THE JOURNAL OF SYMBOLIC LOGIC
Volume 82, Number 3, September 2017

DECIDABILITY AND CLASSIFICATION OF THE THEORY OF INTEGERS WITH PRIMES

ITAY KAPLAN AND SAHARON SHELAH

Abstract. We show that under Dickson's conjecture about the distribution of primes in the natural numbers, the theory $Th(\mathbb{Z}, +, 1, 0, Pr)$ where Pr is a predicate for the prime numbers and their negations is decidable, unstable, and supersimple. This is in contrast with $Th(\mathbb{Z}, +, 0, Pr, <)$ which is known to be undecidable by the works of Jockusch, Bateman, and Woods.

§1. Introduction. It is well known that Presburger arithmetic $T_{+,<} = Th(\mathbb{Z}, +, 0, 1, <)$ is decidable and enjoys quantifier elimination after introducing predicates for divisibility by n for every natural number $n > 1$ (see e.g., [9, Corollary 3.1.21]). The same is true for $T_{+} = Th(\mathbb{Z}, +, 0, 1)$. This is, of course, in contrast to the situation with the theory of Peano arithmetics or $Th(\mathbb{Z}, +, \cdot, 0, 1)$ which is not decidable.

If we are interested in classifying these theories in terms of stability theory, quantifier elimination gives us that T_{+} is superstable of U-rank 1, while $T_{+,<}$ is dp-minimal (a subclass of dependent, or NIP, theories, see e.g., [5, 10, 15]).

Over the years there has been quite extensive research on structures with universe $\mathbb{Z}$ or $\mathbb{N}$ and some extra structure, usually definable from Peano. A very good survey regarding questions of decidability is [2] and a list of such structures defining addition and multiplication is available in [8].

Less research was done on classifying these structures stability-theoretically. For instance, in [12, Theorem 25] and also in [11] it is proved that $Th(\mathbb{Z}, +, 0, P_q)$ is superstable of U-rank ω, where P_q is the set of powers of q.

In this paper we are interested in adding a predicate Pr for the primes and their negations and we consider $T_{+,Pr} = Th(\mathbb{Z}, +, 0, 1, Pr)$ and $T_{+,Pr,<} = Th(\mathbb{Z}, +, 0, 1, Pr, <)$. The language $\{+, 0, 1, Pr\}$ allows us to express famous number-theoretic conjectures such as the twin prime conjecture (for every n, there are at least n pairs of primes/negation of primes of distance 2), and a version of Goldbach's conjecture (all even integers can be expressed as a difference or a sum of primes). Adding the order allows us to express Goldbach's conjecture in full.

Up to now, the only known results about the theory are under a strong number-theoretic conjecture known as Dickson conjecture (D) (see below), which is also the assumption in the works of Jockusch, Bateman, and Woods. In [1, 19], they proved

Received February 14, 2016.
2010 *Mathematics Subject Classification.* 03C45, 03F30, 03B25, 11A41.
Key words and phrases. model theory, decidability, primes, Dickson's conjecture.

0022-4812/17/8203-0011
DOI:10.1017/jsl.2017.16

that assuming Dickson conjecture, $Th(\mathbb{N}, +, 0, Pr)$ is undecidable and even defines multiplication. It follows immediately that $T_{+,Pr,<}$ is undecidable and as complicated as possible in the sense of stability theory. This also explains why we need Pr to include also the negation of primes: by relatives of the Goldbach Conjecture (which are proved, see e.g., [17]), every positive integer greater than N is a sum of K primes for some fixed K, N, and hence the positive integers themselves are also definable from the positive primes.

Conjecture 1.1 (D) (Dickson, 1904 [6]). *Let $k \geq 1$ and $\bar{f} = \langle f_i \mid i < k \rangle$ where $f_i(x) = a_i x + b_i$ with a_i, b_i non-negative integers, $a_i \geq 1$ for all $i < k$. Assume that the following condition holds*:

- $\star_{\bar{f}}$ *There does not exist any integer $n > 1$ dividing all the products $\prod_{i<k} f_i(s)$ for every (non-negative) integer s.*

Then there exist infinitely many natural numbers m such that $f_i(m)$ is prime for all $i < k$.

Note that in fact the condition $\star_{\bar{f}}$ follows easily from the conclusion that there are infinitely many m's with $f_i(m)$ prime for all $i < k$. See also Remark 2.6.

Dickson's conjecture is the linear case of Schinzel's Hypothesis, see [13, pg. 292] for a discussion.

Our main result is the following.

Theorem 1.2. *Assuming* (D), *the theory $T_{+,Pr}$ is decidable, unstable and supersimple of U-rank* 1.

In essence (D) implies that the set of primes is generic up to congruence conditions (while it is not generic in the sense of [3]), and this allows us to get quantifier elimination in a suitable language. Forking then turns out to be trivial: forking formulas are algebraic (Theorem 3.2).

To show that $T_{+,Pr}$ is unstable we show that it has the independence property (see Proposition 3.6). This turns out to follow from the proof of the Green-Tao theorem about arithmetic progressions in the primes [7] (i.e., without using (D)), as was told to us in a private communication by Tamar Ziegler (but we also show that this follows from (D)).

Acknowledgments. We would like to thank Tamar Ziegler for telling us about (D) and for her input on the Green-Tao theorem (see Proposition 3.6).

