Composition with algebra at the background*

on a question by Gurevich and Rabinovich on the monadic theory of linear orderings

Thomas Colcombet
thomas.colcombet@liafa.univ-paris-diderot.fr

Liafa/Cnrs/Université Denis Diderot - Paris 7

Abstract. Gurevich and Rabinovich raised the following question: given a property of a set of rational numbers definable in the real line by a monadic second-order formula, is it possible to define it directly in the rational line? In other words, is it true that the presence of reals at the background do not increase the expressive power of monadic second-order logic?
In this paper, we answer positively this question. The proof in itself is a simple application of classical and more recent techniques. This question will guide us in a tour of results and ideas related to monadic theories of linear orderings.

1 In which the legacy is acknowledged

Büchi, Elgot, Kleene, Rabin, Scott, Shelah, Schützenberger, Trakhtenbrot and others shaped the notion of regular languages as we know it. In less than two decades, a beautiful theory involving computability, logic, algebra and topology has emerged. Today's researcher still walk on this path and are far from its end.

Most of the attention in this theory is put on **monadic second-order logic**. Monadic logic (we will drop the second-order from now) is the extension of first-order logic with the possibility to quantify over sets of elements. For comparison, full second-order logic would allow to quantify over relations of all arities, while in monadic logic, only 1-ary relations are allowed, and these can be interpreted as sets. 1-ary relations are called **monadic**.

Monadic logic allows for instance to express properties of directed graph. In this case, we assume that the elements are vertices of the graph and the only available symbol $\texttt{edge}(x, y)$ expresses the existence of an edge from vertex x to vertex y. The existence of a path starting from a node i and reaching a node f can be expressed in monadic logic, as follows:

> All sets of vertices that contain i and are closed under the edge relation also contain f.

* The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n 259454.

Of course, this is to be translated in mathematical symbols, as follows:

$$\begin{aligned}\texttt{path}(i,f) \quad &\stackrel{\texttt{def}}{=} \quad \forall Z(i \in Z \wedge \forall x \forall y\, (x \in Z \wedge \texttt{edge}(x,y) \to y \in Z)) \\ &\qquad\qquad\qquad \to f \in Z\ .\end{aligned}$$

The important convention here is that lower case letter (here i, f, x, y) implicitly range over elements, while upper case letters (here Z) range over sets of such elements. We also see that a membership relation is available, such as in $x \in Z$, with an obvious meaning. In some works, emphasizing on the fact that sets are 1-ary relations, this is denoted as $Z(x)$.

(In this example, the power of monadic logic compared to first-order logic, is already transparent. Indeed, it is a classical fact that reachability, i.e., the existence of a path, is not expressible in first-order logic. This is usually the first application that is given to Ehrenfeucht-Fraïssé games.)

The view we have on monadic logic is the one of **algorithmic model theory**, *i.e.*, the aim is to develop algorithms that can "solve" the logic. Solving has several meanings depending on the situation. The first is **satisfiability** which means, given a formula, to determine the existence of a structure for which it holds (a model). The second is **model-checking** which means, given a formula and some description of a structure (possibly infinite), to determine if the formula holds on this specific structure. Grand expectations have to be immediately lowered: satisfiability of monadic logic is undecidable as such. This is inherited from first-order logic, that it is extending, which is already undecidable [11].

The interest of monadic logic appears when it is considered on structures of specific shapes, namely **word**s and trees or resemblant models. We will not develop the tree-like structures in this paper. By words, we mean structures that are linearly ordered, and in which each element is decorated by some finite local information. These are also referred to as **chain**s. These two terminologies describe properties of exactly the same nature, but differ concerning many choices of notations. In particular the notations for chains are consistent with the view of logic, while the notations for words are consistent with language theory. We adopt in this paper the *word terminology* and stick to it, even if it is rightfully arguable.

Let us explain how a word can be encoded as a relational structure: a **word** can be seen as a linearly ordered set of positions enriched with unary predicates that describe what letter is carried by each position. Formally, a word over an alphabet $\mathbb{A}$ is a structure $(L, \leq, (a)_{a\in\mathbb{A}})$ where $\leq$ equips L of a total order – we call $(L, \leq)$ the **domain** of the word, and refer to $(L, \leq)$ as a **linear ordering** – and for all elements x of L there exists one and exactly one letter $a \in \mathbb{A}$ such that $a(x)$ holds. Remark here that we did not make any assumptions concerning the finiteness of the linear ordering. This is important since we will eventually be entering the realm of infinite structures. We will refer to **countable words** or **ω-words** if the underlying linear ordering is countable or isomorphic to ω respectively.

A language of finite words over the alphabet $\mathbb{A}^*$ is **monadic definable** if there is a monadic sentence φ such that $L = \{u \in \mathbb{A}^* \ : \ u \models \varphi\}$ (where, as is usual $u \models \varphi$ is pronounced "u **models** φ" and means that φ holds over the relational structure encoding the word u). If $\varphi(X_1, \dots, X_k)$ is a monadic formula of free monadic variables $X_1, \dots, X_k$, and $A_1, \dots, A_k$ are subsets of the domain of a word u, then $u \models \varphi(A_1, \dots, A_k)$ is true if φ is satisfied on u when its free variables $X_1, \dots, X_k$ take $A_1, \dots, A_k$ as respective values.

