The computational content of Milliken's tree theorem

Benoit Monin

Université Paris-Est Créteil

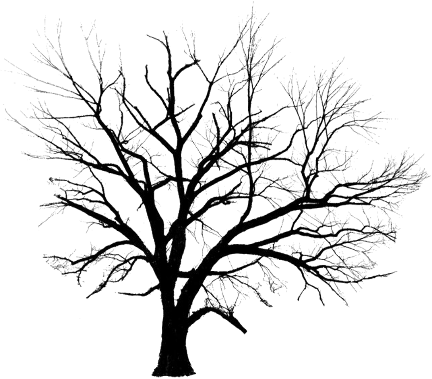
UNIVERSITÉ PARIS-EST CRÉTEIL VAL DE MARNE

Connaissance - Action

Joint work with:

- Paul-Elliot Angles d'Auriac
- Peter Cholak
- Damir Dzhafarov
- Ludovic Patey

During the "Research in Paris program"

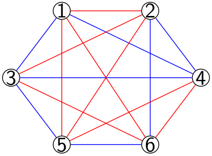


Section 1

Motivations

Ramsey's theorem for pairs

Frank Ramsey, 1903–1930

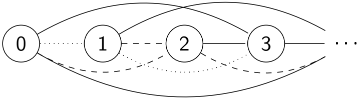


At a gathering of six people, at least three of them all known each other, or all don't know each other.

$[A]^n$: the subsets of A of size n.

Theorem (Ramsey's theorem for pairs)

$\mathbf{RT}^2_k$: *For every infinite set X, for every function $f : [X]^2 \to \{0, \ldots, k-1\}$, there is an infinite set $Y \subseteq X$ and an integer $i < k$ such that $f([Y]^2) = \{i\}$.*



f is called a **coloring**
X is called **monochromatic**

A naive question

What of Ramsey's theorem for functions $f : \mathbb{N} \times \mathbb{N} \to \{0, \ldots, k-1\}$?

Let f be the following function:

$$\begin{aligned} f((n,m)) &= 0 \quad \text{if } n > m \\ f((n,m)) &= 1 \quad \text{if } n < m \\ f((n,m)) &= 2 \quad \text{if } n = m \end{aligned}$$

For any set X of size at least 2:

$\exists (n,m) \in X \times X$ such that $f((n,m)) = 0$

$\exists (n,m) \in X \times X$ such that $f((n,m)) = 1$

$\exists (n,m) \in X \times X$ such that $f((n,m)) = 2$

But, given $f : \mathbb{N} \times \mathbb{N} \to \{0, \ldots, k-1\}$ for any $k > 2$:

1. Define $g_0 : [\mathbb{N}]^2 \to \{0, \ldots, k-1\}$ by $g_0(\{n,m\}) = f((n,m))$ with $n < m$. Apply Ramsey's theorem to get $X_0 \subseteq \mathbb{N}$ on which g_0 is monochromatic.
2. Define $g_1 : [X]^2 \to \{0, \ldots, k-1\}$ by $g_1(\{n,m\}) = f((n,m))$ with $n > m$. Apply Ramsey's theorem to get $X_1 \subseteq X_0$ on which g_1 is monochromatic.
3. Apply the pigeonshole principle to find an infinite subset $X_2 \subseteq X_1$ such that $f((n,n))$ is always of the same color.

Conclusion : we can always reduce the number of color from k to 3.

Taking a step back

Given:

- an infinite mathematical structure G,
- a collection $\mathcal{S}(G)$ of finite substructures of G,

does there exists $l \in \mathbb{N}$ such that for any $k > l$ and any coloring

$$g : \mathcal{S}(G) \to \{0, \dots k-1\}$$

we can find an infinite substructure $G' \subseteq G$ with $G' \cong G$, such that $|g(\mathcal{S}(G'))| \leqslant l$?

Definition (Zucker)

Given $\mathcal{S}(G)$, the minimum such number l, if it exists, is the **big Ramsey degree** of $\mathcal{S}(G)$ in G.

- The big Ramsey degree of any sets of size 2 in any infinite set X is 1.
- The big Ramsey degree of any pair of integers in any product $X \times X$ for X infinite is 3.

Other natural big Ramsey degrees ?

