# The reverse mathematics of inductive inference

Rupert Hölzl
National University of Singapore
Joint work with Frank Stephan and Sanjay Jain

# Purpose of this work

*"Analyze algorithmic learning using the methods of reverse mathematics."*

# 1 Preliminaries

# Reverse mathematics

1. Mathematicians commonly have intuitions about theorems implying each other, theorems being stronger or weaker than others, etc.
2. But all true theorems are logically equivalent, so how can we formalize this intuition?
3. Idea by Friedman, developed by Simpson:
   - Use weak base theory that only allows dumb, purely mechanical manipulations.
   - Then add one theorem, and see if it can be used to prove the other.
   - This allows comparing their strengths.
4. Results from many areas of mathematics have been classified in this way.

# The weak system

1. The weak system will be a fragment of second order arithmetic, that is, we only assume a subset of the usual axioms.
2. Very different models of such a reduced axiom set can exist.
3. For example, in such a model not all subsets of the natural numbers need to exist.
4. Use permitted manipulations to derive existence of new objects from objects known to exist.
5. Depending on what axioms we assume, different manipulations are permitted.

# Basic reverse mathematics principles

1. $\mathrm{I}\Sigma_n$ stipulates that induction holds for sets definable via $\Sigma_n$ formulas.
2. $\mathrm{RCA}_0$ stipulates that a model $(M, +, \cdot, <, 0, 1, \mathscr{S})$ consists of…
   - some model $M$ of the natural numbers satisfying $\mathrm{I}\Sigma_1$,
   - a set $\mathscr{S}$ of subsets of the natural numbers, *at least* containing all computable sets and closed under Turing reduction and join.
3. **Intuition.** $\mathrm{RCA}_0$ corresponds to the dumb manipulations mentioned above.
4. $\mathrm{ACA}_0$ adds closure of $\mathscr{S}$ under Turing jump.

# Inductive inference

1. An algorithm tries to identify an object from a known family.
2. Partial information is sequentially presented to the algorithm.
3. Based on this, the algorithm makes a guess about the object.
   (Strictly speaking, a finite representation of the object is guessed.)
4. Every time the algorithm is presented with a new piece of information, it may revise its guess.
5. Eventually the guess should converge to one of the correct representations of the object (*learning in the limit*).

# Inductive inference

1. Not generally possible for all families of objects, as some families will contain multiple objects that cannot be distinguished using this approach.
2. Variations are possible in …
   - the way the information is presented to the learner,
   - the required convergence behaviour of the learner,
   - the hypothesis space,
   - further restrictions on the behaviour of the learner.
3. Fixing all these conditions defines a *learning environment*.
4. In different environments, different families of objects can be learned.
5. **Topic of this work.** What influence do the different strengths of different fragments of second order arithmetic have?

# Representing families of sets and functions

1. We will study learnability of families of sets.
2. We need to talk about families of sets and of functions existing in a given model, so we represent them by single sets in $\mathscr{S}$.
3. **Uniform representation.** Represent families $\{F_e\}_{e\in M}$ of functions by a set $F \in \mathscr{S}$ via

$$F_e(x) = y \iff \langle e, x, y \rangle \in F.$$

4. $F$ is the *representation set* of the *uniformly represented family* $\{F_e\}_{e\in M}$ *of functions*.
5. If the functions are binary-valued then we can interpret this as a *uniformly represented family of sets*, and write $A$ and $\{A_e\}_{e\in M}$.