We would also like to thank Carl Jockusch, Philipp Hieronymi, Lou van den Dries and Alexis Bès for reassuring us that the results stated here are new.

We would also like to thank the anonymous referee for his report.

§2. Quantifier elimination.

In this section we will prove quantifier elimination in $T_{+,Pr}$ assuming (D) in a suitable language.

Let us first note some useful facts about (D).

Remark 2.1. Suppose that $\langle f_i \mid i < k \rangle$ is as in (D) and $f_i(x) = a_i x + b_i$. Let

$$N = \max(\{a_i \mid i < k\} \cup \{k\}) + 1.$$

Then $\star_{\bar{f}}$ holds iff for every prime $p < N$, p does not divide $\prod_{i<k} f_i(s)$ for all $s \in \mathbb{Z}$ where

Proof. If $\star_{\bar{f}}$ fails, then there is some prime p such that p divides $\prod_{i<k} f_i(s)$ for all s. Let $P(X) \in \mathbb{Z}[X]$ be the polynomial $\prod_{i<k} f_i(X)$. Let $P_p = P \pmod{p} \in \mathbb{F}_p[X]$ (where $\mathbb{F}_p$ is the prime field of size p). It follows that $P_p(a) = 0$ for all $a \in \mathbb{F}_p$. So either $P_p = 0$ or $k \geq \deg(P_p) \geq p$, hence $p \leq k$ or $\prod_{i<k} a_i \equiv 0 \pmod{p}$ (as the leading coefficient) which means that for some $i < k$, $a_i \geq p$, so $p < N$ and we are done. ⊣

Lemma 2.2. *Assume* (D). *Then* (D) *holds also when we allow b_i to be negative.*

Proof. Suppose that $\langle f_i \mid i < k \rangle$ is a sequence of linear maps $f_i(x) = a_i x + b_i$ where $a_i \geq 1$ and $b_i \in \mathbb{Z}$, and assume that $\star_{\bar{f}}$ holds. Let N be as in Remark 2.1. Let $K = N!$ (enough to take the product of the primes below N). Suppose that $l \in \mathbb{N}$ is such that $lK + b_i > 0$ for all $i < k$. Let $f'_i(x) = a_i x + a_i lK + b_i$. Then $a_i \geq 1$, $b'_i = a_i lK + b_i > 0$, so let us show that $\star_{\bar{f}'}$ holds (where $\bar{f}' = \langle f'_i \mid i < k \rangle$). Note that when we compute N in Remark 2.1, we only use k and a_i which haven't changed, so by that remark, it is enough to check that for no prime $p < N$, $\prod_{i<k} f'_i(s) \equiv 0 \pmod{p}$ for all s. But for such p's, $f'_i(s) = f_i(s) + a_i lK \equiv f_i(s) \pmod{p}$, so $\prod_{i<k} f'_i(s) \equiv \prod_{i<k} f_i(s) \not\equiv 0 \pmod{p}$.

By (D), there are infinitely many integers m such that $f'_i(m)$ is prime for all $i < k$. But $f'_i(m) = a_i m + a_i lK + b_i = a_i(m + lK) + b_i$. Hence substituting $m + lK$ for m we get what we wanted. ⊣

Lemma 2.3. *Assume* (D). *Suppose that* $k, k' \in \mathbb{N}$ *and* $\langle a_i, b_i \mid i < k \rangle$, $\langle c_j, d_j \mid j < k' \rangle$ *are two tuples of integers with* $a_i, c_j \geq 1$ *for all* $i < k$, $j < k'$.

For $i < k$, *let* $f_i(x) = a_i x + b$ *and for* $j < k'$, *let* $g_j(x) = c_j x + d_j$.

Suppose that $\star_{\bar{f}}$ *holds and that* $(a_i, b_i) \neq (c_j, d_j)$.

Then there are infinitely many natural numbers m *for which for all* $i < k$ *and* $j < k'$, $f_i(m)$ *is prime and* $g_j(m)$ *is composite.*

Before giving the proof, we note that this lemma generalizes Lemma 1 from [1], which was key in the proof there of the undecidability of $T_{+,Pr,<}$.

Corollary 2.4 ([1, Lemma 1]). (*Assuming* (D)) *Let* $b_0, \ldots, b_{n-1}$ *be an increasing sequence of natural numbers, and assume that there is no prime* p *such that* $\{b_i \pmod{p} \mid i < n\} = p$. *Then there are infinitely many natural numbers* x *such that* $x + b_0, \ldots, x + b_{n-1}$ *are* ***consecutive*** *primes.*

Proof of Corollary. This is immediate from Lemma 2.3 by taking $f_i(x) = x + b_i$ and $g_j(x) = x + c_j$ where c_j run over all numbers between the b_j's. ⊣

Proof of Lemma. By induction on k'. For $k' = 0$ there is nothing to prove by (D) and Lemma 2.2.

Suppose the lemma is true for k' and prove it for $k' + 1$. It is enough to prove that for any n, there is some $m > n$ such that $f_i(m)$ is prime for all $i < k$ and $g_j(m)$ is not prime for all $j \leq k'$.

Fix n. We may assume by enlarging it that for no $m > n$ is it the case that $f_i(m) = g_j(m)$ for $i < k$, $j \leq k'$.