Devlin's theorem for singletons

Proposition (Pigeonhole's principle for rationals)

$\mathbf{DT}^1_k$: *For any $f : \mathbb{Q} \to \{0, \dots, k-1\}$, there is an infinite set $X \subseteq \mathbb{Q}$ order isomorphic to $\mathbb{Q}$ and an integer $i < k$ such that $f(X) = \{i\}$.*

Remark : the proposition remains true starting with any $R \cong \mathbb{Q}$ in place of $\mathbb{Q}$.

Lemma

For $k \geqslant 2$, $\mathrm{DT}^1_k \to \mathrm{DT}^1_{k+1}$

Lemma's proof : Given $f : \mathbb{Q} \to \{0, \dots, k\}$ we define $g : \mathbb{Q} \to \{0, \dots, k-1\}$ by $g(q) = \min(f(q), k-1)$. We apply DT^1_k to find a monochromatic set $X_0 \cong \mathcal{Q}$. The set X_0 has at most two colors with f. We then apply DT^1_2 to find a monochromatic set $X_1 \subseteq X_0$ with $X_1 \cong X_0 \cong \mathcal{Q}$ on which f is monochromatic.

Proposition's proof : Either there is $q_0 < q_1$ such that $f(\mathbb{Q} \cap (q_0, q_1)) = \{0\}$ or for every $q_0 < q_1$ there is $q \in (q_0, q_1)$ such that $f(q) = 1$. In this case we compute $X \cong \mathbb{Q}$ which is monochromatic for f.

Generalizing the singleton case

Definition

$\mathbf{DT}^n_k$: For any $f : [\mathbb{Q}]^n \to \{0, \dots, k-1\}$, there is an infinite set $X \subseteq \mathbb{Q}$ order isomorphic to $\mathbb{Q}$ and an integer $i < k$ such that $f([X]^n) = \{i\}$.

DT^1_2 is a theorem. Is DT^2_2 ?

Let $\{q_n\}_{n\in\mathbb{N}}$ be any enumeration of rationals. Let

$$\begin{aligned} f(\{q_n, q_m\}) &= 0 \text{ if } n < m \\ &= 1 \text{ otherwise} \end{aligned}$$

For any $X \cong \mathbb{Q}$ and any $q_n \in X$ there must exists m_1, m_2 sufficiently large such that $q_{m_1} < q_n < q_{m_2}$: DT^2_2 is false.

The big Ramsey degree of pairs of rational is at least 2.

Big ramsey degrees for finite subsets of $\mathbb{Q}$

Definition

$\mathbf{DT}^n_{k,l}$: For any $f : [\mathbb{Q}]^n \to \{0, \dots, k-1\}$, there is an infinite set $X \subseteq \mathbb{Q}$ order isomorphic to $\mathbb{Q}$ such that $|f([X]^n)| \leqslant l$.

Theorem (Devlin)

For any n, there exists t_n such that $\mathrm{DT}^n_{t_n+1,t_n}$ is true.

Remark : for $k > l$ we have $\mathrm{DT}^n_{k,l} \to \mathrm{DT}^n_{k+1,l}$, by grouping two colors in one and applying $\mathrm{DT}^n_{k,l}$ twice if needed.

The big Ramsey degrees t_n such that $\mathrm{DT}^n_{t_n+1,t_n}$ is true are known as the *odd tangent numbers*:

t_1	t_2	t_3	t_4	t_5	t_6	...
1	2	16	272	7936	353792	...

t_n is the number of increasing labeled full binary trees with $2n-1$ vertices. To see that, we are going to use the **Milliken's tree theorem**.

Trees

A **tree** is merely a set of strings.

Definition (Trees)

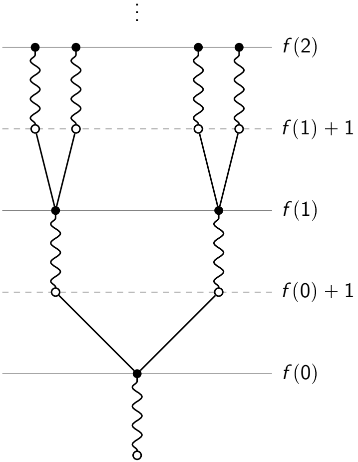
A tree $T \subseteq 2^{<\mathbb{N}}$ is **meet-closed** if $\sigma, \tau \in T$, their longest common prefix is in T. We write $T^\wedge$ for the meet-closure of T.

