Tameness in least fixed-point logic and McCollm’s conjecture

Siddharth Bhaskar ¹  Alex Kruckman ²

¹Datalogisk Institut, Københavns Universitet
²Department of Mathematics and Computer Science, Wesleyan University
Motivating questions

- What does classification theory say about finite model theory/descriptive complexity theory, and vice versa?
- What is the interaction of model-theoretic tameness with classical open problems in FM/DCT?
- To what extent are there robust dividing lines among classes of finite models, and not just for first-order definability?
- Main motivating question: difference between FO and LFP logic on classes of finite structures.
Least fixed-point logic

- Consider first-order $L$-formulas $\varphi(\vec{x}, S)$ with second-order (relation) variables $S$.
- If the length of $\vec{x}$ and the arity of $S$ agree, then $\varphi$ defines a functional $\mathcal{P}(A^n) \to \mathcal{P}(A^n)$ for any $L$-structure $A$.
- The stages of an operative positive elementary $L$-formula $\varphi$ are defined by
  - $\varphi^0 = \emptyset$
  - $\varphi^{n+1} = \{x : \varphi(x, \varphi^n)\}$
  - $\varphi^\alpha = \bigcup_{\beta < \alpha} \varphi^\beta$
- If $S$ occurs positively, then this functional is monotone, so has a least fixed-point, over every $L$-structure.
Least fixed-point logic

• For an $L$-structure $A$, the closure ordinal $\|\varphi\|_A$ is the least $\Gamma$ such that $A \models \varphi^\Gamma = \varphi^{\Gamma+1}$.

• LFP formulas are obtained by extending FO logic by the LFP quantifier,

$$(\text{lfp}_{\vec{x},S} \varphi)(\vec{t}) := \varphi^\Gamma$$

for any operative, positive elementary $\varphi(\vec{x}, S)$

• For $x_1, x_2 \in A$, we say $x_1 \prec x_2$ in case $x_1$ “appears in” $\varphi^\alpha$ before $x_2$. (This is the stage comparison preorder over $\varphi$)

• Moschovakis: the stage comparison preorder of any LFP formula $\varphi(x, S)$ is itself LFP-definable, uniformly over all structures.
A family of finite structures $C$ is proficient in case there exists some $\varphi$ such that $\|\varphi\|_A$ is unbounded in $\omega$ as $A$ ranges over $C$.\footnote{For any finite structure $A$, $\|\varphi\|_A < \omega$.}

Observation: if $C$ is not proficient, then $\text{FO} = \text{LFP}$ over $C$.

McColm, 1986: conjecture that if $C$ is proficient, then $\text{FO} \neq \text{LFP}$ over $C$.

Slogan: non-proficiency is a finite-variable version of countable categoricity!

Roughly speaking, $C$ is non-proficient if for each $m \leq n$, it realizes at most finitely many types consisting of of $\text{FO}^n$ formulas of arity $m$. 
A brief history

- **McColm, 1986:** If $C$ is proficient, then $\text{FO} \neq \text{LFP}$ over $C$.
- **Gurevich, Immerman, Shelah, 1994:** Not so. There are proficient families of structures over which $\text{FO} = \text{LFP}$.
- **Kolaitis and Vardi, 1992:** If $C$ is ordered, then $\text{FO} \neq \text{LFP}$ over $C$.
- A resolution either way of the ordered conjecture would resolve a major open problem in computational complexity!
  - Positive resolution: $\text{LH} \preceq \text{ETIME}$ (Dawar, Hella, 1995)
  - Negative resolution: $\text{PTIME} \preceq \text{PSPACE}$ (Dawar, Lindell, Weinstein, 1995)
Our contribution

**Theorem (BK, 2017)**

*McColm’s conjecture holds for tame classes of finite structures $C$.*

- The proof involves investigating LFP analogues of FO dividing lines like OP, IP, SOP, TP2.
- We also completely classify the implications among the LFP versions of these properties, namely
  \[
  \text{LFP} - \text{SOP} \implies \text{LFP} - \text{TP2} \implies \text{LFP} - \text{IP} \iff \text{LFP} - \text{OP}
  \]
- Note the surprising equivalence between the order and independence property!
**Definition**
The elementary limit theory $\text{Th}^\infty(C)$ of $C$ is the set of sentences which hold in cofinitely many structures of $C$.

