Consistency and games—in search of new combinatorial principles

Pavel Pudlák*
Mathematical Institute of the Academy of Sciences
and
Institute of Theoretical Computer Science†
Prague, Czech Republic

June 1, 2004

Abstract

We show that a semantical interpretation of Herbrand's disjunctions can be used to obtain Π_2 independent sentences whose nature is more combinatorial than the nature of the usual consistency statements. Then we apply this method to Bounded Arithmetic and present $\forall\Sigma_1^b$ combinatorial sentences that characterize all $\forall\Sigma_1^b$ sentences provable in S_2. We use the concept of a two player game to describe these sentences.

1 Introduction

One of the main problems in proof complexity is to prove that certain theories are not equal. The theories that we would like to separate correspond to complexity classes that are conjectured to be distinct. The only tool for such results that we currently have is diagonalization, the same as in computational complexity. In first order logic diagonalization is essentially Gödel's incompleteness theorem. Most researchers believe that conjectures such as $\mathbf{P} \neq \mathbf{NP}$ cannot be proved by diagonalization, and the same opinion is shared about separations of subtheories of Bounded Arithmetic. Yet it is worthwhile to study the sentences used in Gödel's theorem. Recall that the Gödel sentence for a theory T is equivalent to the consistency of T, provided that T is sufficiently strong (this is the content of the second incompleteness theorem). For weak theories, such as Bounded Arithmetic, or Robinson's Arithmetic, these sentences must also be relatively weak, hence their combinatorial meaning should not be very complicated. If we could find some combinatorial interpretations of them, we would learn a lot about provability in Bounded Arithmetic.

*Partially supported by grants EAA1019401 of the AV ČR, and grants 201/01/1195 401/03/H047 of the GA ČR.

†Project No. LN00A056 of the Ministry of Education of the Czech Republic.

The usual consistency statements are not very transparent because the concept of a proof in the usual Hilbert style proof system is fairly complicated. A much more combinatorial concept is the Herbrand proof (the proof based on the characterization of the first order provability given by Herbrand's Theorem). Therefore we have concentrated on studying Herbrand proofs. In this article we shall present a sort of semantical interpretation of such proofs. Using the incompleteness theorem and this interpretation we obtain an independent sentence which is more "mathematical" than the usual consistency statements. However, we have to pay for it by increasing the complexity to Π_2. Independent combinatorial sentences in Peano Arithmetic and stronger theories have been studied since 1977 when Paris and Harrington published the seminal paper [24]. All the known such sentences are Π_2. In section 5, we shall describe a sentence independent of Peano Arithmetic. It is by far not so nice as those of Paris and Harrington and others, but the advantage of our method is that it is universally applicable. What is, perhaps, more important is that the reason why we have to increase the complexity to Π_2 is more apparent in our proof.

The main result of this article is an application of this framework to Bounded Arithmetic. Since it is unlikely that the separation problems could be solved using diagonalization, we consider a related problem of characterizing low complexity theorems of subtheories of Bounded Arithmetic. We shall present $\forall\Sigma^b_1$ sentences $\mathbf{A}_k$ such that $\mathbf{A}_k$ is provable in S^k_2 and all $\forall\Sigma^b_1$ theorems of S^k_2 are provable in $S^1_2 + \mathbf{A}_{k+4}$. The sentences are still fairly complicated, so it will be necessary to simplify them before we will be able to use them to conditional separations of theories. The problem of finding simpler versions of these sentences is not a purely technical point. It requires a better understanding of the dependencies of terms on variables in Herbrand disjunctions. In this article we have restricted ourselves to Herbrand proofs, but the same can be achieved using the Mid Sequent Theorem. Therefore a better understanding of the combinatorial principles $\mathbf{A}_k$ is also closely connected with better understanding of the structure of proofs in the sequent calculus.

In the presentations of the combinatorial principles we shall use the concept of a two player game. This is a natural way to interpret a sequence of alternating quantifiers. When the concept of a game is applied to Herbrand disjunctions new phenomena arise. The game corresponding to the original formula splits into several copies. What is the most interesting fact is that the games are not played sequentially. If we insisted that they should be played sequentially we would get only very weak principles; thus the power is in the parallelism.

As the principle $\mathbf{A}_k$ is too complicated to be used to prove separations (based on oracles or on assumptions from complexity theory), we suggest a more uniform version $\mathbf{B}_k$ in Section 7. It has some nice properties, but, so far, we are neither able to prove that these sentences characterize the $\forall\Sigma^b_1$ consequences of S_2, nor to prove the independence of the oracle versions of them. In the last section we shall suggest some potential applications of our results.

Related problems were studied by several authors. Buss, Chiari, F. Ferreira, Hanika, Impagliazzo and Krajíçek characterized low complexity theorems of S^i_2 for some small i and proved some conditional separations [8, 10, 12, 13, 17, 18, 19]. Adamowicz and Kolodziejczyk [3] used Herbrand's theorem in a similar way as in this paper, but they did not interpret the sentences in terms of games.

2 Games and matches

A game of finite length k is

$$G \subseteq M_1 \times M_2 \times \ldots \times M_k.$$

M_i are sets of possible moves in the i-th step of the game. If all sets M_i are finite, we say that the game is finite. The elements of M_i are legal *moves* in the i-th step. The game is played by two players who alternate in choosing elements form sets M_i. G is the set of winning positions of the first player. We shall also consider games of infinite lengths in which k is ω. More generally, a game of length k is a pair (M, G) where M is an arbitrary nonempty set of strings of length k and $G \subseteq M$. A legal position $(x_1, \ldots, x_i)$ is an initial segment of a string from M, and a legal move from this position is an x_{i+1} such that $(x_1, \ldots, x_i, x_{i+1})$ is a legal position.

Every game of finite length is determined, which means that there exists a winning strategy either for the first player or for the second. However, the strategy may be very difficult to describe even if the whole game is finite and it is described very explicitly. From the point of view of computational complexity, the natural setting is a game G defined by a boolean circuit C. Here we assume that the elements of each M_i are strings of 0s and 1s of length m_i and the size of the circuit C is polynomial in the input size $\sum_i m_i$. In such a situation the number of plays (ie., legal final positions) is exponential in the size of the description of the game. It is well-known that then it is **PSPACE**-complete to determine which player has a wining strategy. Consequently, to describe the moves of a winning strategy is **PSPACE**-hard.

Thus if we restrict the computational power of the players to polynomial size boolean circuits, for some games there may be no winning strategies in this restricted sense. It seems that some of the difficult games that people play (chess, go, checkers) are of this nature. To show that this is a very interesting field of research let us mention a simple example based on randomized strategies:

> *It is conceivable that there exists a game G described by a polynomial size boolean circuit and a strategy σ for the first player such that σ is defined by a probabilistic polynomial size circuit and with probability tending to 1 strategy σ beats all strategies for the second player which are defined by probabilistic polynomial size circuits. In the absolute sense, however, it is the second player who has a winning strategy.*

Such an example can be constructed assuming that one-way functions exist, which is the basic conjecture of cryptography. Let $F : \{0,1\}^n \to \{0,1\}^n$ be a one way function. Define $G = \{(y, z, x);\ F(x) = y \wedge F(z) \neq y\}$. The strategy σ for the first player is to choose randomly uniformly an x, and play $F(x)$ in the first move and x in the third move. The assumption about F guarantees that the second player has a very small chance to find a z such that $F(z) = y$.

In the rest of the paper we shall consider only deterministic strategies.

We shall also consider systems of games, which we shall call *matches*. As in the real world, playing a match means that we play several games and then the winner is determined by the results of the games. Unlike in the real world, in the matches that we shall consider the players may start next game without finishing the previous one (however, they have to finish all games eventually). Formally a match is determined by:

1. a set of boards with a finite length game assigned to every board; the games assigned to different boards may be the same;

2. a schedule that determines a linear order in which the moves of the games will be played; the only restriction is that the moves in every game are played in the order given by this game;

3. a rule to determine the winner from the information about who won the games.

A match, like a game, is played by two players which we shall call *Alice* and *Bob*. Given a game G in a match, it is necessary to define who will start the game (the person who starts will also be called *the first player*). Therefore we shall assume that the rules of every game include also a rule that Alice starts or that Bob starts. Then for every game G we have the *complementary game*, denoted by $G^{\perp}$, in which the roles of Alice and Bob are switched.

We can treat a match also as a finite length game. In particular, we can talk about winning strategies for matches. However it turns out that the most interesting is the structure of a match, which often determines who has a winning strategy without our knowing who has the winning strategies for the games.

3 Herbrand's proofs as winning strategies

We shall use first order logic without equality. As usual, it means that our theories may contain equality, but it is treated as an ordinary binary relation. Consider a sentence of the form

$$\exists x_1 \forall y_1 \exists x_2 \forall y_2 \ \dots \ \gamma(x_1, y_1, x_2, y_2, \dots), \tag{1}$$

where γ is open. We shall use the convention that $a_1, a_2, \dots$ represents a *finite* sequence, unless stated otherwise. (This is especially useful when denoting quantifier prefixes such as above, because the last quantifier may be any of the two.) According to Herbrand's theorem this sentence is provable in the predicate calculus if and only if a disjunction of the form

$$\bigvee_{i=1}^{\ell} \gamma(t_{i1}, f_1(t_{i1}), t_{i2}, f_2(t_{i1}, t_{i2}), \dots), \tag{2}$$

is provable in the propositional calculus, where $t_{i1}, t_{i2}, \dots$ are terms in the language of γ augmented with new function symbols $f_1, f_2, \dots$.

Let such a disjunction be given. For every term $f_j(t_{i1}, \dots, t_{ij})$ we introduce a new variable z_{ij}. Notice that some pairs of terms $f_j(t_{i1}, \dots, t_{ij})$, $f_{j'}(t_{i'1}, \dots, t_{i'j'})$ with $(i, j) \neq (i', j')$ may be equal; in such a case we will use the same variable. Then starting with the longest term $f_j(t_{i1}, \dots, t_{ij})$, we replace gradually all these terms by the corresponding variables in all terms of the disjunction. Substitute z_{11} for all variables other than z_{ij} and for all remaining subterms of the form $f_i(t)$. Thus we eliminate all occurrences of the new function symbols. The resulting disjunction has the form

$$\bigvee_{i=1}^{\ell} \gamma(\tilde{t}_{i1}, z_{i1}, \tilde{t}_{i2}, z_{i2}, \ldots), \tag{3}$$

where $\tilde{t}_{ij}$ are terms resulting from t_{ij} when we apply the substitution described above. Let us stress that z_{ij} with distinct indices do not necessarily represent distinct variables. We have to substitute the same variable for all occurrences of the same term, otherwise the logical validity of the formulas will not be always preserved. This disjunction has two properties:

- (1) logically follows from the universal closure of the formula (3);
- if (2) is a propositional tautology, then so is (3).

The first property is a consequence of the following syntactical property of (3): we can order the terms and variables $\tilde{t}_{ij}, z_{pk}$ in such a way that

1. the terms and variables of the i-th disjunct occur in the following order $\tilde{t}_{i1} \prec z_{i1} \prec \tilde{t}_{i2} \prec z_{i2} \prec \ldots$;
2. $\tilde{t}_{ij}$ may contain only variables which in the ordering $\prec$ occur before it.

The second property follows from the fact that for every term t in (2) we have substituted the same term for all occurrences of t in (2). The latter condition can be expressed by saying that (3) satisfies the following syntactical condition:

3. If $\tilde{t}_{i1} = \tilde{t}_{i'1}, \ldots, \tilde{t}_{ij} = \tilde{t}_{i'j}$, then $z_{ij+1} = z_{i'j+1}$.

We shall not prove it as all this is well-known. We only remark that it can also be proved using the Mid Sequent Theorem (see [30]). We shall call (3) a *Herbrand disjunction* for (1). If the Herbrand disjunction is a propositional tautology we shall call it a *Herbrand proof* of the sentence (1).

Our aim is to interpret this syntactical object in combinatorial way. To this end we shall use games. We shall interpret γ as a definition of the game G in which the first player wins in a finial position $a_1, b_1, a_2, b_2, \ldots$ iff $\gamma(a_1, b_1, a_2, b_2, \ldots)$ is true. Let us first consider one term of the disjunction $\gamma(\tilde{t}_{i1}, z_{i1}, \tilde{t}_{i2}, z_{i2}, \ldots)$. We shall think of terms $\tilde{t}_{ij}$ as defining a strategy for Alice to play the game G as the first player. The variables z_{ij} are moves of Bob who plays as the second.

Now having ℓ formulas of the form $\gamma(\tilde{t}_{i1}, z_{i1}, \tilde{t}_{i2}, z_{i2}, \ldots)$, we shall think of it as a match in which Alice and Bob play the same game γ simultaneously on ℓ boards. At each moment one player makes one move on one board. The order of moves is given by the ordering $\prec$. Condition 1. above says that they play in the correct order on each board, condition 2. says that the terms $\tilde{t}_{ij}$ define a strategy for Alice.

A strange thing in this interpretation is that Alice's moves on one board may depend on what Bob played on other boards. However, recall that we shall consider situations in which the players do not have access to the winning strategies. In the situation where they cannot play the best strategy, any information about the game may be useful. A player may learn a good move from his opponent and use it on another board. Let us view it from the viewpoint of Alice. If Bob plays well, then in every move he reveals nontrivial information. On the other hand, if he plays poorly, then it should not be difficult to beat him. Thus Alice has a good chance to win on at least one board.

Let us now interpret the logical validity of disjunction (3). It means that the strategy defined by the terms ensures Alice to win on at least one board. Hence if we define that this is the goal of the match, the terms define a winning strategy for the match.

