How Many First-order Variables are Needed on Finite Ordered Structures?

ANUJ DAWAR

1 Introduction

Let L^k denote the fragment of first-order logic in which each formula contains at most k variables, free or bound. We are interested in interpreting this language on finite ordered structures. That is, the interpretation is in finite structures $\mathbb{A}$ with a distinguished binary relation symbol $<$ which is to be interpreted as a linear order of the universe. The question that concerns us is whether, for every k, L^{k+1} is strictly more expressive than L^k. That is, is the hierarchy of expressive power formed by the number of first-order variables strict? The question is easy enough to state but turns out to be quite difficult to answer. Indeed, it is not even known if the hierarchy formed is *infinite*. Unlikely as it may seem, to the best of our current knowledge, it is conceivable that there is a k (depending on the signature) such that L^k is as expressive as all of first-order logic in this case.

This paper offers an exposition of the problem and its background illustrating many of the key issues in the subject of finite model theory. Finite variable logics have played an important role in the subject, and the question we consider is central to understanding the power of limited numbers of variables. Ordered structures have also been of crucial importance in finite model theory and it has been particularly difficult to obtain any inexpressiveness results on ordered structures as is again highlighted by this simple example. Finally, the problem also illustrates some of the important connections of finite model theory with other fields of investigation: temporal logic on the one hand and computational complexity on the other.

In addition to establishing the background, the paper also presents a few previously unpublished observations and constructions which explore the precise boundary between cases where the hierarchy of expressive power formed by the number of variables (which we call the L^k hierarchy for short) collapses and those where it is not known to. It has long been known that three first-order variables suffice to express all properties on *temporal* structures, i.e. ordered structures with unary predicates. As we shall see, the question of whether more than three variables are needed when the signature contains one binary relation in addition to the order is equivalent to the general question. In the case when infinite structures are permitted, it is known that the L^k hierarchy is strict and this can be shown on structures where the only relation in addition to the order is an equivalence relation. This makes equivalence relations of particular interest. We show that on finite, ordered equivalence relations, three variables suffice to express all *order-invariant* first-order properties. However, dropping the order-invariance requirement and allowing unary relations in addition to the one equivalence relation leads us to a version that is equivalent to the general question. We also look at a version of the hierarchy question obtained by restricting quantifier alternation. We show that, for each k, there is an *existential* formula in L^{k+1} that is not equivalent to any *existential* formula with only k variables (even an infinitary one) on finite ordered graphs. Indeed, it is not equivalent to any existential formula with k variables, even if infinitary connectives are allowed.

1.1 The Main Question

We only consider first-order relational languages. That is, our vocabularies do not, in general, contain function or constant symbols. Let σ be a relational vocabulary that does not contain the symbol $<$. So, σ is a finite collection of relation symbols R each with an associated arity r. By a (finite) structure over σ, or simply a (finite) σ-structure $\mathbb{A}$, we mean a (finite) set A (the universe of the structure) together with an interpretation $R^{\mathbb{A}} \subseteq A^r$ of each R in σ by a relation over A of the appropriate arity. We write $\mathcal{S}_\sigma$ for the collection of finite σ-structures. We also write $\mathcal{O}_\sigma$ for the collection of finite *ordered* σ-structures. These are structures in the vocabulary $\sigma \cup \{<\}$ which interpret $<$, a binary relation symbol, as a linear order of the universe.

We write FO_σ for the collection of first-order formulas (formed in the usual way) over the vocabulary σ. Where it is clear from context, we will drop the subscript σ. We write L^k for the fragment of FO consisting of formulas with no more than k variables (free or bound). Without loss of generality, we may assume that these formulas use only the variables $x_1, \ldots, x_k$ though we will still use x, y, z, etc. as variables to range over this set. We are mainly concerned with issues of the expressive power of the language rather than its syntactic manipulation so the precise choice of the set of variables is not of concern. In order to have the logic be closed under the renaming of bound variables one may prefer to think of L^k as those formulas φ of FO such that no subformula of φ contains more than k free variables.

For any class $\mathcal{C}$ of σ-structures, we say that two formulas φ and ψ are equivalent *over* $\mathcal{C}$ if, for any structure $\mathbb{A} \in \mathcal{C}$, $\mathbb{A} \models \varphi$ if, and only if, $\mathbb{A} \models \psi$. We can now state the main conjecture we are concerned with as follows:

CONJECTURE 1. For any vocabulary σ which contains at least one relation of arity two or more and for each k, there is a sentence φ of $\mathrm{FO}_{\sigma\cup\{<\}}$ for which there is no sentence ψ of L^k such that φ is equivalent to ψ over $\mathcal{O}_\sigma$.

Clearly, $\mathcal{O}_\sigma$ is a subclass of $\mathcal{S}_{\sigma\cup\{<\}}$ and there are pairs of formulas that may be equivalent over $\mathcal{O}_\sigma$ without being equivalent over $\mathcal{S}_{\sigma\cup\{<\}}$. This may be seen as the crux of the conjecture: to what extent can order be used to reduce the number of variables required to express properties of finite structures?

Conjecture 1 asserts that the hierarchy formed by the number of variables is infinite. There is also a stronger version of this conjecture which would require φ to be in L^{k+1} rather than just in FO. We sometimes use the terms 'the L^k hierarchy conjecture' to refer to Conjecture 1 and 'the strict L^k hierarchy conjecture' to refer to the stronger form. And, more generally, we say that the L^k hierarchy is strict (or infinite) over a class of structures $\mathcal{C}$, if for each k there is a formula of L^{k+1} (resp. FO) that is not equivalent *over* $\mathcal{C}$, to any formula of L^k. We also say the L^k hierarchy *collapses* on $\mathcal{C}$ if it is not infinite over $\mathcal{C}$.

The above question is formulated as a conjecture because intuition strongly suggests that the answer is positive. As we shall see, the failure of the conjecture would have many unexpected consequences. However, there is little in the way of concrete results supporting it. Indeed, it is not known if there is any vocabulary σ for which the conjecture is true. Thus, a weaker form of the conjecture would be the following.

CONJECTURE 2. For some vocabulary σ containing at least one relation of arity two or more and for each k, there is a sentence φ of $\mathrm{FO}_{\sigma\cup\{<\}}$ for which there is no sentence ψ of L^k such that φ is equivalent to ψ over $\mathcal{O}_\sigma$.

This is the version of the conjecture we will primarily work with. The simplest case of the conjecture is when σ consists of a single binary relation E. In this case, it is an open question whether there is any formula of FO_σ that is not equivalent to one of L^3.

1.2 Background

A natural question that may arise in the reader's mind is why one would care about the number of variables a formula uses. Why, further, should one restrict consideration to ordered structures, and finite ones? We look at some of the reasons why these questions are of interest.

Temporal Logic Propositional temporal logics are usually interpreted over flows of time. These are (partially) ordered structures $(T, <)$ which interpret an infinite set of propositional symbols. That is, we associate with each element of T the set of propositional symbols that it interprets as true. Using our notation, flows of time are structures over the vocabulary $\sigma \cup \{<\}$ where σ is a collection of unary predicate symbols corresponding to propositional symbols. A temporal logic is formed from the set of propositional symbols and a set of *temporal connectives*. In general, for most temporal logics considered, there is a translation into $\mathrm{FO}_{\sigma\cup\{<\}}$. That is, for each temporal formula φ, there is a formula φ^* of $\mathrm{FO}_{\sigma\cup\{<\}}$ which is satisfied in a flow of time if, and only if, φ is. Moreover, as Gabbay [1981] observed, if $\mathcal{T}$ is a temporal logic formed from a finite set of temporal connectives, each of which is itself first-order definable, then there is a k such that every formula of $\mathcal{T}$ is equivalent to one of L^k. In other words, standard temporal logics can be translated into first-order logic using only a bounded number of variables. Gabbay introduces the notion of *functional completeness* of temporal logics. A logic $\mathcal{T}$ is functionally complete over a given class of flows of time if every first-order formula (with one free variable) is equivalent over the class to a formula of $\mathcal{T}$. One consequence is that if a flow of time T admits a functionally-complete finite set of temporal connectives then it has what Gabbay termed finite H-dimension. In particular, this implies that there is a k such that every first-order sentence is equivalent to one of L^k. Conversely, Gabbay also shows that any flow of time with finite H-dimension admits a finite set of temporal connectives that is functionally complete. The notion of H-dimension is intimately tied to the bounded variable fragments L^k of first-order logic (see [Hodkinson and Simon, 1997] for a detailed discussion of the relationship).

Kamp [1968] showed that the temporal connectives *until* U and *since* S are functionally complete for Dedekind-complete linearly ordered flows of time but not for a rationally ordered flow of time. Stavi showed that U and S along with two new connectives U' and S' are complete over all linearly ordered flows of time (see [Gabbay *et al.*, 1994] for a proof). As it can be easily checked that the temporal logic with U, S, U' and S' translates into L^3, it follows that any first-order sentence is equivalent over linear flows of time to one of L^3. In other words, Conjecture 1 is false if we drop the requirement that σ contain at least one relation of arity two or more. If σ has only unary symbols, three variables suffice to express all first-order properties of ordered σ-structures. In Section 2 we see an alternative proof of this collapse, due to Poizat [1982].

Descriptive Complexity In his work on descriptive complexity Immerman (see [1999]) has advocated the use of first-order logic for the study of computational complexity. In this view, the logic functions essentially as a model of computation, in which measures such as the number of variables and quantifier alternation become the resources through which we measure complexity. We consider decision problems on σ-structures for some signature σ. In order to simulate computations on structures it is also useful to assume that the structures are additionally equipped with an order $<$ on the universe. It is also common to assume that the first-order language has access not only to the relation $<$ but to arithmetic relations $+$ and $\times$ that it induces. Thus, for instance, the ternary relation $+(x, y, z)$ holds of the ith, jth and kth elements of the linear order $<$ if, and only if, $i + j = k$.