Definition (Levels)

Given a tree T and $\sigma \in T$, **the level** of σ is the number of prefixes of σ in T. We denote by $T(n)$ the set of nodes of T of level n.

Definition (Strong subtrees)

A set $T \subseteq 2^{<\mathbb{N}}$ is a **strong tree** if it is meet-closed and nodes of the same level in T are on the same level in $2^{<\mathbb{N}}$.



Milliken's tree theorem for singletons

Proposition (Milliken's tree theorem for singletons)

$\mathbf{MT}^1_k$: *For any $f : 2^{<\mathbb{N}} \to \{0, \dots, k-1\}$, there is a strong subtree $S \subseteq 2^{<\mathbb{N}}$ with no dead ends such that $|f(S)| = 1$.*

Remark : the theorem remains true starting with any strong tree T in place of $2^{<\mathbb{N}}$.

Proof : Let $f : 2^{<\mathbb{N}} \to \{0, 1\}$. Either there exists a string σ and infinitely many n such that $f(\sigma\tau) = 0$ for every τ of length n, or for every σ and almost every n there is a string τ of length n such that $f(\sigma\tau) = 1$. In any case we can computably build a monochromatic strong subtree.

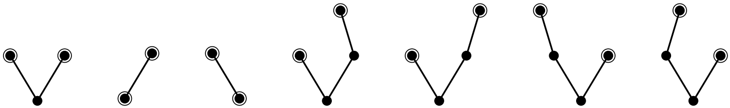
This does not work for a function $f : [2^{<\mathbb{N}}]^2 \to \{0, \dots, k-1\}$ because we can define the function $f(\{\sigma, \tau\}) = 0$ if σ, τ are incomparable and $f(\{\sigma, \tau\}) = 1$ otherwise. Any strong subtree necessarily have comparable and incomparable strings.

Definition (ACDMP, The strong generalized tree principle)

$\mathbf{SGTT}^n_{k,l}$: For any $f : [2^{<\mathbb{N}}]^n \to \{0, \dots, k-1\}$, there is a strong subtree $S \subseteq 2^{<\mathbb{N}}$ with no dead ends such that $|f(S)| \leqslant l$.

The strong generalized tree principle for pairs

We can force at least **seven** colors in any strong subtree with no dead ends:



These seven pictures above each represent an **embedding type**.

Definition (level-closure)

A set of strings S is **level closed** if for any $\sigma, \tau \in S$ with $|\sigma| < |\tau|$, the prefix of τ of length $|\sigma|$ is in S. We write S^{cl} for the smallest meet-closed and level-closed tree generated by S (the smallest strong tree containing S).

Definition (embedding types)

Two finite strong trees S_0, S_1 are **strongly isomorphic** if there is a bijection $f : S_0 \to S_1$ such that $\sigma i \preceq \tau \leftrightarrow f(\sigma)i \preceq f(\tau)$ for any $\sigma, \tau \in S_0$. **Embedding type** are equivalence classes of strongly isomorphic strong trees.

Milliken's tree theorem

Definition

Let T be a strong tree and $\mathfrak{e}$ an embedding type. We denote by $\mathcal{S}_{\mathfrak{e}}(T)$ the set of strong subtrees of T whose embedding type is $\mathfrak{e}$.

Let $\mathfrak{e}$ be an embedding type.

Theorem (Milliken's tree theorem)

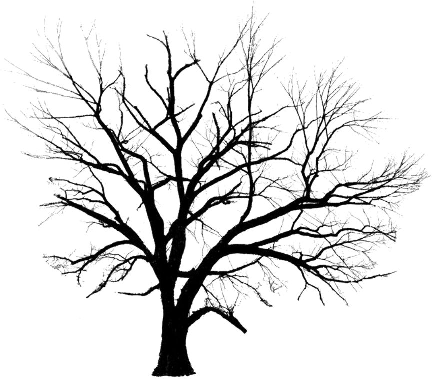
$\mathbf{MTT}^{\mathfrak{e}}_{\mathbf{k}}$: *For any $f : \mathcal{S}_{\mathfrak{e}}(2^{<\mathbb{N}}) \to \{0, \ldots, k-1\}$ be any coloring. Then there is a strong subtree $S \subseteq 2^{<\mathbb{N}}$ with no dead ends such that $|f(\mathcal{S}_{\mathfrak{e}}(S))| = 1$.*

Remark : Milliken's tree theorem is true starting with any strong tree T in place of $2^{\mathbb{N}}$.