For instance, over finite words, a property $\varphi(X)$ expressing that "between any two occurrences of the letter a, at least one point belongs to X" can be written as $\forall x \forall y\, (x < y \wedge a(x) \wedge a(y)) \rightarrow \exists z (z \in X \wedge x < z < y)$.

The starting point in the description of monadic logic is its equivalence over finite words with regular languages, as it has been found independently by Trakhtenbrot on the one side, and Elgot and Büchi on the other side:

Theorem 1 (Büchi, Elgot and Trakhtenbrot [1, 4, 10]). *A language of finite words is definable in monadic logic if and only if it is regular*[1]*.*
Furthermore, the translations are effective, and as a consequence, satisfiability of monadic logic over finite words is decidable.

The decidability result means that we can symbolically test the formula toward an infinite number of potential inputs, here words. But Büchi made a further step by showing that in such results, even the input can be infinite. Here, an ω-word is a word such that the underlying linear ordering is isomorphic to ω, or equivalently $(\mathbb{N}, \leq)$.

Theorem 2 (Büchi [2]). *A language of ω-words is definable in monadic logic if and only if it is recognized by a Büchi-automaton*[2]*.*
Furthermore, the translations are effective, and as a consequence, satisfiability of monadic logic over ω-words is decidable.

This result was the first success in the decidability of the monadic theory of some infinite models.

Remark 1. The original result of Büchi establishes the decidability of the monadic theory of $(\mathbb{N}, \leq)$. Of course, $(\mathbb{N}, \leq)$ can be seen as an instance of a word over a unary alphabet. Thus, from Theorem 2, we can deduce the decidability of $(\mathbb{N}, \leq)$. The converse also holds. Indeed, a word over the finite alphabet $\mathbb{A}$ can be encoded, *e.g.,* as the linear ordering of its domain together with sets $(X_a)_{a \in \mathbb{A}}$ such that a position i belongs to X_a if and only if i carries the letter a. Using this encoding, a monadic formula over ω can guess an ω-word using existential quantifiers over sufficiently many monadic variables. Using this technique, the decidability of the satisfiability of monadic logic over ω-words can be deduced from the decidability of the monadic theory of ω. These kind of encodings are doable for all linear orderings.

Of course, we can go further. The two landmark linear orderings are the **rational line** $(\mathbb{Q}, \leq)$ (sometimes denoted η of $\mathcal{Q}$), and the **real line** $(\mathbb{R}, \leq)$ (sometimes denoted λ or $\mathcal{R}$).

[1] Say, recognized by a finite state automaton.
[2] We do not present this very classical model here.

Theorem 3 (Rabin [8]). *The monadic theory of the rational line is decidable, or equivalently the satisfiability of monadic logic over words of domain the rational line is decidable.*

In fact, the proof of Rabin concerns the infinite binary tree, which is a richer structure than the rational line. The proof of Rabin establishes that, *over infinite trees*, monadic logic is equivalent to a certain form of automata. The decidability of the rational line is then obtained by interpreting the rational line inside the infinite binary tree. This allows to decide the monadic theory of the rational line, but it is not at all informative concerning the expressive power of monadic logic over the rational line. In particular, one cannot deduce, say, a model of automata that would have the expressive power of monadic logic over the rational line.

Example 1. In order to illustrate the above theorem, let us try to give some intuition of what can be defined using monadic logic over linear orderings in general. We do not try to be exhaustive, and merely list some classical constructions and examples.

Relativisation. Given a monadic formula Ψ and a set X, the **relativisation** of Ψ to X is the formula Ψ^X in which all first-order quantifiers are required to range over X and all monadic quantifiers are required to range over subsets of X. This can be done by a simple syntactic transformation of Ψ. It is routine to check that the formula Ψ^X holds over a structure if and only if the formula Ψ holds over the structure restricted to the set X.
As a consequence, Theorem 3 can be used to decide monadic logic over the class of all countable linear orderings. Indeed, remark that any countable linear ordering is isomorphic to a sub-ordering of the rational line. As a consequence, Theorem 3 can be used to decide the existence of a countable linear ordering that satisfy a formula Ψ: such a linear ordering exists if and only if $(\mathbb{Q}, \leq) \models \exists X\, \Psi^X$.

Finiteness. Remark that a non-empty linear ordering that has no maximal point is infinite. Furthermore, this property is expressible in monadic logic (in fact in first-order logic). Now, it is easy to see that conversely, if a linear ordering is infinite, then either it has a non-empty sub-ordering that has no maximal point, or a non-empty sub-ordering that has no minimal point. This is expressible in monadic logic.
Using relativisation, this allows to express properties such as "X is finite" for X a monadic variable. As a consequence we can express properties such as "every finite sub-words belong to L", where L is a regular language of finite words.
Digression. Let us recall that finiteness is not first-order definable in linear orderings. This is a straightforward consequence of compactness, as follows. For the sake of contradiction, assume that the property "the linear ordering is finite" is expressible in first order, then the property P_n = "the linear ordering is finite and contains at least n distinct elements" would be expressible in first-order logic too. But any finite sets of properties P_n has a model (the finite linear ordering of length m where m is the maximal of the n's involved in the P_n's under consideration). Thus by compactness, there exists a linear ordering that satisfies all P_n's simultaneously. This is obviously impossible since this would be a linear ordering that is at the same be finite and contains more than n points for all n.

