

On the Rationality of Escalation

Pierre Lescanne, Perrinel Matthieu

▶ **To cite this version:**

Pierre Lescanne, Perrinel Matthieu. On the Rationality of Escalation. 2009. ensl-00439911v1

HAL Id: ensl-00439911

https://ens-lyon.hal.science/ensl-00439911v1

Preprint submitted on 8 Dec 2009 (v1), last revised 25 Apr 2010 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the Rationality of Escalation

Pierre Lescanne and Matthieu Perrinel
University of Lyon, ENS de Lyon,
CNRS (LIP), U. Claude Bernard de Lyon
46 allée d'Italie, 69364 Lyon, France

December 8, 2009

Abstract

Escalation is a typical feature of infinite games. Therefore tools conceived for studying infinite mathematical structures, namely those deriving from *coinduction* are essential. Here we use coinduction, or backward coinduction (to show its connection with the same concept for finite games) to study carefully and formally the infinite games especially those called *dollar auctions*, which are considered as the paradigm of escalation. Unlike what is commonly admitted, we show that, provided one assumes that the other agent will always stop, bidding is rational, because it results in a subgame perfect equilibrium. We show that this is not the only rational strategy profile (the only subgame perfect equilibrium). Indeed if an agent stops and will stop at every step, we claim that he is rational as well, if one admits that his opponent will never stop, because this corresponds to a subgame perfect equilibrium. Amazingly, in the infinite dollar auction game, the behavior in which both agents stop at each step is not a Nash equilibrium, hence is not a subgame perfect equilibrium, hence is not rational.

Keyword: escalation, rationality, extensive form, backward induction.
JEL Code: C72, D44, D74.

1 Introduction

Escalation takes place in specific sequential games in which players continue although their payoff decreases on the whole. The *dollar auction* game has been presented by Shubik [1971] as the paradigm of escalation. He noted that, even though their cost (the opposite of the payoff) basically increases, players may keep bidding. This attitude is considered as inadequate and when talking about escalation, Shubik [1971] says this is a paradox, O'Neill [1986] and Leininger [1989] consider the bidders as irrational, Gintis [2000] speaks of *illogic conflict of escalation* and Colman [1999] calls it *Macbeth effect* after Shakespeare's play. In contrast with these authors, in this paper, we prove that escalation is logic

and that agents are rational, therefore this is not a paradox and we are led to assert that Macbeth is somewhat rational.

This escalation phenomenon occurs in infinite sequential games and only there. Therefore it must be studied with adequate tools, i.e., in a framework designed for mathematical infinite structures. Like Shubik [1971] we limit ourselves to two players only. In auctions, this consists in the two players bidding forever. This statement is based on the common assumption that a player is rational if he adopts a strategy which corresponds to a *subgame perfect equilibrium*. To characterize this equilibrium the above cited authors consider a finite restriction of the game for which they compute the subgame perfect equilibrium by *backward induction*[1]. In practice, they add a new hypothesis on the amount of money the bidders are ready to pay, also called the limited bankroll. In the amputated game, they conclude that there is a unique subgame perfect equilibrium. This consists in both agents giving up immediately, not starting the auction and adopting the same choice at each step. In our formalization in infinite games, we show that extending that case up to infinity is not a Nash equilibrium, hence not a subgame perfect equilibrium and we found two subgame perfect equilibria, namely the cases when one agent continues at each step and the other leaves at each step. Those equilibria which correspond to rational attitudes account for the phenomenon of escalation.

The origin of the misconception that concludes the irrationality of escalation is the belief that properties of infinite mathematical objects can be extrapolated from properties of finite objects. This does not work. As Fagin [1993] recalls, "most of the classical theorems of logic [for infinite structures] fail for finite structures" (see Ebbinghaus and Flum [1995] for a full development of the finite model theory). The reciprocal holds obviously "most of the results which hold for finite structures, fail for infinite structures". This has been beautifully evidenced in mathematics, when Weierstrass [1872] has exhibited his function:

$$f(x) = \sum_{n=0}^{\infty} b^n \cos(a^n x \pi).$$

Every finite sum is differentiable and the limit, i.e., the *infinite* sum, is not. To give another picture, infinite games are to finite games what fractal curves are to smooth curves [Edgar, 2008]. In game theory the error done by the nineteenth century mathematicians (Weierstrass quotes Cauchy, Dirichlet and Gauss) would lead to the same issue: wrong assumptions. With what we are concerned, a result that holds on finite games does not hold necessarily on infinite games and vice-versa. More specifically equilibria on finite games are not preserved at the limit on infinite games. In particular, we cannot conclude that, whereas the only rational attitude in finite dollar auction would be to stop immediately, it is irrational to escalate in the case of an infinite auction. We have to keep in mind that in the case of escalation, the game is infinite, therefore reasoning made for finite objects are inappropriate and tools specifically conceived for infinite

[1] What is called "backward induction" in game theory is roughly what is called "induction" in logic.

objects should be adopted. Like Weierstrass' discovery led to the development of series, logicians have invented methods for deductions on infinite structures and the right framework for reasoning logically on infinite mathematical objects is called *coinduction*.