Corollary (ACDMPT)

$\mathbf{SGTT}^2_{8,7}$ *is true and 7 is the big Ramsey degree of $[2^{<\mathbb{N}}]^2$ in strong trees.*

proof : We iterate Milliken's tree theorem seven times to make it monochromatic on the seven possible embedding types.

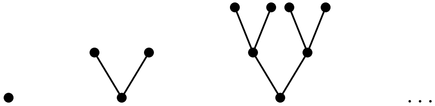


Section 2

Reverse mathematics

The computational content of the Milliken tree theorem

We now focus on the following embedding types:



Definition

Let $\mathcal{S}_n(T)$ denote $\mathcal{S}_{\mathfrak{e}_n}(T)$ where $\mathfrak{e}_n$ is the embedding type of the full tree of height n. Let $\mathrm{MTT}^{\mathrm{n}}_{\mathrm{k}}$ denotes $\mathrm{MTT}^{\mathfrak{e}_n}_{\mathrm{k}}$.

Proposition

Let $\mathfrak{e}$ be an embedding type of height n. Then $\mathrm{RCA}_0 \vdash \mathrm{MTT}^{\mathrm{n}}_{\mathrm{k}} \to \mathrm{MTT}^{\mathfrak{e}}_{\mathrm{k}}$.

proof : Given a color $f : \mathcal{S}_{\mathfrak{e}}(T) \to \{0, \dots, k-1\}$ we define a color $g : \mathcal{S}_{\mathfrak{e}_n}(T) \to \{0, \dots, k-1\}$ by $g(F) = f(F')$ where F' is the unique subtree of F of embedding type $\mathfrak{e}$. Apply $\mathrm{MTT}^{\mathrm{n}}_{\mathrm{k}}$.

Upper bound and lower bound of MTT^n_2

Upper bound:

Theorem (ACDMP)

For every n, $\mathrm{MTT}^{\mathrm{n}}_2$ has a Δ^0_{2n-1} solution and is then provable in ACA_0.

Lower bound:

Proposition

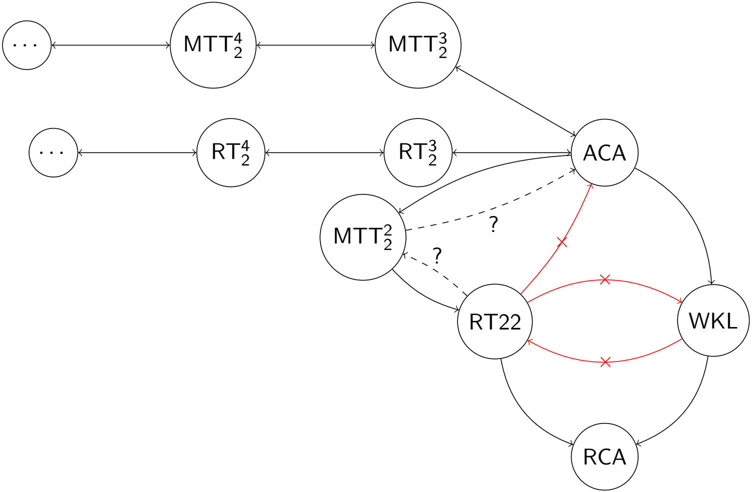
Let $\mathfrak{e}$ be an embedding type of height n. Then $\mathrm{MTT}^{\mathfrak{e}}_{\mathrm{k}}$ implies $\mathrm{RT}^{\mathrm{n}}_{\mathrm{k}}$. In particular for $n = 3$ we have a computable coloring of $\mathcal{S}_{\mathfrak{e}}(2^{<\mathbb{N}})$ every solution of which computes the halting problem.

Proof : We can transform a coloring of $[\mathbb{N}]^n$ into a coloring of $\mathcal{S}_{\mathfrak{e}}(2^{\mathbb{N}})$ identifying nodes with their levels.

Corollary

For any embedding type $\mathfrak{e}$ of height $\geqslant 3$, $\mathrm{MTT}^{\mathfrak{e}}_2$ implies ACA_0.