Density. A linear ordering is said **dense** if for any two points $x < y$ there exists another point $z \in (x, y)$. The rational line is dense, while ω (and more generally any ordinal) is not dense. As an extra example, the integer line $(\mathbb{Z}, \leq)$ is also not dense, and it is not isomorphic to an ordinal either.
In particular, assuming that a linear ordering is countable, then being isomorphic to $(\mathbb{Q}, \leq)$ is expressible in monadic logic (in fact first-order). Indeed, up to isomorphism, $(\mathbb{Q}, \leq)$ is the only countable linear ordering that is dense, contains at least two points, and has no minimal nor maximal point.

Scatteredness. A linear ordering is said **scattered** if none of its sub-orderings are dense. This is definable directly in monadic logic using relativisation.
Remark that it may happen that a linear ordering is neither dense nor scattered: imagine for instance a copy of the rational line in which every point is replaced by a copy of $(\mathbb{Z}, \leq)$.

In his seminal paper [9], Shelah uses another technique – the composition method – and uses it to prove the decidability of the monadic theory of the rational line. However, the most impressive result he obtains is the undecidability of the monadic theory of the real line.

Theorem 4 (Shelah [9]). *The monadic theory of the real line is undecidable*[3].

This is an very deep result which, in some sense, contradicts the intuition.

At this point, we have seen the key results concerning the decidability of the monadic theory of linear orderings. Though both the result of decidability and of undecidability can be improved, these improvements do not help understanding the picture better.

2 In which the problem is exposed and some of its intriguing characteristics appear

As we have seen, the central problems of decidability are solved for most of them. The questions we are really interested in this paper are more related to expressivity.

Question 1. Given a monadic formula $\varphi(X_1, \ldots, X_k)$, does there exists another formula $\varphi^*(X_1, \ldots, X_k)$ such that for all sets of rationals $A_1, \ldots, A_k \subseteq \mathbb{Q}$,

$$(\mathbb{R}, \leq) \models \varphi(A_1, \ldots, A_k) \quad \text{if and only if} \quad (\mathbb{Q}, \leq) \models \varphi^*(A_1, \ldots, A_k)\ .$$

In other words, the question is whether the ability to use all points of the real line does give more expressive power for stating properties of predicates over the rational line. Notice here that implicitely, we use the fact that there is a fixed embedding of $(\mathbb{Q}, \leq)$ into $(\mathbb{R}, \leq)$ (the usual one).

[3] Originally under a weak version of the continuum hypothesis, which has then be removed in collaboration with Gurevich [7].

Gurevich and Rabinovich use the nice and suggestive terminology that the formula φ has access to the reals **at the background** [6]. The open above question is thus a rephrasing of the following open question in [6]:

> "Is it true that a family of point-sets is definable in the chain $(\mathbb{Q}, \leq)$ of rationals if and only if it is definable in $(\mathbb{Q}, \leq)$ with the chain of reals at the background?"

Gurevich and Rabinovich solved this question for the integers.

Theorem 5 (Theorem 1 in [6]). *For all monadic formula $\varphi(X_1, \dots, X_k)$, there exists a formula $\varphi^*(X_1, \dots, X_k)$ such that for all $A_1, \dots, A_k \subseteq \mathbb{N}$,*

$$(\mathbb{R}, \leq) \models \varphi(A_1, \dots, A_k) \quad \textit{if and only if} \quad (\mathbb{N}, \leq) \models \varphi^*(A_1, \dots, A_k)\ .$$

We will not go into this proof which is superseded by what follows. However, already in this case a very interesting phenomenon occurs: the existence of the formula is inherently non-effective. Even if one allows the produced formula to use extra predicates of decidable monadic theory.

Theorem 6 (variant of Theorem 2 in [6]). *Let $Q_1, \dots \subseteq \mathbb{N}$ be monadic predicates such that $(\mathbb{N}, \leq, \bar{Q})$ has a decidable monadic theory. There exists no algorithm which, given a monadic formula φ, constructs φ^* such that:*

$$(\mathbb{R}, \leq) \models \varphi(X_1, \dots, X_k) \quad \textit{if and only if} \quad (\mathbb{N}, \leq, \bar{Q}) \models \varphi^*(X_1, \dots, X_k)\ .$$

Proof. Assume that such an algorithm exists, and consider some monadic sentence φ. We apply the algorithm to φ, and obtain a sentence φ^* such that $(\mathbb{R}, \leq) \models \varphi$ if and only if $(\mathbb{N}, \leq, \bar{Q}) \models \varphi^*$. Since the monadic theory of $(\mathbb{N}, \leq, \bar{Q})$ is decidable, we could decide $(\mathbb{R}, \leq) \models \varphi$. This contradicts Theorem 4. $\square$

Despite this inherent difficulty due to the non-effectiveness of the construction, we shall positively answer Question 1.

3 In which the composition theorem is introduced

In the same seminal paper as the one showing the undecidability of the monadic theory of the real line [9], Shelah develops a technique referred to as the **composition method**. This is a variant for monadic logic of techniques developed originally by Feferman and Vaught for first-order logic [5]. It becomes particularly relevant in the context of monadic logic. The versatility of this approach lies in its central theorem (known as the composition theorem, Theorem 7 below) which is not in itself a decidability result, but provides the skeleton for a decision procedure.