Let $m > n$ be so that $f_i(m)$ is prime for all $i < k$ and $g_j(m)$ is composite for all $j < k'$. If it happens that $g_{k'}(m)$ is composite, then we are done, so suppose that $q = g_{k'}(m)$ is prime. Let $f'_i(x) = a_i(qx + m) + b_i$ and $g'_j(x) = c_j(qx + m) + d_j$ for $i < k$ and $j < k' + 1$. Then $g'_{k'}(x) = c_{k'} q x + q$ is composite for all $x \geq 1$

https://doi.org/10.1017/jsl.2017.16 Published online by Cambridge University Press

(so that $c_j x + 1 \geq 2$). Hence it is enough to find m' large enough so that $f_i'(m')$ is prime for all $i < k$ and $g_j'(m')$ is composite for all $j < k'$.

By the induction hypothesis, it is enough to check that $\star_{\bar{f}'}$ holds (because $(a_i q, a_i m + b_i) \neq (c_j q, c_j m + d_j)$). Suppose that $p > 1$ is a prime which divides $\prod_{i<k} f_i'(s)$ for all s. Hence $\prod_{i<k} f_i'(s) \equiv 0 \pmod p$, and if $p \neq q$, it follows (as q is invertible modulo p) that $\prod_{i<k} f_i(s) \equiv 0 \pmod p$ for all s — a contradiction. If $p = q$, then $f_i'(x) \equiv a_i m + b_i \equiv f_i(m) \pmod q$ for all x, hence for some $i < k$, $f_i(m) = q = g_{k'}(m)$, contradicting our choice of m. ⊣

Expand the language $L = \{+, Pr, 0, 1\}$ to include the Presburger predicates P_n for $2 \leq n < \omega$ interpreted as $P_n(x) \Leftrightarrow x \equiv 0 \pmod n$, and also the predicates Pr_n for $2 \leq n < \omega$ interpreted as $Pr_n(x) \Leftrightarrow P_n(x) \wedge Pr(x/n)$. We need the latter predicate in order to eliminate the quantifiers from $\varphi(x) = \exists y\,(ny = x \wedge Pr(y))$. We also add negation (as a unary function). We need negation because of formulas of the form $\varphi(x, y) = Pr(x - y) = \exists w\,(w + y = x \wedge Pr(w))$.

Let L^* be the resulting language $\{+, -, 1, 0, Pr, Pr_n, P_n \mid 2 \leq n < \omega\}$, and let $T^*_{+,Pr}$ be the complete theory of M^* — the structure with universe $\mathbb{Z}$ in L^*. Note that all the new predicates are definable from L.

Remark 2.5. The condition $\star_{\bar{f}}$ of Dickson's conjecture is first-order expressible in L^*. This means that for every tuple $\langle a_i \mid i < k\rangle$ of positive integers, there is a formula $\varphi_{\bar{a}}(y_0, \ldots, y_{k-1})$ such that for any choice of $b_i \in \mathbb{Z}$ for $i < k$, $M^* \models \varphi_{\bar{a}}\left(\bar{b}\right)$ iff $\star_{\bar{f}}$ holds where $f_i(x) = a_i x + b_i$ for $i < k$. It has the form $\bigwedge_{p<N \text{ prime}} \bigvee_{r<p} \bigwedge_{i<k} \neg P_p(a_i r + y_i)$ for some $N \in \mathbb{N}$.

Proof. Recall Remark 2.1 and the choice of N from there (which depends only on $\langle a_i \mid i < k\rangle$ and k). Let $\varphi_{\bar{a}}(\bar{y})$ be as described in the remark: for every prime $p < N$, for some $0 \leq x < p$, for all $i < k$, $\neg P_p(a_i x + y_i)$. Note that $\varphi_{\bar{a}}$ is quantifier-free in L^* (as it contains 1). ⊣

Remark 2.6. Given $\bar{f} = \langle f_i \mid i < k\rangle$ a tuple of linear maps as above, if there are more than $2k$ integers m such that $f_i(m)$ is prime or a negation of a prime, then $\star_{\bar{f}}$ holds.

Proof. Indeed, otherwise there is some prime p which witnesses this, but then for some i and three different m's, $|f_i(m)| = p$ — a contradiction. ⊣

Lemma 2.7. *$T^*_{+,Pr}$ eliminates quantifiers in L^* provided* (D).

Proof. We start with the following observation.

◇ By Remark 2.5 and Lemma 2.3, our assumption that Dickson's conjecture holds translates into a scheme of first-order statements:

For every n and every choice of positive integers $\langle a_i \mid i < k\rangle$ and $\left\langle a_j' \mid j < k'\right\rangle$ and for all $\langle b_i \mid i < k\rangle$ and $\left\langle b_j' \mid j < k'\right\rangle$, if $\varphi_{\bar{a}}\left(\bar{b}\right)$ holds and for all $i < k$, $j < k'$, $(a_i, b_i) \neq \left(a_j', b_j'\right)$ then there are at least n elements x with

$$\bigwedge_{i<k} Pr(a_i x + b_i) \wedge \bigwedge_{j<k'} \neg Pr\left(a_j' x + b_j'\right).$$

https://doi.org/10.1017/jsl.2017.16 Published online by Cambridge University Press

Conversely, by Remark 2.6, if there are more than $2k$ such elements x, then $\varphi_{\bar{a}}\left(\bar{b}\right)$ holds. In particular, $\varphi_{\bar{a}}\left(\bar{b}\right) \wedge \bigwedge_{i,j}\left(a_i, b_i\right) \neq \left(a'_j, b'_j\right)$ holds iff there are more than $2k$ elements x with

$$\bigwedge_{i<k} Pr\left(a_i x + b_i\right) \wedge \bigwedge_{i<k'} \neg Pr\left(a'_j x + b'_j\right).$$

(Recall that Pr contains the primes and their negations.)