Now, given a decision problem D (i.e. an isomorphism-closed class) of finite σ-structures and a positive integer n, it is possible to write down a first-order formula φ_n which is true of a σ-structure $\mathbb{A}$ with n or fewer elements if, and only if, $\mathbb{A}$ is in D. Indeed, if we allow ourselves the use of the order $<$ in φ_n, we can write the formula so that the number of variables used is bounded, independently of n, and this will be significant in our later discussions. For now, we note that this implies that any decision problem D (whether decidable or not) can be characterised by a family $\mathcal{F} = (\varphi_n \mid n \in \omega)$ of first-order formulas. Clearly, the decidability and complexity of D are related to the regularity of $\mathcal{F}$. By results of Barrington, Immerman and Straubing [1990], it is known that the decision problems obtained when $\mathcal{F}$ consists of a single formula (i.e. all φ_n are the same) are exactly those in uniform AC_0. That is, they are the properties decidable by a uniform family of circuits of polynomial size and constant depth. Immerman considers families $\mathcal{F}$ which are obtained by iterating a single formula using a bounded number of variables and thus obtains a tight connection between the number of variables used and the space complexity of D [Immerman, 1991]. Families with a bounded number of variables and a polynomial bound on their size correspond to polynomial-time computation [Immerman, 1982].

Thus, we obtain a tight connection between the number of variables and more conventional measures of computational complexity. Moreover, the presence of an ordering on the structures is crucial to the connection. As mentioned above, the family $\mathcal{F}$ characterising a problem D can always be chosen so that the number of variables is bounded (by some constant that depends only on σ). However, this may be at the cost of the regularity of $\mathcal{F}$ otherwise. For instance, it is a consequence of Immerman's results that there are families $\mathcal{F}$ in which the φ_n are obtained by iterating a single formula with $k+1$ variables that cannot be obtained by iterating a single formula with k variables. What if $\mathcal{F}$ is defined by a single formula, i.e. D is in uniform AC_0? Is there a bound, depending only on σ on the number of variables needed? If Conjecture 1 fails, then the answer is yes. We suspect that the answer is no, that is, there is a real trade-off between number of variables and quantifier depth beyond the constant, but showing this requires us to prove Conjecture 1.

Fixed-Point Logics Research in descriptive complexity has also led to an interest in the study of fixed-point logics. These are logics that extend first-order logic by means of operators for forming fixed points of relational operators. The most extensively studied such operator is the least-fixed-point operator. We write LFP for the extension of first-order logic by means of such an operator (a more formal definition is given in Section 4 below). It was proved by Immerman [1986] and Vardi [1982] that a property of finite, ordered, relational structures is definable in LFP if, and only if, it is decidable in polynomial time. Similarly, PFP, an extension of first-order logic by means of a so-called partial fixed-point operator, characterises exactly the properties of finite, ordered, structures that are decidable in polynomial space [Vardi, 1982; Abiteboul and Vianu, 1991]. The presence of a linear order in each structure is crucial to these results. In the absence of such an order, the logics are strictly weaker than the respective complexity classes. Nevertheless, as Abiteboul and Vianu [1995] showed, the relationship between LFP and PFP mirrors exactly the relationship between the complexity classes PTime and PSpace even without assuming that there is an order present. That is to say, the logics are equivalent on finite structures in general if, and only if, the complexity classes coincide. A proof of this result (see [Dawar *et al.*, 1995]) relies on an analysis of these logics in terms of bounded-variable types.

To make things more precise, let $L_{\infty\omega}$ denote the closure of first-order logic with infinitary conjunctions and disjunctions, which take arbitrary (rather than just finite) sets of formulas. $L^k_{\infty\omega}$ is the fragment of $L_{\infty\omega}$ which uses only k distinct variables and $L^\omega_{\infty\omega} = \bigcup_{k\in\omega} L^k_{\infty\omega}$, i.e. $L^\omega_{\infty\omega}$ is the collection of formulas of $L_{\infty\omega}$ in which each formula has only finitely many variables. Kolaitis and Vardi [1992b] showed that, when restricted to finite structures, any formula φ of LFP is equivalent to a formula of $L^\omega_{\infty\omega}$. Indeed, there is a k (which may depend on φ) such that φ is equivalent to a single infinite disjunction of formulas of L^k. This led to an investigation of McColm's conjectures (see [Kolaitis and Vardi, 1992a]) which essentially stated the equivalence of three conditions for any class $\mathcal{C}$ of finite structures: (1) that LFP and FO have equivalent expressive power on $\mathcal{C}$; (2) that $L^\omega_{\infty\omega}$ and FO have equivalent expressive power on $\mathcal{C}$; and (3) that every first-order definable induction closes in a bounded number of stages on $\mathcal{C}$. While it was shown [Kolaitis and Vardi, 1992a] that (2) and (3) are equivalent, Gurevich et al. [1994] constructed a class of structures on which (1) holds but (3) fails. Nevertheless, Kolaitis and Vardi [1992a] had put a forth a weaker version of the conjecture, namely that (1) and (3) are equivalent on any *ordered* class $\mathcal{C}$. That is, any class $\mathcal{C}$ that is contained in $\mathcal{O}_\sigma$ for some σ. This is known as the *ordered conjecture* and it was shown in [Dawar *et al.*, 1996] that either a positive or negative resolution to this conjecture would have significant consequences in complexity theory. It also follows from results in [Dawar *et al.*, 1996] that the failure of Conjecture 2 would imply a positive resolution to the ordered conjecture. We return to this topic in Section 4.

2 Variables, Quantifiers and Games

We write $\mathbb{A} \cong \mathbb{B}$ to indicate that two structures $\mathbb{A}$ and $\mathbb{B}$ are isomorphic. $\mathbb{A} \equiv \mathbb{B}$ denotes that $\mathbb{A}$ and $\mathbb{B}$ are elementarily equivalent, i.e. they are not distinguished by any sentence of FO. We also write $\mathbb{A} \equiv^k \mathbb{B}$ to denote that $\mathbb{A}$ and $\mathbb{B}$ agree on all sentences of L^k and $\mathbb{A} \equiv_p \mathbb{B}$ to denote that they agree on all sentences of FO with quantifier rank at most p. The quantifier rank of a formula is simply the maximal depth of nesting of quantifiers (irrespective of alternations) occurring within it. We combine these last two notations in $\mathbb{A} \equiv^k_p \mathbb{B}$ to denote that $\mathbb{A}$ and $\mathbb{B}$ are not distinguished by any sentence of L^k with quantifier rank at most p. More generally, if $\mathbf{s}$ is a tuple of elements from $\mathbb{A}$ and $\mathbf{t}$ a tuple from $\mathbb{B}$, we write $(\mathbb{A}, \mathbf{s}) \equiv^k (\mathbb{B}, \mathbf{t})$ to denote that, for any formula φ of L^k, $\mathbb{A} \models \varphi[\mathbf{s}]$ if, and only if, $\mathbb{B} \models \varphi[\mathbf{t}]$ and similarly for the other equivalences. We now briefly examine the relation between these equivalences and definability.

It is an easy observation that when the structures involved are finite, elementary equivalence coincides with isomorphism. Suppose σ is a *finite* relational vocabulary and $\mathbb{A}$ a σ-structure with n elements. We can construct a sentence

$$\delta_{\mathbb{A}} = \exists x_1 \ldots \exists x_n (\psi \wedge \forall y \bigvee_{1 \leq i \leq n} y = x_i)$$

where, $\psi(x_1, \ldots, x_n)$ is the conjunction of all atomic and negated atomic formulas that hold in $\mathbb{A}$. Now, for any structure $\mathbb{B}$, $\mathbb{B} \models \delta_{\mathbb{A}}$ if, and only if, $\mathbb{A} \cong \mathbb{B}$. This implies that any isomorphism-closed class of finite σ- structures $\mathcal{C}$ is defined by a formula $\varphi_{\mathcal{C}}$ of $L_{\infty\omega}$. Indeed a single infinitary disjunction of first-order formulas, or a single infinitary conjunction, will suffice. Note, however, that the number of variables required for $\varphi_{\mathcal{C}}$ is potentially infinite. This cannot be avoided as there are structures for every k (for instance a pure set with $k+1$ elements) that cannot be characterised by a formula of L^k (or indeed $L^k_{\infty\omega}$)

However, in the presence of order, we can do better. For each σ, there is a k such that any structure $\mathbb{A}$ in $\mathcal{O}_\sigma$ can be characterised up to isomorphism by a formula of L^k [Dawar *et al.*, 1996]. To see this, we first note that for every element a of $\mathbb{A}$, there is a formula $\beta_a(x)$ of L^2 such that a is the unique element of $\mathbb{A}$ satisfying $\mathbb{A} \models \beta_a[a]$. This is established by inductively defining the following class of formulas.

$$\begin{array}{lcl} \alpha_0(x) & \equiv & \neg(x = x) \\ \alpha_{n+1}(x) & \equiv & \forall y((y < x) \to (\alpha_n(y)))) \end{array}$$

where $\alpha_n(y)$ is the formula obtained from $\alpha_n(x)$ by *interchanging* all occurrences, free or bound, of x and y. It is clear that $\mathbb{A} \models \alpha_n[a]$ if and only if there are at most n elements less than or equal to a in the linear order $<^{\mathbb{A}}$. Thus, the formula $\beta_n \equiv \neg\alpha_{n-1} \wedge \alpha_n$ identifies the nth element of the order uniquely.

Using these formulas, it is clear that any m-tuple can be uniquely identified by a formula of L^m, and we can therefore construct a sentence $\psi_{\mathbb{A}}$ of L^m that determines the structure $\mathbb{A}$ up to isomorphism among structures in $\mathcal{O}_\sigma$, where m is at least 3 and at least as great as the maximum arity of any relation in σ. This means that any isomorphism-closed subclass of $\mathcal{O}_\sigma$ is definable by a sentence of $L^m_{\infty\omega}$ for some large enough m, depending only on σ.