In the same way we can analyze other forms of quantifier prefixes, namely, prefixes in which the first quantifier is universal and prefixes in which there are consecutive occurrences of the same quantifier. Let us consider one special case, the quantifier prefix $\forall\exists\forall$. (We shall always associate Alice with $\exists$ and Bob with $\forall$, thus the order in which they play is Bob-Alice-Bob.) Then the disjunction (3) has the form:

$$\bigvee_{i=1}^{\ell} \gamma(z_0, \tilde{t}_i, z_i). \tag{4}$$

(Notice that Bob starts with the same move on all boards. This is a special case of Rule $\mathcal{B}$ that will be defined in Section 4.) To get an ordering satisfying the conditions above, we shall first consider the partial ordering of terms t_i by their size, then we extend it to a total ordering and order the terms $\tilde{t}_i$ in this way. W.l.o.g. assume that we get $\tilde{t}_1 \prec \tilde{t}_2 \prec \ldots \prec \tilde{t}_\ell$. Then we extend it to

$$z_0 \prec \tilde{t}_1 \prec z_1 \prec \tilde{t}_2 \prec z_2 \prec \ldots \prec \tilde{t}_\ell \prec z_\ell. \tag{5}$$

In general, z_i may not occur in $\tilde{t}_{i+1}$, if so, then it is possible to move z_i behind $\tilde{t}_{i+1}$, but such an ordering may not be always possible. The ordering in (5) is the most general one. In terms of games it means that we allow Alice to use as much information as possible. The Herbrand disjunction has the form

$$\bigvee_{i=1}^{\ell} \gamma(z_0, \tilde{t}_i(z_0, z_1, \ldots, z_{i-1}), z_i), \tag{6}$$

where $\tilde{t}_i(z_0, z_1, \ldots, z_{i-1})$ means that only variables $z_0, z_1, \ldots, z_{i-1}$ may occur in $\tilde{t}_i$.

For $\forall\exists\forall\exists$ we get a very similar formula:

$$\bigvee_{i=1}^{\ell} \bigvee_{j=1}^{k_i} \gamma(z_0, \tilde{t}_i(z_0, z_1, \ldots, z_{i-1}), z_i, \tilde{s}_{ij}(z_0, z_1, \ldots, z_\ell)).$$

An explanation in terms of games is as follows. The most general ordering is an ordering in which Alice has as much advantage as possible. If she is to play the last move in the game γ, then for her it is the best if she can postpone her last moves on all boards after all moves of Bob.

The fact that there are the most general orderings for the prefixes contained in $\forall\exists\forall\exists$ case is well-known and it is often used. For $\exists\forall\exists\forall$ and prefixes with more quantifier alternations there does not seem to exist the most general ordering. This is the traditional stumbling block to the attempts to generalize results to longer quantifier prefixes. However, if we do not insist on the same number of terms in the disjunction, then we can find a trivial normal form: simply take the disjunction of all Herbrand disjunctions of a given length.

4 Π_1 and Π_2 unprovable sentences.

Our main goal is to understand the statements expressing consistency in such a way that we will be able to find a combinatorial interpretation of them. Herbrand's theorem alone transforms the consistency statement, which speaks about proofs in first order logic, into a more combinatorial one, which only refers to propositional provability. In the previous section we have suggested an interpretation of Herbrand disjunctions as statements about games. Now we shall explain it in greater detail.

Suppose we have a consistent theory T axiomatized by a sentence Φ whose prenex form

$$\forall x_1 \exists y_1 \forall x_2 \exists y_2 \ \ldots \ \phi(x_1, y_1, x_2, y_2, \ldots), \tag{7}$$

has k existential quantifiers. The inconsistency of T means that the negation of this sentence, which is

$$\exists x_1 \forall y_1 \exists x_2 \forall y_2 \ \ldots \ \neg\phi(x_1, y_1, x_2, y_2, \ldots), \tag{8}$$

is provable in the predicate calculus. Provability in the predicate calculus is equivalent to the Herbrand provability, thus if (8) were provable, there would be a Herbrand disjunction based on this formula which would be a tautology. Let

$$\Delta(z_1, \ldots, z_m)$$

be such a Herbrand disjunction, where all free variables are displayed. As T is consistent, it is not a propositional tautology. We want to find a semantical interpretation of this disjunction. Thus we interpret it in terms of games and we obtain a match. In this match Alice and Bob play $\neg\phi(x_1, y_1, x_2, y_2, \ldots)$ on several boards and Alice wins the match if she wins on at least one board. The terms of the disjunction define a strategy for Alice. Now instead of saying that $\Delta(z_1, \ldots, z_m)$ is not a propositional tautology we would like to say that Bob can beat Alice's strategy. What is the difference between these two statements? The first one is:

the formula $\Delta(z_1, \ldots, z_m)$ is not a propositional tautology.

The second one is

$$\exists z_1 \ldots \exists z_m \ \neg\Delta(z_1, \ldots, z_m).$$

Clearly, the second one is stronger, because it says that not only $\Delta(z_1, \ldots, z_m)$ is not a propositional tautology, but one can find a falsifying assignment to the atomic formulas *by a suitable substitution of values for $z_1, \ldots, z_m$*. In fact, it is not obvious that the latter should be true, it depends on the model in which we interpret it.

Let us see now what happens if we modify the statement that T is consistent in this way. Let $HD(\Xi, \Delta)$ denote the formalization of '*Δ is a Herbrand disjunction for Ξ*'. We get the following:

$$\forall \Delta (HD(\neg\Phi, \Delta) \to \exists z_1 \ldots \exists z_m \ \neg\Delta(z_1, \ldots, z_m)).$$

This is not a first order formula; in order to write this formula in the first order logic, we need a formula that expresses the truth of existential formulas. Let us assume that we can formalize it and denote such a formula by $Tr_{\exists}$. Then a more precise statement reads:

$$\forall \Delta (HD(\neg\Phi, \Delta) \rightarrow Tr_{\exists}(\exists z_1 \ldots \exists z_m \neg\Delta(z_1, \ldots, z_m))). \tag{9}$$

Now we would like to give a game-theoretical interpretation of this sentence. The Herbrand disjunction defines a match in which Alice and Bob play several copies of the game defined by $\neg\phi(x_1, y_1, x_2, y_2, \ldots)$. The order of moves is determined (in general, not uniquely) by the dependences of terms on variables. Furthermore, we have to satisfy the syntactical condition 3. from Section 3. (Recall that the condition expressed the fact that the same variable z_{ij} was substituted for all occurrences of the same term.) As our aim is to have a semantical interpretation, we replace it by the following condition, which, in general, is stronger than the corresponding syntactical one.

Rule $\mathcal{B}$. *If the same position occurs on several boards, then Bob must play the same move on all these boards.*

A special instance of this rule is the following. If Bob is the first player in the game on which the match is based, then he must start with the same move on all boards. By imposing this rule we make Bob's task harder, hence the statement stronger. But if Φ is true, then he has a winning strategy, so using the same winning strategy on all boards, he can obey Rule $\mathcal{B}$ and beat Alice. Let us stress that we will not consider such a restriction for Alice, it would make the principle too weak. (Remember: $\mathcal{B}$ stands for Bob.)

Thus we get the following principle.

The Match Principle for Φ. *For every match based on the game defined by $\neg\phi(x_1, y_1, x_2, y_2, \ldots)$ and for every strategy of Alice computable by terms of the language of T, one can find moves of Bob which satisfy Rule $\mathcal{B}$ and which beat Alice (ie., Bob wins on all boards).*

By interpreting the syntactical object, the Herbrand proof, semantically, we have gained a statement that looks more natural, more "mathematical". But we have also lost something: we have transformed the Π_1 sentence expressing the consistency of T into a Π_2 sentence. The fact that, for a sufficiently strong theory, the complexity of this statement cannot be reduced to Π_1 is an immediate corollary of the following proposition.

Proposition 4.1 *Suppose that T is axiomatized by the sentence Φ and contains at least the subtheory $I\Sigma_1$ of PA. Then, provably in T, the match principle for Φ is equivalent to the Σ_1 reflection principle for T.*

Proof. Recall that the Σ_1 reflection principle for T is the formalization of the sentence: *'For all Σ_1 sentences ψ, if T proves ψ, then ψ is true.'* A little bit more formally the sentence is

$$\forall x \in \Sigma_1 (Pr_T(x) \rightarrow Tr_{\Sigma_1}(x)).$$

Since every bounded formula is equivalent to an existential formula, provably in $I\Sigma_1$ (as Matiyasevich's Theorem is provable in $I\Sigma_1$), it suffices to consider only existential sentences.

1. First we prove that the match principle for Φ implies the Σ_1-reflection principle for T. Since T contains $I\Sigma_1$, the provability of a sentence Ψ in T is equivalent to Herbrand provability of $\Phi \to \Psi$.

We shall argue in T. Suppose that Ψ is an arbitrary sentence of the form $\exists x_1 \exists x_2 \ldots \exists x_q$ $\psi(x_1, x_2, \ldots, x_q)$ with ψ open. Consider a Herbrand proof of $\Phi \to \Psi$. We can present it in the form

$$\bigvee_j \psi(t_{1j}, t_{2,j}, \ldots, t_{q,j}) \vee \Delta,$$

where Δ is a Herbrand disjunction for $\neg\Phi$. Assuming the match principle for Φ we can find elements such that after substituting them for the free variables of Δ, formula Δ becomes false. Hence

$$\bigvee_j \psi(\tilde{t}_{1j}, \tilde{t}_{2,j}, \ldots, \tilde{t}_{q,j}),$$

must be true, where $\tilde{t}_{ij}$ denote t_{ij} after this substitution. Thus $\exists x_1 \exists x_2 \ldots \exists x_q\ \psi(x_1, x_2, \ldots, x_q)$ is true.

2. Let us now prove the opposite. The match principle for Ψ has the form $\forall n \exists m\ \Gamma_1(n, m)$, where n is a number that codes a match and a strategy of Alice, and n codes moves of Bob. Again, we can replace this Σ_1 sentence by an equivalent existential sentence

$$\forall n \exists m_1 \ldots \exists m_k \Gamma(n, m_1, \ldots, m_k). \tag{10}$$

Lemma 4.2 *(1) For every n, $\exists m_1 \ldots \exists m_k\ \Gamma(\mathbf{n}, m_1, \ldots, m_k)$ is provable in T ($\mathbf{n}$ denotes the n-th numeral). (2) This proof can be formalized in $I\Sigma_1$.*

Assuming the Σ_1 reflection principle for T, we get immediately (10), so it only remains to prove the lemma. The idea of the proof of (1) is based on the observation that Φ expresses the fact that Bob has a winning strategy for the game defined by ϕ. Since n is an actual number, we have a concrete match and we can use this strategy to define moves of Bob one by one on all boards so that Bob wins on all boards. This gives a construction of a proof of $\exists m_1 \ldots \exists m_k \Gamma(\mathbf{n}, m_1, \ldots, m_k)$ in T from the match coded by the number n.

More formally, let n be given. The number n determines a match and a strategy, which is defined by terms. Using this match and the terms we construct a proof of $\exists m_1 \ldots \exists m_k$ $\Gamma(\mathbf{n}, m_1, \ldots, m_k)$ as follows.

First suppose that the match has a single board. Since the game is defined by ϕ and the axiom of theory T is Φ, T proves that Bob can always win. Thus, if we plug in the moves of Alice we get gradually

$$\exists y_1 \ldots \exists y_i \forall x_{i+1} \exists y_{i+1} \forall x_{i+2} \exists y_{i+2}\ \ldots\ \phi(a_1, y_1, a_2, y_2, \ldots, a_i, y_i, x_{i+1}, y_{i+1}, x_{i+2}, y_{i+2}, \ldots), \tag{11}$$

where $a_1, \ldots, a_i$ are moves of Alice determined by the given strategy. Eventually we get

$$\exists y_1 \exists y_2 \ldots \phi(a_1, y_1, a_2, y_2, \ldots),$$

which is (equivalent to) the instance of the principle $\exists m_1 \ldots \exists m_k \Gamma(\mathbf{n}, m_1, \ldots, m_k)$.

When the match has more boards we shall keep one formula of the form (11) for each board and gradually replace the universal quantifiers by the moves of Alice determined by her strategy. The order in which we change the formulas follows the order of the moves of the match. (Recall that the match is given by a concrete number, so also the number of boards is a concrete number.)

Notice that what we are essentially doing is showing the well-known fact that the every prenex formula implies the existential closure of any of its Herbrand disjunctions.

To prove (2) just observe that the above proof is described very explicitly. To prove its existence we only refer to syntactical operations with proofs and Matiyasevich's Theorem (the latter can be avoided with some more work).

■

This proof actually shows more: we can add arbitrary universal sentences to T without changing the independent sentence. In particular this applies also to the equality axioms, thus we can ignore them completely.

The following proposition is an application of the above semantical interpretation of Herbrand disjunctions which shows that certain orderings are not universal for Herbrand's theorem.

Proposition 4.3 *There exists a provable $\exists\forall\exists$ sentence which does not have a Herbrand proof of the form*

$$\phi(t_1, z_1, s_1(z_1)) \ \vee \ \phi(t_2(z_1), z_2, s_2(z_1, z_2)) \ \vee \dots \\ \vee \ \phi(t_\ell(z_1, \dots, z_{\ell-1}), z_\ell, s_\ell(z_1, \dots, z_{\ell-1}, z_\ell)). \tag{12}$$

Proof. Suppose the contrary, namely, that every $\exists\forall\exists$ provable sentence has a Herbrand proof of this form. Let Ψ be a Π_2 sentence expressing this fact. Consider the theory $T = I\Sigma_1 + \Psi$. This theory is finitely axiomatizable and has a Π_3 axiomatization. If we expand the language of arithmetic with suitable function symbols, we get a $\forall\exists\forall$ axiomatization.[1] So let Γ be a $\forall\exists\forall$ sentence axiomatizing T. In T we know that if $\neg\Gamma$ is provable, then it also has a Herbrand proof of the form (12). But in T we can also prove that Bob can beat any strategy of Alice in every match of the form (12). The point is that in T we have Γ which is $\exists x \forall z \exists y \neg\phi(x, z, y)$. It says that Bob has a winning strategy for the game used in the match. Then, using Σ_1 induction, we can prove that in fact Bob can win all games in such a sequence of games. In more detail: by induction on $k = 1, \dots, \ell$, we prove

$$\exists z_1 \dots \exists z_k \ \neg\phi(t_1, z_1, s_1(z_1)) \wedge \dots \wedge \neg\phi(t_k(z_1, \dots, z_{k-1}), z_k, s_k(z_1, \dots, z_{k-1}, z_k)).$$

Thus T would prove its own consistency, which means that T would be inconsistent. Hence Ψ must be false. ■

[1]We cannot reduce the complexity of T further, because if we skolemize the induction axiom we only get induction for Σ_1 arithmetical formulas, but we do not get the induction for Σ_1 formulas in the language expanded by the Skolem function.