**Lemma (Lindell)**
The proficiency of $C$ and whether or not $\text{FO} = \text{LFP}$ over $C$ are both properties of $\text{Th}^\infty(C)$.

**Corollary**
McColm’s conjecture is a property of $\text{Th}^\infty(C)$.
Lemma

$(\text{lfp} \varphi)(t)$ is first-order definable over $C$ iff there is a first-order formula $\theta(t)$ such that the following sentences are in $\text{Th}^\infty(C)$:

\[
\varphi(x, \theta) \iff \theta(x) \quad (1)
\]

\[
\psi(x, \neg \theta) \iff \neg \theta(x) \quad (2)
\]

where $\psi$ is complementary to $\varphi$.

Proof.

$(\iff)$ (1) says that $\theta$ is a fixed point of $S \mapsto \varphi(x, S)$, hence $(\text{lfp} \varphi)(t) \rightarrow \theta(t)$. Similarly, (2) says that $(\text{lfp} \psi)(t) \rightarrow \theta(t)$.

Since they are complementary, $(\text{lfp} \varphi)(t) \leftrightarrow \theta(t)$.
Let $\varphi(x; y)$ be any formula (FO or LFP), $n \in \mathbb{N}$, and $M$ be a structure.

- $\varphi$ has an \textit{n-instance of the order property} ($\text{OP}(n)$) in $M$ if there exist tuples $a_1, \ldots, a_n \in M^{|x|}$ and $b_1, \ldots, b_n \in M^{|y|}$ such that $M \models \varphi(a_i; b_j)$ if and only if $i \leq j$.

- $\varphi$ has an \textit{n-instance of the independence property} ($\text{IP}(n)$) in $M$ if there exist tuples $a_i \in M^{|x|}$ for all $i \in \{1, \ldots, n\}$ and $b_X \in M^{|y|}$ for all $X \subseteq \{1, \ldots, n\}$ such that $M \models \varphi(a_i; b_X)$ if and only if $i \in X$.

- $\varphi$ has an \textit{n-instance of the strict order property} ($\text{SOP}(n)$) in $M$ if there exist tuples $b_1, \ldots, b_n \in M^{|y|}$ such that $\varphi(M; b_i) \subseteq \varphi(M; b_j)$ if and only if $i \leq j$. 
Let $\varphi(x; y)$ be any formula ($\text{FO}$ or $\text{LFP}$), $n \in \mathbb{N}$, and $M$ be a structure.

- $\varphi(x; y)$ has an \textit{n-instance of the tree property of the second kind} ($\text{TP}_2(n)$) in $M$ if there are tuples $b_{i,j} \in M^{|y|}$ for $1 \leq i, j \leq n$ such that for any $i$ and any $j \neq k$, $\varphi(M; b_{i,j}) \cap \varphi(M; b_{i,k}) = \emptyset$, but for any function $f : \{1, \ldots, n\} \to \{1, \ldots, n\}$,

\[
\bigcap_{i=1}^{n} \varphi(M; b_{i,f(i)}) \neq \emptyset.
\]
Tame classes of theories

- A family $\mathcal{C}$ of finite structures has property (FO- or LFP-) $P$, for any $P \in \{OP, IP, SOP, TP2\}$, in case there is a (FO- or LFP-) formula $\varphi$ with arbitrarily large instances of $P(n)$ in $M$, as $M$ varies over structures in $\mathcal{C}$.
- Therefore, $\mathcal{C}$ does not have $P$, in case the $n$-instances of $P$, for any formula $\varphi$, are uniformly bounded as we vary over all $M \in \mathcal{C}$.
- $\mathcal{C}$ having FO-$P$ is equivalent to $\text{Th}^\infty(\mathcal{C})$ having $P$.
- We study these four, because:
  - Most commonly studied combinatorial dividing lines imply either NSOP or NTP2, and
  - IP and OP are otherwise the “most important.”
LFP-SOP and proficiency