In order to prove quantifier elimination we will use a back-and-forth criteria. Namely, suppose that $\mathfrak{C} \models T^*_{+,Pr}$ is a monster model (very large, saturated model) and that $h : A \to B$ is an isomorphism of small substructures A, B. Given $a \in \mathfrak{C}\backslash A$ we want to extend h so that its domain contains a.

We may assume, by our choice of language (which includes Pr_n and $-$), that both A and B are groups such that if $c \in A$ and $\mathfrak{C} \models P_n(a)$ then $c/n \in A$ and similarly for B. Why? For such a c, elements of the group generated by adding c/n to A have the form $m(c/n) + b$ for $m \in \mathbb{Z}$ and $b \in A$. We have to show that the map taking c/n to $h(c)/n$ and extends h is an isomorphism. For instance, we have to show that if $\mathfrak{C} \models Pr(m(c/n) + b)$ then $\mathfrak{C} \models Pr(m(h(c)/n) + h(b))$. But $\mathfrak{C} \models Pr(m(c/n) + b)$ iff $\mathfrak{C} \models Pr_n(mc + nb)$. Similarly we deal with Pr_k and P_k.

Let $p^a(x) = \text{tp}^{\text{qf}}(a/A)$, and let $q^a(x) = h(p^a)$. Let $p^a_{\equiv} = p^a \upharpoonright L^*_{\equiv}$ and $p^a_{Pr} = p^a \upharpoonright L^*_{Pr}$, where $L^*_{\equiv} = L^* \backslash \{Pr, Pr_n \mid 2 \leq n < \omega\}$ and $L^*_{Pr} = L^* \backslash \{P_n \mid 2 \leq n < \omega\}$, so that $p^a = p^a_{\equiv} \cup p^a_{Pr}$, and we have to realize q^a.

Claim 2.8. *It is enough to prove that we can realize $q^a_{Pr} = h(p^a_{Pr})$ for all a as above.*

Proof. Easily, as we included 1 in the language, $q^a_{\equiv}$ is isolated by $\{x \neq c \mid c \in B\}$ and equations of the form $x \equiv k \pmod{n}$ for $k < n$, and for every $n < \omega$ there is exactly one $k < n$ with such an equation appearing in q^a. Also, every finite set of such equations is implied by one such equation (e.g., if the equations are $\{x \equiv k_i \pmod{n}_i \mid i < s\}$ then take $x \equiv k \pmod{\prod)_{i<s} n_i}$ where k is such that this equation is in q^a). Hence it is enough to show that $x \equiv k \pmod{n} \cup q^a_{Pr}(x)$ is consistent (q^a_{Pr} already contains $\{x \neq c \mid c \in B\}$). As $a \equiv k \pmod{n}$, $b = (a-k)/n \in \mathfrak{C}$. Let $p^b = \text{tp}^{\text{qf}}(b/A)$ so by our assumption there is some $d \in \mathfrak{C}$ such that $d \models h\left(p^b\right)_{Pr}$. Then $nd + k \models q^a_{Pr}(x)$ and of course satisfies the equation $x \equiv k \pmod{n}$. ⊣

Let $p^a_{Pr_0} = p^a \upharpoonright L_{Pr_0}$ where $L_{Pr_0} = L_{Pr} \backslash \{Pr_n \mid 2 \leq n < \omega\}$.

Claim 2.9. *It is enough to prove that we can realize $q^a_{Pr_0} = h\left(p^a_{Pr_0}\right)$ for a as above.*

Proof. This is similar to Claim 2.8. It is enough to show that $q^a_{Pr_0}(x) \cup \Sigma(x)$ is consistent where Σ is a finite set of formulas from $q^a_{Pr} \backslash q^a_{Pr_0}$. So Σ consists of formulas of the form $Pr_n(mx + c)$ or its negation for $m \in \mathbb{Z}$, $1 < n \in \mathbb{N}$ and $c \in B$. Without loss of generality, by replacing the n's with their product N and $Pr_n(mx + c)$ by $Pr_N((N/n)(mx + c))$, we may assume that all the n's appearing in Σ are equal to $n > 1$. Let $b = (a-k)/n$ where $a \equiv k \pmod{n}$ and $k < n$. Let $p^b = \text{tp}^{\text{qf}}(b/A)$. By our assumption there is some $d \in \mathfrak{C}$ such that $d \models h\left(p^b_{Pr_0}\right)$. Let us check that $nd + k \models q^a_{Pr_0}(x) \cup \Sigma(x)$.

First, if $\varphi(x, c) \in q^a_{Pr_0}(x)$ (c a tuple from B) then $\mathfrak{C} \models \varphi\left(a, h^{-1}(c)\right)$ so that $\mathfrak{C} \models \varphi\left(nb + k, h^{-1}(c)\right)$ so $d \models \varphi(nx + k, c)$ so $nd + k \models \varphi(x, c)$.