In terms of games this is an example that shows that Alice can really get an advantage if the games in a match are not played sequentially, one after the other, even if they are only of length 3. As we have already noted, for Alice it is better to postpone the last moves on all boards to the very end of the match.

The most intriguing question is whether we can obtain a Π_1 sentence in a similar way. When trying to preserve the property of being Π_1 we have to use essentially the structure of the formulas in the Herbrand disjunction, we cannot simply say that they are some arithmetical formulas. But then we can easily get into troubles, because a single equation of the form $p(\bar{x}) = 0$, with p a multivariate polynomial with integer coefficients, can express very complex facts. For instance, the existence of a solution may be equivalent to the inconsistency of the theory. Thus our feeling is that in order to get an independent Π_1 sentence we need an essentially new idea.

5 A sentence independent of Peano Arithmetic

We shall apply the ideas above to the specific case of Peano Arithmetic. We shall assume that it is axiomatized by open axioms plus induction axioms for all formulas. We assume that the relation $<$ is present in the language. Thus on top of the usual operations we shall need the operation of the predecessor and subtraction truncated to nonnegative numbers. We already know that the particular choice of the language and open axioms is not important, thus we shall not specify it more.

We shall use the induction axiom for a formula $\Phi(x)$ in the form

$$\forall x(\Phi(\mathbf{0}) \wedge \forall y(\Phi(y) \rightarrow \Phi(S(y))) \rightarrow \Phi(x)).$$

It is well known that one does not need parameters in Φ, so the only free variable is x. The negation of this sentence can be written as

$$\exists x \forall y(\Phi(\mathbf{0}) \wedge (\Phi(y) \rightarrow \Phi(S(y))) \wedge \neg\Phi(x)).$$

We shall further simplify this formula by assuming that $\Phi(\mathbf{0})$ and $\neg\Phi(x)$ is ensured by a syntactical condition on Φ. More precisely, we shall use formulas with two free variables $\Phi(x,y)$ that have the form

$$(\Phi'(y) \vee y = \mathbf{0}) \wedge y \neq x.$$

Hence the negation of the induction axiom attains the following simple form

$$\exists x > \mathbf{0} \forall y(\Phi(x,y) \rightarrow \Phi(x,S(y))). \tag{13}$$

Suppose $\Phi(x,y)$ is $\mathbf{Q}\bar{z}\phi(x,y,\bar{z})$, where $\mathbf{Q}$ denotes a quantifier prefix with variables $\bar{z}$. Thus a prenex form of (7) is

$$\exists x > \mathbf{0} \forall y \mathbf{Q}^{\perp}\bar{z}\mathbf{Q}\bar{u}\ (\phi(x,y,\bar{z}) \rightarrow \phi(x,S(y),\bar{u})), \tag{14}$$

where $\mathbf{Q}^{\perp}$ denotes the quantifier prefix dual to $\mathbf{Q}$. Let us interpret it in terms of games. We shall assume that the quantifier prefix starts with $\forall$. Let us call the games determined by

$\phi(x, y, \bar{z})$ *the small games*. Thus the game defined by formula (14) has two small games. It can be described in words as follows.

Games of type G_{IND}. *Small games indexed by two parameters x and y are given such that if $0 = y \neq x$ then the first player always wins, and if $y = x$ then the second player always wins. First Alice chooses an $x > 0$ and Bob chooses a y. Then there are two rounds. In the first round they play the small game with index (x, y) and with Bob as the first player. In the second round they play the small game with index $(x, y + 1)$ and Alice is the first player. Bob wins if he beats Alice in both rounds, otherwise Alice wins.*

Now we can apply the framework developed in the previous section to formula (14) and we get a sentence independent from PA. The sentence is:

The match principle for induction axioms. *For every match based on a game of type G_{IND} and every strategy of Alice, Bob can beat Alice's strategy on at least one board.*

This principle is independent from PA. It is not much more than a restatement of Herbrand's theorem for PA, but we can make it nicer. So far we have tacitly assumed that the small games are defined by a quantifier free arithmetical formula and the strategies of Alice are defined by arithmetical terms. It is clear that the principle is valid without such restrictions, but it seems that in order to formalize it in PA we have to use definable strategies and games. Yet there is a way out. The idea is to let Alice *pretend* to know the small games and to know a strategy to play the match, while letting Bob know nothing. Now Bob can win in two ways: either by beating Alice, or by catching Alice lying, which means that he will find her answers inconsistent with her claim that she uses particular small games and plays a fixed strategy.

Let us state it more precisely. Let a schedule R to play a match based on a game of type G_{IND} be given. Alice and Bob play the match without knowing the small games. For all boards, as soon as they finish a round, Alice decides who won in the small game. We shall call it *the virtual match* for R.

There is an easy strategy for Alice to win the virtual match, but the point is that they will play the match many times and Alice must pretend that the small games are the same and that she always uses the same strategy. More precisely this means that if Bob plays an initial part of the match in the same way as before, she must respond in the same way as before, and whenever the final position in a small game is the same as before, she must decide about the winner in the same way as before. Recall that Bob has to obey Rule $\mathcal{B}$. This rule is sort of the opposite to the first restriction on Alice: Bob must play in the same way in the same positions in each match, but in different matches he can use different moves.

Now we can state the principle.

The strong match principle for induction axioms. *For every schedule R to play a match based on a game of type G_{IND} and for every virtual match for R, there exist an m and a strategy for Bob to win at least once if they play the virtual match m times and Alice plays consistently.*

Proposition 5.1 *The strong match principle for induction axioms is a true principle, unprovable in Peano Arithmetic.*

Proof. 1. The proof of the principle is a simple application of König's lemma. Suppose that the strong game principle fails. Bob's strategy will be to try systematically all possible combinations of moves in the match. We fix an enumeration of all infinitely many possible moves of Bob. Given an m he will play the first m moves of this enumeration. We assume that for every m Alice has a winning strategy for playing the match m times. By König's lemma, Alice has a strategy to consistently play all infinitely many games that Bob wants to try so that she never loses. Let us fix also this strategy and let us play these two strategies against each other. Let S be the sequence of plays thus obtained. S contains, in particular, the information about who wins in small games. Let an index (x, y) be given. We define a small game for this index by taking all strings that occurred as the final positions with index (x, y) in S as the legal end positions $M_{x,y}$, and we take as $G_{x,y}$ those that Alice proclaimed to be the winning ones. The game $(M_{x,y}, G_{x,y})$ is determined, thus Bob can use a winning strategy either when playing as the first player or when playing as the second player. Hence, if Alice plays x on board j, he can take the largest $y < x$ such that the first player has a winning strategy and win on this board. Thus he can win on all boards. We need only to check that if Bob plays in such a way, all the resulting end positions in small games will occur in S. Indeed, since S was obtained by Bob's systematic trying all possible combinations of moves, also the play based on the optimal strategy must occur in S. Thus it is not possible that Alice has a winning strategy for every m.

2. It is clear that the strong match principle is stronger than principle with concrete games and strategies. Since the latter is unprovable the former is unprovable too. ∎

6 Games in Bounded Arithmetic

Bounded Arithmetic started with the seminal papers of Parikh [23] and Cook [9]. The currently most studied systems are $I\Delta_0 + \Omega_1$ of Paris and Wilkie [31], and S_2 and T_2 of Buss [6]. S_2 and T_2 are equivalent theories and they are conservative extensions of $I\Delta_0 + \Omega_1$. The importance of these theories stems from the fact that they are proof complexity counterparts of the Polynomial Hierarchy. These theories are too weak to prove the equivalence of the provability in the standard calculi (such as the sequent calculus) and Herbrand's provability.[2] However, for suitable formalizations, Gödel's second incompleteness theorem can be proved for Herbrand's proofs too. This was shown by Adamowicz [1]; further results of this type were proved by Adamowicz and Zbierski [4], Adamowicz [2], Willard [32], and Salehi [29].

In Bounded Arithmetic the semantical interpretation of Herbrand's disjunctions does not produce a Π_1 independent sentence, but for reasons different from those in PA. The formulas in the induction schema are bounded, thus the games are finite. The problem is, however, that terms define functions that we cannot interpret as functions. Provably total functions of S_2 increase the lengths of numbers at most polynomially, but the values of terms constructed from # grow faster. Therefore we cannot define the values of terms, i.e., expressed in the language of games, we cannot define Alice's moves. One may propose to restrict the concept of Herbrand's provability so that only terms whose values grow polynomially be

[2] There exist sequences of first order tautologies which have Hilbert style proofs of polynomial length, but the sizes of the shortest Herbrand proofs grow superexponentially.

allowed in Herbrand disjunctions, but then we are not be able to prove the incompleteness theorem. This route was studied by Krajíček and Takeuti [22] who defined a concept of a restricted consistency. One of the main restrictions was a bound on the terms occurring in proofs. They were able use these statements to prove certain separation results, but failed to prove the Second Incompleteness Theorem for such proofs. Thus it seems essential for the diagonalization used in the proof of the Second Incompleteness Theorem that we can speak about objects that we cannot interpret. In the previous sections we have avoided this problem by using Π_2 sentences. These sentences say that some fast growing function is total. Such sentences are not interesting in Bounded Arithmetic. Therefore, in this section we shall consider a related, but different problem: the problem of characterizing low complexity consequences of subtheories of S_2. Having nice characterizations may eventually help us to prove nonconservativity results for subtheories of S_2, which is an important problem in Bounded Arithmetic (we will say more about this problem in Section 8). We shall present a combinatorial principle based on games that axiomatizes all $\forall\exists$ theorems of S_2. The principle has a parameter k which roughly corresponds to the fragments S_2^k.

Before we start we have to recall some basic definitions and results. S_2 contains the standard arithmetical operations plus several others. In particular, it contains the length function $|x|$ and the smash operation $x\#y$ whose interpretations in the standard model are $\lceil \log_2(x+1) \rceil$ and $2^{|x|\cdot|y|}$. The smash function is included to enable definitions of operations that increase the lengths of numbers polynomially. The classes of bounded formulas Σ_i^b are defined in analogy with the arithmetical hierarchy except that occurrences of sharply bounded quantifiers are ignored. Sharply bounded quantifiers are bounded quantifiers whose bounds are of the form $x \leq |t|$. The intuition is that we can search all elements less than or equal to $|t|$ using a polynomial time algorithm.

Several types of induction schemas are used in Bounded Arithmetic. One of these is called $LIND$:

$$\forall x, p(\Phi(\mathbf{0}, p) \wedge \forall y(\Phi(y,p) \rightarrow \Phi(S(y),p)) \rightarrow \Phi(|x|,p)).$$

T_2 is axiomatized by a finite set of basic open axioms called $BASIC$ together with the usual induction schema (denoted IND) for bounded formulas. S_2 is the same except that $LIND$ is used in place of IND. The subtheories T_2^i and S_2^i, $i = 1, 2, \ldots$, are defined by restricting the IND and $LIND$ schemas to Σ_i^b formulas. Having a parameter p in the schemas is important for the definition of subtheories. The relation between the sets of provable sentences in these subtheories is:

$$S_2^i \subseteq T_2^i \subseteq S_2^{i+1}.$$

In particular $S_2 \equiv T_2$.

PV, introduced by Cook [9], is an equational theory whose function symbols define (in the standard model) all polynomial time computable functions and no other functions. (Of course, different function symbols may define the same function.) It is often convenient to have all these function symbols; therefore, we enlarge all the subtheories by PV. In such theories quantifier free formulas will define all relations computable in polynomial time and only those. In S_2^1 one can define all polynomial time computable functions using Σ_1^b formulas. S_2^1 is then shown to be a $\forall\Sigma_1^b$ conservative extension of PV (more precisely, of the first order theory axiomatized by the equalities of PV). Therefore, by adding PV to the fragments above we obtain conservative extensions of the original systems. The enlarged systems are denoted by $T^i_2(PV)$ and $S^i_2(PV)$ (cf. [21, 16]). It is well-known that in these theories Σ^b_i and Π^b_i formulas are equivalent to *strict* Σ^b_i, respectively Π^b_i, formulas, which means that no sharply bounded quantifier is allowed to precede a quantifier that is not sharply bounded. In the presence of all PV terms we can eliminate also the remaining sharply bounded quantifiers. Blocks of quantifiers of the same type can be replaced by a single one (using the pairing function, as usual) already in the weakest subtheory. Moreover, it is possible to axiomatize the subtheories T^i_2 and S^i_2 by the corresponding schemas restricted to strict Σ^b_i, or strict Π^b_i formulas.

We shall use the basic result known as *Parikh's Theorem,* which holds for all the theories mentioned above. It says that if T proves a sentence $\forall x \exists y\ \varphi(x,y)$, where φ is a bounded formula, then T proves $\forall x \exists y \leq t(x)\ \varphi(x,y)$ for some term $t(x)$.

Furthermore we need the following result of Buss [7]. It says that the *Strong Replacement Schema* for Σ^b_i formulas

$$\exists z \forall x \leq |\tau|((\exists y \leq \sigma\ \alpha(x,y)) \equiv ((z)_x \leq \sigma \wedge \alpha(x,(z)_x))),$$

is provable in S^i_2. Here τ and σ denote terms and $(z)_x$ denotes the x-th element of the sequence coded by z.