It was shown by Kolaitis and Vardi [1992b] that a class of finite structures is definable in $L^k_{\infty\omega}$ if, and only if, it is closed under the relation $\equiv^k$. It is also not difficult to see that, provided the signature σ is finite, a class of structures is definable in FO if, and only if, it is closed under $\equiv_p$ for some p. Indeed, p is the quantifier rank of the defining sentence. This is because there are, up to logical equivalence, only finitely many sentences of quantifier rank p and the equivalence relations $\equiv_p$ therefore have finite index (see [Ebbinghaus and Flum, 1999, Sec. 2.2] for details). In particular, this implies that any infinitary sentence of finite quantifier rank is equivalent to a finitary sentence. Similar arguments show that a class of σ-structures is definable in L^k if, and only if, it is closed under $\equiv^k_p$ for some p. It is worth remarking here that the restriction to finite σ is important. For instance, if $\sigma = (P_i \mid i \in \omega)$ is an infinite collection of unary relation symbols and $S \subseteq \omega$ is an infinite and co-infinite set, then the sentence $\exists x(\bigwedge_{i\in S} P_i(x))$ is of quantifier rank 1 but not equivalent to any finitary sentence.

A natural question to ask might be, if a class of structures is closed under $\equiv^k$ and $\equiv_p$ for some k and some p, do there exist k' and p' such that it is also closed under $\equiv^{k'}_{p'}$? This question is discussed in some detail in [Dawar *et al.*, 1996]. Here we simply observe that the answer is yes on *ordered* structures if, and only if, Conjecture 2 fails. Thus, we have reformulated the main question as a question about suitable equivalence relations. We next look at characterisations of these relations by means of games.

Games The relation $\equiv_p^k$ has an elegant characterisation in terms of Ehrenfeucht-Fraïssé style two player games, essentially given by Barwise [1977] though versions were also independently presented by Immerman [1982] and Poizat [1982]. The game board consists of two structures $\mathbb{A}$ and $\mathbb{B}$ and a supply of k pairs of pebbles $(a_i, b_i), 1 \leq i \leq k$. The pebbles $a_1, \ldots, a_l$ are initially placed on the elements of an l-tuple $\mathbf{s}$ of elements in $\mathbb{A}$, and the pebbles $b_1, \ldots, b_l$ on a tuple $\mathbf{t}$ in $\mathbb{B}$. There are two players, Spoiler and Duplicator. At each move of the game, Spoiler picks up a pebble (either an unused pebble or one that is already on the board) and places it on an element of the corresponding structure. For instance he[1] might take pebble b_i and place it on an element of $\mathbb{B}$. Duplicator must respond by placing the other pebble of the pair in the other structure. In the above example, she must place a_i on an element of $\mathbb{A}$. If at the end of the move the partial map $f : \mathbb{A} \to \mathbb{B}$ given by $a_i \mapsto b_i$ is not a partial isomorphism, then Spoiler has won the game, otherwise it can continue for another move.

The result that links this game with the equivalence relation $\equiv_p^k$ is the following:

> If Duplicator has a strategy for playing the k-pebble game for p moves starting with the position $(\mathbb{A}, \mathbf{s}), (\mathbb{B}, \mathbf{t})$, then $(\mathbb{A}, \mathbf{s}) \equiv_p^k (\mathbb{B}, \mathbf{t})$.

If Duplicator has a strategy for playing indefinitely without losing, then $(\mathbb{A}, \mathbf{s}) \equiv^k (\mathbb{B}, \mathbf{t})$.

Finite Coloured Orders As an application of the game, we reproduce the argument given by Poizat [1982] showing that on coloured linear orders every sentence of FO is equivalent to one of L^3.

Let σ be a *finite* vocabulary consisting solely of unary relations and let $\mathbb{A}$ and $\mathbb{B}$ be structures in $\mathcal{O}_\sigma$. We refer to such structures as *coloured orders.* The key combinatorial argument in establishing the collapse of the L^k hierarchy on such structures is the following lemma.

LEMMA 3. *Let* $\mathbf{s} = (a_1, \ldots, a_m)$ *be a tuple of elements from* $\mathbb{A}$ *and* $\mathbf{t} = (b_1, \ldots, b_m)$ *a tuple of elements from* $\mathbb{B}$ *with* $a_i <^{\mathbb{A}} a_{i+1}$ *and* $b_i <^{\mathbb{B}} b_{i+1}$ *for all* i. *If, for each* $i < m$, $(\mathbb{A}, a_i, a_{i+1}) \equiv_p^3 (\mathbb{B}, b_i, b_{i+1})$ *then* $(\mathbb{A}, \mathbf{s}) \equiv_p (\mathbb{B}, \mathbf{t})$.

Proof. The proof is by induction on p. For the base case, $p = 0$, we only need to establish that the map $a_i \mapsto b_i$ is a partial isomorphism preserving $<$ and all relations in σ. The order is preserved by the hypothesis that the tuples are in increasing order and the unary relations are preserved since $(\mathbb{A}, a_i, a_{i+1}) \equiv_0^3 (\mathbb{B}, b_i, b_{i+1})$ for all i.

[1] By convention, Spoiler is male and Duplicator female.

Assume now that the claim is true for p and that $(\mathbb{A}, a_i, a_{i+1}) \equiv^3_{p+1} (\mathbb{B}, b_i, b_{i+1})$ for all $i < m$. We wish to use this to show that Duplicator can win the $p+1$-move game (with an unlimited number of pebbles) played from the starting position $(\mathbb{A}, \mathbf{s}), (\mathbb{B}, \mathbf{t})$. Suppose, without loss of generality, that Spoiler plays his first move in the structure $\mathbb{A}$, choosing element a. The case when $a = a_i$ for some i is trivial as Duplicator plays on b_i, so we assume that a is distinct from all elements of $\mathbf{s}$. If there is an i such that $a_i < a < a_{i+1}$, then by hypothesis Duplicator has a strategy to win the *3-pebble*, $p+1$-move game played from $(\mathbb{A}, a_i, a_{i+1}), (\mathbb{B}, b_i, b_{i+1})$. Let b be the response dictated by this strategy to Spoiler moving on a. Then, $(\mathbb{A}, a_i, a, a_{i+1}) \equiv^3_p (\mathbb{B}, b_i, b, b_{i+1})$ and so, in particular, $(\mathbb{A}, a_i, a) \equiv^3_p (\mathbb{B}, b_i, b)$ and $(\mathbb{A}, a, a_{i+1}) \equiv^3_p (\mathbb{B}, b, b_{i+1})$. Therefore the tuples $a_1, \ldots, a_i, a, a_{i+1}, a_m$ and $b_1, \ldots, b_i, b, b_{i+1}, b_m$ satisfy the hypothesis of the lemma and by the inductive hypothesis, we have that $(\mathbb{A}, a_1, \ldots, a_i, a, a_{i+1}, a_m) \equiv_p (\mathbb{B}, b_1, \ldots, b_i, b, b_{i+1}, b_m)$, as required. This leaves the cases when either $a < a_1$ or $a_m < a$. In these cases, by a similar argument, we are guaranteed a winning response using the winning strategies in the 3-pebble games for $(\mathbb{A}, a_1, a_2), (\mathbb{B}, b_1, b_2)$ and $(\mathbb{A}, a_{m-1}, a_m)$, $(\mathbb{B}, b_{m-1}, b_m)$. In all cases, we have established that $(\mathbb{A}, \mathbf{s}) \equiv_{p+1} (\mathbb{B}, \mathbf{t})$. ■

The result for coloured orders now follows.

THEOREM 4. *If σ is a finite vocabulary with only unary predicates then every sentence of* FO_σ *is equivalent, over $\mathcal{O}_\sigma$ to one of L^3.*

Proof. Let φ be a sentence of FO and p be its quantifier rank. Thus, the class of structures $\mathcal{C} \subseteq \mathcal{O}_\sigma$ defined by φ is closed under $\equiv_p$. By Lemma 3, if $\mathbb{A}, \mathbb{B} \in \mathcal{O}_\sigma$ are equivalent under $\equiv^3_p$, they are equivalent under $\equiv_p$. Thus, $\mathcal{C}$ is also closed under $\equiv^3_p$ and therefore, since σ is finite, definable by a sentence of L^3 of quantifier rank p. ■

It is worth remarking that the proofs of Lemma 3 and Theorem 4 do not rely on the structures being finite and work equally well with infinite structures. The restriction to finite vocabularies is required for the construction of the sentence of L^3 in the proof of Theorem 4. Of course, any sentence of FO can only refer to finitely many symbols, so a simple re-phrasing would establish it for arbitrary vocabularies.

3 Binary Relations

Theorem 4 establishes that the hierarchy formed by the number of variables on ordered finite structures collapses to L^3 when all relations other than the order are unary. The next step up in considering Conjecture 2 would be to consider the case when σ consists of a single, binary relation symbol. This case turns out to be equivalent to the general form of Conjecture 2. That is, we can make the following observation.

OBSERVATION 5. If the L^k hierarchy collapses on $\mathcal{O}_\sigma$ in the case where σ consists of a single binary relation, it collapses on $\mathcal{O}_\tau$ for all vocabularies τ.