Now, suppose that $Pr_n(mx + c) \in \Sigma$.

https://doi.org/10.1017/jsl.2017.16 Published online by Cambridge University Press

Then $\mathfrak{C} \models Pr_n(ma + h^{-1}(c))$, so $\mathfrak{C} \models Pr_n(m(nb+k) + h^{-1}(c))$. Hence $m(nb+k) + h^{-1}(c)$ is divisible by n which means that $mk + h^{-1}(c)$ is divisible by n, and as h is an isomorphism (and the language includes 1), so is $mk + c$, hence $m(nd+k) + c$ is also divisible by n. Moreover the quotient $e = [mk + h^{-1}(c)]/n \in A$ maps to $e' = [mk + c]/n \in B$. As $\mathfrak{C} \models Pr(mb + e)$, it follows that $\mathfrak{C} \models Pr(md + e')$, so that $\mathfrak{C} \models Pr_n(m(nd+k)+c)$. The same logic works if $\neg Pr_n(mx + c) \in \Sigma$. ⊣

Divide into cases.

Case 1: There are infinitely many solutions to $p^a_{Pr_0}$.

Given any finite set $\Sigma \subseteq q^a_{Pr_0}$, it has the form

$$\{Pr(m_i x + c_i) \mid i < k\} \cup \{\neg Pr(m'_j x + c'_j) \mid j < k'\}$$

where $m_i, m'_j \in \mathbb{Z}$ and $c_i, c'_j \in B$ (it also includes formulas of the form $x \neq c$). As[1] $\mathfrak{C} \models \forall x Pr(x) \leftrightarrow Pr(-x)$, we may assume that $m_i, m'_j \geq 1$. Also, it is of course impossible that $(m_i, c_i) = (m'_j, c'_j)$. By $\diamondsuit$, it is enough to check that $\mathfrak{C} \models \varphi_{\bar{m}}(\bar{c})$ where $\bar{m} = \langle m_i \mid i < k\rangle$ and $\bar{c} = \langle c_i \mid i < k\rangle$ and $\varphi_{\bar{m}}$ is from Remark 2.5. As $\varphi_{\bar{m}}$ is quantifier-free, and as $\mathfrak{C} \models \varphi_{\bar{m}}(h^{-1}(\bar{c}))$ (because $h^{-1}(\Sigma)$ has infinitely many solutions and by $\diamondsuit$), we are done.

Case 2: There are only finitely many solutions to p_{Pr_0}.

By $\diamondsuit$, and as $\mathfrak{C} \models \forall x Pr(x) \leftrightarrow Pr(-x)$, there are some $m_i \geq 1, e_i \in A$ such that $\{Pr(m_i x + e_i) \mid i < k\}$ already has finitely many solutions. Hence $\varphi_{\bar{m}}(\bar{e})$ fails. Let N be from Remark 2.5. We get that for some $p < N$, there is some $i < k$ such that $P_p(m_i a + e_i)$. But as $Pr(m_i a + e_i)$, it must be that $\pm p = m_i a + e_i$. As $\pm p, e_i \in A$, and as A is closed under dividing by m_i, it follows that $a \in A$ and this cannot happen by assumption. ⊣

§3. Decidability and classification. We start with the decidability result that is now almost immediate.

Corollary 3.1. *The theory $T^*_{+,Pr}$ is decidable and hence so is $T_{+,Pr}$ provided that Dickson's conjecture holds.*

Proof. Observing the proof of Lemma 2.7, we see that we can recursively enumerate the axioms that we used. Let us denote this set by Σ. Let Σ' be the complete quantifier-free theory of $\mathbb{Z}$ in L^*. Then Σ' is recursive and contained in $T^*_{+,Pr}$.

Now $\Sigma \cup \Sigma'$ is consistent and complete (every sentence is equivalent to a quantifier free sentence which is decided by Σ'). Hence it is decidable. ⊣

Now we turn to classification in the sense of [14], where one is interested in classifying first-order theories by finding "dividing lines" between them, inducing classes with interesting properties both inside and outside. The most studied such class is that of stable theories, which is a very well-behaved and well-understood class. Containing it is the class of simple theories, and among them the "simplest" simple theories are supersimple of U-rank 1. For the definitions of simple and supersimple theories as well as of forking and dividing, we refer the reader to e.g., [18, Chapter 7, Definition 8.6.3].

[1] Here we use the fact that Pr contains both the primes and their negations.

https://doi.org/10.1017/jsl.2017.16 Published online by Cambridge University Press

THEOREM 3.2. *Assuming* (D), $T^*_{+,Pr}$ *(and* $T_{+,Pr}$*) is supersimple of* U*-rank* 1: *working in the monster model* $\mathfrak{C}$, *if* $\varphi(x,a)$ *forks over* $\emptyset$ *where* x *is a singleton and* a *is some tuple from* $\mathfrak{C}$ *then* φ *is algebraic (i.e.,* $\varphi \vdash \bigvee_{i<k} x = c_i$*).*

PROOF. The proof is similar to that of Lemma 2.7.

Let N be an ω-saturated model. Suppose that φ forks over $\emptyset$ but is not algebraic. Extend φ to a type $p(x) \in S(N)$ which is nonalgebraic over N. So p forks over $\emptyset$, and hence it divides over $\emptyset$ by saturation. By quantifier elimination we may assume that p is quantifier free.