We need some notations first. Let α be a linear ordering and u_i for all $i \in \alpha$ be words. Then, we denote by:

$$\prod_{i \in \alpha} u_i$$

the word consisting of copies of the u_i's arranged according to the linear ordering α. Formally, assume that $\alpha = (L, \leq)$ and that the domain of u_i is $(K_i, \leq_i)$ for all $i \in \alpha$, then the domain of the word $\prod_{i\in\alpha} u_i$ is the set of pairs (i, x) where $i \in L$ and $x \in K_i$, ordered by $(i, x) \leq (j, y)$ if $i < j$ or $i = j$ and $x \leq_i y$. Furthermore, the letter at position (i, x) is the letter at position x in u_i. In the terminology of Shelah, this is denoted as a sum. The product notation reflects the fact that it extends the concatenation product used for words.

Example 2. Before stating it, let us try to give an intuitive meaning to the composition theorem. Consider that you are interested in the following property of a word u: "there is an even number of occurrences of the letter a". For simplicity, we denote from now by $|u|_a$ the number of occurrences of the letter a in the word u. Let us distinguish four formulae:

- `none` holds over u if $|u|_a = 0$,
- `even` holds over u if $|u|_a > 0$ is finite and even,
- `odd` holds over u if $|u|_a > 0$ is finite and odd,
- `infinite` holds over u if $|u|_a$ is infinite.

Over any word, exactly one of these four formulae holds, and the property we are interested in is a disjunction of such formulae, namely `none` $\vee$ `even`.

Now consider a word $u = \prod_{i\in\alpha} u_i$, where the u_i's are themselves words. It is easy to check that:

- $u \models$ `none` if and only if $u_i \models$ `none` for all $i \in \alpha$,
- $u \models$ `even` if and only if
  1. $u_i \not\models$ `infinity` for all $i \in \alpha$,
  2. either $u_i \models$ `even` or $u_i \models$ `odd` for some $i \in \alpha$, and;
  3. there is a finite number of $i \in \alpha$ such that $u_i \models$ `even`,
  4. there is a finite and even number of $i \in \alpha$ such that $u_i \models$ `odd`,
- $u \models$ `odd` if and only if [...as above, replacing even by odd in 4...].
- $u \models$ `infinite` if and only if either $u_i \models$ `infinite` for some $i \in \alpha$ or there are infinitely many $i \in \alpha$ such that $u_i \models$ `even` or there are infinitely many $I \in \alpha$ such that $u_i \models$ `odd`.

For all words u there is one and only one formula φ among `none`, `even`, `odd` and `infinity` such that $u \models \varphi$: let us call this formula the type of u, and denote it `type`(u). What we have seen in this example is that in order to know the type of $\prod_{i\in\alpha} u_i$, it is sufficient to know the type of each of the u_i's. What is crucial is that there are only finitely many types that are sufficient. The composition theorem (Theorem 7) generalizes this example. It states that in order to know the truth value of any monadic formula it is sufficient to consider only finitely many types that have properties similar to this example.

Let us fix now a constant $k \in \mathbb{N}$. A monadic formula has **quantifier rank** k if there are at most k nested quantifiers.

Proposition 1. *Over a fixed finite signature with only relational symbols (the case of words over a fixed finite alphabet), there are only finitely many formulas up to syntactic equivalence. Here, the syntactic equivalence involves the usual associativity, commutativity, idempotency, and distributivity of conjunctions and disjunctions, the renaming of bound variables, and*

$$\exists s(\varphi \vee \psi) \equiv (\exists s\, \varphi) \vee (\exists s\, \psi) \qquad \textit{and} \qquad \forall s(\varphi \wedge \psi) \equiv (\forall s\, \varphi) \wedge (\forall s\, \psi)\ ,$$

for t a first-order variable or a monadic variable.

From now, all formulae will be considered modulo this syntactic equivalence, and since we fix k, by the above fact, we only have to consider finitely many formulae.

Now, given a word u, its **(k-)type** $\texttt{type}_k(u)$ is the set of sentences φ of quantifier rank at most k such that $u \models \varphi$.

Theorem 7 (composition [9]). *If* $\texttt{type}_k(u_i) = \texttt{type}_k(v_i)$ *for all* $i \in \alpha$*, then*

$$\texttt{type}_k\left(\prod_{i\in\alpha} u_i\right) = \texttt{type}_k\left(\prod_{i\in\alpha} v_i\right)\ .$$

This completely reflects the intuition of Example 2, in which only four possible types were distinguished for simplicity.

Remark 2. In fact the real composition theorem is more precise in that it expresses how $\texttt{type}_k(\prod_{i\in\alpha} u_i)$ can be defined from $\prod_{i\in\alpha}\texttt{type}_k(u_i)$. Formally, it states that for all type t of quantifier rank k, there exists a monadic formula t^* such that:

$$\texttt{type}_k\left(\prod_{i\in\alpha} u_i\right) = t \qquad \text{iff} \qquad \prod_{i\in\alpha}\texttt{type}_k(u_i) \models t^*\ .$$

This implies Theorem 7 as if $\texttt{type}_k(u_i) = \texttt{type}_k(v_i)$ for all $i \in \alpha$, this means that $\prod_{i\in\alpha}\texttt{type}_k(u_i) = \prod_{i\in\alpha}\texttt{type}_k(v_i)$. However, this more complete presentation is quite misleading since the quantifier rank of t^* may be (much) higher than k. In practice, the decision procedures using the composition method do not make use of this formula t^*.