Like induction, coinduction is based on a fixpoint, but whereas induction is based on the least fixpoint, coinduction is based on the greatest fixpoint, for an ordering we are not going to describe here as it would go beyond the scope of this paper. Attached to induction is the concept of inductive definition, which characterizes objects like finite lists, finite trees, finite games, finite strategy profiles, etc. Similarly attached to coinduction is the concept of coinductive definition which characterizes streams (infinite lists), infinite trees, infinite games, infinite strategy profiles etc. An inductive definition yields the least set that satisfies the definition and a coinductive definition yields the greatest set that satisfies the definition. Associated with these definitions we have inference principles. For induction there is the famous *induction principle*. On coinductively defined sets of objects there is a principle like induction principle which uses the fact that the set satisfies the definition (proofs by case or by pattern) and that it is the largest set with this property. Since coinductive definitions allow us building infinite objects, one can imagine constructing a specific category of objects with "loops", like the infinite word $(abc)^\omega$ (i.e., *abcabcabc...*) which is made by repeating the sequence *abc* infinitely many times (other examples with trees are given in Section 2 and with strategy profiles in Section 3). Such an object is a fixpoint, this means that it contains an object like itself. For instance $(abc)^\omega = abc(abc)^\omega$ contains itself. We say that such an object is defined as a cofixpoint. To prove a property P on a cofixpoint $\mathfrak{o} = f(\mathfrak{o})$, one assumes P holds on $\mathfrak{o}$ (the $\mathfrak{o}$ in $f(\mathfrak{o})$), considered as a sub-object of $\mathfrak{o}$. If one can prove P on the whole object (on $f(\mathfrak{o})$), then one has proved that P holds on $\mathfrak{o}$. This is called the *coinduction principle* a concept which comes from Park [1981] and Milner and Tofte [1991] and was introduced in the framework we are considering by Coquand [1993], see Sangiorgi [2009] for a historical account. To be sure not be entangled, it is advisable to use a proof assistant that implements coinduction to build and check the proof, but reasoning with coinduction is sometimes so counter-intuitive that the use of a proof assistant is not only advisable but compulsory. For instance, we were, at first, convinced that the strategy profile consisting in both agents stopping at every step was a Nash equilibrium and only failing in proving it mechanically convinced us of the contrary and we were able to prove the opposite. In our case we have checked every statement using Coq[2] and in what follows a sentence like "we have prover that ..." means that we have succeeded in performing a formal proof in Coq.

The mathematical development presented here corresponds to a Coq script[3] which can be found on the following url's:

`http://perso.ens-lyon.fr/pierre.lescanne/COQ/EscRat/`
`http://perso.ens-lyon.fr/pierre.lescanne/COQ/EscRat/SCRIPTS/`

[2] Coq is a proof assistant built by The Coq development team [2007], see Bertot and Castéran [2004] for a good introduction.

[3] A *script* is a list of commands of a proof assistant.

Backward coinduction as an invariant proved back from the future

In infinite games, the proof method we develop can be seen as follows. A property which holds on an infinite extensive game is an invariant along the temporal line and to prove that this is an invariant one proceeds backward the time, from the future to the past. Therefore the name *backward coinduction* is appropriate.

Structure of the paper

The paper is structured as follows. In Section 2 we present coinduction illustrated by the example of infinite binary trees. In Section 3, we introduce the core concept of infinite strategy profile which allows us presenting equilibria in Section 4. The dollar auction game is presented in Section 5 and the escalation is discussed in Section 6. Readers who want to have a quick idea about the results of this paper about the rationality of escalation are advised to read sections 5, 6 and 7.

Related works

Coinduction is a relatively new concept and its application to specific domains has just started. Probably the most connected work is this of Coupet-Grimal [2003] on temporal logic, indeed "time" is a component of extensive games. Other applications are on representation of real numbers by infinite sequences (Bertot [2007], Julien [2008]) and implementation of streams (infinite lists) in electronic circuits (Coupet-Grimal and Jakubiec [2004]). An ancestor of our description of infinite games and infinite strategy profiles is the constructive description of finite games, finite strategy profiles, and equilibria by Vestergaard [2006]. But clearly there is a gap between this approach and ours, namely the difference between finiteness and infiniteness. Nevertheless, this paper was inspiring. Lescanne [2009] introduces the framework of infinite games with more detail. Infinite games are introduced in Osborne and Rubinstein [1994] and Osborne [2004] but this is not algorithmic and therefore not amenable to formal proofs. Many authors have studied infinite games (see for instance Martin [1998], Mazala [2001]), but except the name "game" (an overloaded name), those games have nothing to see with infinite extensive games as presented in this paper. The infiniteness of Blackwell games for instance is derived from a topology, by adding real numbers and probability. Moreover those games are examined from a model or a set theory point of view, not from the point of view of what can be proved in them, i.e., from the proof theory point of view or of the rationality point of view. Said otherwise, those approaches consider an actual infinity when we consider a potential infinity.

2 Coinduction and infinite binary trees

As an example of a coinductive definition consider this of *lazy binary trees*, i.e., finite and infinite binary trees.

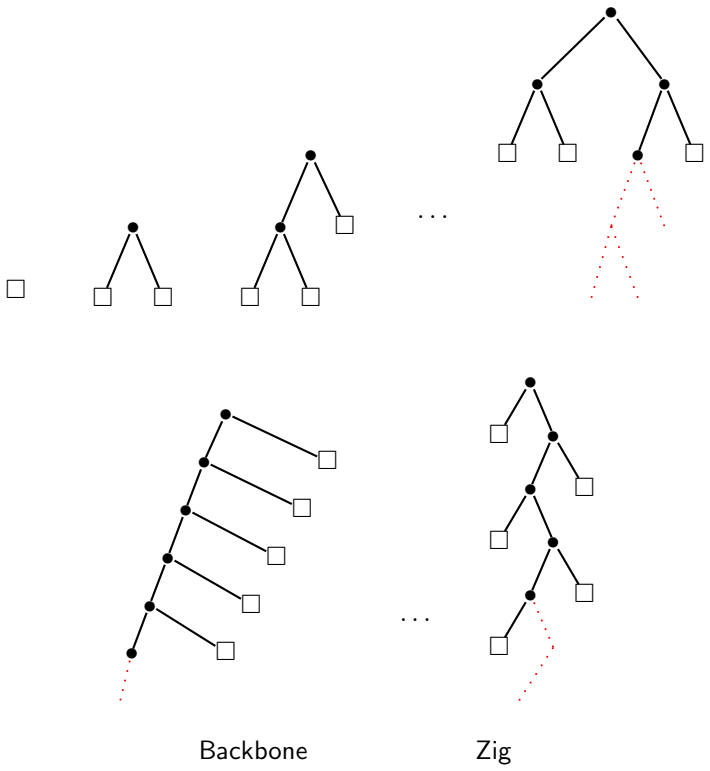


Figure 1: Coinductive binary trees

A coinductive *binary tree* (or a lazy binary tree or a finite-infinite binary tree) is

- either the empty binary tree $\Box$,
- or a binary tree of the form $t \cdot t'$, where t and t' are binary trees.