# Representing families of sets and functions

1. **Weak representation.** Assume some $F \in \mathscr{S}$ satisfies
   - $\forall e,x,y,z,y',z' : \ \langle e,x,y,z\rangle, \langle e,x,y',z'\rangle \in F \ \Rightarrow \ y = y' \wedge z = z'$,
   - $\langle e,x,y,z\rangle \in F \wedge x' < x \implies$
   $$\exists y',z' \colon \ \langle e,x',y',z'\rangle \in F \ \wedge \ \langle e,x',y',z'\rangle < \langle e,x,y,z\rangle.$$

   Then
   - let $D = \{e \colon \ \forall x \exists y,z [\langle e,x,y,z\rangle \in F]\}$;
   - call $\{F_e\}_{e \in D}$ the *weakly represented family of functions* defined by the *representation set* $F$ where
   $$\forall e \in D \ [F_e(x) = y \iff \exists z \ \langle e,x,y,z\rangle \in F];$$
   - call $D$ the *index set* of $\{F_e\}_{e \in D}$;
   - if $e \in D$ it is called *valid*, otherwise *invalid*.
2. If the functions are binary-valued then we can interpret this as a *weakly represented family of sets*, and write $\{A_e\}_{e \in D}$ and $A$.

# Uniform versus weak representation

1. Weakly represented families of functions are much more general than uniformly represented families of functions.
   - **Example.** For $A \in \mathscr{S}$ the family of all $A$-computable functions is weakly representable but in general not uniformly representable.
2. Uniformly represented families of functions can be thought of as weakly represented families with $D = M$.
3. In general, the index set $D$ of a weakly represented family need not be in $\mathscr{S}$.

# Sets versus functions

1. All learnability results in this talk will concern families of *sets* given as uniformly or weakly represented families. The words "of sets" are usually omitted.
2. Weakly represented families *of functions* will appear only in reverse mathematics axioms.

# Texts

1. As mentioned above, a possible learning target is presented to the learner in an infinite sequences of data, so-called *texts*.
2. Their definition should be compatible with reverse maths.
3. That is, when $M$ is the standard natural numbers then the definition should coincide with the traditional one, but in case of non-standard models it will differ.
4. **Definition.** We call # the pause symbol.
5. **Definition (Angluin, Gold).** A *text* for an $A \in \mathscr{S}$ is a function $T \in \mathscr{S}$ with $T\colon M \to M \cup \{\#\}$ such that

$$\{T(n) : n \in M \wedge T(n) \neq \#\} = A.$$

6. Write $M^*$ for the set of finite sequences over $M \cup \{\#\}$.
7. $M^*$ can be understood as the prefixes of texts.

**Definition (Angluin, Gold).**

1. A *learner* is a function $L : M^* \to M$ in $\mathscr{S}$.
2. A learner $L$ *learns* $\{A_e\}_{e \in D}$ *with hypothesis space* $\{B_e\}_{e \in E}$ *in the limit* if for every $e \in D$ and every text $T$ for $A_e$ there is an $n$ such that for all $m \geq n$,
   - $L(T(0)\dots T(m)) = L(T(0)\dots T(m+1))$, (convergence)
   - $L(T(0)\dots T(m)) \in E$, (validity)
   - $B_{L(T(0)\dots T(m))} = A_e$. (correctness)
3. $\{A_e\}_{e \in D}$ *is learnable in the limit with hypothesis space* $\{B_e\}_{e \in E}$ if there exists an $L$ as above.
4. If $E = D$ and for all $e \in D$ we have $B_e = A_e$, we will just say that $L$ *learns* $\{A_e\}_{e \in D}$ *in the limit* and that $\{A_e\}_{e \in D}$ *is learnable in the limit*, respectively.

# 2 The tell-tale criterion

# The tell-tale criterion

1. Angluin studied *indexed families* of sets, that is, $\{A_e\}_{e\in\mathbb{N}}$ such that there is a computable $A$ with $A(e,x) = 1 \Leftrightarrow x \in A_e$.
2. Indexed families are a special case of uniformly represented families as they are *actually* uniformly computable.
3. Angluin identified the *tell-tale criterion* for their learnability: $\{A_e\}_{e\in\mathbb{N}}$ is learnable in the limit iff uniformly in $e \in \mathbb{N}$ one can computably enumerate a finite *tell-tale set* $B_e \subseteq A_e$; that is,

$$\nexists d \in \mathbb{N}\colon\ B_e \subseteq A_d \subset A_e.$$

4. **Intuition.** Once we have seen the tell-tale set for $e$ we can conjecture $e$ without danger of overgeneralizing.
5. Of course it is typically still a possibility that we will later have to discard $e$ to conjecture a *super*set of $A_e$.