Recall that S^{i+1}_2 is $\forall\Sigma^b_{i+1}$ conservative over T^i_2. For these and other theorems that we shall use see also [6, 16, 11].

Now we turn our attention to games. The games in the following principles will be defined by boolean circuits, as described in Section 2. Thus by saying *'a game is given'* we mean that a boolean circuit for the game is given. For example, a PV quantifier-free formula of the form $\varphi(x, y_1, z_1, y_2, z_2, \ldots)$ defines a game for every fixed x. Because $\varphi(x, y_1, z_1, y_2, z_2, \ldots)$ is a polynomial time relation, there exists a circuit whose size is polynomial in $n = |x|$ for every number x. The circuit has $2k$ inputs for n-bit strings which correspond to the moves in a game of length $2k$. The transformation of a polynomial time relation into a series of circuits whose size grows polynomially is essentially Cook's Theorem. This theorem is provable in S^1_2; the proof is implicitly contained in the proofs of theorems about translations of arithmetical formulas into propositional formulas, see [20, 16]. Similarly, by saying that *'a strategy for Alice is given'* we shall mean that boolean circuits computing the moves of Alice are given.

In the principle we will have several games represented by circuits. We do not insist that different circuits define different sets of winning positions. In general the number of boards will be larger than the number of games, so one game will be assigned to several boards. For every game, Rule $\mathcal{B}$ will apply to the boards on which the game is played. The complementary game $G^\perp$ of a game G will be determined by the same circuit, except that the roles of who starts will be switched.

Formally such a match is given by numbers k, ℓ and n, a string w of numbers of length $k\ell$ in which every number $1 \leq j \leq \ell$ occurs exactly k times, circuits $C_1, \ldots, C_n$ defining games of length k, and an assignment $h : \{1, \ldots, \ell\} \rightarrow \{1, \ldots, n\} \times \{A, B\}$. k is the length of the games, ℓ is the number of boards, w is the schedule (the order of moves), h is the assignment of games to boards, where $h(j) = (i, A)$ (respectively $h(j) = (i, B)$) means that game i is assigned to board j and Alice (respectively Bob) starts.

Principle $\mathbf{A}_k$. *Let a match be given in which games of length k are played and in which Alice wins if she wins two complementary games. Let a strategy for Alice to play this match be given. Then it is possible to find moves of Bob which beat this strategy.*

The principle says, roughly speaking, that in every such match Bob is the person who has a winning strategy, but it is important how this statement is formalized. We do not say explicitly that Bob has a winning strategy, because we do not know how to express it in Bounded Arithmetic. Instead, we say that every given strategy of Alice can be beaten. We shall illustrate this concept on a couple of examples.

Example 1. Consider the match based on a single game G of length 3 and its dual $G^\perp$ in which the order of moves is as follows:

$$\begin{array}{lccccccccccc}
G: & a_1 & & & & b_3 & & & & a_5 & & \\
G^\perp: & & b_1 & & & & & a_4 & & & b_5 & \\
G: & & & & a_2 & & & & b_4 & & & a_6 & \\
G^\perp: & & & b_2 & & & a_3 & & & & & & b_6
\end{array}$$

Let γ be a formula defining the game G, let $s_1, \ldots, s_6$ be functions defining Alice's strategy for the match. Then the special case of Principle $\mathbf{A}_3$ determined by this match is the following sentence:

$$\exists b_1 \ldots \exists b_6 \, ((\neg\gamma(s_1, b_3, s_5(b_1, \ldots, b_4)) \wedge \neg\gamma(s_2(b_1, b_2), b_4, s_6(b_1, \ldots, b_5))) \vee$$

$$(\gamma(b_1, s_4(b_1, b_2, b_3), b_5) \wedge \gamma(b_2, s_3(b_1, b_2, b_3), b_6)) \wedge$$

$$b_1 = b_2 \wedge (s_1 = s_2(b_1, b_2) \to b_3 = b_4) \wedge (s_4(b_1, b_2, b_3) = s_3(b_1, b_2) \to b_5 = b_6))\,.$$

The first two lines express that there exist moves of Bob such that he either wins on the two boards with G or on the two boards with $G^\perp$. The last line expresses the Rule $\mathcal{B}$ for this match (in order to make the rule quite explicit we use two variable b_1 and b_2 instead of just one).

Example 2. Let us now consider a similar match with two games G_1 and G_2 of length 3.

$$\begin{array}{lccccccccccc}
G_1: & a_1 & & & & b_3 & & & & a_5 & & \\
G_1^\perp: & & b_1 & & & & & a_4 & & & b_5 & \\
G_2: & & & & a_2 & & & & b_4 & & & a_6 & \\
G_2^\perp: & & & b_2 & & & a_3 & & & & & & b_6
\end{array}$$

The corresponding special case of Principle $\mathbf{A}_3$ is:

$$\exists b_1 \ldots \exists b_6 \, ((\neg\gamma(s_1, b_3, s_5(b_1, \ldots, b_4) \vee \gamma(b_1, s_4(b_1, b_2, b_3), b_5))) \wedge$$

$$(\neg\gamma(s_2(b_1, b_2), b_4, s_6(b_1, \ldots, b_5)) \vee \gamma(b_2, s_3(b_1, b_2, b_3), b_6)))\,.$$

Notice that since every game occurs only once, the Rule $\mathcal{B}$ is empty in this case. It is important to realize that the distinction of games G_1 and G_2 is purely formal. If the sets of winning positions of G_1 and G_2 are the same we still get a sentence different from the one in the Example 1.

Theorem 6.1

(i) For all $k \geq 1$, Principle $\mathbf{A}_k$ is a $\forall\Sigma_1^b$ formula and it is provable in $S_2^k(PV)$;

(ii) For all $k \geq 2$, $S_2^1(PV)$ plus $\mathbf{A}_{k+4}$ proves all $\forall\exists$ theorems of $S_2^k(PV)$.

Proof. (i) The proof is similar to the proof of Lemma 4.2. As our means are more restricted now, we have to describe the proof more explicitly. First, let us recall the basic idea. We want to define Bob's moves as follows. If there exists a winning strategy from the position on a board that he should play, then we let him play a move that preserves this property. Furthermore, when he encounters the same position in the same game as it was on another board before, he will play in the same way, which will ensure Rule $\mathcal{B}$. Otherwise he plays arbitrarily.

To define the moves formally, we need the following generalization of the Strong Replacement, the *Schema of Dependent Choices.* We shall use the following notation: z_u denotes the $u+1$-st element of the sequence coded by z; $(z_0 \dots z_u)$ denotes the number that codes the sequence coded by z truncated to the first $u+1$ elements.

$$\forall x\, \{\forall u \leq |x| \exists v \leq x\ \alpha(x,u,v) \to \exists z\, [\forall u \leq |x|\ \alpha(x,u,z_u) \wedge \tag{15}$$
$$\forall u < |x| (\exists v \leq x(\alpha(x,u+1,v) \wedge \beta(x,(z_0 \dots z_u),v)) \to \beta(x,(z_0 \dots z_u),z_{u+1}))]\}\,.$$

This is provable for every Σ_k^b formulas α and β in T_2^k. It is essentially the definition of Σ_{k+1}^b computations, thus the proof of this schema is implicitly contained in the proof of definability of these computations.

First we shall use the Schema of Dependent Choices to define moves Alice and Bob. The match and Alice's strategy are coded by the number x. The formula $\alpha(x,u,z_u)$ expresses that z_u is a legal move in the match x in the step $u+1$.

The formula $\beta(x,(z_0 \dots z_u),v)$ expresses that after playing $z_0, \dots, z_u$,

1. if it is Alice's turn, then v is a move played according to her strategy;
2. if it is Bob's turn, then v is a "winning move" in the step $u+1$, or that it is the same move as he played on some board with the same game and same position before, in case Rule $\mathcal{B}$ applies to this move.

The condition that v is a winning move can be expressed as the following formula:

$$\forall b \leq |x| \bigvee_{i=1}^{k} \forall x_{i+2} \exists x_{i+3} \forall x_{i+4} \exists x_{i+5} \dots \gamma(x,b,(z_0 \dots z_u),v,x_{i+2},x_{i+3},\dots),$$

where $\gamma(x,b,(z_0 \dots z_u),v,x_{i+2},x_{i+3},\dots)$ expresses that in the match coded by x after playing $z_0 \dots z_u$ it is Bob's turn and he should play on the board b, and if $z_{\ell_1}, \dots, z_{\ell_{i-1}}$ are the previous moves on board b, then $(z_{\ell_1}, \dots, z_{\ell_{i-1}}, v, x_{i+2}, x_{i+3}, \dots, x_k)$ is a winning position for Bob on board b.

The formula α is Σ_1^b, the formula β is Σ_k^b (Π_{k-1}^b if $k > 2$). Hence this instance of the schema is provable in S_2^k.

Now, having defined the moves $z_0, \dots, z_m$, we prove the following. Let b be a fixed board on which they play a game G such that 'Bob has a winning strategy for G'. The last condition means that the sentence

$$\exists x_1 \forall x_2 \exists x_3 \forall x_4 \dots \gamma(x_1, \dots, x_k),$$

in case Bob starts, and

$$\forall x_1 \exists x_2 \forall x_3 \exists x_4 \dots \neg\gamma(x_1, \dots, x_k),$$

in case Alice starts, is satisfied (where γ defines G). Then at every stage of the play Bob has a winning strategy in the current position on b. Since k is a fixed number and they play only k moves on the board b, we do not need any induction to this end. Thus we get that Bob wins on all such boards. Since for every pair of complementary games G and $G^{\perp}$, Bob has a winning strategy for G or for $G^{\perp}$ (in the above sense), the sequence of moves $z_0, \dots, z_m$ proves that Bob can beat Alice's strategy.

(ii) We need a parameter-free version of the schema $LIND$. If we simply omitted parameters, we would get a very weak sentence, because all quantifiers in $\Phi(|x|)$ would be sharply bounded. Thus instead we introduce the following schema, which we shall call LI.

$$\forall x \left\{ \Phi(\mathbf{1}) \wedge \forall y \leq x (|y| < |x| \wedge \Phi(2^{|y|}) \to \Phi(2^{S(|y|)})) \to \Phi(2^{|x|}) \right\}. \tag{16}$$

In the sequel we shall omit the bound $y \leq x$ which is in the above formula only in order to formally satisfy the definition of a bounded formula. We shall use this schema without parameters which means that x is the only free variable of $\Phi(x)$. The function $2^{|x|}$ is definable, for instance, by $x\#1$. We also need the above schema with logarithmic number of possible parameters. This will be formalized by:

$$\forall x, p \left\{ \Phi(\mathbf{1}, 2^{|p|}) \wedge \forall y (|y| < |x| \wedge \Phi(2^{|y|}, 2^{|p|}) \to \Phi(2^{S(|y|)}, 2^{|p|})) \to \Phi(2^{|x|}, 2^{|p|}) \right\}.$$

In the following three lemmas we shall assume PV to be the base theory in which the implications are provable, and $i \geq 1$.

Lemma 6.2 *Parameter-free schema LI for strict Π^b_{i+2} formulas implies LI for strict Π^b_i formulas with logarithmic number of parameters. Moreover, for every finite set of strict Π^b_i formulas there exists a single strict Π^b_{i+2} formula with only one free variable such that the LI axioms with logarithmic number of parameters for these strict Π^b_i formulas are provable from LI for the Π^b_{i+2} formula.*

Proof. Let a Π^b_i formula $\Psi(x, p)$ be given. Define a Π^b_{i+2} formula $\Phi(z)$ by:

$$\forall w \forall v \left\{ \langle v, w \rangle \leq |z| \wedge \Psi(\mathbf{1}, 2^v) \wedge \forall y (|y| < |w| \wedge \Psi(2^y, 2^v) \to \Psi(2^{S(y)}, 2^v)) \to \Psi(2^w, 2^v) \right\},$$

where $\langle v, w \rangle$ denotes the usual pairing function. Now apply LI to $\Phi(z)$. $\Phi(\mathbf{1})$ is trivially true because $\langle 0, 0 \rangle = 0$. The implication $\Phi(2^{|z|}) \to \Phi(2^{S(|z|)})$ follows from the properties of the pairing function. Indeed, if $\langle v, w \rangle = z$ and we increase z to $z + 1$, then we have either $z + 1 = \langle v + 1, 0 \rangle$, in which case we use $\Psi(\mathbf{1}, v + 1)$ in the antecedent to derive it in the consequent, or $z+1=\langle v-1, w+1\rangle$ in which case we use the induction assumption and $\Phi(2^w, 2^{v-1}) \to \Phi(2^{S(w)}, 2^{v-1})$ to derive $\Psi(2^{w+1}, 2^{v-1})$. Thus we get $\forall z \Phi(2^{|z|})$. To get

$$\forall x, p \left\{ \Psi(\mathbf{1}, 2^{|p|}) \wedge \forall y(|y| < |x| \wedge \Psi(2^{|y|}, 2^{|p|}) \to \Psi(2^{S(|y|)}, 2^{|p|})) \to \Psi(2^{|x|}, 2^{|p|}) \right\}.$$

take $z = 2^{|\langle p,x\rangle|}$.

To prove the stronger version of the lemma just observe that we can encode Π_i^b formulas $\Psi_1(2^{|x|}, 2^{|p|}), \ldots, \Psi_\ell(2^{|x|}, 2^{|p|})$ by a single Π_i^b formula $\Psi(2^{|x|}, 2^{|p|})$ of the same complexity such that for $i = 1, \ldots, \ell$ and S_2^1 proves $\Psi(2^{|x|}, 2^{\langle |p|, |\mathbf{i}|\rangle}) \equiv \Psi_i(2^{|x|}, 2^{|p|})$, and $\exists q(|q| = \langle |p|, |\mathbf{i}|\rangle)$. ■

Lemma 6.3 *The schema LI for strict Π_{i+1}^b formulas with logarithmic number of parameters implies parameter-free induction IND for strict Π_i^b formulas.*

Proof. This is only a modification of a well-known proof that $\Pi_{i+1}^b LIND$ implies $\Pi_i^b IND$, but we must check that logarithmic number of parameters suffices for the parameter-free version of the induction.