By the keyword coinductive we mean that we define a coinductive set of objects, hence we accept infinite objects. Some coinductive binary trees are given on Fig. 1. We define on a coinductive binary tree a predicate which has also a coinductive definition:

A binary tree is *infinite* if (coinductively)

- either its left subtree is *infinite*

- or its right subtree is *infinite*.

We define two trees that we call zig and zag.

zig and zag are defined together as *cofixpoint*s as follows:

- zig has □ as left subtree and zag as right subtree,
- zag has zig as left subtree and □ as right subtree.

This says that zig and zag are the greatest solutions[4] of the two simultaneous equations:

$$\begin{aligned} \mathsf{zig} &= \square \cdot \mathsf{zag} \\ \mathsf{zag} &= \mathsf{zig} \cdot \square \end{aligned}$$

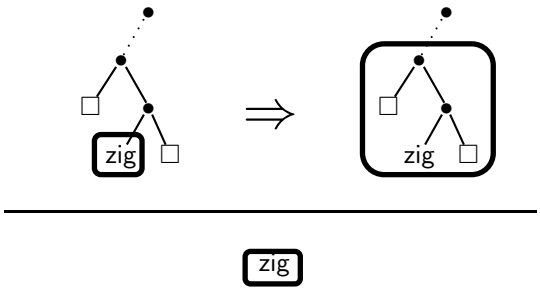


Figure 2: How cofix works on zig for *is infinite*?

It is common sense that zig and zag are infinite, but to prove that *"zig is infinite"* using the cofix tactic[5], we do as follows: assume *"zig is infinite"*, then zag is infinite, from which we get that *"zig is infinite"*. Since we use the assumption on a strict subtree of zig (the direct subtree of zag, which is itself a direct subtree of zig) we can conclude that the cofix tactic has been used properly and that the property holds, namely that *"zig is infinite"*. This is pictured on Fig.2, where the square box represents the predicate *is infinite*. Above the rule, there is the step of coinduction and below the rule the conclusion, namely that the whole zig is infinite. We let the reader prove that backbone is infinite, where backbone is the greatest fixpoint of the equation:

$$\mathsf{backbone} = \mathsf{backbone} \cdot \square$$

[4]In this case, the least solutions are uninteresting as they are objects nowhere defined. Indeed there is no basic case in the inductive definition.

[5]The cofix tactic is a method proposed by the proof assistant Coq which implements coinduction on cofixpoint objects. Roughly speaking, it attempts to prove that a property is an *invariant*, by proving it is preserved along the infinite object. Here *" is infinite"* is such an invariant on zig.

and is an infinite tree that looks like the skeleton of a centipede game as shown on Fig.1.

Interested readers may have a look at Coupet-Grimal [2003], Coupet-Grimal and Jakubiec [2004], Lescanne [2009], Bertot [2005, 2007] and especially Bertot and Castéran [2004, chap. 13] for other examples of cofix reasoning.

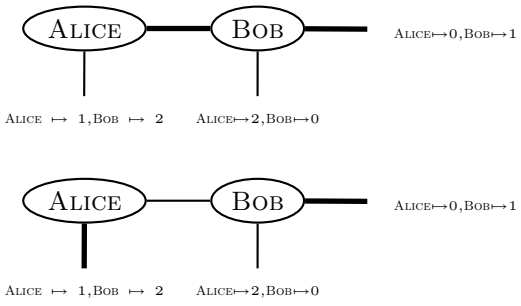
3 Finite or infinite strategy profiles

In this section we define *finite or infinite binary strategy profiles* or *StratProf*s in short. They are based on games which are extensive (or sequential) games and in which each agent has two choices: ℓ (left) and r (right).[6] In addition these games are infinite, we should say "can be infinite", as we consider both finite and infinite games. We do not give explicitly the definition of a finite or infinite extensive game since we do not use it in what follows, but it can be easily obtained by removing the choices from a strategy profile. To define finite or infinite strategy profiles, we suppose given a *utility* and a *utility function*. As said, we define directly strategy profiles as they are the only concept we are interested in. Indeed an equilibrium is a strategy profile.

The type of *StratProf*s is defined as a **coinductive** as follows:

- a *Utility function* makes a *StratProf*.
- an *Agent*, a *Choice* and two *StratProf*s make a *StratProf*.

Basically[7] an infinite strategy profile which is not a *leaf* is a *node* with four items: an agent, a choice, two infinite strategy profiles. A strategy profile is a game plus a choice at each node. Strategy profiles of the first kind are written $\ll f \gg$ and strategy profiles of the second kind are written $\ll a, c, s_l, s_r \gg$. In other words, if between the "$\ll$" and the "$\gg$" there is one component, this component is a utility function and the result is a leaf strategy profile and if there are four components, this is a compound strategy profile. In what follows, we say that s_l and s_r are strategy subprofiles of $\ll a, c, s_l, s_r \gg$. For instance, here are the drawing of two strategy profiles (s_0 and s_1):



[6] In pictures, we take a subjective point of view: *left* and *right* are from the perspective of the agent.