The reason for this is that structures in an arbitrary fixed vocabulary can be encoded as graphs (i.e. structures with one binary relation E). Such an encoding is given, for instance, by a bi-interpretation in [Hodges, 1993, Sec. 5.5]. It is an easy adaptation of that construction to allow a bi-interpretation for ordered structures. This provides, for each σ, a translation Γ that maps a σ-structure $\mathbb{A}$ to a graph $\Gamma\mathbb{A}$ along with translations of formulas both ways. That is, there is for each σ-formula φ a formula $\gamma^{\rightarrow}(\varphi)$ in the language of graphs such that $\Gamma\mathbb{A} \models \gamma^{\rightarrow}(\varphi)$ if, and only if, $\mathbb{A} \models \varphi$. Conversely, there is for each formula ψ in the language of graphs a σ-formula $\gamma^{\leftarrow}(\psi)$ such that $\mathbb{A} \models \gamma^{\leftarrow}(\psi)$ if, and only if, $\Gamma\mathbb{A} \models \psi$. The translations $\gamma^{\leftarrow}$ and $\gamma^{\rightarrow}$ are themselves obtained as simple substitutions, i.e. $\gamma^{\rightarrow}(\varphi)$ is obtained from φ by replacing each occurrence of a relation symbol R from σ by an (E)-formula defining R. One consequence is that any increase in the number of variables in going from φ to $\gamma^{\rightarrow}(\varphi)$ (or from ψ to $\gamma^{\leftarrow}(\psi)$) is bounded by a multiplicative constant. Now, suppose there were a k such that every FO formula over ordered graphs was equivalent to one of L^k. Then, for an arbitrary σ-formula φ, $\gamma^{\rightarrow}(\varphi)$ would be equivalent, over ordered graphs, to a formula θ of L^k. Hence $\gamma^{\leftarrow}(\theta)$ would be a formula, equivalent to φ with at most ck variables for some constant c, establishing that any formula over σ-structures is equivalent to one of L^{ck}.

Infinite Structures While in the case of vocabularies with unary relations, Theorem 4 settles the question for both finite and infinite structures, in the presence of binary relations, the distinction between finite and infinite becomes important. Indeed, it is known that on infinite ordered structures with one binary relation, the L^k hierarchy is strict. This is established by an argument due to Venema [1990].

THEOREM 6. *There is a sentence φ of L^{k+1} in a vocabulary with one binary relation that is not equivalent, on* infinite *ordered structures to any sentence of $L^k_{\infty\omega}$.*

Proof. Let Q_k be a structure whose universe is the set of rational numbers. It interprets $<$ as the usual order on the rationals and interprets E as an equivalence relation with k equivalence classes, all of which are dense in the order. That is, between any two elements we can find representatives of all k equivalence classes. It is not difficult to see that $Q_k \equiv^k_{\infty\omega} Q_{k+1}$. That is to say, Duplicator can play an infinitely long k-pebble game on this pair of structures. Indeed, in response to any move by Spoiler, she only needs to find, in the suitable interval, a representative of a suitable equivalence class, i.e. one that bears the same relation of equivalence to the $k-1$ currently pebbled elements as the element selected in Spoiler's move. With at least k equivalence classes on each side, the existence of such a suitable element is guaranteed.

Thus, if φ is the sentence that asserts the existence of $k+1$ pairwise inequivalent elements, $Q_{k+1} \models \varphi$, $Q_k \not\models \varphi$ and we can conclude that φ is not equivalent to any sentence of $L^k_{\infty\omega}$ ■

Theorem 6 clearly does not hold on finite structures for, as previously observed (see [Dawar *et al.*, 1996]) *any* property of ordered finite graphs is definable in $L^3_{\infty\omega}$. Thus, if a version of this theorem is to settle Conjecture 2, we must replace $L^k_{\infty\omega}$ by L^k. This is to say that the game argument must also take quantifier rank into account. For any candidate sentence φ of L^{k+1} it will be the case that the finite ordered graphs satisfying φ are closed under $\equiv^k$ but we would like to show that for any p they are not closed under $\equiv^k_p$. However, as we see next, for the particular sentences used in the proof of Theorem 6, this is not possible.

Finite Equivalence Relations Let $\mathcal{E} \subseteq \mathcal{O}_{(E)}$ be the class of finite ordered equivalence relations. It is easily seen that the usual axioms for an equivalence relation give a sentence θ of L^3 that defines this class. We now define a sentence η_k using just two variables x and y that, on $\mathcal{E}$, defines the relations with at least k distinct equivalence classes. We first define the formula $\mu(x)$ as:

$$\mu(x) \equiv \forall y (E(x,y) \to x \leq y)$$

and write $\mu(y)$ to denote the formula obtained by interchanging in $\mu(x)$ all occurrences of x and y. Clearly, for any $\mathbb{A} \in \mathcal{E}$, $\mathbb{A} \models \mu[a]$ if, and only if, a is $<$-minimal in its equivalence class. We now define the following formulas δ_k inductively.

$$\begin{aligned} \delta_1(x) &\equiv \mu(x) \\ \delta_{k+1}(x) &\equiv \mu(x) \wedge \exists y\, (x < y \wedge \delta_k(y)), \end{aligned}$$

where, once again, $\delta_k(y)$ denotes the formula obtained from $\delta_k(x)$ by interchanging all occurrences of x and y. The formula $\delta_k(x)$ asserts that x is minimal in its equivalence class and there are at least $k-1$ elements greater than x in the order $<$, each of which is minimal in its equivalence class. The sentence $\eta_k \equiv \exists x\, \delta_k$ therefore defines exactly those structures in $\mathcal{E}$ which have at least k equivalence classes. $\eta_k \wedge \theta$ is then a sentence of L^3 defining equivalence relations of index at least k.

The above construction shows that the formulas used in the proof of Theorem 6 cannot be used to establish the L^k hierarchy on finite ordered structures. Indeed, the above construction can be generalised to show that any order-invariant property of finite ordered equivalence relations is definable in L^3.

We say that a formula φ of $\mathrm{FO}_{(E,<)}$ is order invariant on finite equivalence relations if, for any finite equivalence relation $\mathbb{A}$ and any two orderings $<_1$ and $<_2$ of the universe of $\mathbb{A}$, $(\mathbb{A}, <_1) \models \varphi$ if, and only if, $(\mathbb{A}, <_2) \models \varphi$. For instance, the sentence η_k constructed above is order invariant as it asserts a property of the equivalence relation and not of the order.

THEOREM 7. *Any sentence φ of $\mathrm{FO}_{(E,<)}$ that is order invariant on the class of finite equivalence relations is equivalent, on $\mathcal{E}$, to a sentence of L^2.*

Proof. Let the quantifier rank of φ be p. Our aim is to show that the models of φ in $\mathcal{E}$ are closed under $\equiv_q^2$ for some suitable value of q.

We associate with each equivalence relation $\mathbb{A}$ a (2^p+1)-tuple of integers which we call its *p-footprint*. To be precise, the p-footprint of $\mathbb{A}$ is the tuple $(n_1, \ldots, n_{2^p+1})$ where:

- for $1 \leq i \leq 2^p$, $n_i = \min(2^p, N_i)$ where N_i is the number of equivalence classes of $\mathbb{A}$ with exactly i elements.
- $n_{2^p+1} = \min(2^p, N)$ where N is the number of equivalence classes of $\mathbb{A}$ with more than 2^p elements.

We say an ordered equivalence relation $(\mathbb{A}, <)$ is in *normal order* if:

- the equivalence classes occupy contiguous blocks along the order, i.e. $(\mathbb{A}, <)$ satisfies the axiom $\forall x, y\, (E(x,y) \wedge x < y) \Rightarrow (\forall z\, x < z < y \Rightarrow E(x,z))$; and
- the equivalence classes occur in non-decreasing order of size, i.e. if $x < y$ then the equivalence class of y contains at least as many elements as that of x.

We now observe that if $\mathbb{A}$ and $\mathbb{B}$ are two equivalence relations with the same p-footprint, they cannot be distinguished by any order-invariant sentence φ of quantifier rank p or less. As φ is order-invariant, it suffices to show that if $(\mathbb{A}, <)$ and $(\mathbb{B}, <')$ are normal-order orderings of $\mathbb{A}$ and $\mathbb{B}$, then $(\mathbb{A}, <) \equiv_p (\mathbb{B}, <')$. This can be established by a game argument. Essentially, Duplicator's strategy in the p-move game on this pair of structures is to ensure that if, after r rounds have been played, pebbles have been placed on the elements $a_1 < \cdots < a_r$ and $b_1 <' \cdots <' b_r$ then the following conditions hold:

- for each i, the number of elements in the equivalence class of a_i is the same as that in the equivalence class of b_i *or* both numbers are greater than 2^p;
- for each i, the number of elements between a_i and a_{i+1} is either equal to the number of elements between b_i and b_{i+1} or both numbers are greater than 2^{p-r};
- the number of elements in the equivalence class of a_i that precede it in the order is equal to the corresponding number for b_i or both numbers are greater than 2^{p-r}; and
- the number of elements in the equivalence class of a_i that follow it in the order is equal to the corresponding number for b_i or both numbers are greater than 2^{p-r}.

The definition of p-footprint essentially guarantees that these conditions can be maintained. The proof is an adaptation of the proof that shows Duplicator can win a p-move game on a pair of linear orders of length greater than 2^p (see [Ebbinghaus and Flum, 1999, Example 2.3.6]) and we omit the details.