- Suppose a family of structures is proficient. Then the stage comparison relation witnesses LFP-SOP.
- Conversely, if $\varphi(\vec{x}, \vec{y})$ witnesses LFP-SOP, then $\psi(\vec{y}_1, \vec{y}_2) \equiv (\forall \vec{x})(\varphi(\vec{x}, \vec{y}_1) \rightarrow \varphi(\vec{x}, \vec{y}_2))$ defines a partial order with arbitrarily long chains.
- Given a partial order with arbitrarily long chains, we can define a linear order with arbitrarily long chains (roughly, by comparing rank).
- Given a linear order with arbitrarily long chains, we can define a proficient formula.
LFP-SOP implies all other properties

- LFP-SOP $\Rightarrow$ LFP-OP, under general considerations.
- Consider the family $\mathcal{N}$ of finite initial segments of $(\mathbb{N}, <)$.
- Over $\mathcal{N}$, the BIT and FACTOR predicates are LFP-definable
  - BIT($x, y$) $\iff$ the $x$-th bit of $y$ base 2 is 1.
  - FACTOR($x, y, z$) $\iff$ $y^z$ is the largest power of $y$ dividing $x$.
- BIT has IP, FACTOR has TP2
- Let $b_{i,j} = (p_i, j)$, where $(p_i)_{i \in \omega}$ is an enumeration of the primes: for any function $f : n \to n$, let $a_f = \prod_{i < n} p_i^{f(i)}$. Then,
  - FACTOR($\mathbb{N}; p_i, j$) and FACTOR($\mathbb{N}; p_i, k$) are disjoint, but
  - $a_f \in \bigcap_{i < n} \text{FACTOR}(\mathbb{N}; p_i, f(i))$
- Hence, LFP – SOP $\Rightarrow$ LFP – IP and LFP – TP2 over any $\mathcal{C}$. 
First-order characterizations of $\text{LFP-P}$

**Lemma**  
For any $P \in \{\text{OP, IP, SOP, TP2}\}$, $C$ has $\text{LFP-P}$ iff $C$ is proficient or $C$ has $\text{FO-P}$.

**Corollary**  
Whether or not $C$ has $\text{LFP-P}$ is a property of $\text{Th}^\infty(C)$.

**Corollary**  
For any property $P$, McColm’s conjecture holds for any $C$ that fails any $\text{FO-P}$.

**Proof.**  
If $C$ is proficient, then it satisfies $\text{LFP-P}$, but fails $\text{FO-P}$.  □
Implications among LFP-P

• We know LFP-SOP $\implies$ LFP-TP2 $\implies$ LFP-IP $\implies$ LFP-OP. What about conversely?

• There are countably categorical theories with the finite model property that have:
  - IP but NTP2 (e.g., theory of the random graph)
  - TP2 but NSOP (e.g., generic theory of parameterized equivalence relations)

  which show the first two implications are strict.

• If $C$ has LFP-OP, but is not proficient, then FO = LFP, so by OP $\iff$ IP $\lor$ SOP (Shelah), it must have IP.

• Hence, LFP-OP $\implies$ LFP-IP!
Future work

- Investigate combinatorial dividing lines for other fixed-point logics (e.g., transitive closure logic).
- Can we obtain some sort of asymptotic structure theory for tame, non-proficient classes of structures?
- To what extent can FMT assumptions of bounded cliquewidth and bounded treewidth be assimilated into model-theoretic tameness considerations?
- Chen and Flum, 2012: The ordered conjecture is true for families of finite structures of bounded cliquewidth and treewidth.
- Open question: Does the ordered conjecture hold for any tame family of finite structures?
Acknowledgements

• Thanks to Cameron Hill, Greg McColm, Steve Lindell, and Scott Weinstein for their help and encouragement.

• Thanks to Larry Moss and the logic group at Indiana University for their support.

• Thanks for listening! If you want to talk more about this, please contact us by email.