[7] The formal definition in the Coq vernacular is given in appendix A.

which correspond to the expressions

$s_0 = \ll$Alice,ℓ,$\ll$Bob, ℓ, $\ll$Alice $\mapsto 0$, Bob $\mapsto 1 \gg$, $\ll$Alice $\mapsto 2$, Bob $\mapsto 0 \gg \gg$,
$\ll$ Alice $\mapsto 1$,Bob $\mapsto 2 \gg \gg$

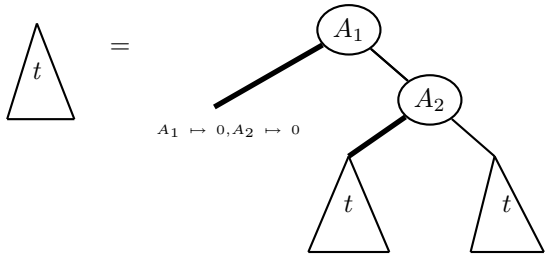
and

$s_1 = \ll$Alice,r,$\ll$Bob,ℓ,$\ll$Alice $\mapsto 0$, Bob $\mapsto 1 \gg$, $\ll$Alice $\mapsto 2$, Bob $\mapsto 0 \gg$,
$\ll$ Alice $\mapsto 1$,Bob $\mapsto 2 \gg, \gg$.

To describe a specific infinite strategy profile one uses most of the time a fixpoint equation like:

$$t \quad = \quad \ll \text{Alice}, \mathsf{r}, \ll \text{Alice} \mapsto 0, \text{Bob} \mapsto 0 \gg, \ll \text{Bob}, \mathsf{r}, t, t \gg \gg$$

which corresponds to the pictures:



Other examples of infinite strategy profiles are given in Section 5. Usually an infinite game is defined as a cofixpoint, i.e., as the solution of an equation, possibly a parametric equation.

Whereas in the finite case one can easily associate with a strategy profile a utility function, i.e., a function which assigns a utility to an agent, as the result of a recursive evaluation, this is no more the case with infinite strategy profiles. One reason is that it is no more the case that the utility function can be computed since the strategy profile may run for ever. This makes the function partial[8] and it cannot be defined as an inductive or a coinductive. Therefore we make $s2u$ (an abbreviation for *Strategy-profile-to-Utility*) a relation between a strategy profile and a utility function and we define it coinductively; $s2u$ appears in expression of the form[9] $(s2u\ s\ a\ u)$ where s is a strategy profile, a is an agent and u is a utility. It reads "u is a utility of the agent a in the strategy profile s".

$s2u$ is a predicate defined **inductively** as follows:

- $s2u \ll f \gg\ a\ (f(a))$ holds,
- if $s2u\ s_l\ a\ u$ holds then $s2u\ \ll a', \ell, s_l, s_r \gg\ a\ u$ holds,
- if $s2u\ s_r\ a\ u$ holds then $s2u\ \ll a', \mathsf{r}, s_l, s_r \gg\ a\ u$ holds.

This means the utility of a for the leaf strategy profile $\ll f \gg$ is $f(a)$, i.e., the value delivered by the function f when applied to a. The utility of a for the

[8] Assigning arbitrarily (i.e., not algorithmically) a utility function to an infinite "history", as it is made sometimes in the literature, is artificial and not really handy for formal reasoning.

[9] Notice the lighter notation $(f\ x\ y\ z)$ for what is usually written $f(x)(y)(z)$.

strategy profile $\ll a', \ell, s_l, s_r \gg$ is u if the utility of a for the strategy profile s_l is u. In the case of s_0, the first above strategy profile, one has $s2u\ s_0$ ALICE 2, which means that, for the strategy profile s_0, the utility of ALICE is 2.

4 Subgame perfect equilibria and Nash equilibria

4.1 Convertibility

An important binary relation on strategy profiles is *convertibility*. We write $\vdash^a \dashv$. the convertibility of agent a.

The relation $\vdash^a \dashv$ is defined **inductively** as follows:

- $\vdash^a \dashv$ is reflexive, i.e., for all s, $s \vdash^a \dashv s$.
- If the node has the same agent as the agent in $\vdash^a \dashv$ then the choice may change, i.e.,

$$\frac{s_1 \vdash^a \dashv s_1' \qquad s_2 \vdash^a \dashv s_2'}{\ll a, c, s_1, s_2 \gg \ \vdash^a \dashv \ \ll a, c', s_1', s_2' \gg}$$

- If the node does not have the same agent as in $\vdash^a \dashv$, then the choice has to be the same:

$$\frac{s_1 \vdash^a \dashv s_1' \qquad s_2 \vdash^a \dashv s_2'}{\ll a', c, s_1, s_2 \gg \ \vdash^a \dashv \ \ll a', c, s_1', s_2' \gg}$$

Roughly speaking two strategy profiles are convertible for a if they change only for the choices for a. Since it is defined inductively, this means that those changes are finitely many. We feels that this makes sense since an agent can only conceive finitely many issues.

4.2 Nash equilibria

The notion of Nash equilibrium is translated from the notion in textbooks. The concept of Nash equilibrium is based on a comparison of utilities; this assumes that an actual utility exists and therefore this requires convertible strategy profiles to *"lead to a leaf"*. s is a *Nash equilibrium* if the following implication holds:

> If s *"leads to a leaf"* and for all agent a and for all strategy profile s' which is convertible to s, i.e., $s \vdash^a \dashv s'$, and which *"leads to a leaf"*, if u is the utility of s for a and u' is the utility of s' for a, then $u' \preceq u$.

Roughly speaking this means that a Nash equilibrium is a strategy profile in which no agent has interest to change his choice since doing so he cannot get a better payoff.

4.3 Subgame Perfect Equilibria

In order to insure that $s2u$ has a result we define an operator *"leads to a leaf"* that says that if one follows the choices shown by the strategy profile one reaches a leaf, i.e., one does not go forever.