Recalling the notation from the proof of Lemma 2.7, we have the following claim.

CLAIM 3.3. *It is enough to prove that for every type* $q(x) \in S(N)$, *if* $q_{Pr} = q \restriction L^*_{Pr}$ *divides over* $\emptyset$, *then* q_{Pr} *is algebraic.*

PROOF. We want to show that p is algebraic, thus getting a contradiction. Let $\langle N_i \mid i < \omega \rangle$ be an indiscernible sequence starting with $N_0 = N$ in $\mathfrak{C}$, which witnesses that p divides.

By indiscernibility, all the congruent conditions in $p(x, N_i)$ (i.e., equations such as $mx + c \equiv d$) are implied by the congruent conditions in $p|_\emptyset$. It follows that $\bigcup \{p_{Pr}(x, N_i) \mid i < \omega\} \cup \Sigma$ is inconsistent for some finite $\Sigma \subseteq p$, which is isolated by a formula of the form $x \equiv k \pmod{n}$ for some $k < n$.

Let $c \models p$. Then $c \equiv k \pmod{n}$ for some $k < n$, and let $d = (c-k)/n$. Then $[\mathrm{tp}(d/N)]_{Pr}$ divides over $\emptyset$ as witnessed by the same sequence $\langle N_i \mid i < \omega \rangle$ (let $r = \mathrm{tp}(d/N)$, then if $d' \models \bigcup \{r_{Pr}(x, N_i) \mid i < \omega\}$ then $nd' + k \models \Sigma \cup \bigcup \{p_{Pr}(x, N_i) \mid i < \omega\}$). Hence, $[\mathrm{tp}(d/N)]_{Pr}$ is algebraic, i.e., $d \in N$, but then so is c. ⊣

CLAIM 3.4. *It is enough to prove that for every type* $q(x) \in S(N)$, *if* $q_{Pr_0} = q \restriction L^*_{Pr_0}$ *divides over* $\emptyset$, *then* q_{Pr_0} *is algebraic.*

PROOF. This is similar to the proof of Claim 3.3.

By Claim 3.3, it is enough to prove that for any $q(x) \in S(N)$, if q_{Pr} divides over $\emptyset$ then q_{Pr} is algebraic. Suppose that q_{Pr} divides over $\emptyset$ and let $\langle N_i \mid i < \omega \rangle$ be as in the proof of Claim 3.3. There is some finite set of formulas $\Sigma(x, N) \subseteq q_{Pr} \backslash q_{Pr_0}$ such that $\bigcup \{q_{Pr_0}(x, N_i) \cup \Sigma(x, N_i) \mid i < \omega\}$ is inconsistent. As in the proof of Lemma 2.7, we may assume that for some $n \in \mathbb{N}$, Σ consists of formulas of the form $Pr_n(mx + c)$ for $c \in N$ and $m \in \mathbb{Z}$. Let $d \models q$, and assume that $d \equiv k \pmod{n}$ for $k < n$. Then for some $e \in \mathfrak{C}$, $d = ne + k$, and $[\mathrm{tp}(e/N)]_{Pr_0}$ divides over $\emptyset$ (let $r = \mathrm{tp}(e/N)$, then if $e' \models \bigcup \{r_{Pr_0}(x, N_i) \mid i < \omega\}$ then $ne' + k \models \bigcup \{q_{Pr_0}(x, N_i) \cup \Sigma(x, N_i) \mid i < \omega\}$, as in the proof of Lemma 2.7). Hence this type is algebraic and hence so is q. ⊣

CLAIM 3.5. *It is enough to prove that if* $\Sigma(x, \bar{c})$ *is a finite set of formulas of the form* $Pr(mx + c)$ *or* $\neg Pr(mx + c)$ *for* $m \in \mathbb{Z}$ *and* $c \in N$, *which has infinitely many solutions, then for any indiscernible sequence* $\langle \bar{c}_i \mid i < \omega \rangle$ *starting with* $\bar{c}$, $\{\Sigma(x, \bar{c}_i) \mid i < \omega\}$ *has infinitely many solutions.*

PROOF. Use Claim 3.4. We have to prove that if q_{Pr_0} divides over $\emptyset$ then it is algebraic. Suppose it is not, and let $\Sigma(x, \bar{c}) \subseteq q_{Pr_0}$ be a finite set of formulas of the form $Pr(mx + c)$ or $\neg Pr(mx + c)$ for $m \in \mathbb{Z}$ and $c \in N$, and let $S \subseteq N$ be finite such that $\Delta\left(x, \bar{c}, \bar{d}\right) = \Sigma(x, \bar{c}) \cup \{x \neq d \mid d \in S\}$ divides over $\emptyset$. Let $\left\{\left(\bar{c}_i, \bar{d}_i\right) \,\middle|\, i < \omega\right\}$ be an indiscernible sequence witnessing dividing. But then $\bigcup \{\Sigma(\bar{x}, \bar{c}_i) \mid i < \omega\}$ has

https://doi.org/10.1017/jsl.2017.16 Published online by Cambridge University Press

infinitely many solutions by assumption, so by saturation (of $\mathfrak{C}$) there is a solution which is distinct from $\bigcup \left\{ \bar{d}_i \,\middle|\, i < \omega \right\}$, contradicting dividing. ⊣

Let $\Sigma(x)$ be as in Claim 3.5.

