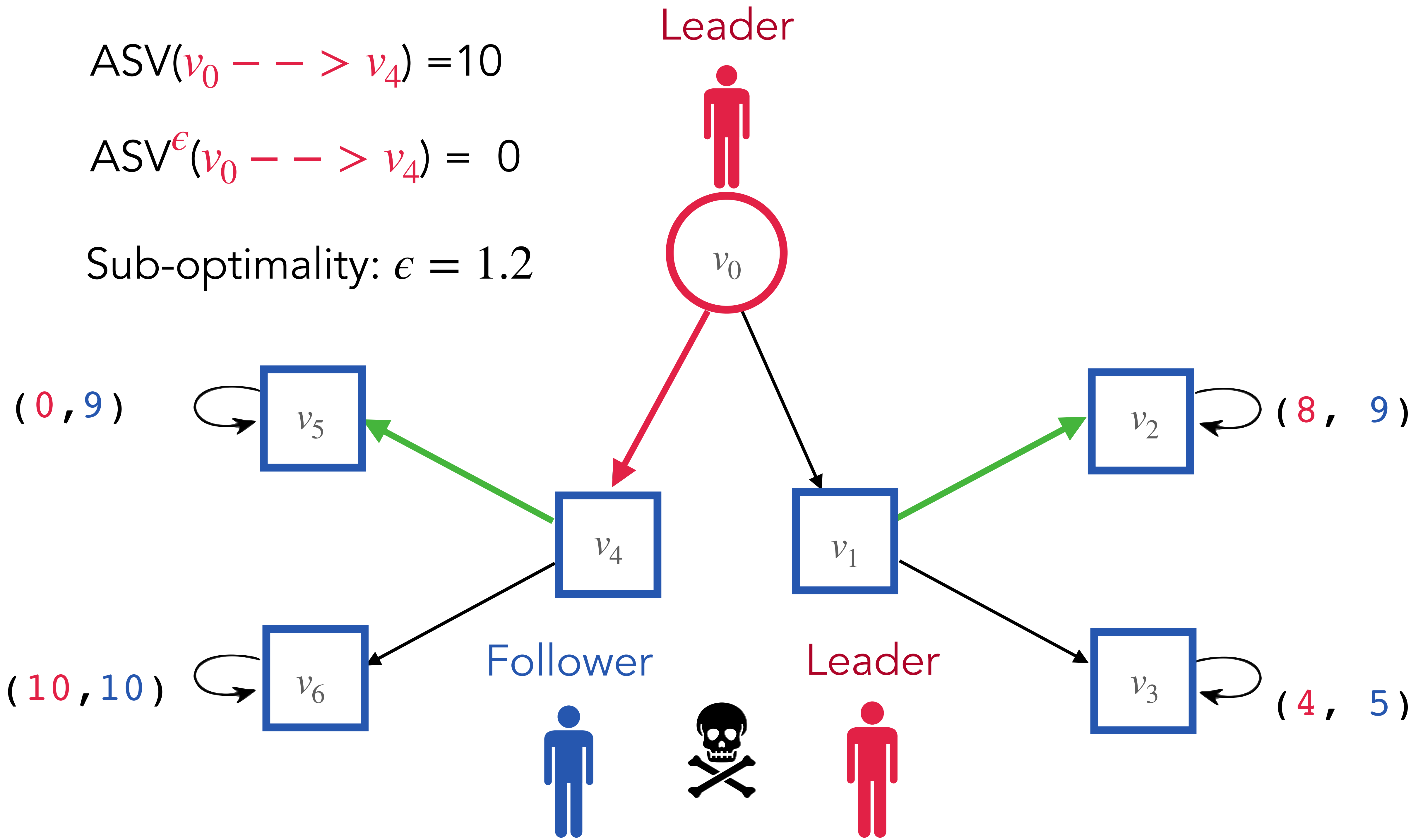


Robustness in zero-sum games

$\mathrm{ASV}(v_0 - - > v_4) = 10$

$\mathrm{ASV}^{\epsilon}(v_0 - - > v_4) = 0$

Sub-optimality: $\epsilon = 1.2$



Strategy in zero-sum games are robust against perturbations and ϵ-best responses of Follower.

Robustness against perturbation

We suggest the solution concept ASV^{ϵ} instead of ASV.

$$\mathbf{ASV}^{\epsilon}(\sigma_0)(v) = \inf_{\sigma_1 \in \mathbf{BR}^{\epsilon}(\sigma_0)} \text{Mean-Payoff}_L\left[\text{Outcome}(\sigma_0, \sigma_1)\right]$$

$$\mathbf{ASV}^{\epsilon}(v) = \sup_{\sigma_0} \mathbf{ASV}^{\epsilon}(\sigma_0)(v)$$

Robustness against sub-optimal responses ($\mathsf{ASV}_G^{2\delta}$) implies robustness against perturbation of δ.

Theorem: $\forall H \, \forall \sigma_0 \in G^{\pm\delta} \; : \mathsf{ASV}_H(\sigma_0)(v) > \mathsf{ASV}_G^{2\delta}(\sigma_0)(v) - \delta$

Combined robustness of ASV$^{\epsilon}$

Combined robustness against perturbation of δ, and sub-optimal response of Player 1.

$$\forall H \in G^{\pm\delta} \; : \mathsf{ASV}_{H}^{\epsilon}(v) > \mathsf{ASV}_{G}^{2\delta+\epsilon}(v) - \delta$$

While adversarial Stackelberg value (ASV) is fragile against perturbation and suboptimal responses of Player 1, the ASV$^{\epsilon}$ is robust against both.

Combined robustness of ASV$^\epsilon$

Combined robustness against perturbation of δ, and sub-optimal response of Player 1.

$$\forall H \in G^{\pm\delta} \ : \mathsf{ASV}^{\epsilon}_{H}(v) > \mathsf{ASV}^{2\delta+\epsilon}_{G}(v) - \delta$$

While adversarial Stackelberg value (ASV) is fragile against perturbation and suboptimal responses of Player 1, the ASV$^\epsilon$ is robust against both.

Given a threshold c, we can compute in EXPTime the largest ϵ such that ASV$^\epsilon > c$.

ASV$^\epsilon$ is achievable unlike ASV.