To complete the proof we show that for each p-footprint t, there is a sentence φ_t of L^2 that defines exactly the equivalence relations that have p-footprint t. These formulas cannot rely on a normal order, they need to work for any possible ordering of the equivalence relation. Observe that, since a p-footprint is a (2^p+1)-tuple of integers each between 0 and 2^p, there are only finitely many p-footprints for each p. To construct φ_t we first define formulas $\alpha_i(x)$ which assert that the equivalence class of x contains at least i elements that are greater than or equal to x in the order. They can be defined inductively as follows.

$$\begin{array}{lcl} \alpha_1(x) & \equiv & x = x \\ \alpha_{i+1}(x) & \equiv & \exists y\,(x < y \wedge E(x,y) \wedge \alpha_i(y)). \end{array}$$

Let $\beta_i(x) \equiv \mu(x) \wedge \alpha_i(x)$ be the formula that asserts that x is minimal in its equivalence class and that equivalence class contains at least i elements. We can now use this (in a manner by now familiar) to define formulas that state that there exist at least n elements satisfying $\beta(x)$. We first define formulas $\xi_{i,n}(x)$ inductively as follows:

$$\begin{array}{lcl} \xi_{i,1}(x) & \equiv & \beta_i(x) \\ \xi_{i,n+1}(x) & \equiv & \beta_i(x) \wedge \exists y\,(x < y \wedge \xi_{i,n}(y)) \end{array}$$

It is now easily seen that the sentence $\chi_{i,n} \equiv \exists x\, \xi_{i,n}(x)$ is true in $\mathbb{A}$ if, and only if, $\mathbb{A}$ contains at least n equivalence classes with i or more elements. It is then easy to see that a finite Boolean combination of these sentences will yield the desired φ_t. Finally, our original sentence φ, since it is characterised by a finite set of p-footprints, is equivalent to a finite disjunction of the sentences φ_t. ■

One reason why equivalence relations are significant is because they appear to be a rather weak excursion from the case of unary relations into the realm of the binary. An equivalence relation is really the coding in a single binary relation of a collection of sets, i.e. a unary vocabulary. As the number of equivalence classes is unbounded, this really codes an infinite unary vocabulary but, as Theorem 7 demonstrates, the methods of Theorem 4 do partially extend to this case. As long as our sentence is order-invariant, it can only talk about the sizes of the equivalence classes and then only up to some fixed finite limit (depending on the sentence) and, in the presence of order, all of this can be done with two variables. The proof of Theorem 7 is easily adapted to the case where, in addition to the one binary relation interpreted as an equivalence relation, σ contains any number of unary relations.

THEOREM 8. *Let σ be a vocabulary with one binary relation E in addition to any number of unary relations. Let $\mathcal{E} \subseteq \mathcal{O}_\sigma$ be the class of structures that interpret E as an equivalence relation. Then, any order-invariant* FO_σ *sentence φ is equivalent, over $\mathcal{E}$, to a sentence of L^2.*

We shall next show that, if we could drop the restriction to order-invariant sentences in Theorem 8, this would imply the failure of Conjecture 2.

Coloured Grids For the rest of this section, let σ be the vocabulary consisting of a binary relation E and two unary relations D and G. Also, let τ be the vocabulary consisting of a single binary relation *edge*. We write $\mathcal{E}$ for the class of structures in $\mathcal{O}_\sigma$ which interpret E as an equivalence relation. Our aim is to show that the question of whether the L^k hierarchy collapses on $\mathcal{E}$ is equivalent to asking whether the L^k hierarchy collapses on $\mathcal{O}_\tau$. We already know (see Observation 5) that the latter is equivalent to settling Conjecture 2.

LEMMA 9. *If every sentence of* FO *is equivalent to one of L^k on $\mathcal{E}$, then every sentence of* FO *is equivalent to one of L^{2k} on $\mathcal{O}_\tau$.*

Proof. The proof proceeds, like Observation 5, by establishing a bi-interpretation from $\mathcal{O}_\tau$ into $\mathcal{E}$. That is, we construct a map $\Gamma : \mathcal{O}_\tau \to \mathcal{E}$ such that for each $\mathbb{A} \in \mathcal{O}_\tau$, $\Gamma\mathbb{A}$ is first-order interpretable in $\mathbb{A}$ and $\mathbb{A}$ is first-order interpretable in $\Gamma\mathbb{A}$.

To be precise, for $\mathbb{A} = (A, \mathit{edge}, <^{\mathbb{A}})$ we define $\Gamma\mathbb{A} = (A^2, E, D, G, <^{\Gamma\mathbb{A}})$ to be the structure defined on the universe A^2 with the following relations:

- $E((a_1, a_2), (b_1, b_2))$ if, and only if, $a_2 = b_2$;
- $D(a_1, a_2)$ if, and only if, $a_1 = a_2$;
- $G(a_1, a_2)$ if, and only if, $edge(a_1, a_2)$; and
- $(a_1, a_2) <^{\Gamma\mathbb{A}} (b_1, b_2)$ if, and only if, $a_1 <^{\mathbb{A}} a_2$ or $a_1 = a_2$ and $b_1 <^{\mathbb{A}} b_2$.

In other words, $\Gamma\mathbb{A}$ interprets $<$ as the lexicographical order induced on A^2 by $<^{\mathbb{A}}$; E as the equivalence relation of equality in the second component; D as the diagonal through the square A^2 and G encodes the edge relation of the original graph $\mathbb{A}$. It is clear that $\Gamma\mathbb{A}$ is interpreted in $\mathbb{A}$ by first-order formulas. Indeed, the above definition of Γ explicitly gives *quantifier-free* formulas defining the relations in $\Gamma\mathbb{A}$. Thus, any $\sigma \cup \{<\}$-formula φ can be translated into a $\tau \cup \{<\}$-formula $\gamma^{\leftarrow}(\varphi)$ by substituting the above definitions for the relation symbols and replacing quantifiers $\exists x$ by pairs of quantifiers $\exists x_1, x_2$. The result is that $\Gamma\mathbb{A} \models \gamma^{\leftarrow}(\varphi)$ if, and only if, $\mathbb{A} \models \varphi$ and it is easily seen that if φ uses k distinct variables then $\gamma^{\leftarrow}(\varphi)$ requires at most $2k$ variables.

We now show how $\mathbb{A}$ can be interpreted inside $\Gamma\mathbb{A}$. As a first step, we show how the relation on A^2 of equality in the first component is definable from the relations E, D and $<^{\Gamma\mathbb{A}}$. Note that $\Gamma\mathbb{A}$ satisfies the axiom

$$\forall x \exists! y \, (E(x, y) \wedge D(y)).$$

That is, every element has a unique companion on the diagonal with which it shares the second component. We write $\delta(x)$ for the companion of x with the understanding that its use in formulas can be replaced with a proper first-order definition. We now define the following formulas.

$$\begin{array}{rcl} U(x) & \equiv & x < \delta(x) \\ L(x) & \equiv & \delta(x) < x \\ Ze(x, y) & \equiv & x < y \wedge \forall z \, (x < z \wedge z < y) \Rightarrow \neg D(z) \\ Si(x, y) & \equiv & x < y \wedge \exists! z \, (x < z \wedge z < y \wedge D(z)). \end{array}$$

In other words, $U(x)$ defines in $\Gamma\mathbb{A}$ the region of A^2 above the diagonal, i.e. those elements (a_1, a_2) with $a_1 <^{\mathbb{A}} a_2$ and $L(x)$ defines the region below the diagonal. U, L and D form a partition of A^2. $Ze(x, y)$ is a binary predicate asserting that no elements of the diagonal occur between x and y in the order $<^{\Gamma\mathbb{A}}$ and $Si(x, y)$ asserts that there is a unique diagonal element between them. Note that if two elements x and y in A^2 share the same first component then there is at most one diagonal element that can occur between them in the order $<$. The following formula $\eta(x, y)$ now defines all cases where $x < y$ and x and y share the same first component.

$$\begin{aligned}\eta(x,y) \quad \equiv \quad & [(U(x) \vee D(x)) \wedge U(y) \wedge Ze(x,y)] \vee \\ & [L(x) \wedge (D(y) \vee L(y)) \wedge Ze(x,y)] \vee \\ & [U(x) \wedge L(y) \wedge Si(x,y)] \vee \\ & [L(x) \wedge U(y) \wedge Si(x,y)].\end{aligned}$$

Thus, the formula $F(x,y) \equiv \eta(x,y) \vee \eta(y,x) \vee x = y$ defines the relation of equality in the first component.

We can now define (a structure isomorphic to) $\mathbb{A}$ inside $\Gamma\mathbb{A}$ by taking the universe to be the diagonal $\{a \mid D(a)\}$, the order to be the order induced on this by $<^{\Gamma\mathbb{A}}$ and defining the edge relation by the following formula

$$\epsilon(x,y) \quad \equiv \quad \exists z\, E(x,z) \wedge F(y,z) \wedge G(z).$$

Thus, we can translate any $(\mathit{edge}, <)$-formula φ into a formula $\gamma^{\rightarrow}(\varphi)$ by relativising all quantifiers to D and replacing all occurrences of *edge* by the formula ϵ. The result is that $\mathbb{A} \models \varphi$ if, and only if, $\Gamma\mathbb{A} \models \gamma^{\rightarrow}(\varphi)$.

To complete the proof note that if the hypothesis of the theorem is true, then for an arbitrary formula φ, $\gamma^{\rightarrow}(\varphi)$ is equivalent over $\mathcal{E}$ (and, in particular, over $\{\Gamma\mathbb{A} \mid \mathbb{A} \in \mathcal{O}_\tau\}$) to a formula θ of L^k. Thus, $\gamma^{\leftarrow}(\theta)$ is a formula of L^{2k} that is equivalent to φ over $\mathcal{O}_\tau$. ∎

A similar, and simpler, argument is used to establish a converse statement.

LEMMA 10. *If every sentence of* FO *is equivalent to one of L^k on $\mathcal{O}_\tau$, then every sentence of* FO *is equivalent to one of L^{k+2} on $\mathcal{E}$.*

Proof. The idea once again is to establish a bi-interpretation, i.e. a map $\Delta : \mathcal{E} \to \mathcal{O}_\tau$ which is definable by first-order formulas, as is its inverse. If $\mathbb{A} = (V, E, D, G, <)$, we define $\Delta\mathbb{A}$ to be the ordered graph whose universe is $V \uplus \{0,1\}$, the extension of V with two new elements with the order $<$ extended so that $0 < 1 < v$ for all $v \in V$ and the relation *edge* to include the pairs

- $(0,v)$ if $v \in V$ and $D(v)$;
- $(1,v)$ if $v \in V$ and $G(v)$; and
- (u,v) if $u,v \in V$ and $E(u,v)$.