Let a Π_i^b formula $\Psi(x)$, with a single free variable x, be given. Suppose

$$\Psi(\mathbf{0}) \wedge \forall y(\Psi(y) \to \Psi(S(y))) \tag{17}$$

is true. Let $\Psi'(x) := \forall y \le x \Psi(y)$. Then (17) is true also for Ψ'. Suppose that $\Psi(a)$ is false for some a. Then we also have $\neg\Psi'(2^{|a|})$. Let

$$\Phi(y, z) \ := \ \forall x \le z(\Psi'(x) \to \Psi'(x+y)).$$

We have $\Phi(\mathbf{1}, 2^{|a|})$ by (17). Suppose that $\Phi(y, 2^{|a|})$ is true. Then, for every x such that $\Psi'(x)$, we have $\Psi'(x+y)$, hence also $x+y < 2^{|a|}$. Thus applying $\Phi(y, 2^{|a|})$ we get $\Psi'(x+y+y)$. This proves $\forall y(|y| < |a| \wedge \Phi(2^{|y|}, 2^{|a|}) \to \Phi(2^{S(|y|)}, 2^{|a|}))$. Hence, using LI for Φ we get $\Phi(2^{|a|}, 2^{|a|})$. Since we assume $\Psi'(\mathbf{0})$ and $\neg\Psi'(2^{|a|})$, this is a contradiction. So we have shown $\forall x \Psi(x)$. ■

Lemma 6.4 *Parameter-free induction IND for strict Π_{i+2}^b formulas implies induction IND for strict Π_i^b formulas (with parameters).*

Proof. Let a Π_i^b formula $\Psi(x, p)$, with x and p the only free variables, be given. Define a Π_{i+2}^b formula $\Phi(z)$ by:

$$\forall w \le z \forall v \le z(\langle v, w\rangle \le z \wedge \Psi(\mathbf{0}, v) \wedge \forall y < w(\Psi(y, v) \to \Psi(S(y), v)) \to \Psi(w, v)),$$

where $\langle v, w\rangle$ denotes the usual pairing function. Now apply induction to $\Phi(z)$. $\Phi(\mathbf{0})$ is trivially true. The implication $\Phi(z) \to \Phi(S(z))$ follows from the properties of the pairing function in the same way as in the proof of Lemma 6.2. Thus we get $\forall z \Phi(z)$. To get

$$\Psi(\mathbf{0}, p) \wedge \forall y(\Psi(y, p) \to \Psi(S(y), p)) \to \Psi(x, p),$$

take $z = \langle p, x\rangle$. ■

Now we can start proving (ii) of the theorem. Let $k \geq 2$. Suppose that $S_2^k(PV)$ proves a sentence $\Psi := \forall x \exists y\ \psi(x, y)$, where ψ is open. Then Ψ is provable:

1. in $T_2^{k-1}(PV)$, because by Parikh's Theorem we can add a bound to the existential quantifier and $S_2^k(PV)$ is $\forall\Sigma_k^b$ conservative over $T_2^{k-1}(PV)$,

2. in parameter-free IND for strict Π_{k+1}^b formulas, by Lemma 6.4 and the well-known fact that $T_2^{k-1}(PV)$ can be axiomatized by IND for strict Π_{k-1}^b formulas,

3. from LI for Π_{k+2}^b formulas with logarithmic number of parameters, by Lemma 6.3,

4. from a single instance of parameter-free LI for a strict Π_{k+4} formula Φ, by Lemma 6.2.

Thus we have in PV

$$\Lambda_\Phi \to \forall x \exists y\ \psi(x, y),$$

where Λ_Φ is (16). By renaming the variable x to u, Λ_Φ gets the form $\forall u\ \lambda_\Phi(u)$ with $\lambda_\Phi(u)$ a bounded formula. By Parikh's Theorem, we have in PV

$$\forall x \exists u \leq s(x) \exists y (\lambda_\Phi(u) \to \psi(x, y)),$$

for some term $s(x)$. Hence in the predicate calculus we have

$$\Theta \to \forall x \exists u \leq s(x) \exists y (\lambda_\Phi(u) \to \psi(x, y)).$$

where Θ is a conjunction of some axioms of PV and some equality axioms. We move Θ to ψ; thus we get

$$\forall x \exists u \leq s(x) \exists y \exists y_1' \ldots \exists y_h' (\lambda_\Phi(u) \to (\theta(y_1', \ldots, y_h') \to \psi(x, y))),$$

where $\theta(y_1' \ldots y_h')$ is the open part of Θ. To prove $\exists y\ \psi(x, y)$ for a given x, it suffices to prove $\exists y \exists y_1' \ldots \exists y_h' (\theta(y_1', \ldots, y_h') \to \psi(x, y))$, because $\theta(y_1' \ldots y_h')$ is a conjunction of some equality axioms and axioms of PV. Since it is a formula of the same type as $\exists y\ \psi(x, y)$, we shall omit Θ in the rest of the proof.

Furthermore, we can assume that Φ is in the prenex form with $k+4$ bounded quantifiers and with no other quantifiers. Let $\Phi(z)$ have the form $\mathbf{Q}\bar{w}\ \varphi(z, \bar{w})$, with φ quantifier-free. Thus we assume that the following formula is provable in the predicate calculus:

$$\forall x \exists u \forall z \exists y \mathbf{Q}\bar{w} \mathbf{Q}^\perp \bar{w}' \mathbf{Q}\bar{w}'' \mathbf{Q}^\perp \bar{w}'''$$

$$\left\{u > s(x) \vee \left[\varphi(\mathbf{1}, \bar{w}) \wedge (|z| < |u| \wedge \varphi(2^{|z|}, \bar{w}') \to \varphi(2^{S(|z|)}, \bar{w}'')) \wedge \neg\varphi(2^{|u|}, \bar{w}''')\right] \vee \psi(x, y)\right\}.$$

Now we apply Herbrand's theorem and group the formulas that come from the induction axiom together and those that come from the formula $\forall x \exists y\ \psi(x, y)$ together. Thus the Herbrand disjunction gets the form:

$$\Delta(x, |z_1|, \ldots, |z_\ell|, \bar{q}) \vee \bigvee_i \psi(x, t_i), \tag{18}$$

where $z_1, \ldots, z_\ell$ are variables that come from z, $\bar{q}$ are variables that come from the universally quantified variables in $\bar{w}, \bar{w}', \bar{w}'', \bar{w}'''$, and t_i are terms which may depend on all variables.

Now we shall describe a proof which can be formalized in $S^1_2(PV)$. First we shall present the proof and then explain how it can be formalized.

Let n be given, let $m = |s(n)|$. First we substitute $\mathbf{n}$ for x in $\Delta(x, z_1, \ldots, z_\ell, \bar{q})$. Then for $j = 0, \ldots, m-1$ we consider substitutions that substitute $\mathbf{j}$ simultaneously for all $|z_1|, \ldots, |z_\ell|$. Recall that m is logarithmic in n, thus we can list all these formulas. If we find an assignment $\bar{b}$ to the free variables $\bar{q}$ such that formula $\Delta(\mathbf{n}, \mathbf{j}, \ldots, \mathbf{j}, \bar{\mathbf{b}})$ is falsified for some j, then $\bigvee_i \psi(\mathbf{n}, t_i[\mathbf{n}, \mathbf{j}, \bar{\mathbf{b}}])$ will be true, where $t_i[\mathbf{n}, \mathbf{j}, \bar{\mathbf{b}}]$ denotes the result of substituting n for x, j for $z_1, \ldots, z_\ell$ and $\bar{b}$ for $\bar{q}$ in t_i. Thus we get $\exists y\ \psi(\mathbf{n}, y)$. So it remains only to prove that we can find such an assignment using Principle $\mathbf{A}_{k+4}$.

The disjunction $\Delta(\mathbf{n}, \mathbf{j}, \ldots, \mathbf{j}, \bar{b})$ has the following form:

$$\bigvee_{i=1}^{\ell} \{\phi(\mathbf{1}, W_{i1}) \wedge (\mathbf{j} \geq r_i(\mathbf{n}, \mathbf{j}, \bar{q}) \vee \neg\phi(2^{\mathbf{j}}, W_{i2}) \vee \phi(2^{S(\mathbf{j})}, W_{i3})) \wedge$$
$$r_i(\mathbf{n}, \mathbf{j}, \bar{q}) \leq s(\mathbf{n}) \wedge \neg\phi(2^{|r_i(\mathbf{n}, \mathbf{j}, \bar{q})|}, W_{i4})\}\,. \tag{19}$$

In this formula, $W_{i1}, \ldots, W_{i4}$ denote strings of terms and variables corresponding to the existentially, respectively, universally quantified variables of the quantifier prefixes $\mathbf{Q}\bar{w}$, $\mathbf{Q}^{\perp}\bar{w}'$, $\mathbf{Q}\bar{w}''$, $\mathbf{Q}^{\perp}\bar{w}'''$ of the occurrences of Φ. The terms $r_i(\mathbf{n}, \mathbf{j}, \bar{q})$ come from u. We put the whole string of variables $\bar{q}$ there, but in fact r_i does not depend on the variables of $W_{i1}, W_{i2}, W_{i2}, W_{i4}$.

Let us interpret the disjunction (19) in terms of games. The game G_e will be the game defined by $\phi(2^{|e|}, \bar{w})$. Thus the terms in (19) define a strategy for Alice to play games G_e and $G_e^{\perp}$ on 4ℓ boards. There is a problem, however, because the games corresponding to the formulas $\neg\phi(2^{|r_i(\mathbf{n}, \mathbf{j}, \bar{q})|}, W_{i4})$ are not fixed, they are determined by Alice's moves $|r_i(\mathbf{n}, \mathbf{j}, \bar{q})|$. Recall that we have $|r_i(\mathbf{n}, \mathbf{j}, \bar{q})| \leq |s(\mathbf{n})| = \mathbf{m}$. So we resolve this problem by introducing $m+1$ boards with games $G_0^{\perp}, \ldots, G_m^{\perp}$ for every such formula. Then we interpret the term $|r_i(\mathbf{n}, \mathbf{j}, \bar{q})|$ as Alice's decision on which board of these $m+1$ boards she will concentrate, while ignoring the others. Formally, she will play according to the terms in W_{i4} on the chosen board and she will play 0 in every move on all m other boards.

A satisfying assignment for $\Delta(\mathbf{n}, \mathbf{j}, \ldots, \mathbf{j}, \bar{q})$ is an assignment $\bar{b}$ that makes one of the formula in the disjunction true. If we interpret the assignment $\bar{b}$ as Bob's moves, it means that Alice wins the following games on some boards:

1. G_0, and
2. $G_{r(j)}^{\perp}$, for some $r(j) \leq m$, and
3. if $j < r(j)$, then she also wins $G_j^{\perp}$ or G_{j+1}.

The match and the strategy of Alice to which we apply Principle $\mathbf{A}_{k+4}$ is determined by the conjunction

$$\bigwedge_{j=0}^{m-1} \Delta(\mathbf{n}, \mathbf{j}, \ldots, \mathbf{j}, \bar{q}). \tag{20}$$

This match is the union of the matches described above for all $j = 0, \ldots, m-1$. (The same is true about the strategies, they use only Bob's moves on the components. This fact is not relevant for what we want to prove, but it might be used to obtain some strengthening of the theorem. It would also suffice to require Rule $\mathcal{B}$ to be satisfied only locally on each block.)

Suppose that Alice wins this match. Let $r = \min_j r(j)$. Then there are boards on which she wins

1. G_0, and
2. $G_j^\perp$ or G_{j+1} for all $j = 0, \ldots, r-1$, and
3. $G_r^\perp$.

Thus Alice always wins a pair of complementary games. Hence Principle $\mathbf{A}_{k+4}$ implies that there is a falsifying assignment to (20), and we get Ψ.

It remains to show that the above argument can be done in $S^1_2(PV)$.

First we should realize that the Herbrand proof (18) is of constant size. For formulas of constant size (with terms of constant size) we can define satisfiability and prove the Tarski conditions without using any induction. In the proof we used the notation

$$\alpha(\mathbf{n}_1, \ldots, \mathbf{n}_h)$$

which means that numerals for $n_1, \ldots, n_h$ are substituted in α. This is convenient if we need to write complicated formulas, but for formalizing the proof it is better to use a relation of satisfiability. So the above formula should be

$$Sat(\alpha; n_1, \ldots, n_h),$$

where Sat is a suitable formula defining satisfiability for a class of open formulas. Since we have to introduce $m+1$ boards for the last term in (19), and then in (20) we have m terms, we need to extend the satisfaction relation to formulas that are unbounded disjunctions and unbounded conjunctions of open formulas of a fixed size. Since m is of logarithmic size, this presents no problems: the Tarski conditions are proved using $\Sigma^b_1 LIND$. The construction of the match uses only elementary operations with strings and graphs and all objects are of polynomial size in the $\log n$. It is well-known that all such elementary operations can be formalized in $S^1_2(PV)$ in such a way that the basic properties of these constructions are provable. The transition from $\phi(2^{|e|}, \bar{w})$ to a boolean circuit for the game G_e is done by the formalization of the proof of Cook's Theorem. It is well-known that $S^1_2(PV)$ proves the existence of the minimum of a set of logarithmic size (the set is assumed to be coded by a number). Hence we can prove the existence of $r = \min_j r(j)$. To find a pair of complementary games in which Alice wins we only need to find the largest j, $j < r$, such that she wins only the second game from the pair $G_j^\perp, G_{j+1}$. This follows from the same principle as the previous thing.