The predicate *"leads to a leaf"* is defined **inductively** as

- the strategy profile $\ll f \gg$ *"leads to a leaf"*,
- if s_l *"leads to a leaf"*, then $\ll a, \ell, s_l, s_r \gg$ *"leads to a leaf"*,
- if s_r *"leads to a leaf"*, then $\ll a, \mathsf{r}, s_l, s_r \gg$ *"leads to a leaf"*.

This means that a strategy profile, which is itself a leaf, *"leads to a leaf"* and if the strategy profile is a node, if the choice is ℓ and if the left strategy subprofile *"leads to a leaf"* then the whole strategy profile *"leads to a leaf"* and similarly if the choice is r.

If s is a strategy profile that satisfies the predicate *"leads to a leaf"* then the utility exists and is unique, in other words:

- For all agent a and for all strategy profile s, if s *"leads to a leaf"* then there exists a utility u which "is a utility of the agent a in the strategy profile s".
- For all agent a and for all strategy profile s, if s *"leads to a leaf"*, if "u is a utility of the agent a in the strategy profile s" and "v is a utility of the agent a in the strategy profile s" then $u = v$.

This means $s2u$ works like a function on strategy profiles which *lead to a leaf*. We also consider a predicate *"always leads to a leaf"* which means that everywhere in the strategy profile, if one follows the choices, one leads to a leaf. This property is defined everywhere on an infinite strategy profile and is therefore coinductive.

The predicate *"always leads to a leaf"* is defined **coinductively** by saying:

- the strategy profile $\ll f \gg$ *"always leads to a leaf"*,
- for all choice c, if $\ll a, c, s_l, s_r \gg$ *"leads to a leaf"*, if s_l *"always leads to a leaf"*, if s_r *"always leads to a leaf"*, then $\ll a, c, s_l, s_r \gg$ *"always leads to a leaf"*.

This says that a strategy profile, which is a leaf, *"always leads to a leaf"* and that a composed strategy profile inherits the predicate from its strategy subprofiles provided itself *"leads to a leaf"*.

Let us consider now *subgame perfect equilibria*, which we write $SGPE$. $SGPE$ is a property of strategy profiles. It requires the strategy subprofiles to fulfill coinductively the same property, namely to be a $SGPE$, and to insure that the strategy profile with the best utility for the node agent to be chosen. Since both the strategy profile and its strategy subprofiles are potentially infinite, it makes sense to define $SGPE$ coinductively.

$SGPE$ is defined **coinductively** as follows:

- $SGPE \ll f \gg$,
- if $\ll a, \ell, s_l, s_r \gg$ *"always leads to a leaf"*, if $SGPE(s_l)$ and $SGPE(s_r)$, if $s2u\ s_l\ a\ u$ and $s2u\ s_r\ a\ v$, if $v \preceq u$ then $SGPE \ll a, \ell, s_l, s_r \gg$,
- if $\ll a, \mathsf{r}, s_l, s_r \gg$ *"always leads to a leaf"*, if $SGPE(s_l)$ and $SGPE(s_r)$, if $s2u\ s_l\ a\ u$ and $s2u\ s_r\ a\ v$, if $u \preceq v$ then $SGPE \ll a, \mathsf{r}, s_l, s_r \gg$,

This means that a strategy profile, which is a leaf, is a subgame perfect equilibrium (condition *SGPE_Leaf*). Moreover if the strategy profile is a node, if the strategy profile *"always leads to a leaf"*, if it has agent a and choice ℓ, if both strategy subprofiles are subgame perfect equilibria and if the utility of the agent a for the right strategy subprofile is less than this for the left strategy subprofile then the whole strategy profile is a subgame perfect equilibrium and vice versa (condition *SGPE_left*). If the choice is r (condition *SGPE_right*) this works similarly.

Notice that since we require that the utility can be computed not only for the strategy profile, but for the strategy subprofiles and for the strategy subsubprofiles and so on, we require these strategy profiles not only to *"lead to a leaf"* but to *"always lead to a leaf"*.

Proposition 1 *A subgame perfect equilibrium is a Nash equilibrium.*

Proof: See the Coq script. □

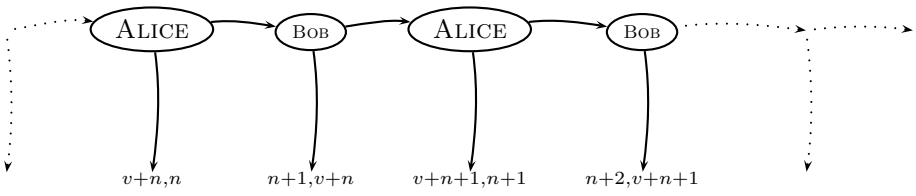


Figure 3: The *dollar auction* game

5 Dollar auction games and Nash equilibria

The dollar auction has been presented by Shubik [1971] as the paradigm of escalation, insisting on its paradoxical aspect. It is a sequential game presented as an auction in which two agents compete to acquire an object of value v ($v > 0$) (see Gintis [2000, Ex. 3.13]). Suppose that both agents bid \$1 at each turn. If one of them gives up, the other receives the object and both pay the amount of their bid.[10] For instance, if agent Alice stops immediately, she pays

[10] In a variant, each bidder, when he bids, puts a dollar bill in a hat or in a piggy bank and their is no return at the end of the auction. The last bidder gets the object.

nothing and agent Bob, who acquires the object, has a payoff v. In the general turn of the auction, if Alice abandons, she looses the auction and has a payoff $-n$ and Bob who has already bid $-n$ has a payoff $v - n$. At the next turn after Alice decides to continue, bids \$1 for this and acquires the object due to Bob stopping, Alice has a payoff $v-(n+1)$ and Bob has a payoff $-n$. In our formalization we have considered the *dollar auction* up to infinity. Since we are interested only by the "asymptotic" behavior, we can consider the auction after the value of the object has been passed and the payoffs are negative. The dollar auction game can be summarized by Fig. 3. Notice that we assume that Alice starts. We have recognized three classes of infinite strategy profiles, indexed by n:

1. The strategy profile *always give up*, in which both Alice and Bob stop at each turn, in short $\mathsf{dolAsBs}_n$.
2. The strategy profile Alice *stops always and* Bob *continues always*, in short $\mathsf{dolAsBc}_n$.
3. The strategy profile Alice *continues always and* Bob *stops always*, in short $\mathsf{dolAcBs}_n$.