Then $\Sigma(x, \bar{c}, \bar{c}') = \{Pr(m_i x + c_i) \mid i < k\} \cup \left\{\neg Pr\left(m'_j x + c'_j\right) \,\middle|\, j < k'\right\}$, for $m_i, m'_j \in \mathbb{Z}$ and $c_i, c'_j \in N$. Now take an indiscernible sequence $\langle \bar{c}_\alpha, \bar{c}'_\alpha \mid \alpha < \omega \rangle$ starting with $\langle c_i \mid i < k \rangle \frown \left\langle c'_j \,\middle|\, j < k' \right\rangle$. Consider a finite union of the form $\bigcup \{\Sigma(x, \bar{c}_\alpha, \bar{c}'_\alpha) \mid \alpha < l\}$. Then by indiscernibility it cannot be that $(m_i, c_{i,\alpha}) = \left(m'_j, c'_{j,\beta}\right)$ for some $\alpha, \beta < l$, $i < k$, and $j < k'$. Hence by (D), it is enough to show that $\star_{\bar{f}}$ holds for $\bar{f} = \langle f_{i,\alpha} \mid i < k, \alpha < l \rangle$ where $f_{i,\alpha}(x) = m_i x + c_{i,\alpha}$, and by Remark 2.5 we have to show that $\varphi_{\bar{m}}(\langle \bar{c}_\alpha \mid \alpha < l \rangle)$ holds.

Let $N \in \mathbb{N}$ be from Remark 2.5 (it depends only on $\bar{m}$, k and l). We have to check that if $r < N$ is a prime, for some $0 \le t < r$, for all $i < k$ and $\alpha < l$, $m_i t + c_{i,\alpha} \not\equiv 0 \pmod r$. If this does not happen for r, then, as (by indiscernibility) $c_{i,\alpha} \equiv c_i \pmod r$, we get that for all $0 \le t < r$, for some $i < k$, $m_i t + c_i \equiv 0 \pmod r$. But this means that Σ cannot have infinitely many solutions by Remark 2.6 — contradiction. ⊣

We move to NIP. We will show that $T_{+,Pr}$ has the independence property IP (and thus the theory is not NIP), and even the n-independence property. This shows in particular that $T_{+,Pr}$ is unstable. We will recall the definition in the proof of Theorem 3.7, but the interested reader may find more in [16] (about NIP) and [4] (on n-dependence).

We will use the following proposition.

Proposition 3.6. *For all $n < \omega$ and $s \subseteq n$ there is an arithmetic progression $\langle at + b \mid t < n \rangle$ of natural numbers such that $at + b$ is prime iff $t \in s$.*

Proof. As we said in the introduction, according to a private communication with Tamar Ziegler, this follows from the proof of the Green-Tao theorem about arithmetic progression of primes [7].

We give a very detail-free explanation of why this should be true. Heuristically, the primes below N behave like a random set of density $1/\log N$, so the number of $x, d \le N$ such that $x + d$, $x + 2d, \ldots, x + kd$ are all primes is $N^2/(\log N)^k$. If we skip the i'th element in the sequence (i.e., we do not ask it to be prime), then the number is $N^2/(\log N)^{k-1}$. Hence, we may remove all the prime arithmetic progressions and still find some sequence where the i'th element is not prime.

We will however give a proof that relies on (D). Fix n and s. Let $b = n! + 1$. Use Lemma 2.3, with the linear maps $x + b, 2x + b, \ldots, nx + b$. By Remark 2.1, it is enough to check that for all primes $p \le n$, for some $t < p$, $kt + b \not\equiv 0 \pmod p$ for all $1 \le k \le n$. But $b \equiv 1 \pmod p$ so this holds for $t = 0$. ⊣

Theorem 3.7 (Without assuming Dickson's conjecture). *$T_{+,Pr}$ has the independence property and even the n-independence property. Hence so does $T^*_{+,Pr}$.*

Proof. We use only Proposition 3.6. To prove that T is n-independent, we have to find a formula $\varphi(x, y_1, \ldots, y_n)$ such that for all $k < \omega$, there are tuples $a_{i,j}$ for $i < n, j < k$ inside some model $M \models T$ such that for every subset $s \subseteq k^n$, there is some tuple $b_s \in M$ with $M \models \varphi\left(b_s, a_{0,j_0}, \ldots, a_{n-1,j_{n-1}}\right)$ iff $(j_0, \ldots, j_{n-1}) \in s$. This of course implies the independent property.

https://doi.org/10.1017/jsl.2017.16 Published online by Cambridge University Press

The formula we take is $\varphi(x, y_1, \ldots, y_n) = Pr(x + y_1 + \cdots + y_n)$, and we work in $\mathbb{Z}$.

Given k, by Proposition 3.6 there is an arithmetic progression of length $k^n \cdot 2^{(k^n)}$, which we write as $\langle \bar{c}_s \mid s \subseteq k^n \rangle$ where $\bar{c}_s = \langle c_{s,l} \mid l < k^n \rangle$, such that for each subset $s \subseteq k^n$ and $l < k^n$, $Pr(c_{s,l})$ iff $(j_0, \ldots, j_{n-1}) \in s$ where $j_i < k$ are (unique) such that $l = \sum_{i<n} j_i k^i$.