MTT^n_2 in reverse mathematics



What about MTT^2_2 ?

The case of MTT^2_2

Theorem (Following from a result of Patey)

RT^2_2 *does not imply* MTT^2_k.

Theorem (ACDMP, strong cone avoidance of MTT^1_2)

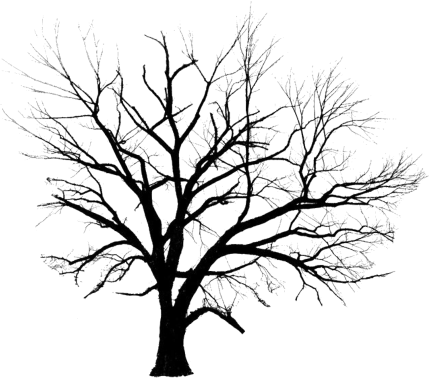
For any non-computable set C, any arbitrary instance of MTT^1_2 admits a solution which does not compute C.

Theorem (ACDMP, cone avoidance of MTT^2_2)

For any non-computable set C, any computable instance of MTT^2_k admits a solution which does not compute C.

Corollary (ACDMP)

MTT^2_k *does not imply* ACA_0.



Section 3

Devlin's theorem

Coming back to Devlin's theorem

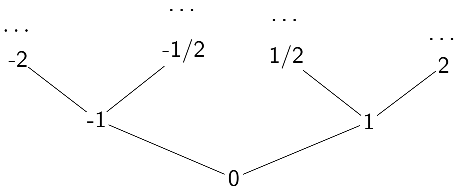
We saw that DT^2_2 is not a theorem. Given any enumeration $\{q_n\}_{n\in\mathbb{N}}$ of the rationals. Let

$$\begin{aligned} f(\{q_n, q_m\}) &= 0 \text{ if } n < m \\ &= 1 \text{ otherwise} \end{aligned}$$

For any $X \cong \mathbb{Q}$ and any $q_n \in X$ there must exists m_1, m_2 sufficiently large such that $q_{m_1} < q_n < q_{m_2}$: DT^2_2 is false.

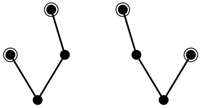
> Can we show that $\mathrm{DT}^2_{3,2}$ is a theorem ?

We equip $2^{<\mathbb{N}}$ with a total order isomorphic to $\mathbb{Q}$: $\sigma <_{\mathbb{Q}} \tau$ if there is a prefix $\tau' \preceq \tau$ such that $\tau'0 \npreceq \tau$ and $\tau'0 \preceq \sigma$.



Devlin embedding types

We are now interested in the two following embedding types:



Proposition (Devlin)

Given a strong tree T with no leaves, we can compute a countable anti-chain $A \subseteq T$ or order type $\mathbb{Q}$ and whose leaves generate only one of the two embedding types listed above.

This gives rise to the concept of Devlin embedding types:

Definition (Devlin)

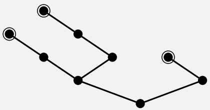
A **Devlin embedding type** of size n is the equivalence class of a finite strong tree with n leaves $\overline{\sigma}$ such that:

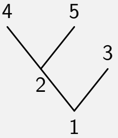
1. Every element of $\overline{\sigma}^{\wedge}$ is of different size.
2. Every node which is not a leaf and not branching "goes at the left".

Devlin embedding types and Joyce trees

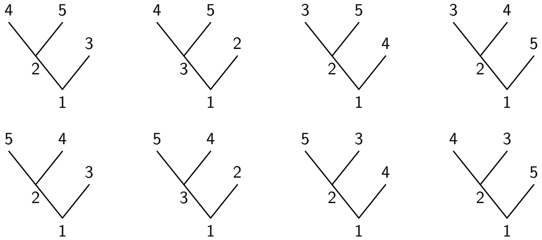
Devlin types of size n can be put in bijections with **Joyce trees** of height n : increasing labeled full binary trees with $2n-1$ vertices:

Here is a Devlin type of size 3 and its corresponding Joyce tree:





Here are 8 among the 16 Joyce trees of size 3 (the remaining cases are symmetric):



Devlin's theorem from Milliken's tree theorem

Theorem (Devlin)

Given a strong tree T with no leaves, we can compute a countable anti-chain $A \subseteq T$ or order-type $\mathbb{Q}$, among which each anti-chain of n strings always generates a Devlin embedding type, and such that each Devlin embedding type of size n is realized by any anti-chain $B \subseteq A$ isomorphic to $\mathbb{Q}$.