It is then an easy exercise to see that the map Δ admits translations of first-order formulas between $\mathbb{A}$ and $\Delta\mathbb{A}$. In particular, in translating an $(\mathit{edge}, <)$ formula we replace occurrences of $\mathit{edge}(x,y)$ by the formula $(x =$

$0 \wedge D(y)) \vee (x = 1 \wedge G(y)) \vee E(x, y)$, where "$x = 0$" and "$x = 1$" are abbreviations for the formulas that state x is minimal in the order and the successor of the element minimal in the order respectively. It can be verified that this translation can be carried out so as to require at most two additional variables. ■

Lemmas 9 and 10 immediately yield the following theorem.

THEOREM 11. *There is a k such that every sentence of* FO *is equivalent to one of L^k on $\mathcal{O}_\tau$ if, and only if, there is a k such that every sentence of* FO *is equivalent to one of L^k on $\mathcal{E}$.*

Theorem 11 should be contrasted with Theorem 8. While the latter shows that the L^k hierarchy collapses on the class $\mathcal{E}$ if only order-invariant sentences are concerned, the former shows that such a collapse for non-order-invariant sentences would refute Conjecture 2.

We conclude this section with the observation that the equivalence relations used in the proof of Lemma 9 are of a special kind. Call a structure $\mathbb{A}$ in the vocabulary with one binary relation E a *grid-equivalence* if its universe is A^2 for some set A and E is interpreted as equality on the second component of the pairs. Let $\mathcal{GE}$ be the class of ordered grid-equivalences where the order is a lexicographic order of pairs. As any such structure is characterised up to isomorphism by the cardinality of A, we write G_n to denote a canonical such structure with $A = \{1, \ldots, n\}$.

OBSERVATION 12. Every first-order sentence is equivalent, on $\mathcal{GE}$ to one of L^2.

Proof. We aim to show that if $m, n > 2^p$ $G_m \equiv_p G_n$. That is, Duplicator can win the p-move Ehrenfeucht game played on this pair of structures. Duplicator's strategy is essentially an adaptation of the strategy used in the game on two linear orders (see the proof of Theorem 7). Recall that there Duplicator's strategy was to ensure that after r moves, the corresponding intervals formed by pebbled points in the two structures were either of equal length or both larger than 2^{p-r}. In the case of grid-equivalences the strategy is to maintain this condition simultaneously on both projections of the grid. Thus, suppose that after r moves have been played, pebbles are on $a_1, \ldots, a_r \in G_m$ and $b_1, \ldots, b_r \in G_n$. We write a_i^1 and a_i^2 for the two elements of $\{1, \ldots, m\}$ that make up the pair a_i and similarly b_i^1 and b_i^2 for the two components of b_i. Duplicator's strategy in the game is to ensure that after r moves have been played, the following condition holds: for each $i, j \in 1, \ldots, r$ and each $k \in \{1, 2\}$ either $b_j^k - b_i^k = a_j^k - a_i^k$, or $b_j^k - b_i^k > 2^{p-r}$ and $a_j^k - a_i^k > 2^{p-r}$. An easy inductive argument shows that this can be done. ■

While Theorem 8 shows that in the proof of Theorem 11 we cannot remove the use of non-order-invariant sentences, Observation 12 shows that we cannot remove the use of unary relations from that proof.

4 Fixed-Point Logics

In this section, we look at the relationship between the L^k hierarchy conjecture and certain questions on fixed-point logics. These logics, much studied in finite model theory, are extensions of first-order logic by means of operators for recursive or iterative definition of relations. In analysing their expressive power, finite variable logics have played an important role. Here we examine specifically the connection between the L^k hierarchy conjecture and what is known as Kolaitis and Vardi's ordered conjecture and some of the complexity-theoretic implications of it. We begin with a brief introduction to fixed-point logics. For more background, the reader is referred to [Ebbinghaus and Flum, 1999] or [Dawar and Gurevich, 2002].

Let $\varphi(R, \mathbf{x})$ be a formula, where R is a relation symbol, and $\mathbf{x}$ is a tuple of first order variables whose length r is the same as the arity of R. If $\mathbb{A}$ is a structure, with universe A, interpreting all symbols in φ other than those displayed, we think of φ as defining a map Φ from $\mathrm{Pow}(A^r)$—the set of all r-ary relations on A—to $\mathrm{Pow}(A^r)$ given by $\Phi(P) = \{\mathbf{a} \mid (\mathbb{A}, P, \mathbf{a}) \models \varphi\}$.

This view of a formula defining an operator on the space of relations gives a natural formalisation of inductive definitions. For instance, if Φ is a monotone map, we can speak of the least relation R such that $R(\mathbf{x}) \leftrightarrow \varphi(R, \mathbf{x})$. This relation is the *least fixed point* of the operator defined by φ. We can now define the logic LFP, obtained by closing first-order logic simultaneously under all the formula forming operations of first-order logic along with the rule:

> if R is a k-ary relation variable, $\mathbf{x}$ is a k-tuple of first order variables, $\mathbf{t}$ is a k-tuple of terms and φ is a formula in which R occurs only positively, then
>
> $$[\mathbf{lfp}[R, \mathbf{x}]\varphi](\mathbf{t})$$
>
> is a formula, in which all occurrences of R are bound, and all occurrences of the variables in $\mathbf{x}$ except those occurring in $\mathbf{t}$ are bound.

The intended semantics of this formula formation rule is that, for any structure $\mathbb{A}$, $\mathbb{A} \models [\mathbf{lfp}[R, \mathbf{x}]\varphi](\mathbf{t})$ if, and only if, $\mathbf{t}^{\mathbb{A}}$—the tuple of elements of $\mathbb{A}$ defined by the terms $\mathbf{t}$—is in the least fixed point of the monotone operator defined by $\varphi(R, \mathbf{x})$ on A^k.

The least fixed point of a monotone operator on a structure $\mathbb{A}$ can be obtained by iterating the operator. That is, we define a sequence of stages R^n where $R^0 = \emptyset$ and $R^{n+1} = \Phi(R^n)$, the relation defined by the formula φ when R is interpreted by R^n. If $\mathbb{A}$ is infinite, the iteration may have to be carried into the transfinite, but this is a case that need not concern us here. The sequence of relations R^n is non-decreasing and there is, therefore, a stage N at which $R^N = R^{N+1}$. The relation defined at this point is the least fixed point and we call the least N for which this is true the *closure ordinal* of φ on $\mathbb{A}$.

If the formula φ is not positive in the relation R, it does not define a monotone operator. It is, nonetheless, possible to iterate the operator it defines, starting with the empty relation. One may do this in an inflationary or a non-inflationary fashion. Adding operators for the fixed-points so formed yields the logic IFP (of inflationary fixed points) and PFP (of partial fixed points) respectively. We omit the precise definitions and refer the reader to [Ebbinghaus and Flum, 1999] or [Dawar and Gurevich, 2002] for details. In terms of expressive power, it is known that IFP and LFP are equivalent [Gurevich and Shelah, 1986; Kreutzer, 2002] and that LFP and PFP are equivalent if, and only if, the complexity classes PTime and PSpace coincide [Abiteboul and Vianu, 1995].

As noted in Section 1.2, Kolaitis and Vardi [1992b] showed that any formula of LFP (and, indeed, PFP) is equivalent, over finite structures, to a formula of $L^{\omega}_{\infty\omega}$. In particular, for any formula $\varphi(R, \mathbf{x})$ in L^k, there is a formula φ^n of L^{2k} that defines the relation R^n in all structures. Thus, $[\mathbf{lfp}[R, \mathbf{x}]\varphi](\mathbf{t})$ is equivalent to $\bigvee_{n\in\omega} \varphi^n(\mathbf{t})$. Of course, as we have earlier observed, for every vocabulary σ, there is a k such that every isomorphism-closed property of ordered σ-structures is expressible in $L^k_{\infty\omega}$. Thus, in the presence of order, the observation that the fixed-point logics translate into $L^{\omega}_{\infty\omega}$ is not very informative. Still, the form of the formula $\bigvee_{n\in\omega} \varphi^n(\mathbf{t})$ suggests an infinitary formula of a particularly uniform kind, as the formulas φ^n are generated from a single pattern. It is true that the number of variables in this formula depends on the formula φ and is not determined solely by the vocabulary σ. We may use the order to reduce the number of variables we use but this may result in a loss of uniformity. It is this trade-off between the number of variables and the uniformity of the first-order formula that we wish now to explore (for a related discussion of issues of uniformity in connection with descriptive complexity see [Barrington *et al.*, 1990]).

McColm [1990a; 1990b] considered the relative expressive power of LFP and FO in relation to a property he called *proficiency*. A class of structures $\mathcal{C}$ is said to be proficient if there is some FO formula $\varphi(R, \mathbf{x})$ such that the closure ordinals of φ on structures in $\mathcal{C}$ are unbounded. As McColm observed, the existence of unbounded inductions allows one to define arithmetic relations and allow certain coding and diagonalisation constructions. This led to the conjecture that on any proficient class of structures LFP is more expressive than FO. Though this was refuted in [Gurevich *et al.*, 1994], Kolaitis and Vardi [1992a] put forward a weaker form of the conjecture.

CONJECTURE 13 (Kolaitis-Vardi). On any class of ordered structures containing arbitrarily large finite structures, there is a formula of LFP that is not equivalent to any formula of FO.

The reason this is a weak form of McColm's conjecture is that it is trivial to define, on a linearly ordered structure, a formula whose closure ordinal is equal to the size of the structure. Thus, a class of structures satisfying the hypothesis of Conjecture 13 is necessarily proficient. The ordered conjecture of Kolaitis and Vardi remains open and it was shown in [Dawar *et al.*, 1996] that its resolution either way would solve significant complexity theoretic questions. In particular, a proof of the conjecture would show that the complexity classes LIN-H and ETIME are different. Here, LIN-H is the linear time hierarchy and ETIME is exponential time, i.e. the class of problems solvable in time $2^{O(n)}$. On the other hand, a refutation of Conjecture 13 would yield a separation of PTIME from PSPACE.