4 In which algebraic recognizability for countable words is defined

In this section, we introduce a notion of recognizability that is suitable for capturing monadic logic over countable words [3]. It is highly related to the composition method as shall be shown below.

Let us denote by M° the set of words of countable length over the alphabet M. We call them **M°-words** from now. Consider an application $\otimes : M^\circ \to M$.

We will often use, for α a countable linear ordering, and $a_i \in M$ for all $i \in \alpha$, the notation

$$\bigotimes_{i\in\alpha} a_i$$

to denote $\otimes u$ where u is the word of domain α that carries at position i the letter a_i for all $i \in \alpha$. This operation $\otimes$ is **associative**, or **a product**, if

- $\otimes(a) = a$ for all $a \in M$, and;
- for all countable linear orderings α and all families of M°-words $(u_i)_{i\in\alpha}$,

$$\otimes\left(\prod_{i\in\alpha} u_i\right) = \bigotimes_{i\in\alpha} \otimes(u_i)\ .$$

A **$\circ$-monoid** $\mathbf{M} = (M, \otimes)$ is a set M equipped with a product $\otimes$.

Remark 3. A consequence of the above definition is that for a, b, c in M:

$$\otimes(a \otimes (bc)) = \otimes(\otimes(a) \otimes (bc)) = \otimes(abc) = \otimes(\otimes(ab) \otimes (c)) = \otimes(\otimes(ab)c)\ .$$

Thus, if you denote $\otimes(ab)$ as $a \cdot b$, this means $a \cdot (b \cdot c) = (a \cdot b) \cdot c$. Hence this generalizes the usual notion of associativity. Also, notice that the empty-word ε belonging to M°, it has a value $\otimes(\varepsilon)$ under $\otimes$. Let us denote by 1 this value. Still using associativity, we get for all $a \in M$,

$$a \cdot 1 = \otimes(a1) = \otimes(\otimes(a) \otimes (\varepsilon)) = \otimes(a\varepsilon) = \otimes(a) = a\ ,$$

and similarly $1 \cdot a = a$. Thus, every $\circ$-monoid is in particular a monoid.

Of course, the **free $\circ$-monoid** generated by a set M is simply $(M^\circ, \prod)$. The notion of a morphism of $\circ$-monoids is also natural. Given two $\circ$-monoids $\mathbf{M} = (M, \otimes)$ and $\mathbf{M}' = (M', \otimes')$, a **morphism** from $\mathbf{M}$ to $\mathbf{M}'$ is a function f from M to M' such that for all countable linear orderings α and $(a_i)_{i\in\alpha}$ elements of M,

$$f\left(\bigotimes_{i\in\alpha} a_i\right) = \bigotimes_{i\in\alpha}{}' f(a_i)\ .$$

We are now ready to define the notion of a recognizable set of countable words. A set of words $L \subseteq \mathbb{A}^\circ$ is **recognizable** if there exist a finite $\circ$-monoid $\mathbf{M} = (M, \otimes)$, a morphism f from $\mathbb{A}^\circ$ to $\mathbf{M}$ and a set $F \subseteq M$ such that for all words $u \in \mathbb{A}^\circ$,

$$u \in L \qquad \text{if and only if} \qquad f(u) \in F\ .$$

Remark here that the finiteness refers to the carrier M of $\mathbf{M}$ (even if M is finite, $\otimes$ is a mapping of infinite domain). The fact that the product $\otimes$ is “infinite” means that, *as such*, a finite $\circ$-monoid cannot be represented in a computer.

In fact, the very strong properties of associativity of the product have as effect that only a finite quantity of information is sufficient for representing a finite $\circ$-monoid. Concretely, it is sufficient to know the value of $\otimes$ over a finite number of words for being able to reconstruct $\otimes$ uniquely over all countable words [9, 3]. This is very similar to the fact that once the product of two elements is known in a monoid, the product can be extended uniquely to arbitrarily long finite sequences of elements. Decidability results and effective constructions are done manipulating this finite presentation of the product.

Example 3. Let us consider the set of $\{a,b\}^\circ$-words that have a finite and even number of occurrences of the letter a (this corresponds to Example 2), and show that it is recognizable.

We consider as carrier of the monoid the set $M = \{1, \mathtt{e}, \mathtt{o}, 0\}$. We start by giving the morphism f that sends $\{a,b\}^\circ$ to M:

$$f(u) = \begin{cases} 1 & \text{if } |u|_a = 0, \\ \mathtt{e} & \text{if } |u|_a > 0 \text{ is finite and even,} \\ \mathtt{o} & \text{if } |u|_a > 0 \text{ is finite and odd,} \\ 0 & \text{otherwise.} \end{cases}$$

With this morphism in mind, it is simple to uniquely define the product over M. Consider for instance the M°-word $\mathtt{oo}$, since $\mathtt{o} = f(a)$, this means $\otimes(\mathtt{oo}) = \otimes(f(a)f(a)) = \otimes(f(aa)) = \mathtt{e}$. Using similar arguments, we can complete the product as:

- $\otimes(u) = 1$ if u contains only 1-letters,
- $\otimes(u) = \mathtt{e}$ if u does not contain the letter 0, contains finitely many occurrences of $\mathtt{e}$- and $\mathtt{o}$-letters, and at least one $\mathtt{e}$- or one $\mathtt{o}$-letter, and an even number of $\mathtt{o}$-letters,
- $\otimes(u) = \mathtt{e}$ if u does not contain the letter 0, contains finitely many occurrences of $\mathtt{e}$- and $\mathtt{o}$-letters, and at least one $\mathtt{e}$- or one $\mathtt{o}$-letter, and an odd number of $\mathtt{o}$-letters,
- $\otimes(u) = 0$ otherwise.