# The tell-tale criterion in reverse mathematics

1. Tell-tale sets are classically finite.
2. To obtain results analogous to the tell-tale criterion for uniformly and weakly represented families it suffices to consider *tell-tale bounds* $b_e$ for each $e \in M$ such that there is no $d \in M$ with $A_e \cap \{0, 1, \ldots, b_e\} \subseteq A_d \subset A_e$.

# Levels of effectivity of tell-tale bounds

**Definition.** A weakly represented family $\{A_e\}_{e \in D}$ satisfies Angluin's criterion…

1. … *in general* if for all $e \in D$ there is a $b_e$ …
2. … *in the limit* if there is a function $g \in \mathscr{S}$ such that for all $e \in D$ the sequence $i \mapsto g(\langle e, i \rangle)$ converges from below to $b_e$ …
3. … *effectively* iff there is a function $e \mapsto b_e$ in $\mathscr{S}$ …

…such that there is no $d \in D$ with $A_e \cap \{0, \ldots, b_e\} \subseteq A_d \subset A_e$.

For uniformly represented families these definitions hold with $D = M$.

# 3 Main results

# The principle DOM

1. As mention above, the indexed families studied by Angluin correspond to uniformly represented families.
2. It is therefore not surprising that to obtain results for weakly represented families we need assumptions exceeding $\mathsf{RCA}_0$.
3. **The principle** DOM**.** For every weakly represented family in $\mathscr{S}$ of functions there is a function in $\mathscr{S}$ which grows faster than any function in the family.

# Learning weakly represented families

**Theorem.** T.f.a.e. over $\mathsf{RCA}_0$.

1. DOM
2. The index set of every weakly represented family of functions can be approximated in the limit.
3. Every weakly represented family satisfying Angluin's criterion effectively is learnable in the limit.
4. Every weakly represented family satisfying Angluin's criterion in the limit is learnable in the limit.
5. Every weakly represented family satisfying Angluin's criterion generally is learnable in the limit.

# Learning uniformly represented families

1. **Theorem.** Over $\mathsf{RCA}_0$, a uniformly represented family is learnable in the limit if and only if it satisfies the tell-tale criterion in the limit.
2. **Theorem.** Over $\mathsf{RCA}_0$, DOM holds iff every uniformly represented family that satisfies the tell-tale criterion in general is learnable in the limit.

# 4 Learning in alternate environments

# Conservative learning

1. Angluin introduced the notion of *conservative learning*: conservative learners avoid mind-changes until forced by data.
2. Before conjecturing a set $L$ they need to ensure that no proper subset of $L$ can explain the data seen so far.
3. **Reverse mathematics version.** If $n < m$ and

$$L(T(0)\dots T(n)) \neq L(T(0)\dots T(m))$$

then $L(T(0)\dots T(n)) \notin D$ or $\exists k \leq m\colon\ T(k) \in M - A_{L(T(0)\dots T(n))}$.

4. **Theorem.** Over $\mathrm{RCA}_0$, a uniformly represented family $\{C_e\}_{e\in M}$ is conservatively learnable with hypothesis space $\{A_e\}_{e\in M}$ if and only if $\{C_e\}_{e\in M}$ is contained in a uniformly represented family $\{B_e\}_{e\in M}$ which satisfies the tell-tale criterion effectively.

# Learning with $\mathsf{ACA}_0$

**Corollary.** T.f.a.e. over $\mathsf{RCA}_0$.

1. $\mathsf{ACA}_0$
2. Every uniformly represented family that satisfies the tell-tale criterion in general is conservatively learnable.
3. Every weakly represented family that satisfies the tell-tale criterion in general is conservatively learnable.