The relevance of this to the present discussion is that a refutation of Conjecture 2 would yield a proof of Conjecture 13. This is because McColm [1990a] showed, by a diagonalisation argument, that on any proficient class of structures, for every k, there is a sentence of LFP that is not expressible in LFP^k, the fragment of LFP that uses only k first-order variables. Thus, in particular, if $\mathcal{C} \subseteq \mathcal{O}_\sigma$ is a class of ordered σ-structures containing structures of arbitrarily large finite size, the number of first-order variables required to express all LFP formulas is unbounded. If Conjecture 2 were refuted, it would follow that there is a k, depending only on σ such that every formula of FO is equivalent on $\mathcal{C}$ to a formula of L^k. Hence any LFP formula that witnesses the fact that LFP is not equivalent to LFP^k would not be equivalent to any formula of FO.

One way of understanding McColm's parametrisation result, which shows that the number of variables hierarchy formed inside LFP is infinite on any infinite class of ordered structures is that it shows a trade-off between uniformity and the number of variables in infinitary formulas. While there is a k, depending only on σ such that every sentence of LFP is equivalent on $\mathcal{O}_\sigma$ to one of $L^k_{\infty\omega}$, it is not the case that it can be expressed simultaneously in LFP and using only k variables. The uniformity of the infinitary formula obtained as a translation of an LFP sentence requires an unbounded number of variables. Conjecture 2 asks for a similar trade-off in the case of first-order logic. However, this comparison also raises the possibility of looking for weaker trade-offs. Can we force the number of variables to be bounded by looking at weaker notions of uniformity? For instance, is it possible that every sentence of FO is equivalent on $\mathcal{O}_\sigma$ to one of LFP^k? This would not refute Conjecture 2 but would be sufficient to establish Conjecture 13.

It is worth noting here that we can show that there is a k, depending only on σ, such that any sentence of FO is equivalent, on $\mathcal{O}_\sigma$ to one of PFP^k, the fragment of PFP that uses only k variables. This follows from the fact that the satisfaction problem: 'given a finite structure $\mathbb{A}$ and first-order sentence φ, is it the case that $\mathbb{A} \models \varphi$?' is itself decidable in PSPACE. Since every property computable in PSPACE on ordered structures is definable in PFP, there is a coding of sentences of FO as natural numbers and a PFP formula $\mu(x)$ such that for $\mathbb{A} \in \mathcal{O}_\sigma$, $\mathbb{A} \models \mu[n]$ where n denotes the nth element of the linear order in $\mathbb{A}$ if, and only if, $n = \overline{\varphi}$ and $\mathbb{A} \models \varphi$, where $\overline{\varphi}$ denotes the number coding the formula φ. The formula μ has some fixed number c of variables and replacing all occurrences of n by its L^2 definition gives us a formula of PFP^{c+2} that is equivalent to φ on all sufficiently large structures (i.e. those with n or more elements). A disjunction of this with the finitely many sentences $\psi_{\mathbb{A}}$ (see Section 2) describing the smaller structures satisfying φ yields the desired sentence of PFP (see [Dawar and Hella, 1995] for a similar argument). This observation essentially establishes an upper bound on the amount of uniformity that needs to be sacrificed in order to turn all FO sentences on $\mathcal{O}_\sigma$ into infinitary ones with a bounded number of variables. We can do it in PFP. If we could do it in LFP, we would establish Conjecture 13. If we could do it in FO itself, we would refute Conjecture 2.

5 Clique and Circuit Complexity

We stated in Section 1.1 that Conjecture 1 was formulated in its positive form because intuition strongly suggests it is true. Nonetheless, in Sections 2 and 3 we concentrated on cases where the L^k-hierarchy collapses. This was intended to explore the boundary between the cases where a collapse occurs and those where we expect the hierarchy to be strict. However, another reason for concentrating on the cases where the hierarchy collapses is that we have no results which actually exhibit an interesting L^k hierarchy. Such a result would prove Conjecture 2. In this section we pose the question of what a proof of the main conjecture might look like. We will examine in particular the case of ordered graphs, i.e. the case when the vocabulary σ consists of a single binary relation E.

In order to prove that the L^k hierarchy is strict on $\mathcal{O}_\sigma$ the class of ordered graphs, we need two ingredients. We need a family of candidate formulas $(\varphi_k \mid k \in \omega)$ such that φ_k requires k variables to express and we need a technique for proving that φ_{k+1} is not expressible in L^k (or, for the weaker version, that for each k there is a φ_l that is not expressible in L^k).

For the first part—the family of formulas—there is a natural candidate that suggests itself. Let φ_k be the first-order sentence that asserts the existence of a k-clique. This sentence can clearly be written using k variables. Indeed, it only requires k *existential* quantifiers in prenex normal form. Moreover, not only is the sentence order-invariant, it does not require the use of the order at all. It is also straightforward to show, using Ehrenfeucht-Fraïssé games that, in the absence of order, φ_k is not expressible by any sentence (even an infinitary one) with fewer than k variables. The question then is whether we can show that the presence of an order does not help us in reducing the number of variables required to express the property. Or, at least, that as k increases, an unbounded number of variables is required.

This leads us to the question of the second ingredient, a technique for proving that φ_{k+1} cannot be expressed in L^k even when using an order. The first technique that suggests itself is the Ehrenfeucht-Fraïssé game which succeeds so well in the absence of order. Once we are dealing with ordered structures, we can no longer show that φ_{k+1} is not expressible in $L^k_{\infty\omega}$ so we must exploit the finite quantifier rank of formulas of L^k. That is to say, we would have to demonstrate, for each p, a pair of ordered graphs $\mathbb{A}_p$ and $\mathbb{B}_p$, one of which contains a $k+1$-clique and the other does not, such that $\mathbb{A}_p \equiv^k_p \mathbb{B}_p$. Unfortunately, constructing such Ehrenfeucht-Fraïssé games on ordered structures is notoriously difficult and so far such a construction has eluded us. Using the maxim that it is sometimes easier to aim to prove a more general result than the one you are interested in, we will consider a more powerful method, that of circuit complexity.

Methods based on circuit complexity have proved extremely useful in establishing inexpressibility results for first-order logic on ordered structures. A paradigmatic example is the result due to Furst, Saxe and Sipser [1984] that shows that the parity of a unary relation is not decidable by any family of bounded-depth, polynomial-size circuits. This implies that it is not definable in first-order logic on ordered structures, even in the presence of arbitrary numerical predicates. We now make these notions a little more precise.

As mentioned in Section 1.2, Barrington et al. [1990] showed that any first-order formula φ on ordered graphs can be turned into a family of circuits. Such a family is a sequence $(C_n \mid n \in \omega)$ of circuits, made of (unbounded fan-in) `and` and `or` gates and single input `not` gates. C_n takes at its inputs a binary string representing a structure $\mathbb{A}$ with universe $\{0,\ldots,n-1\}$ and produces a single bit output which is 1 if $\mathbb{A} \models \varphi$ and 0 if $\mathbb{A} \not\models \varphi$. Moreover, there is a constant d and a polynomial p depending only on φ such that the depth of C_n is bounded by d and the size of C_n is bounded by $p(n)$.

In addition to the size and depth of a family of circuits, there is one other crucial resource to consider for such a family and that is its *uniformity.* This refers to the complexity of the function that maps n to C_n. The family of circuits generated by a first-order formula is highly uniform. However, we can take an arbitrary family of constant depth and polynomial size and produce an equivalent first-order sentence if we allow ourselves arbitrary numerical predicates. A numerical predicate (of arity a) is a map from n to an a-ary relation on $\{0,\ldots,n-1\}$. Thus, the family of circuits is translated into a formula with additional predicate symbols for the numerical predicates. A cogent presentation of the relationship between the uniformity of circuits and the presence of numerical predicates can be found in [Denenberg *et al.*, 1986].

Circuit complexity, in particular the method of restrictions, has provided some of the most powerful techniques for establishing complexity lower bounds. The question we wish to consider is whether such a technique could be adapted to the problem at hand, of showing that the number of variables required to express the property of a graph having a clique is unbounded. In order to do this, we need to consider what measure on a circuit corresponds to the number of variables in a first-order formula.

A circuit of depth d can be thought of as being formed of d layers of gates with wires going from lower layers to higher ones. Define the width of a layer to be the number of gates in it and the *width* of the circuit to be the maximum width of any of its layers. It is then clear that the size s of a circuit (i.e. the total number of gates) is bounded by $d \cdot w$ where w is the width. The important point is that a careful analysis of the translation of a formula to a family of circuits (such as in [Barrington *et al.*, 1990]) shows that a formula with k variables and quantifier rank d can be translated into a family of circuits of depth d and width n^k. To understand why this is the case, consider a subformula ψ of φ with l free variables $x_1,\ldots,x_l$. Just as a circuit corresponding to the sentence φ has one output bit for the truth value of the formula φ on the structure on its input so, inductively, we construct a circuit C_ψ which has n^l output bits. These give the truth value of ψ for each of the n^l possible interpretations of its free variables in the universe $\{0,\ldots,n-1\}$. Thus, if the sentence φ has no more than k variables to begin with, none of its subformulas has more than k free variables and we can therefore limit the number of wires leaving any layer of the circuit to n^k. This gives us the required bound on the width.

It follows that in order to prove that φ_{k+1} cannot be expressed in L^k it would suffice to show that the problem of detecting a $k+1$-clique in an ordered graph cannot be solved by any family of circuits of constant depth and size $O(n^k)$. The clique problem has been much studied in the context of circuit complexity and a trade-off between the depth and size of circuits needed to solve it has been established. In particular, Lynch [1986] has established that there is a constant c such that any family of circuits of depth d that solves the k-clique problem must have size $\Omega(n^{ck/d})$. This shows a trade-off between depth and size but still leaves open the possibility that, given sufficient (but still constant) depth, a family of circuits could be found of size $O(n^k)$ for solving the $k+1$-clique problem.