Thus we have proved (ii). ∎

We believe that the gap of size 4 between the upper and the lower bounds on the Principle $\mathbf{A}_k$ should be possible to reduce, maybe, even to zero. There are several possible avenues to achieve it, but the solution may also be their combination. The first one is to extend the techniques of [14, 5] to get a more suitable conservativity result between the parameter free and general induction schemas. The results of Bloch [5] about the conservativity of general theories over the parameter-free versions cannot be applied directly, because they use general Σ_k^b formulas, whereas we need strict Σ_k^b formulas. One can also try to extend the simulation of proofs by the principle to an induction schema with (logarithmic number of) parameters. Another approach is to use the quantified propositional sequent calculus. In [20] we showed that for $2 \leq j \leq i$, the $\forall\Sigma_j^b$ theorems of T_2^i (which are the same as those of S_2^{i+1}) can be characterized by certain reflection principles for subsystems of this calculus. The deadline for submitting the paper does not provide us enough time to optimize our result.

We can make this principle more uniform by taking universal schedules. We shall not do it in general, instead we shall consider only the two simplest instances. For such small values of k one can state this principle in a fairly explicit way by taking the order of moves that is the most favorable for Alice. The principle says that Alice must always loose, thus an ordering that is the most favorable for Alice produces the strongest principle. Due to the gap between the theories considered in (i) and (ii) of Theorem 6.1, we do not know if $\mathbf{A}_1$ and $\mathbf{A}_2$ imply anything interesting; they just may be all provable in S_2^1. However, the explicit presentation of them below may help us to clarify their status. In particular, we may compare them with other characterizations of $\forall\exists$ theorems of subtheories of S_2.

The general idea used here is to schedule the moves of Bob as early as possible. Thus first to go are all Bob's moves on the boards on which he starts. As soon as Alice plays a move on a board B, then immediately Bob answers on B. In both instances we will assume that we have n pairs of of complementary games $G_i, G_i^\perp$; in all G_i Bob is the player who starts. Every G_i will be played on n boards and every $G_i^\perp$ will be played on n boards too. However, we shall state the principles without talking about games and strategies. Let $Circ(n,m)$ denote the set of all circuits computing boolean function with n input bits and m output bits.

Principle $\mathbf{A}_1$. *Suppose that circuits $C_1,\dots,C_n \in Circ(n,1)$, and $\alpha_{ij} \in Circ(n^3,n)$, for $i,j = 1,\dots,n$, are given. Then there exists an assignment $\bar{b} = (b_1,\dots,b_n)$, $b_i \in \{0,1\}^n$, such that for all $i,j \in \{1,\dots,n\}$*

$$C_i(b_i) \geq C_i(\alpha_{ij}(\bar{b})). \tag{21}$$

Let us note that the condition on the assignment $\bar{b}$ can also be stated as follows. For all $i \in \{1,\dots,n\}$,

1. either $C_i(b_{ij}) = 1$,
2. or $C_i(\alpha_{ij}(\bar{b})) = 0$ for all $j \in \{1,\dots,n\}$.

We already know that $\mathbf{A}_1$ is provable in S_2^1 from the general Theorem 6.1, but it is instructive to prove this special case. The proof is very simple: the Σ_1^b Strong Replacement gives us immediately an assignment $\bar{b}$ such that

$$\forall i, j((\exists x(C_i(x) = 1)) \equiv C_i(b_i) = 1). \tag{22}$$

This $\bar{b}$, clearly, satisfies (21).

Recall that Buss's Witnessing Theorem for S^1_2 implies that there is a polynomial time algorithm to construct an assignment $\bar{b}$ satisfying (21) from given circuits $C_1, \ldots, C_n$ and α_{ij}. To find this algorithm is fairly simple and we leave it to the reader. But it is interesting to observe that the proof above gives no hint at all how to do it—an assignment satisfying (22) cannot be constructed in polynomial time, unless $\mathbf{P} = \mathbf{NP}$!

Principle $\mathbf{A}_2$. *Suppose that circuits $C_1, \ldots, C_n \in Circ(2n, 1)$, $\alpha_{i,j} \in Circ(n^3 + (j-1)n^2, n)$, and $\alpha'_{ij} \in Circ(2n^3, n)$, for $i, j = 1, \ldots, n$ are given. Then there exists an assignment $(\bar{b}, \bar{b'})$, $\bar{b} = (b_i)^n_{i=1}$, $\bar{b'} = (b'_{ij})^n_{i,j=1}$, $b_i, b'_{ij} \in \{0,1\}^n$, such that for all $i, j, k \in \{1, \ldots, n\}$,*

$$C_i(b_i, \alpha'_{ij}(\bar{b}, \bar{b'})) \geq C_i(\alpha_{ik}(\bar{b}, \bar{b'}_1, \ldots, \bar{b'}_{k-1}), b'_{ik}),$$

and

$$\alpha_{ij}(\bar{b}, \bar{b'}_1, \ldots, \bar{b'}_{j-1}) = \alpha_{ik}(\bar{b}, \bar{b'}_1, \ldots, \bar{b'}_{k-1}) \Rightarrow b'_{ij} = b'_{ik},$$

where $\bar{b'}_j$ denotes $(b'_{1j}, b'_{2j}, \ldots, b'_{nj})$.

Here is an explanation in terms of games. First Bob plays all first moves on the boards with $G_1, \ldots, G_n$, i.e., on the boards where he starts. Then Alice must play. She can postpone her moves in G_i to the very end, because there are no other moves in these games than hers. Thus she has to play some $G_i^\perp$. As soon as she plays Bob responds. If we denote by B_{ij} (respectively B'_{ij}) boards with G_{ij} (respectively $G_{ij}^\perp$), then Alice will play in this order:

$$B'_{11}, B'_{21}, \ldots, B'_{n1}, B'_{12}, B'_{22}, \ldots, B'_{n2}, \ldots\ldots, B'_{1n}, B'_{2n}, \ldots, B'_{nn}.$$

(This order is not forced, but it is universal in the sense that every order on a smaller set can be embedded into it. Here we use the simple principle that given an alphabet $c_1, \ldots, c_m$, every word of length ℓ can be obtained from the word $(c_1c_2\ldots c_m)^\ell$ by erasing certain letters.) Finally Alice plays her second moves on all boards B_{ij}.

F. Ferreira found a characterization of $\forall\exists$ theorems of T^2_2 [10]. As it is also based on Herbrand's theorem, the best way to describe it is to use games. Below we shall assume that 0-1 strings are encoded by numbers in the natural way; all zeros string is the number 0, $00\ldots01$ is 1, $00\ldots10$ is 2 etc.

Let a circuit $C \in Circ(2n, 1)$ be given such that $C(0, y) = 1$ for all $y \in \{0,1\}^n$. Alice and Bob play the following match. Bob chooses $b \in \{0,1\}^n$. Then in several rounds Alice alternatively tries to find

1. an a_i such that $C(b, a_i) = 0$,
2. or an a'_i such that $a'_i > b$ (where we compare the numbers that represent the strings) and Bob cannot respond with a b'_i such that $C(a'_i, b'_i) = 0$.

So the order of moves is $b, a_1, a'_1, b'_1, a_2, a'_2, b'_2, \ldots$. If she succeeds to find such an a_i or a'_i in at most n rounds, she wins. Otherwise Bob wins.

Principle $\mathbf{M}_2$ *For every n, every circuit $C \in Circ(2n, 1)$, and every strategy of Alice to play the above match, it is possible to find moves of Bob that beat Alice's strategy.*

Theorem 6.5 [10] *Principle* $\mathbf{M}_2$ *is a* $\forall\Sigma_1^b$ *sentence provable in* T_2^2 *and every* $\forall\Sigma_1^b$ *theorem of* T_2^2 *is provable in* $S_2^1 + \mathbf{M}_2$.

Due to the $\forall\Sigma_1^b$-conservativity, the theorem is also true with T_2^2 replaced by S_2^3.

7 Matches with synchronized games.

The principle introduced in the previous section is still too complicated. We hope that eventually we will be able to prove that some nice special case of it is as strong as the general case. In this section we propose such a special case, but we are not able to prove that it is as strong as Principles $\mathbf{A}_k$ or that it implies all $\forall\Sigma_1^b$ theorems of S_2^i for some $i > 1$.

We know that principles based on playing games sequentially, ie., one after the other, are too weak. This follows from the proof of Proposition 4.3. The other extreme is to play games in the *synchronized way,* ie., to play the first move on all boards at once, then to play the second move on all boards at once, and so on. Another question is whether we need to have different games in the match or it is enough to have only one game and its complement. We shall adopt the latter plus another restriction in the following principle.

Principle $\mathbf{B}_{\ell,k}$**.** *Let a game G of length k be given in which Bob starts. Let two strategies σ and σ' for Alice be given such that σ is a strategy to play G on ℓ boards in the synchronized way, and σ' is a strategy to play $G^\perp$ on ℓ boards in the synchronized way. Then either there exist Bob's moves to beat σ on all boards or there exist Bob's moves to beat σ' on all boards.*

Principle $\mathbf{B}_k$ *is the sentence* $\forall x\ \mathbf{B}_{|x|,k}$. [3]

To get a match of the type used in $\mathbf{A}_k$, use 2ℓ boards, play G on ℓ boards and $G^\perp$ on the other ℓ boards and think of σ, σ' as a strategy for this match. Thus the additional restriction mentioned above is the splitting of the match in which both G and $G^\perp$ are played into two: one for G and one for $G^\perp$. Hence $\mathbf{B}_k$ is a special case of $\mathbf{A}_k$.

Since we want to get rid of the Rule $\mathcal{B}$, we shall restrict the class of strategies that Alice can use. We shall require that when playing the first moves on the boards on which she starts, she must play different moves on all boards, and, similarly, when she plays the second moves on the boards on which Bob starts, she must play different moves on all boards. Then what remains from Rule $\mathcal{B}$ is only the requirement that Bob must play the same move on all boards on which he starts. *Thus Principle* $\mathbf{B}_k$ *means the above principle with this additional restriction.*

There are indications that $\mathbf{B}_k$ may be weaker $\mathbf{A}_k$. In particular, we only know that $\mathbf{A}_k$ is provable in S_2^k. However, it follows from an unpublished result of J. Håstad that in S_2^k one can prove $\mathbf{B}_{k+3}$. The only hardness result that we have so far is rather weak. We studied an infinite version of this principle in [28]. The infinite principle is exactly the same except that the length of the games is ω, thus we can call it $\mathbf{B}_{\ell,\omega}$. It is well-known that there are infinite games which are not determined, ie., none of the two players has a winning strategy. If the games are not determined, such principles may fail. Indeed, we showed that already $\mathbf{B}_{2,\omega}$ is false [28]. That said, we do not know anything about provability of $\forall y\ \mathbf{B}_{2,|y|}$ in S_2 or its subtheories

[3] We can only code games with logarithmic number of boards, therefore we use $|x|$.

Though we do not know precisely the power of this principle it is interesting to study it because it has interesting properties. The most interesting of these is the fact that, for a given ℓ and k, one can fix the strategies σ and σ'. The idea is that the most general strategies are the *copycat* strategies which simply copy the moves of the opponent. We shall describe it below and show that the principle restricted to copycat strategies is equivalent to the general one.

We need to introduce suitable notation. Once we have it, the proofs will be easy. Let ℓ and k be fixed. We assume that k is a constant, and ℓ can be arbitrary. In $\mathbf{B}_{\ell,k}$ we need a game and two strategies. We shall represent it by the following structure.

$$(M_1, M_2, \ldots, M_k; G, \sigma_1, \ldots, \sigma_k)$$

The sets M_i are possible moves in the step i of the game, G is the set of winning positions, and

$$\sigma_j : M_1^\ell \times \ldots \times M_{j-1}^\ell \to M_j^\ell.$$

We assume that if j is odd, then σ_j depends only on even moves, and if j is even, then σ_j depends only on odd moves. The strategies are $\sigma = (\sigma_2, \sigma_4, \ldots)$ and $\sigma' = (\sigma_1, \sigma_3, \ldots)$.

We shall imagine playing ℓ games of length k as constructing an $\ell \times k$ matrix, where the rows correspond to boards, and the columns correspond to moves. The mapping σ_j tells us how to construct the j-th column from the first $j-1$ columns.

Our first observation is that we can assume w.l.o.g. that $M_1 = \{1, \ldots, \ell\}$. Indeed, by renaming moves we can assume that if Alice starts then she plays i on the i-th board (formally, $\sigma_1 = (1, \ldots, \ell)$; recall that Alice has to play different moves on different boards when she starts). Then we redefine G by prohibiting to start with an element not in $\{1, \ldots, \ell\}$. Since this restriction concerns only Bob, the restricted principle is at least as strong as the unrestricted one.

The sets of moves for the copycat strategy are defined as follows.

1. $N_1 = N_2 = \{1, \ldots, \ell\}$;
2. for $j > 2$, $N_j = N_{j-1}^{\ell-1}$.

The strategies are defined by $\rho_1, \ldots, \rho_k$, such that

1. ρ_1 and ρ_2 simply print i in the i-th board; thus ρ_2 does not depend on the first moves;
2. for $j > 2$, on the i-th board ρ_j prints the $j-1$-st moves from all other boards; in terms of matrices, ρ_j produces the j-th column such that in every row i the entry is the string of all elements from the $j-1$-st column with the i-th element omitted.

These functions determine strategies $\rho = (\rho_2, \rho_4, \ldots)$ and $\rho' = (\rho_1, \rho_3, \ldots)$.

The idea of the copycat strategy is that it records information about all boards on each board, without any redundancy. Thus when Bob starts with the same moves on all ℓ boards, Alice does not have to copy the moves from other boards to the given board. What she does instead is encoding the number of the board into the position on the board. Similarly, the reason for omitting the i-th element when copying the previous moves onto the i-th board is that this information is already present on the i-th board.