The three kinds of strategy profiles are presented in Fig. 4.

We have shown that the second and third kinds of strategy profiles, in which one of the agents always stops and the other continues, are subgame perfect equilibria. The proofs are typical uses of the Coq cofix tactic. For instance, consider the strategy profile $\mathsf{dolAsBc}_n$. Assume $SGPE(\mathsf{dolAsBc}_{n+1})$. It works as follows: if $\mathsf{dolAsBc}_{n+1}$ is a subgame perfect equilibrium corresponding to the payoff $-(v+n+1), -(n+1)$, then

$$\ll \textsc{Bob}, \ell, \mathsf{dolAsBc}_{n+1}, \ll \textsc{Alice} \mapsto n+1, \textsc{Bob} \mapsto v+n \gg\gg$$

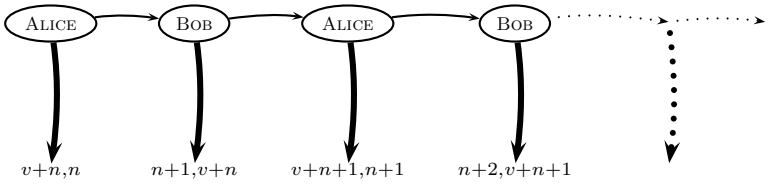
is again a subgame perfect equilibrium (since $v+n \geq n+1$) and therefore $\mathsf{dolAsBc}_n$ is a subgame perfect equilibrium, since again $v+n \geq n+1$.[11] We can conclude that for all n, $\mathsf{dolAsBc}_n$ *is a subgame perfect equilibrium.* In other words, we have assumed that $SGPE(\mathsf{dolAsBc}_n)$ is an *invariant* all along the game and that this invariant is preserved as we proceed backward, through time, into the game.

With the condition $v > 1$, we can prove that $\mathsf{dolAsBs}_0$ is not a Nash equilibrium, then as a consequence not a subgame perfect equilibrium. Therefore, the strategy profile that consists in stopping from the beginning and forever is not a Nash equilibrium, this contradicts what is said in the literature [Shubik, 1971, O'Neill, 1986, Leininger, 1989, Gintis, 2000].

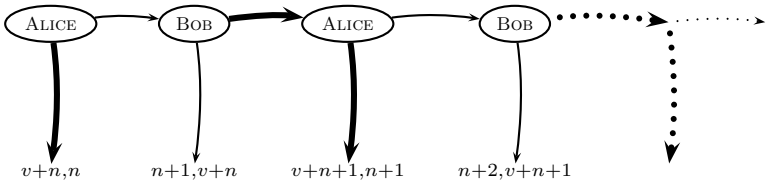
6 Why escalation is rational?

Many authors agree (see however [Halpern, 2001, Stalnaker, 1998]) that choosing a subgame perfect equilibrium is rational [Aumann, 1995]. Let us show that this

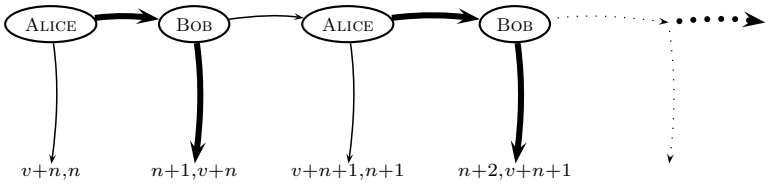
[11] Since the cofix tactic has been used on a strict strategy subprofile, the reasoning is correct.



dolAsBs$_n$ aka **Always give up**



dolAsBc$_n$ aka **Alice abandons always and Bob continues always**



dolAcBs$_n$ aka **Alice continues always and Bob abandons always**

Figure 4: Three strategy profiles

can lead to an escalation. Suppose I am Alice in the middle of the auction, I have two options that are rational: one option is to stop right away, since I assume that Bob will continue always. But the second option says that it could be the case that from now on Bob will stop always (strategy profile $\mathsf{dolAcBs}_n$) and I will always continue which is a subgame perfect equilibrium hence rational. If Bob acts similarly this is the escalation. So at each step an agent can stop and be rational, as well as at each step an agent can continue and be rational; both options make perfect sense. We claim that human agents reason coinductively unknowingly. Therefore, for them, escalation is one of their rational options at least if one considers strictly the rules of the dollar auction game, in particular with no limit on the bankroll. Many experiences (Colman [1999]) have shown that human are inclined to escalate or at least to go very far in the auction when playing the dollar auction game. We propose the following explanation: the finiteness of the game was not explicit for the participants and for them the game was naturally infinite. Therefore they adopted a form of reasoning similar to the one we developed here, probably in an intuitive form and they conclude it was equally rational to continue or to leave according to their feeling on the threat of their opponent, hence their attitude. Actually our theoretical work reconciles experiences with logic,[12] and human with rationality.