The intuition behind the previous can be generalized. A direct consequence of the composition theorem is the recognizability of all monadic definable languages.

Theorem 8. *All monadic definable languages of $\circ$-words are recognizable.*

Proof. Let φ be a monadic formula defining $L \subseteq \mathbb{A}^\circ$. Let k be the quantifier rank of φ.

We shall construct a $\circ$-monoid $\mathbf{M}$ and a morphism f from $\mathbb{A}^\circ$ to $\mathbf{M}$. Define

$$M = \{\mathtt{type}_k(v) \ : \ v \in \mathbb{A}^\circ\} \ .$$

Since $\mathtt{type}_k$ is surjective from $\mathbb{A}^\circ$ onto M, there exists a mapping $g : M \to \mathbb{A}^\circ$ such that $\mathtt{type}_k \circ g$ is the identity over M. We extend it to M°-words letter by letter, yielding a mapping $\tilde{g}$ from M° to $\mathbb{A}^\circ$ (formally $\tilde{g}(\prod_{i\in\alpha} a_i) = \prod_{i\in\alpha} g(a_i)$). We now equip M with an operation $\otimes$ as follows. Let u be a word in M°, define

$$\otimes(u) = \mathtt{type}_k(\tilde{g}(u)) \ .$$

We have for all $\mathbb{A}^\circ$-words $(v_i)_{i\in\alpha}$ indexed by a countable linear ordering α,

$$\begin{aligned}
\mathtt{type}_k\left(\prod_{i\in\alpha} v_i\right) &= \mathtt{type}_k\left(\prod_{i\in\alpha} g(\mathtt{type}_k(v_i))\right) && \text{(Theorem 7)}\\
&= \mathtt{type}_k\left(\tilde{g}\left(\prod_{i\in\alpha} \mathtt{type}_k(v_i)\right)\right) && \text{(definition of } \tilde{g})\\
&= \bigotimes_{i\in\alpha} \mathtt{type}_k(v_i) \ , && \text{(definition of } \otimes)
\end{aligned}$$

in which the first equality is by the composition theorem (Theorem 7) since by construction of g, $\mathtt{type}_k(v_i) = \mathtt{type}_k(g(\mathtt{type}_k(v_i)))$.

We do not know yet that $\otimes$ is a product. However, since $\mathtt{type}_k$ is surjective from $\mathbb{A}^\circ$ onto M and satisfies the properties of a morphism, it follows that the fact that $(\mathbb{A}^\circ, \prod)$ is a $\circ$-monoid is automatically transferred to $(M, \otimes)$. Thus $(M, \otimes)$ is also a $\circ$-monoid, and $\mathtt{type}_k$ is a morphism from $(\mathbb{A}^\circ, \prod)$ to $(M, \otimes)$.

Let now $F = \{\mathtt{type}_k(u) \ : \ u \in L\}$. Let us show that $\mathbf{M}, \mathtt{type}_k, F$ recognize L. Let $u \in \mathbb{A}^\circ$ be a word. We have that $u \in L$ if and only if $u \models \varphi$ if and only if $\varphi \in \mathtt{type}_k(u)$ if and only if $\mathtt{type}_k(u) \in F$. Hence $\mathbf{M}, \mathtt{type}_k, F$ recognize L. □

One of the main contributions in [3] is to provide a form of converse to Theorem 8.

Theorem 9. *All recognizable languages of $\circ$-words are monadic definable.*

This direction relies on completely different techniques. It involves in particular the theory of ideals of the monoid underlying the $\circ$-monoid and special forms of factorizations of the words. In particular, it heavily relies on the fact that the $\circ$-monoids used to recognize languages are finite, an assumption that was not used so far (in fact, it is already crucial in Shelah's work, but for different reasons, for decidability). We will use this result as a black-box.

5 In which the question is answered

Let us consider now Question 1 again. We will see that the situation is not much different from the previous section.

In order to deal with the real line, we need to describe a bit more precisely the relationship between the rational line and the real line. For technical reasons it is not very convenient to work directly with the real line, but rather on the expansion of the rational line with all "Dedekind cuts". The real line itself is obtained from the rational line using a similar expansion, but keeping only the so-called "natural cuts". Nevertheless, as far as logic is concerned, this difference is very minor.