Given a matrix U, we shall denote by $row_i(U)$ (respectively $col_j(U)$) the i-th row (respectively j-th column) of U. Thus $row_i(U)$ represents the position on the i-th board and $col_j(U)$ represents the moves in the j-th step. We shall consider matrices of the dimensions $\ell \times j$, $1 \leq j \leq k$, that will represent positions of playing ℓ copies of a game for j steps. We shall say that U *was played using* ρ (respectively using ρ') if for every even (respectively odd) $j' \leq j$,

$$col_{j'}(U) = \rho_{j'}(col_1(U) \dots col_{j'-1}(U)) \text{ (respectively, } col_{j'}(U) = \rho'_{j'}(col_1(U) \dots col_{j'-1}(U))).$$

If U was played using ρ, we also assume that the entries in the first column are all the same.

Lemma 7.1 *For every $1 \leq j \leq k$ and $w \in N_1 \times \ldots \times N_j$ there exists a unique i and a unique $\ell \times j$ matrix U such that*

1. *$row_i(U) = w$, and*
2. *if j is even then U was played using ρ, if j is odd then U was played using ρ'.*

Proof. For $j = 1, 2$ it follows immediately from definitions. For $j > 2$ we shall use induction. Let $w = w_1 \dots w_j$. By the induction assumption we have a unique i and a unique U' for $w_1 \dots w_{j-2}$. Then w_{j-1} and w_j determine uniquely the $j-1$-st column of U. To get the j-th column of U apply ρ_j. ■

Theorem 7.2 *Let $k \geq 1$ be fixed. Then it is provable in S^1_2 that $\mathbf{B}_{\ell,k}$ follows from its special case where the strategies are fixed to be the copycat strategies.*

Proof. Let $(M_1, M_2, \dots, M_k; G, \sigma_1, \dots, \sigma_k)$ be a general structure. We shall construct a homomorphism μ_k from $(N_1, N_2, \dots, M_k; \rho_1, \dots, \rho_k)$ into $(M_1, M_2, \dots, M_k; \sigma_1, \dots, \sigma_k)$. Then we take the inverse image of G and obtain an $H \subseteq N_1 \times \ldots \times N_k$. Thus we get a game for the match with the copycat strategies. If we find moves for Bob that beat ρ or ρ', then their homomorphic image will be moves for Bob that beat σ or σ'.

The homomorphism μ_k will be defined by defining its initial segments $\mu_j : N_1 \times \ldots \times N_j \to M_1 \times \ldots \times M_j$ by recursion. μ_1 is the identity mapping. Let $j > 1$, let $w \in N_1 \times \ldots \times N_j$. Take the unique i and the unique $\ell \times j$ matrix U with the properties guaranteed by the lemma. Let U' be U less the last column. Then we put

$$\mu_j(w) = \mu_j(row_i(U)) = \mu_{j-1}(row_i(U'))\ \sigma_{ij} \begin{pmatrix} \mu_{j-1}(row_1(U')) \\ \vdots \\ \mu_{j-1}(row_\ell(U')), \end{pmatrix} \tag{23}$$

where σ_{ij} denotes the entry in the i-th row of the column determined by σ_j. We define

$$H = \{w \in N_1 \times \ldots \times N_k;\ \mu_k(w) \in G\}.$$

Now we claim that if U is a $\ell \times k$ matrix played using ρ (respectively ρ'), then its image

$$W = \left(\begin{array}{c} \mu_k(row_1(U)) \\ \vdots \\ \mu_k(row_\ell(U)), \end{array} \right)$$

is a matrix played using σ (respectively σ'). This easily follows by induction from (23). (Notice that, eg., if U was played using ρ, then we shall use initial parts of U when defining *even* columns of W using the formula (23). However, for defining odd columns we shall use completely different matrices.)

Thus we have reduced the task of finding Bob's moves that beat Alice in the general match to the same task in the game with the copycat strategies. It remains to argue that the reduction can be done in S^1_2. Since all the arguments and transformations were completely elementary, the problems boils down to showing that in order to define $\mu_k(w)$ we only need a polynomial number of steps. In (23) we need to know μ_{j-1} only for ℓ values when defining one value of μ_j. Hence to define $\mu_k(w)$ we need only $O(\ell^k)$ steps. Since k is constant this is polynomial.

■

Corollary 7.3 *For every $k \geq 1$, S^1_2 proves $\forall x\ \mathbf{B}_{\lceil \sqrt[k]{\log\log x}\rceil,k}$.*

Proof. If $\ell = \lceil \sqrt[k]{\log\log x}\rceil$, then the size of $N_1 \times \ldots \times N_k$ is polynomial in the length of x. Hence a winning strategy for one of the players can be coded. Then we can use this strategy to defines Bob's moves in the match. ■

This theorem enables us to present Principle $\mathbf{B}_{\ell,k}$ in terms of a problem about coloring a concrete hypergraph. Given ℓ and k we define a hypergraph $\Gamma_{\ell,k}$ as follows. The hypergraph has edges of size ℓ of two types: blue and red. The set of vertices is $V = N_1 \times \ldots \times N_k$ as defined before. A blue edge is an ℓ element subset of V whose elements are rows of a matrix played using ρ; a red edge is an ℓ element subset of V whose elements are rows of a matrix played using ρ'. An $H \subseteq V$ determines a coloring in which v is blue if $v \in H$, otherwise v is red. Now we can state Principle $\mathbf{B}_{\ell,k}$ in the following equivalent form.

An equivalent form of Principle $\mathbf{B}_{\ell,k}$. *There exists no coloring of $\Gamma_{\ell,k}$ such that every blue hyperedge contains a blue vertex and every red hyperedge contains a red vertex.*

Krajíček observed (see [16], Theorem 14.2.3) that given any **NP**-complete problem X and any $\forall\Sigma^b_0$ sentence Φ, we can represent Φ as the statement $\forall n\ s(n) \notin X$ for some concrete sequence $s(n)$, $n = 1, 2, \ldots$. (This can be shown by a more or less straightforward application of Cook's theorem.) Thus we can take the consistency statement for the theory in question and we get a "combinatorial" unprovable sentence. However, the sequence $s(n)$ is some strange sequence with a very complicated definition, thus it is difficult to find any "mathematical" interpretation of such sentences. Our sequence of hypergraphs is fairly explicitly defined. In fact, there is a simple definition of these hypergraphs by recursion on k. But our statement has two drawbacks. Firstly, it is $\forall\Sigma^b_1$, because the hypergraphs are not represented by vertices and hyperedges, they are defined by circuits, and so are the colorings. Secondly, which is much worse, we are not able to prove that our principle is unprovable in any reasonably strong theory.

We conclude this section by noting that the theorem that the infinite principle $\mathbf{B}_{2,\omega}$ is false was proved using the infinite graph $\Gamma_{2,\omega}$. This implies that there is no way to use this result to derive any hardness of the finite principle $\forall x\ \mathbf{B}_{2,|x|}$, because we can show that the latter principle restricted to the copycat strategies is easy (this result will be published elsewhere).

8 The separation problem for subtheories of Bounded Arithmetic

One of the main problems in proof complexity is to prove that the hierarchy of subtheories of S_2 is proper. For example, if we proved that $S_2^{i+1} \neq S_2^i$, we would know that S_2^i does not prove $\Sigma_{i+1}^b = \Sigma_i^b$. The problem of proving $S_2^i \nvdash \Sigma_{i+1}^b = \Sigma_i^b$ is formally weaker than the problem of proving $\Sigma_{i+1}^b \neq \Sigma_i^b$. Yet our experience suggests that it may be as hard as the latter problem. Problems in proof complexity seem to be so closely tight with the corresponding problems in computational complexity that any such separation of theories would be as big a breakthrough as the separation of the corresponding complexity classes.

In this situation we can only hope to prove *conditional results*. These are results based on some complexity theoretical assumptions. The first such result was the result of Krajíček, Pudlák and Takeuti [21] showing that the hierarchy of subtheories S_2^i, $i = 1, 2, \ldots$ is strictly increasing provided that the Polynomial Hierarchy Σ_i^b, $i = 1, 2, \ldots$ is. More precisely, we proved that S_2^{i+1} is stronger than T_2^i if $\Sigma_{i+2}^p \neq \Pi_{i+2}^p$. A similar result was proved by Krajíček [17]. He proved that T_2^i is stronger than S_2^i if polynomial time computations with polynomial number of queries to a Σ_i^p oracle can recognize more languages than polynomial time computations with logarithmic number of queries to a Σ_i^p oracle. This confirms the conjecture that the hierarchy $S_2^1 \subseteq T_2^1 \subseteq S_2^2 \subseteq T_2^2 \subseteq \ldots$ is strict everywhere; still we want to know more. Assuming that the Polynomial Hierarchy is strictly increasing, we know that there exists a $\forall\Sigma_{i+2}^b$ sentence provable in S_2^{i+1}, but not in T_2^i. This is the best possible result, because S_2^{i+1} is $\forall\Sigma_{i+1}^b$ conservative over T_2^i. For pairs T_2^i, S_2^i the complexity of the known separating sentences also increases with i, but no conservativity result is known. In fact, we conjecture that they can be separated already by $\forall\Sigma_0$ sentences (if we consider theories augmented with PV, then they should be separable by purely universal sentences). At present we do not have any results that would support this conjecture.

Separation of subtheories of S_2 by low complexity sentences is an important open problem in proof complexity. Our feeling is that this is more closely related to the separation of the corresponding classes in complexity theory than merely proving that the theories are different. The main reason we think so is that the problem is closely connected with the problem of showing that propositional proof systems associated with these theories form an increasing hierarchy. (It is outside the scope of this article to explain what the proof system associated with a theory is; we refer the reader to [16, 27].)

The first thing that we need in order to solve this problem is to have sentences that could separate the theories. As far as $\forall\Sigma_0$ sentences are concerned, we do have sentences that very likely separate S_2^{i+1} from S_2^i, but we do not like them. The sentences express the consistency of subsystems of the quantified propositional sequent calculus. As usual, we are not able to interpret them in a clear combinatorial way. Apart from them we do not have other $\forall\Sigma_0$ sentences that would be good candidates for separating sentences. The next highest level is $\forall\Sigma_1$ sentences. We have shown here that a semantical interpretation of the Herbrand disjunctions produces such sentences.

For the two lowest levels very explicit sentences were found. We have already mentioned $\forall\Sigma_1^b$ sentences of Ferreira. For T_2^1, $\forall\Sigma_1^b$ sentences which axiomatize all $\forall\Sigma_1^b$ theorems of this theory were found by Buss and Krajíček [8]. Their sentences express the fact that every polynomial local search problem has a solution. We shall state their principle, again, using circuits.

Principle PLS. *For every circuits D, F and C, $D \in Circ(n,1)$, $F, C \in Circ(n,n)$ such that $D(0) = 1$, there exists $b \in \{0,1\}^n$ such that either $C(F(b)) \leq C(b)$ (where we compare the numbers that represent the strings) or $D(F(b)) = 0$.*

Theorem 8.1 [8] *Principle* **PLS** *is a $\forall\Sigma_1^b$ sentence provable in T_2^1 and every $\forall\Sigma_1^b$ theorem of T_2^1 is provable in $S_2^1 +$* **PLS**.

Having such characterizations of $\forall\Sigma_1^b$ theorems we can transform the logical problem of showing (conditional) separations into a purely computational problem. Given a sentence $\forall\Sigma_1^b$ of the form $\forall x \exists y \leq t\ \phi(x,y)$, with ϕ sharply bounded (or simply open, if we include PV into our theories), we can consider the associated *search problem,* which is the problem, from a given x, to construct a y such that $y \leq t \wedge \phi(x,y)$. The reduction is based on the fact that the complexity of the search problem is closely related to the provability of the sentence.

Let us start with the simplest problem: to prove that a $\forall\Sigma_1^b$ sentence is not provable in S_2^1. Recall that Buss's Witnessing Theorem [6] says that if such a sentence is provable in S_2^1, then the associated search problem can be solved in polynomial time. Thus to prove the unprovability in S_2^1, it suffices to show the opposite. Of course, we have to assume something, because a proof that such an algorithm does not exists gives $\mathbf{P} \neq \mathbf{NP}$.

Now consider a general situation. Let T be an arbitrary subtheory of S_2. Suppose we have a $\forall\Sigma_1^b$ sentence $\Phi := \forall x \exists y \leq t\ \phi(x,y)$, with ϕ sharply bounded, such that all $\forall\Sigma_1^b$ theorems of a theory T are provable in $PV + \Phi$. Suppose that we want to prove that another $\forall\Sigma_1^b$ sentence $\Psi := \forall x \exists y \leq s\ \psi(x,y)$, with ψ sharply bounded, is not provable in T. Then we can use the following reduction.

Let Ψ be provable in T. Then we have

$$PV \vdash\ \forall x \exists y \leq t\ \phi(x,y) \rightarrow \forall x \exists y \leq s\ \psi(x,y). \tag{24}$$

Let us assume that the bounds $y \leq t$ and $y \leq s$ are already in the formulas ϕ and ψ. Applying Herbrand's theorem for $\forall\exists\forall$ prenex sentences (see formula (6), or [21]) we get some PV terms $t_1, \ldots, t_\ell$ such that

$$\mathbb{N} \models \bigvee_{i=1}^{\ell} (\phi(t_i, z_i) \rightarrow \psi(z_0, s_i)),$$

where terms t_i and s_i may depend only on variables $z_0, \ldots, z_{i-1}$. We can write it in the following equivalent form:

$$\mathbb{N} \models \bigwedge_i \phi(t_i, z_i) \ \rightarrow \ \bigvee_i \psi(z_0, s_i). \tag{25}$$

Let us write the antecedent with the dependences on variables explicitly displayed.

$$\phi(t_1(z_0), z_1) \wedge \phi(t_2(z_0, z_1), z_2) \wedge \ldots \wedge \phi(t_\ell(z_0, z_1, \ldots, z_{\ell-1}), z_\ell).$$

We shall interpret the terms $t_1, \ldots, t_\ell$ as a polynomial time oracle algorithm that uses an input z_0 and answers $z_1, \ldots, z_\ell$ of an oracle for the search problem associated with Φ. This algorithm makes the antecedent of (25) true. Then the terms $s_1, \ldots, s_\ell$ produce a solution to the search problem associated with Ψ. Thus we can state the reduction to a computational problem as the following lemma.