# Partial learning

1. Osherson, Stob, and Weinstein introduced *partial learning*: successful partial learners output one correct hypothesis infinitely often and all other hypotheses at most finitely often.
2. **Theorem.** Over $\mathrm{RCA}_0$, every uniformly represented family is partially learnable.
3. **Theorem.** Over $\mathrm{RCA}_0$ and $\mathrm{I}\Sigma_2$, every weakly represented family $\{A_e\}_{e\in D}$ is partially learnable with hypothesis space $\{B_{\langle e,b\rangle}\}_{e\in D, b\in M}$ where $B_{\langle e,b\rangle} = A_e$ for all $e \in D$ and $b \in M$.
4. **Theorem.** Over $\mathrm{RCA}_0$, $\mathrm{I}\Sigma_2$ and DOM, every weakly represented family $\{A_e\}_{e\in D}$ is partially learnable.

# 5 Sufficient learning criteria

# Finite thickness

1. **Definition (Angluin; modified).** $\{A_e\}_{e \in D}$ has *finite thickness* iff every $x \in M$ is contained in only finitely many $A_e$, that is, for every $x \in M$ there is $b$ such that for all $e > b$, either $e \notin D$ or $x \notin A_e$ or $A_e = A_d$ for some $d \leq b$.
2. **Theorem.** Over $\mathsf{RCA}_0$, every uniformly represented family which has finite thickness is learnable in the limit.
3. **Theorem.** Over $\mathsf{RCA}_0$, DOM is equivalent to the statement that every weakly represented family which has finite thickness is learnable in the limit.

# Finite elasticity

1. **Definition (Angluin; modified).** $\{A_e\}_{e \in D}$ has *finite elasticity* iff for every
$$T : M \to M \cup \{\#\}$$
in $\mathscr{S}$ there is $\sigma \preceq T$ such that for all $e \in D$,
$$\operatorname{range}(\sigma) \subseteq A_e \implies \operatorname{range}(T) \subseteq A_e.$$
2. **Theorem.** Over $\mathsf{RCA}_0$, every uniformly represented family which has finite elasticity is learnable in the limit.
3. **Theorem.** Over $\mathsf{RCA}_0$, DOM is equivalent to the statement that every weakly represented family which has finite elasticity is learnable in the limit.

# Characteristic subsets

1. **Definition (Angluin, Kobayashi; modified).** $\{A_e\}_{e \in D}$ *admits characteristic subsets* iff for every $e \in D$ there is a $b$ such that for all $d \in D$ it holds that

$$A_e \cap \{0, 1, \ldots, b\} \subseteq A_d \implies A_e \subseteq A_d.$$

2. Stronger than the tell-tale criterion, as the latter merely requires

$$A_e \cap \{0, 1, \ldots, b\} \subseteq A_d \Rightarrow A_d \not\subset A_e.$$

3. **Theorem.** Over $\mathrm{RCA}_0$, every uniformly represented family which admits characteristic subsets is learnable in the limit.
4. **Theorem.** Over $\mathrm{RCA}_0$, DOM is equivalent to the statement that every weakly represented family which admits characteristic subsets is learnable in the limit.

# 6 Conclusion

# Take-away message

*"Fixing different subsystems of second order arithmetic of different strengths has a strong influence on whether algorithmic learning is possible within a given learning environment."*

The relevant principles over $\mathrm{RCA}_0$, that were identified in this talk, are DOM, $\mathrm{I}\Sigma_2$, and $\mathrm{ACA}_0$.

# Take-away message

*"Fixing different subsystems of second order arithmetic of different strengths has a strong influence on whether algorithmic learning is possible within a given learning environment."*

The relevant principles over $\mathsf{RCA}_0$, that were identified in this talk, are DOM, $\mathrm{I}\Sigma_2$, and $\mathsf{ACA}_0$.

*Thank you for your attention.*