Given a linear ordering $(E, \leq)$, a (Dedekind) **cut** is an ordered pair (A, B) of sets $A, B \subseteq E$ such that $A \cup B = E$ and $x < y$ for all $x \in A$ and $y \in B$. Cuts are ordered by $(A, B) \leq (A', B')$ if $A \subseteq A'$. Cuts can also be compared with the elements of E by $x < (A, B)$ if $x \in A$ and $(A, B) < x$ if $x \in B$. Equipped with this relation, the (disjoint) union of the elements of E with the cuts form a linear ordering. A cut (A, B) is **extremal** if either $A = \emptyset$ or $B = \emptyset$. Given a linear ordering $\alpha = (L, \leq)$, denote by $\hat{\alpha}$ the **completion** of α, which is obtained from α by adding to it all non-extremal cuts, and consider the ordering extended as above. Remark that to every element in $x \in E$ corresponds three copies $x^- < x < x^+$ in the completed linear ordering $\hat{\alpha}$, where $x^- = ((-\infty, x), [x, \infty))$ and $x^+ = ((-\infty, x], (x, \infty))$. Cuts that are not of the form x^- or x^+ are called **natural**. As mentioned above, the real line is obtained from the rational line by adding to it the non-extremal natural cuts only.

The completion of a word is done as follows. We fix ourselves a dummy letter ι that is intended to label cuts. The **(cut) completion** $\mathtt{comp}(u)$ of a word u over the alphabet $\mathbb{A}$ is a word over the alphabet $\mathbb{A} \cup \{\iota\}$ defined as $\prod_{i \in \hat{\alpha}} b_i$ where $b_i = a_i$ for all $i \in \alpha$ and $b_i = \iota$ otherwise (*i.e.*, if i is a cut).

A simple, yet important, point is the relationship between the completion and the product $\prod$. The following lemma discloses this point. It essentially states that the completion of the product is equivalent to a variant product of the completion, where the variant product "fills the missing cuts".

Lemma 1. *For all linear orderings α and words $(v_i)_{i \in \alpha}$,*

$$\mathtt{comp}\left(\prod_{i \in \alpha} v_i\right) \equiv \hat{\prod_{i \in \alpha}} \mathtt{comp}(v_i) \ ,$$

where for all words (w_i) indexed by a linear ordering α we set

$$\hat{\prod_{i \in \alpha}} w_i = \prod_{i \in \hat{\alpha}} w'_i$$

with for all $i \in \hat{\alpha}$, $w'_i = \begin{cases} w_i & \text{if } i \in \alpha, \\ \iota & \text{otherwise, i.e., if } i \text{ is a cut.} \end{cases}$

A language L of countable words is called **monadic definable with cuts at the background** if there exists a monadic formula φ such that $u \in L$ if and only if $\mathtt{comp}(u) \models \varphi$.

The following key proposition follows exactly the same proof scheme as the one of Theorem 8.

Proposition 2. *Languages of countable words that are monadic definable with cuts at the background are $\circ$-recognizable.*

Proof. Let φ be a monadic formula defining with cuts at the background a language of countable words $L \subseteq \mathbb{A}^\circ$. Let k be the quantifier rank of φ.

We shall construct a $\circ$-monoid $\mathbf{M}$ and a morphism f from $\mathbb{A}^\circ$ to $\mathbf{M}$. Let first set $\mathtt{typec}_k(v)$ to be $\mathtt{type}_k \circ \mathtt{comp}(v)$ for all $v \in \mathbb{A}^\circ$. Define

$$M = \{\mathtt{typec}_k(v) \ : \ v \in \mathbb{A}^\circ\} \ .$$

Since $\mathtt{typec}_k$ is surjective from $\mathbb{A}^\circ$ onto M, there exists a mapping $g : M \to \mathbb{A}^\circ$ such that $\mathtt{typec}_k \circ g$ is the identity over M. We extend it to M°-words letter by letter, yielding a mapping $\tilde{g}$ from M° to $\mathbb{A}^\circ$ (formally $\tilde{g}(\prod_{i\in\alpha} a_i) = \prod_{i\in\alpha} g(a_i)$). We now equip M with an operation $\hat{\otimes}$ as follows. Let u be a word in M°, define

$$\hat{\otimes}(u) = \mathtt{typec}_k(\tilde{g}(u)) \ .$$

We show now that $\mathtt{typec}_k$ has the properties of a morphism from $\mathbb{A}^\circ$ to M (though we do not know yet that $\hat{\otimes}$ is a product). Consider a family of $\mathbb{A}^\circ$-words $(v_i)_{i\in\alpha}$ indexed by a countable linear ordering α, we have:

$$\begin{aligned}
\mathtt{typec}_k\left(\prod_{i\in\alpha} v_i\right) &= \mathtt{type}_k\left(\hat{\prod_{i\in\alpha}}\mathtt{comp}(v_i)\right) && \text{(Lemma 1)}\\
&= \mathtt{type}_k\left(\hat{\prod_{i\in\alpha}}\mathtt{comp}(g(\mathtt{typec}_k(v_i)))\right) && \text{(Theorem 7)}\\
&= \mathtt{typec}_k\left(\prod_{i\in\alpha} g(\mathtt{typec}_k(v_i))\right) && \text{(Lemma 1)}\\
&= \mathtt{typec}_k\left(\tilde{g}\left(\prod_{i\in\alpha}\mathtt{typec}_k(v_i)\right)\right) && \text{(definition of } \tilde{g})\\
&= \hat{\bigotimes_{i\in\alpha}}\mathtt{typec}_k(v_i) \ , && \text{(definition of } \hat{\otimes})
\end{aligned}$$

in which the equality between the first and second line is by the composition theorem using the fact that $\mathtt{typec}_k(v_i) = \mathtt{typec}_k(g(\mathtt{typec}_k(v_i))$.