Lemma 8.2 *Suppose that $PV \vdash \Phi \rightarrow \Psi$. Then there exists a polynomial time oracle algorithm and a number ℓ such that the algorithm finds a solution of the search problem associated with Ψ using ℓ adaptive queries to the oracle for the search problem associated with Φ.*

(*'Adaptive'* means that queries may depend on on the answers to the previously asked queries.)

To assume that the base theory is only PV may be too restrictive. If we assume instead that all $\forall\Sigma^b_1$ theorems of a theory T are provable in $S^1_2 + \Phi$, then the same is true except that the number of queries is not restricted (hence in every computation it is implicitly bounded by a polynomial). This follows from Theorem 3 of [26]. For more results in this area see [12].

As an example consider the conjecture that S_2 is not $\forall\Sigma^b_1$ conservative over any S^i_2. To prove this conjecture, it suffices to prove that for every k there exists k' such that the search problem associated with $\mathbf{A}_{k'}$ cannot be solved in polynomial time using an oracle for the search problem associated with $\mathbf{A}_k$.

Another context in which we can study the problem about the conservativity of subtheories of S_2 is the realm of *relativized* theories which are theories with an *oracle*. We adopt the terminology of computational complexity, though oracles in proof complexity are different. Given a subtheory T of PA in which the induction axiom schema is restricted to a class of formulas Θ, we define $T[R]$, the relativization of T, to be the theory whose language is augmented by a unary predicate R and whose induction schema is extended to $\Theta[R]$. If T uses a special form of induction, such as $LIND$, then so does $T[R]$. (This is not a precise definition applicable to every theory, but it suffices for all that we need here.) It is also possible to view $T[R]$ as a very weak extension of T to a second order theory in which R acts as a free second order variable.

Let us stress that we do not accept any other axioms about R. That is why such an oracle is different from oracles in computational complexity—when considering a Turing machine with an oracle, the oracle is always a concrete set. Nothing prevents us from adding specific axioms about R; the reason why we do not do it is that we want to show independence results. When we prove that $\Phi[R]$ is not provable in $T[R]$, we know that there exists no *uniform way* of proving all sentences $\Phi[R/\psi]$ for ψ such that $\Phi[R/\psi] \in \Theta$.

As in computational complexity, also in proof complexity we can prove unconditional separation results for relativized theories. When we have a conditional separation of theories and a relativization that makes the condition true is known, we can often use it to obtain an unconditional separation of the relativizations of the theories. Thus an oracle that makes the Polynomial Hierarchy strict enabled us to prove $T^i_2[R] \neq S^{i+1}_2[R]$ for all $i \geq 1$, cf. [21]. In a similar vein, Krajíček proved $S^i_2[R] \neq T^i_2[R]$, cf. [17].

The first result concerning relativized $\forall\Sigma^b_1$ nonconservativity appeared already in the seminal book of Buss [6]. It is the result that $T^1_2[R]$ is not $\forall\Sigma^b_1[R]$ conservative over $S^1_2[R]$. This also follows from Theorem 8.1 and a construction of an oracle such that relativized PLS problems cannot be solved in the relativized polynomial time. Another result is due to Krajíček [15] and Pudlák [26] (see also [18]) who proved that $T^2_2[R]$ is not $\forall\Sigma^b_1[R]$ conservative over $S^2_2[R]$. Impagliazzo and Krajíček [13] proved that $T^i_1[R]$ is not $\forall\Sigma^b_1[R]$ conservative over $S^i_1[R]$ for all $i \geq 1$. The theories S^i_1 and T^i_1 are S^i_2 and T^i_2 without the # function. For further results see [19, 12].

It is well-known that these separation problems can be reduced to lower bounds on the size of propositional proofs in bounded depth Frege systems [25]. For example, to prove that $T^j_2[R]$ is not $\forall\Sigma^b_1[R]$ conservative over $T^i_2[R]$, it would suffice to prove more than quasipolynomial lower bound on the lengths of proofs of the propositional translations of $\mathbf{A}_j$ in the depth i Frege system. As $\mathbf{A}_j$ is rather complicated, one may also try $\mathbf{B}_j$. We defined $\mathbf{B}_j$ hoping that the propositional translations of these principles would be the counterparts of the Sipser functions which separate the levels of the $\mathbf{AC^0}$ hierarchy.

We conclude this section by observing that the provability of a strong version of $\mathbf{B}_k$ in S^1_2 (strong in the same sense as in Section 5) is equivalent to the provability of the relativized principle $\mathbf{B}_k[R]$ in $S^1_2[R]$. Having a possibility to translate relativized independence to unrelativized would be great, provided that the translations would be also Π_1 statements. But our strong version is Π_2; moreover, it concerns only S^1_2.

We shall consider $\mathbf{B}_{\ell,k}$ only with the copycat strategies. Thus given k and ℓ everything is fixed, except the game G. As in Section 5, in the strong principle we do not specify G and instead we let Alice to decide who won after they play k moves on a board. Again, Alice and Bob play several matches one after the other, and Alice ought to play consistently: once she declares a winner in a final position, then if the position occurs again, she must declare the same person the winner.

Strong Principle $\mathbf{B}_k$. *For every virtual match with games of length k there exists an m and a strategy for Bob to play the match m times such that the strategy guarantees Bob's winning at least one of these m matches.*

As before, we assume that Bob's strategy is given by a boolean circuit. The principle is trivially true, because the games are finite and thus there are circuits computing the winning strategies. Let us stress that we do not claim that the circuits are of polynomial size, most likely they are not. In the relativized principle $\mathbf{B}_k[R]$ everything is the same except for the circuits. The circuits get answers from the oracle R for unit cost.

Proposition 8.3 *S^1_2 proves Strong Principle $\mathbf{B}_k$ if and only if $S^1_2[R]$ proves Principle $\mathbf{B}_k[R]$.*

Proof. ($\Rightarrow$) Suppose S^1_2 proves Strong Principle $\mathbf{B}_k$. Let $\mathcal{M}^R$ be a model of $S^1_2[R]$ and let $\mathcal{M}$ be $\mathcal{M}^R$ with the interpretation of R omitted. Let a match $M(s,C)$ based on a schedule s and a circuit C be in $\mathcal{M}^R$. Let a strategy ρ for Alice to play $M(s,C)$ be in $\mathcal{M}^R$. Suppose that the game in $M(s,C)$ defined by C is of length k. Let $M(s)$ be the corresponding virtual match. Then M is in $\mathcal{M}$. Thus in $\mathcal{M}$ we have a strategy σ for Bob to win at least one copy of M in a series of m virtual matches, for some m. We can use this strategy also in $\mathcal{M}^R$, and we can play it against ρ. If σ beats ρ we are done, so suppose not. Thus ρ beats σ. Take the sequence of moves produced in this way. This is a sequence that is coded by a number, so it is also in $\mathcal{M}$. But this is impossible, because we assumed that σ is a winning strategy for the series of virtual matches.

($\Leftarrow$) Suppose $S^1_2[R]$ proves Principle $\mathbf{B}_k[R]$. We shall consider only matches $M(s,G)$ with a schedule s and the game G of length k defined as follows: w is a winning position in G if $\langle s,w\rangle \in R$. By the relativized Buss's Witnessing Theorem, there exists an oracle polynomial time algorithm A which queries R and finds Bob's moves that beat Alice in the match $M(s,G)$. Now we want to use this algorithm as Bob's strategy to win a series of virtual matches. The only problem is that, according to the rules, Bob is not allowed to directly ask Alice who is the winner in a given position w. But due to Lemma 7.1, Bob can play on ℓ boards so that w appears as the final position on one of them. Hence Bob's strategy σ will be to get the answers to the queries in this way and then play the moves that beat Alice. We have to show that σ beats Alice provably in S^1_2. Suppose not. Take a model $\mathcal{M}$ and ℓ in $\mathcal{M}$ such that σ fails. The moves of Alice that beat σ form a set coded by a number. We shall use the moves to define the interpretation of R. We put exactly those pairs $\langle s,w\rangle$ in R for which Alice said that w was a winning position in the game. We have $LIND$ for $\Sigma^b_1[R]$ formulas in this model, because the interpretation of R is a set coded in $\mathcal{M}$. But if we run A in the expanded model, A will not do what it should, which is a contradiction. Hence it is provable in S^1_2 that σ beats Alice. ∎

Acknowledgment. I would like to thank to Jan Krajíček for suggesting some improvements, and to Johan Håstad and Alexander Razborov for useful discussions.

References

[1] Z. Adamowicz, On tableaux consistency in weak theories, Preprint 618 (2001) of the Mathematical Institute of the Polish Academy of Sciences.

[2] Z. Adamowicz, Herbrand consistency and Bounded Arithmetic, *Fundamenta Mathematicae*, to appear.

[3] Z. Adamowicz and L. Kolodziejczyk, Well-behaved principle alternative to bounded induction, to appear in *Theoretical Computer Science.*

[4] Z. Adamowicz and P. Zbierski, On Herbrand consistency in weak arithmetic, *Archive for Mathematical Logic 40*, pp.399-413.

[5] S.A. Bloch, Divide and Conquer in Parallel Complexity and Proof Theory, PhD thesis, University of California, San Diego, 1992

[6] S.R. Buss, Bounded Arithmetic, Bibliopolis, 1986

[7] S.R. Buss, Axiomatizations and conservation results for fragments of bounded arithmetic, in *Logic and Computations,* Contemporary Mathematics 106, Amer. Math. Soc., 1990, pp.57-84.

[8] S.R. Buss and J. Krajíček, An application of boolean complexity to separation problems in bounded arithmetic, *Proc. of the London Math. Soc.* 69(3), 1994, pp.1-21.

[9] S.A. Cook, Feasibly constructive proofs and the propositional calculus, in *Proc. of the 7th Annual Symposium on Theory of Computing,* ACM Press 1975, pp.83-97.

[10] F. Ferreira, What are the $\forall\Sigma^b_1$-consequences of T^1_2 and T^2_2?, *Annals of Pure and Applied Logic 75,* (1995), pp.79-88.

[11] P. Hájek and P. Pudlák, Metamathematics of First-Order Arithmetic, Perspectives in Mathematical Logic, Springer-Verlag, 1993.

[12] J. Hanika, PhD thesis in preparation.

[13] R. Impagliazzo and J. Krajíček, A note on conservativity relations among bounded arithmetic theories, Math. Logic Quarterly 48(3), pp.375-377.

[14] R. Kaye, J. Paris and C. Dimitracopoulos, On parameter schemas, *Journal of Symbolic Logic*, 53, 4, (1988) pp.1082-1097.

[15] J. Krajíček, No counter-example interpretation and interactive computations, in *Logic from Computer Science,* ed. Y.N. Moschovakis, Springer-Verlag 1992, pp.287-293.

[16] J. Krajíček, Bounded Arithmetic, Propositional Logic, and Complexity Theory, Encyclopedia of Mathematics and its Applications 60, Cambridge Univ. Press, 1995.

[17] J. Krajíček, Fragments of bounded arithmetic and bounded query classes, *Transactions of the AMS,* 338(2), 1993, pp.587-98.

[18] J. Krajíček and M. Chiari, Witnessing functions in bounded arithmetic and search problems, *Journal of Symbolic Logic 63(3)*, 1998, pp.1095-1115.

[19] J. Krajíček and M. Chiari, Lifting independence results in bounded arithmetic, *Arch. Math. Logic 38,* 1999, pp.123-138.

[20] J. Krajíček and P. Pudlák, Quantified propositional calculi and fragments of Bounded Arithmetic, *Zeitschrift f. math. Logik und Grundlagen d. Math. 36,* 1990, pp.29-46.

[21] J. Krajíček, P. Pudlák and G. Takeuti, Bounded Arithmetic and the Polynomial Hierarchy, *Annals of Pure and Applied Logic 52,* 1991, pp.143-153.

[22] J. Krajíček and G. Takeuti, On induction-free provability, *Annals of Mathematics and Artificial Intelligence 6*, pp.107-126.

[23] R. Parikh, Existence and feasibility in arithmetic, *Journal of Symbolic Logic 36,* 1971, pp.494-508.

[24] J.B. Paris and L. Harrington, A mathematical incompleteness in Peano Arithmetic, in *Handbook of Mathematical Logic,* North-Holland, 1977, pp.1133-1142.

[25] J.B. Paris and A.J. Wilkie, Counting problems in bounded arithmetic, in *Methods in Mathematical Logic,* Springer-Verlag LNM 1130, 1985, pp. 317-340.

[26] P. Pudlák, Some relations between subsystems of arithmetic and the complexity of computations, in *Logic from Computer Science,* ed. Y.N. Moschovakis, Springer-Verlag 1992, pp.308-317.

[27] P. Pudlák, The lengths of proofs, in *Handbook of Proof Theory*, Elsevier, 1998, pp.547-637.

[28] P. Pudlák, Parallel strategies, *Journal of Symbolic Logic,* to appear.

[29] S. Salehi, Herbrand consistency in arithmetics with bounded induction, PhD thesis, Mathematical Institute of the Polish Academy of Sciences, 2001.

[30] G. Takeuti, Proof Theory, North-Holland, 1975.

[31] A.J. Wilkie and J.B. Paris, On the schema of induction for bounded arithmetical formulas, *Annals of Pure and Applied Logic 35,* 1987, pp.261-302.

[32] D. Willard, The semantic tableaux version of the second incompleteness theorem extends almost to Robinson's Arithmetic Q, in *Automated Reasoning with Semantic Tableaux and Related Methods,* Springer-Verlag LNCS 1847, 2000, pp. 415-430.