Suppose this progression has difference $d > 0$. Now we choose $a_{i,j}$ for $i < n$, $j < k$ and b_s for $s \subseteq k^n$ as follows.

Let $a_{0,j} = j \cdot d$ for $j < k$ and in general, for $i < n$, $a_{i,j} = j \cdot d \cdot k^i$. Let $b_s = c_{s,0}$. Now note that

$$c_{s,0} + \sum_{i<n} (j_i d) k^i = c_{s, \sum_{i<n} j_i \cdot k^i}.$$

And so we are done. ⊣

§4. **Acknowledgements.** The first author would like to thank the Israel Science foundation for partial support of this research (Grant no. 1533/14). The research leading to these results has received funding from the European Research Council, ERC Grant Agreement n. 338821. No. 1082 on the second author's list of publications.

REFERENCES

[1] P. T. Bateman, C. G. Jockusch, and A. R. Woods, *Decidability and undecidability of theories with a predicate for the primes*, this Journal, vol. 58 (1993), no. 2, pp. 672–687.

[2] A. Bès, *A survey of arithmetical definability*. ***Bulletin of The Belgian Mathematical Society-Simon Stevin***, 2001, Suppl., pp. 1–54.

[3] Z. Chatzidakis and A. Pillay, *Generic structures and simple theories*. ***Annals of Pure and Applied Logic***, vol. 95 (1998), no. 1–3, pp. 71–92.

[4] A. Chernikov, D. Palacin, and K. Takeuchi, *On n-dependence*. ***Notre Dame Journal of Formal Logic***, 2014, accepted, arXiv:1411.0120.

[5] A. Dolich, J. Goodrick, and D. Lippel, *Dp-minimality: Basic facts and examples*. ***Notre Dame Journal of Formal Logic***, vol. 52 (2011), no. 3, pp. 267–288.

[6] L. E. Dickson, *A new extension of Dirichlet's theorem on prime numbers*. ***Messenger of Mathematics***, vol. 33 (1904), pp. 155–161.

[7] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*. ***Annals of Mathematics (2)***, vol. 167 (2008), no. 2, pp. 481–547.

[8] I. Korec, *A list of arithmetical structures complete with respect to the first-order definability*. ***Theoretical Computer Science***, vol. 257 (2001), no. 1–2, pp. 115–151.

[9] D. Marker, ***Model Theory: An Introduction***, Graduate Texts in Mathematics, vol. 217, Springer, New York, 2002.

[10] A. Onshuus and A. Usvyatsov, *On dp-minimality, strong dependence and weight,* this Journal, vol. 76 (2011), no. 3, pp. 737–758.

[11] D. Palacin and R. Sklinos, *On superstable expansions of free abelian groups*. ***Notre Dame Journal of Formal Logic***, 2014, accepted, arXiv:1405.0568.

[12] B. Poizat, *Supergénérix*. ***Journal of Algebra***, vol. 404 (2014), pp. 240–270.

[13] P. Ribenboim, ***The Book of Prime Number Records***, second ed., Springer-Verlag, New York, 1989.

[14] S. Shelah, ***Classification Theory and the Number of Nonisomorphic Models***, second ed. Studies in Logic and the Foundations of Mathematics, vol. 92, North-Holland Publishing Co., Amsterdam, 1990.

[15] P. Simon, *On dp-minimal ordered structures*, this Journal, vol. 76 (2011), no. 2, pp. 448–460.

[16] ———, ***A Guide to NIP Theories***, Lecture Notes in Logic, vol. 44, Cambridge University Press, Cambridge, 2015.

[17] T. Tao, *Every odd number greater than 1 is the sum of at most five primes*. ***Mathematics of Computation***, vol. 83 (2013), no. 286, pp. 997–1038.

[18] K. Tent and M. Ziegler, ***A Course in Model Theory***, Lecture Notes in Logic, vol. 40, Association for Symbolic Logic, La Jolla, CA; Cambridge University Press, Cambridge, 2012.

[19] A. R. Woods, *Some problems in logic and number theory, and their connections*, ***New Studies in Weak Arithmetics*** (P. Cégielski, C. Cornaros, and C. Dimitracopoulos, editors), CSLI Lecture Notes, vol. 211, CSLI Publications, Stanford, CA, 2013, pp. 271–388. Dissertation, University of Manchester, Manchester, 1981.

THE HEBREW UNIVERSITY OF JERUSALEM
EINSTEIN INSTITUTE OF MATHEMATICS
EDMOND J. SAFRA CAMPUS, GIVAT RAM
JERUSALEM 91904, ISRAEL
E-mail: kaplan@math.huji.ac.il
URL: https://sites.google.com/site/itay80/

THE HEBREW UNIVERSITY OF JERUSALEM
EINSTEIN INSTITUTE OF MATHEMATICS
EDMOND J. SAFRA CAMPUS, GIVAT RAM
JERUSALEM 91904, ISRAEL
and
DEPARTMENT OF MATHEMATICS
HILL CENTER-BUSCH CAMPUS
RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY
110 FRELINGHUYSEN ROAD
PISCATAWAY, NJ 08854-8019, USA
E-mail: shelah@math.huji.ac.il
URL: http://shelah.logic.at/