We do not know yet that $\hat{\otimes}$ is a product. However, since $\mathtt{typec}_k$ is surjective from $\mathbb{A}^\circ$ onto M and satisfies the properties of a morphism, it follows that the fact that $(\mathbb{A}^\circ, \prod)$ is a $\circ$-monoid is automatically transferred to $(M, \hat{\otimes})$. Thus $(M, \hat{\otimes})$ is also a $\circ$-monoid, and $\mathtt{typec}_k$ is a morphism from $(\mathbb{A}^\circ, \prod)$ to $(M, \hat{\otimes})$.

Let now $F = \{\mathtt{typec}_k(u) \ : \ u \in L\}$. Let us show that $\mathbf{M}, \mathtt{typec}_k, F$ recognize L. Let $u \in \mathbb{A}^\circ$ be a word. We have that $u \in L$ if and only if $\mathtt{comp}(u) \models \varphi$ if and only if $\varphi \in \mathtt{typec}_k(u)$ if and only if $\mathtt{typec}_k(u) \in F$. Hence $\mathbf{M}, \mathtt{type}_k, F$ recognize L. □

Thus, in combination with Theorem 9, we get the following corollary.

Corollary 1. *All languages of countable words that are monadic definable with cuts at the background are monadic definable.*

If we restate this corollary in terms of relational structures, we get:

Corollary 2. *Given a monadic formula* $\varphi(X_1, \ldots, X_k)$*, there exists a formula* $\varphi^*(X_1, \ldots, X_k)$ *such that for all countable linear orderings* α *and all sets of rationals* $A_1, \ldots, A_k \subseteq \alpha$*,*

$$\hat{\alpha} \models \varphi(A_1, \ldots, A_k) \quad \textit{if and only if} \quad \alpha \models \varphi^*(A_1, \ldots, A_k)\ .$$

Indeed, a countable linear ordering α labeled with $A_1, \ldots, A_k$ can be seen as a countable word over the alphabet 2^k.

Now, recall that the reals are the completion of the rationals with natural cuts. The only reason that the above theorem does not exactly solve Question 1 as it stands is that $\hat{\alpha}$ also contains cuts that are not natural. Thus, it is sufficient to remark that given a linear ordering α of domain A, there is a first-order formula $\varphi(X, x)$ such that $\hat{\alpha} \models \varphi(A, a)$ if and only if a is a non-natural cut (recall that the non-natural cuts are the ones that are predecessors or successors of elements of A; a property that makes them easily definable). Thus, using Corollary 2 together with a relativisation to the natural cuts and the original ordering, we finally answer positively Question 1.

Theorem 10. *Given a monadic formula* $\varphi(X_1, \ldots, X_k)$*, there exists a formula* $\varphi^*(X_1, \ldots, X_k)$ *such that for all* $A_1, \ldots, A_k \subseteq \mathbb{Q}$*,*

$$(\mathbb{R}, \leq) \models \varphi(X_1, \ldots, X_k) \quad \textit{if and only if} \quad (\mathbb{Q}, \leq) \models \varphi^*(X_1, \ldots, X_k)\ .$$

Acknowledgments

I am grateful to Alexander Rabinovich who introduced me to the question of reals at the background, and for all the discussions we had on the subject. I also thank Olivier Carton and Gabriele Puppis for many discussions we had together.

References

1. J. Richard Büchi. Weak second-order arithmetic and finite automata. *Z. Math. Logik Grundlagen Math.*, 6:66–92, 1960.
2. J. Richard Büchi. On a decision method in restricted second order arithmetic. In *Logic, Methodology and Philosophy of Science (Proc. 1960 Internat. Congr .)*, pages 1–11. Stanford Univ. Press, Stanford, Calif., 1962.
3. Olivier Carton, Thomas Colcombet, and Gabriele Puppis. Regular languages of words over countable linear orderings. In Luca Aceto, Monika Henzinger, and Jiri Sgall, editors, *ICALP 2011 (2): Automata, Languages and Programming - 38th International Colloquium*, volume 6756 of *Lecture Notes in Comput. Sci.*, pages 125–136. Springer, 2011.
4. Calvin C. Elgot. Decision problems of finite automata design and related arithmetics. *Trans. Amer. Math. Soc.*, 98:21–51, 1961.
5. S. Feferman and R. L. Vaught. The first order properties of products of algebraic systems. *Fund. Math.*, 47:57–103, 1959.
6. Yuri Gurevich and Alexander Moshe Rabinovich. Definability and undefinability with real order at the background. *J. Symb. Log.*, 65(2):946–958, 2000.

7. Yuri Gurevich and Saharon Shelah. Monadic theory of order and topology in zfc. In *Ann. of Math. Logic*, volume 23, pages 179–198, 1982.
8. Michael O. Rabin. Decidability of second-order theories and automata on infinite trees. *Trans. Amer. Math. Soc.*, 141:1–35, 1969.
9. Saharon Shelah. The monadic theory of order. *Ann. of Math. (2)*, 102(3):379–419, 1975.
10. Boris A. Trakhtenbrot. Finite automata and monadic second order logic (russian). *Siberian Math. J*, 3:103–131, 1962.
11. B. A. Trakthenbrot. The impossibility of an algorithm for the decision problem for finite domains (russian). *Doklady Academii Nauk SSSR*, 70:569–572, 1950.