Iterating the Milliken's tree theorem on each Devlin embedding type:

Corollary (Devlin)

Let $\mathrm{dt_n}$ be the number of Devlin type of size n. Then $\mathrm{DT}^{\mathrm{n}}_{\mathrm{dt_n}+1,\mathrm{dt_n}}$ is a theorem and $\mathrm{DT}^{\mathrm{n}}_{\mathrm{dt_n},\mathrm{dt_n}-1}$ is false : $\mathrm{dt_n}$ is the big Ramsey degree of the set of n rationals.

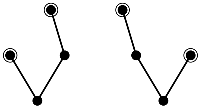
Corollary (Devlin)

For any n, $\mathrm{DT}^{\mathrm{n}}_{\mathrm{dt_n}+1,\mathrm{dt_n}}$ is provable in ACA_0.

Does $\mathrm{DT}^2_{3,2}$ admits cone avoidance ?

The Devlin's theorem in reverse mathematics

$\mathrm{DT}^2_{3,2}$ is a consequence of $\mathrm{MTT}^{\mathfrak{e}}_3$ for $\mathfrak{e}$ among the two following embedding types:



These embedding types are of size three and we can design a computable instance of $\mathrm{MTT}^{\mathfrak{e}}_3$ every solution of which computes $\emptyset'$ for each of them.

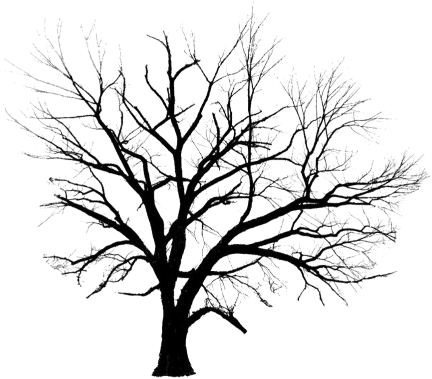
We can do something similar for $\mathrm{DT}^2_{3,2}$

Proposition (ACDMP)

There is a computable instance of $\mathrm{DT}^2_{3,2}$ every solution of which computes the halting problem (it can also be done for $\mathrm{DT}^2_{4,3}$).

Corollary (ACDMP)

For every $n \geqslant 2$, $\mathrm{DT}^{\mathrm{n}}_{\mathrm{dt_n}+1,\mathrm{dt_n}}$ is equivalent to ACA_0.



Section 4

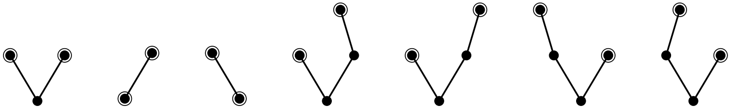
The generalized tree theorem

The generalized tree theorem

We come back to the strong generalized tree principles:

Definition (The strong generalized tree principles)

$\mathbf{SGTT}^n_{k,l}$: For any $f : [2^{<\mathbb{N}}]^n \to \{0, \dots, k-1\}$, there is a strong subtree $S \subseteq 2^{<\mathbb{N}}$ with no dead ends such that $|f(T)| \leqslant l$.



Definition

Let $e_{\mathrm{sTT}}(n)$ be the number of embedding types generated by n strings.

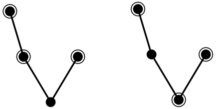
We have $e_{\mathrm{sTT}}(1) = 1$ and $e_{\mathrm{sTT}}(2) = 7$.

Proposition (ACDMP)

$\mathrm{SGTT}^2_{8,7}$ is provable in ACA_0 and $\mathrm{SGTT}^2_{7,6}$ is false. 7 is the big Ramsey degree of pairs of string with respect to strong trees with no leaves.

A refinement of embedding types : the tuple types

In general $\mathrm{SGTT}^n_{e_{\mathrm{sTT}}(n)+1, e_{\mathrm{sTT}}(n)}$ is not a theorem : two set of strings may generate the same embedding type, while **the role** of each string in the generation of this embedding type is different:



Two ways of generating the same embedding type with three strings.