7 Conclusion

We have shown that coinduction is the right tool to study infinite structures, e. g., the infinite *dollar auction game*. This way we get results which contradict forty years of claims that escalation is irrational. We can show where the failure comes from, namely from the fact that authors have extrapolated on infinite structures results obtained on finite ones. Actually in a strategy profile in which one of the agent threatens credibly the other to continue in every case, common sense says that the other agent should abandon at each step (taking seriously the threat), this is a subgame perfect equilibrium. If the threat to continue is not credible, the other agent may think that his opponent bluffs and will abandon at every step from now on, hence a rational attitude for him is to continue. As a matter of fact, coinduction meets common sense. Since our reasoning on infinite games proceeds from future to past, we call *backward coinduction* the new method for proving that a given infinite strategy profile is a subgame perfect equilibrium. This study has also demonstrated the use of a proof assistant in such a development. Indeed the results on infinite objects are sometime so counter-intuitive that a check on a proof assistant is essential. We think that this opens new perspectives to game theory toward a more formal approach based on the last advances in mathematics offered by proof assistants [Harrison, 2008, Dowek, 2007]. In particular, an issue is to study the connection between subgame perfect equilibria (or backward coinduction) and common knowledge [Aumann, 1995]. Should we keep the usual sense of common knowledge or elaborate a new one which takes into account infiniteness? Indeed classical

[12] A logic which includes coinduction.

common knowledge is the least fixed point [Lescanne, 2007, Fagin et al., 1995] of an operator on propositions. Should we switch toward the greatest fixed point?

References

Aumann, R. J., 1995. Backward induction and common knowledge of rationality. Games and Economic Behavior 8, 6–19.

Bertot, Y., 2005. Filters on coinductive streams, an application to Eratostenes' sieve. In: Typed Lambda Calculus and Applications'05. Vol. 3461 of Lecture Notes in Computer Science. Springer-Verlag, pp. 102–115.

Bertot, Y., 2007. Affine functions and series with co-inductive real numbers. Mathematical Structures in Computer Science 17 (1), 37–63.

Bertot, Y., Castéran, P., 2004. Interactive Theorem Proving and Program Development Coq'Art: The Calculus of Inductive Constructions. Springer-Verlag.

Colman, A. M., 1999. Game theory and its applications in the social and biological sciences. London New York : Routledge, second edition.

Coquand, T., 1993. Infinite objects in type theory. In: Barendregt, H., Nipkow, T. (Eds.), TYPES. Vol. 806 of Lecture Notes in Computer Science. Springer, pp. 62–78.

Coupet-Grimal, S., 2003. An axiomatization of linear temporal logic in the calculus of inductive constructions. J Logic Computation 13 (6), 801–813.

Coupet-Grimal, S., Jakubiec, L., 2004. Certifying circuits in type theory. Formal Asp. Comput. 16 (4), 352–373.

Dowek, G., 2007. Les métamorphoses du calcul. Le Pommier, Grand Prix of Philosophy of the French Academy, English title: *The metamorphoses of computation*.

Ebbinghaus, H. D., Flum, J., 1995. Finite Model Theory. Springer-Verlag, New York.

Edgar, G., 2008. Measure, Topology, and Fractal Geometry. Springer-Verlag, 2nd ed.

Fagin, R., 1993. Finite model theory – a personal perspective. Theoretical Computer Science 116, 3–31.

Fagin, R., Halpern, J. Y., Moses, Y., Vardi, M. Y., 1995. Reasoning about Knowledge. The MIT Press.

Gintis, H., 2000. Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Interaction. Princeton University Press.

Halpern, J. Y., 2001. Substantive rationality and backward induction. Games and Economic Behavior 37 (2), 425–435.

Harrison, J., 2008. Formal proof – theory and practice. Notices of the American Mathematical Society 55, 1395–1406.

Julien, N., 2008. Certified exact real arithmetic using co-induction in arbitrary integer base. In: Garrigue, J., Hermenegildo, M. V. (Eds.), FLOPS. Vol. 4989 of Lecture Notes in Computer Science. Springer, pp. 48–63.

Leininger, W., 1989. Escalation and cooperation in conflict situations. J. of Conflict Resolution 33, 231–254.

Lescanne, P., 2007. Mechanizing epistemic logic with Coq. Annals of Mathematics and Artificial Intelligence 48, 15–43.

Lescanne, P., 2009. Deconstruction of infinite extensive games using coinduction. CoRR abs/0904.3528.

Martin, D., 1998. The determinacy of Blackwell games. The Journal of Symbolic Logic 63 (4), 1565–1581.

Mazala, R., 2001. Infinite games. In: Grädel, E., Thomas, W., Wilke, T. (Eds.), Automata, Logics, and Infinite Games. Vol. 2500 of Lecture Notes in Computer Science. Springer, pp. 23–42.

Milner, R., Tofte, M., 1991. Co-induction in relational semantics. Theor. Comput. Sci. 87 (1), 209–220.

O'Neill, B., 1986. International escalation and the dollar auction. J. of Conflict Resolution 30 (33-50).

Osborne, M. J., 2004. An Introduction to Game Theory. Oxford University Press.

Osborne, M. J., Rubinstein, A., 1994. A Course in Game Theory. The MIT Press, Cambridge, Massachusetts.

Park, D., 1981. Concurrency and automata on infinite sequences. In: Deussen, P. (Ed.), Theoretical Computer Science. Vol. 104 of Lecture Notes in Computer Science. Springer, pp. 167–183.

Sangiorgi, D., 2009. On the origins of bisimulation and coinduction. ACM Trans. Program. Lang. Syst. 31 (4).

Shubik, M., 1971. The dollar auction game: A paradox in noncooperative behavior and escalation. Journal of Conflict Resolution 15 (1), 109–111.

Stalnaker, R. C., 1998. Belief revision in games: Forward and backward induction. Math. Soc. SQci. 36, 31–56.

The Coq development team, 2007. The Coq proof assistant reference manual. LogiCal Project, version 8.1.
URL `http://coq.inria.fr`

Vestergaard, R., 2006. A constructive approach to sequential Nash equilibria. Inf. Process. Lett. 97, 46–51.

Weierstrass, K., 1872. Über continuirliche Funktionen eines reellen Arguments, die für keinen Werth des letzteren einen bestimmten Differentialquotienten besitzen. In *Karl Weiertrass Mathematische Werke (1967), Abhandlungen II.* Gelesen in der Königl. Akademie der Wisseschaften am 18 Juli 1872.

A Excerpts of Coq scripts

Infinite binary trees

```
CoInductive LBintree : Set :=
— LbtNil: LBintree
— LbtNode: LBintree → LBintree → LBintree.

CoInductive InfiniteLBT: LBintree → Prop :=
— IBTLeft : ∀ bl br, InfiniteLBT bl → InfiniteLBT (LbtNode bl br)
— IBTRight : ∀ bl br, InfiniteLBT br → InfiniteLBT (LbtNode bl br).

CoFixpoint Zig: LBintree := LbtNode Zag LbtNil
with Zag: LBintree := LbtNode LbtNil Zig.
```

Infinite strategy profiles

```
CoInductive StratProf : Set :=
— sLeaf : Utility_fun → StratProf
— sNode : Agent → Choice → StratProf → StratProf → StratProf.

Inductive s2u : StratProf → Agent → Utility → Prop :=
— s2uLeaf: ∀ a f, s2u (≪f≫) a (f a)
— s2uLeft: ∀ (a a':Agent) (u:Utility) (sl sr:StratProf),
    s2u sl a u → s2u (≪a',l,sl,sr≫) a u
— s2uRight: ∀ (a a':Agent) (u:Utility) (sl sr:StratProf),
    s2u sr a u → s2u (≪a',r,sl,sr≫) a u.

Inductive LeadsToLeaf: StratProf → Prop :=
— LtLLeaf: ∀ f, LeadsToLeaf (≪f≫)
— LtLLeft: ∀ (a:Agent)(sl: StratProf) (sr:StratProf),
    LeadsToLeaf sl → LeadsToLeaf (≪a,l,sl,sr≫)
— LtLRight: ∀ (a:Agent)(sl: StratProf) (sr:StratProf),
```

$LeadsToLeaf\ sr \rightarrow LeadsToLeaf\ (\ll a,r,sl,sr\gg)$.

CoInductive $AlwLeadsToLeaf$: $StratProf \rightarrow Prop :=$
— $ALtLeaf$: $\forall\ (f{:}Utility_fun),\ AlwLeadsToLeaf\ (\ll f\gg)$
— $ALtL$: $\forall\ (a{:}Agent)(c{:}Choice)(sl\ sr{:}StratProf),$
$LeadsToLeaf\ (\ll a,c,sl,sr\gg) \rightarrow AlwLeadsToLeaf\ sl \rightarrow AlwLeadsToLeaf\ sr \rightarrow$
$AlwLeadsToLeaf\ (\ll a,c,sl,sr\gg)$.

SGPE

CoInductive $SGPE$: $StratProf \rightarrow Prop :=$
— $SGPE_leaf$: $\forall\ f{:}Utility_fun,\ SGPE\ (\ll f\gg)$
— $SGPE_left$: $\forall\ (a{:}Agent)(u\ v{:}\ Utility)\ (sl\ sr{:}\ StratProf),$
$AlwLeadsToLeaf\ (\ll a,l,sl,sr\gg) \rightarrow$
$SGPE\ sl \rightarrow SGPE\ sr \rightarrow$
$s2u\ sl\ a\ u \rightarrow s2u\ sr\ a\ v \rightarrow (v \preceq u) \rightarrow$
$SGPE\ (\ll a,l,sl,sr\gg)$
— $SGPE_right$: $\forall\ (a{:}Agent)\ (u\ v{:}Utility)\ (sl\ sr{:}\ StratProf),$
$AlwLeadsToLeaf\ (\ll a,r,sl,sr\gg) \rightarrow$
$SGPE\ sl \rightarrow SGPE\ sr \rightarrow$
$s2u\ sl\ a\ u \rightarrow s2u\ sr\ a\ v \rightarrow (u \preceq v) \rightarrow$
$SGPE\ (\ll a,r,sl,sr\gg)$.

Nash equilibrium

Definition $NashEq\ (s{:}\ StratProf){:}\ Prop :=$
$\forall\ a\ s'\ u\ u',\ s' \vdash^a \dashv s \rightarrow$
$LeadsToLeaf\ s' \rightarrow (s2u\ s'\ a\ u') \rightarrow$
$LeadsToLeaf\ s \rightarrow (s2u\ s\ a\ u) \rightarrow$
$(u' \preceq u)$.

Alice stops always and Bob continues always

Definition $add_Alice_Bob_dol\ (cA\ cB{:}Choice)\ (n{:}nat)\ (s{:}Strat) :=$
$\ll Alice,cA,\ll Bob,\ cB,s,[n{+}1,\ v{+}n]\gg,[v{+}n,n]\gg$.

CoFixpoint $dolAcBs\ (n{:}nat){:}\ Strat := add_Alice_Bob_dol\ l\ r\ n\ (dolAcBs\ (n{+}1))$.

Theorem $SGPE_dol_Ac_Bs$: $\forall\ (n{:}nat),\ SGPE\ ge\ (dolAcBs\ n)$.

Alice continues always and Bob stops always

CoFixpoint $dolAsBc\ (n{:}nat){:}\ Strat := add_Alice_Bob_dol\ r\ l\ n\ (dolAsBc\ (n{+}1))$.

Theorem $SGPE_dol_As_Bc$: $\forall$ (n:nat), $SGPE$ ge ($dolAsBc$ n).

Always give up

CoFixpoint $dolAsBs$ (n:nat): $Strat$:= $add_Alice_Bob_dol$ r r n ($dolAsBs$ (n+1)).

Theorem $NotSGPE_dolAsBs$: (v>1) → ˜($NashEq$ ge ($dolAsBs$ 0)).