Definition (Tuple types)

A **tuple type** is the equivalence class of the following equivalence relation defined on tuples of strings :

$\overline{\sigma}$ is equivalent to $\overline{\tau}$ if there is a strong bijection between the strong tree generated by $\overline{\sigma}$ and the one generated by $\overline{\tau}$, which maps elements of $\overline{\sigma}$ to elements of $\overline{\tau}$.

Let $t_{\mathrm{sTT}}(n)$ be the number of tuple types generated by n strings.

The generalized tree theorem

Theorem (ACDMP)

$\mathrm{SGTT}^n_{t_{\mathrm{sTT}}(n)+1, t_{\mathrm{sTT}}(n)}$ *is provable in* ACA_0 *and* $\mathrm{SGTT}^n_{t_{\mathrm{sTT}}(n), t_{\mathrm{sTT}}(n)-1}$ *is false.*

	0	1	2	3	4	...
$e_{\mathrm{sTT}}(n)$	1	1	7	345	136949	...
$t_{\mathrm{sTT}}(n)$	1	1	7	369	145215	...

These sequences have been obtained via brute force computation and do not appear on OEIS, The On-Line Encyclopedia of Integer Sequences.

$\Rightarrow$ they seem to be new natural combinatorial sequences.

The tree theorem

Definition (Chubb, Hirst, McNichol)

$\mathbf{TT}^n_k$: for any coloring of the n-tuples of comparable strings with k colors, there exists a – not necessarily strong – monochromatic perfect tree.

Theorem (Chubb, Hirst, McNichol)

TT^n_k is provable in ACA_0.

Definition (ACDMP)

$\mathbf{GTT}^n_{k,l}$: for any coloring of the n-tuples of strings with k colors, there exists a – not necessarily strong – perfect tree using at most l colors.

Definition (ACDMP)

An **ACDMP type** is a tuple type generated by a tuple $\overline{\sigma}$ such that:

1. every string of $\overline{\sigma}$ is not in $\overline{\sigma}^\wedge \backslash \overline{\sigma}$.
2. every string of $\overline{\sigma}^\wedge$ is of different length.
3. every node in $\overline{\sigma}^{cl}$ which is not a leaf and not branching "goes at the left".

The generalized tree theorem

Theorem (ACDMP)

Inside every strong perfect tree T we can compute with the help of T a perfect (non-strong) subtree S whose every tuple type is an ACDMP tuple type and such that every perfect subtree $R \subseteq S$ realizes every ACDMP tuple type.

Definition (ACDMP)

Let $t_{\mathrm{TT}}(n)$ be the number of ACDMP tuple type and $e_{\mathrm{TT}}(n)$ be the number of embedding type they belong to.

Theorem (ACDMP)

$\mathrm{GTT}^{\mathrm{n}}_{\mathrm{t_{TT}(n)+1,t_{TT}(n)}}$ *is a theorem provable in* ACA_0 *whereas* $\mathrm{GTT}^{\mathrm{n}}_{\mathrm{t_{TT}(n),t_{TT}(n)-1}}$ *is false.*

Corollary (Chubb, Hirst, McNichol)

$\mathrm{TT}^{\mathrm{n}}_{\mathrm{k}}$ *is a theorem provable in* ACA_0 *for every* n, k*.*

The reason is that there is only one ACDMP tuple type of size n generated by comparable strings.

Some open questions

The first values of our combinatorial sequences are:

	0	1	2	3	4	...
$e_{\mathrm{sTT}}(n)$	1	1	7	345	136949	...
$t_{\mathrm{sTT}}(n)$	1	1	7	369	145215	...
$e_{\mathrm{TT}}(n)$	1	1	7	27	561	...
$t_{\mathrm{TT}}(n)$	1	1	7	29	635	...

None of them appear on The On-Line Encyclopedia of Integer Sequences.

Question

Can any of these sequence be defined inductively by a simple closed formula ? what is the computational complexity of computing any of them ?

Question

Does every instance of $\mathrm{MTT}^{\mathrm{n}}_{\mathrm{k}}$ admits a Δ^0_{n+1} solution (we only have Δ^0_{2n-1}) ?

Question

Does MTT^2_2 implies WKL_0 ?