Showing that no constant-depth family of circuits of size $O(n^k)$ could detect cliques of size $k+1$ would require a major breakthrough in circuit complexity. Moreover, such a result would establish something much stronger than we seek. Not only would it establish Conjecture 2, it would show that the problem of finding a $k+1$-clique could not be expressed in L^k even if we allow, in addition to the order, arbitrary arithmetic and numerical predicates. In the absence of such a breakthrough, might it be possible at least to use the methods of circuit lower bounds to establish the result in the restricted case where numerical predicates are not permitted? Barrington et al. [1990] are primarily concerned with uniform circuit, using a uniformity condition that is equivalent to having first-order logic with additional predicates for *order*, *addition* and *multiplication*. As Conjecture 2 only requires order, it can be seen as an extremely uniform version of a plausible complexity-theoretic conjecture, namely that there is no $O(n^k)$ size, constant-depth family of circuits for $k+1$-clique. The question then is, can we exploit the uniformity to give a proof of Conjecture 2 that does not involve solving the harder question? While circuit complexity has produced powerful methods for proving lower bounds, these methods do not typically rely on any assumptions of uniformity. It is not therefore clear what techniques could exploit the uniformity requirement embodied in Conjecture 2. Of course, one could think of dropping the order altogether as the ultimate uniformity requirement and, in this case, we do have a technique that works—that of Ehrenfeucht-Fraïssé games. Whether that can be pushed to work in the case when order is present remains open.

6 Existential Fragments

We noted that the formula expressing the existence of a $k+1$-clique in an ordered graph is one obvious candidate for showing a separation between L^{k+1} and L^k. This is a particularly simple property in that it is expressible in L^{k+1} using a purely existential formula (i.e. one that involves no universal quantification). While we are not able to show that there is no equivalent formula of L^k, we can show that no *existential* formula of L^k can express the existence of a $k+1$-clique. Thus, if there is a formula of L^k that expresses this property, it must make essential use of quantifier alternation. Indeed, we show that there isn't even an infinitary existential formula using only k variables that can express the existence of a $k+1$-clique. We write $\exists L^k$ for the fragment of L^k which involves no universal quantifiers and in which no existential quantifier appears within the scope of a negation. $\exists L^k_{\infty\omega}$ is the similarly defined fragment of $L^k_{\infty\omega}$. The logic $\exists L^k_{\infty\omega}$ of infinitary existential formulas with k variables has been studied (see [Kolaitis and Vardi, 1995; Kolaitis and Vardi, 2000]) as a tool for analysing the database language Datalog and in the context of constraint satisfaction problems.

For a pair of structures $\mathbb{A}$ and $\mathbb{B}$ we write $\mathbb{A} \Rrightarrow^k \mathbb{B}$ to denote that every sentence of $\exists L^k$ that is true in $\mathbb{A}$ is also true in $\mathbb{B}$. Note that, as the logic $\exists L^k$ is not closed under negation, this relation is not symmetric. This relation is characterised by a non-symmetric version of the k-pebble game, namely one in which Spoiler is confined to placing pebbles on elements of $\mathbb{A}$ and Duplicator must respond with elements of $\mathbb{B}$ (see [Kolaitis and Vardi, 1995] for details). If Duplicator can play this game forever without losing, then $\mathbb{A} \Rrightarrow^k \mathbb{B}$ and, indeed, any sentence of $\exists L^k_{\infty\omega}$ that is true in $\mathbb{A}$ is true in $\mathbb{B}$.

We now aim to define a single pair of ordered graphs $\mathbb{A}$ and $\mathbb{B}$ such that $\mathbb{A}$ contains a $k+1$ clique, $\mathbb{B}$ does not and Duplicator wins the infinite k-pebble existential game in which Spoiler only plays in $\mathbb{A}$.

Let $\mathbb{A}$ be the structure with universe $A = \{1, \ldots, k+1\}$ with $<$ interpreted as the natural order and $E(i,j)$ for all $i, j \in A$. $\mathbb{B}$ is a structure on the universe $B = \{1, \ldots, k+1\} \times \{1, \ldots, k\}$ with $<$ being the natural lexicographic order on pairs and $E((i,j),(i',j'))$ if, and only if, $i \neq i'$ and $j \neq j'$. It is easy to see that $\mathbb{B}$ does not contain a $k+1$-clique as any set of elements forming a clique must have pairwise disjoint second components, but there are only k distinct possibilities for the second component.

Now, it remains to show that Duplicator can play an infinitely long k-pebble game. Her strategy is to respond to any move by Spoiler on an element $i \in A$ not currently pebbled by playing on $(i,j) \in B$ where j is such that there is no pebble currently on any element (i',j). This is possible as there are k possible values of j and only $k-1$ pebbles currently on the board. To see that this strategy maintains a partial isomorphism, note that as the order on B is lexicographic, $(i,j) < (i',j')$ if $i < i'$, so the order is maintained. Furthermore the relation E holds between all pairs of elements in A and by the choice of Duplicator's move it holds on all pairs of pebbled elements in B.

This argument immediately yields the following result.

THEOREM 14. *For each k, there is a formula of $\exists L^{k+1}$ that is not equivalent to any formula of $\exists L^{k}_{\infty\omega}$ on ordered graphs.*

7 Other Connections

Throughout this paper, in considering the ramifications of Conjecture 2, we have established connections with a number of different areas of investigation ranging from temporal logics through fixed-point logics to complexity theory. The apparently simple question posed turns out to be closely linked to a number of other open problems. In the present section we present a few more questions from the areas of logic and computational complexity which have links to the main conjecture considered in the paper but which we have only space to consider briefly. The links mainly occur in the form of implausible (but not known to be false) consequences of the failure of Conjecture 2. Since it is much harder to find such dramatic consequences for a confirmation of the conjecture, this reinforces our belief in the conjecture.

Circuit Complexity We discussed in Section 5 the possibility that methods from circuit complexity might resolve the main conjecture. We noted that it would require a breakthrough to show that there is no constant-depth circuit of size $O(n^k)$ for detecting a $k+1$-clique (or for that matter, an l clique for some $l > k$). There is, of course, a flip side to that observation. If Conjecture 2 were to fail, this would imply that there is, indeed, a fixed k such that for any l, the l-clique problem is solvable by a family of constant depth circuits of size $O(n^k)$. This would be more than a breakthrough, it would be a stunning result. Moreover, this would hold not just for the problem of detecting an l-clique but for any first-order definable property of ordered graphs. This seems a most unlikely scenario.

Evaluating First-Order Logic Closely related to the above observation on the size of circuits corresponding to a first-order sentence is the fact that any formula of L^k can be turned into an algorithm for deciding the property it defines that runs in time $O(n^k)$ (see [Dawar *et al.*, 1996] for an argument). Thus, the failure of Conjecture 2 would imply that there is a k such that any property of ordered graphs definable in first-order logic is, in fact, decidable in time $O(n^k)$. This would settle (in a most unexpected fashion) an open question posed by Stolboushkin and Taitslin [1995].

Fixed-Parameter Tractability Of course, if it were the case that there is a k such that every first-order definable property is decidable in time $O(n^k)$, this would also be true of the existence of an l-clique for arbitrary l. This does not imply that there is a polynomial time algorithm for solving the CLIQUE problem as the L^k formulas (and hence the constants involved in the running time of the $O(n^k)$ algorithms) for the l-clique problem would depend on l. But, we can make two observations. Suppose Conjecture 2 fails and there is a k such that every first-order sentence is equivalent on ordered graphs to one of L^k. Suppose furthermore that this collapse of the L^k hierarchy is *effective*. That is, there is a computable function that maps any sentence of FO to an equivalent sentence of L^k. This would imply that there is an algorithm that, given a graph G and a parameter l, determines whether or not G contains an l-clique and runs in time $f(l)n^k$ where f is some computable function and k is independent of n and l. That is to say, that Clique would be fixed-parameter tractable. This is considered very unlikely as Clique is $W[1]$-complete (see [Downey and Fellows, 1999] for more on fixed-parameter tractability and $W[1]$-completeness). The second observation is that, if the collapse of the L^k-hierarchy were not only effective but *feasible*, i.e. there was a polynomial-time computable function that mapped any sentence of FO to an equivalent L^k sentence, then we would have a polynomial time algorithm for solving Clique and this implies that P = NP.

Second-order Hierarchy A well-known result of Fagin [1974] shows that a property of finite relational structures is decidable in the complexity class NP if, and only if, it is expressible in existential second-order logic. Since it is also known that monadic second-order quantifiers alone do not suffice to express all properties in NP, Fagin [1990] also asks whether one can show that the hierarchy formed by the arities of second-order quantifiers is strict. Indeed, it is an open question whether more than one binary second-order quantifier is ever needed to express any NP property. We can show that if Conjecture 2 were to fail then one (or indeed any fixed finite number) of binary quantifiers (or again, quantifiers of any fixed finite arity) would not suffice to express all NP properties of ordered graphs. The reason is that, if we fix a finite number and a maximum arity of second-order variables, then the first-order matrix of any sentence of existential second-order logic can be seen as being in a fixed finite vocabulary σ. If Conjecture 2 were to fail, there would be a k such that each such matrix could be replaced by a sentence of L^k. Just as any sentence of L^k can be evaluated in time $O(n^k)$ on structures of size n, one could show that, with the number and arity of second-order quantifiers bounded and the first-order matrix limited to L^k, any existential second-order formula can be evaluated by a nondeterministic algorithm running in time $O(n^l)$ for some fixed l. This would, of course, contradict the nondeterministic time hierarchy theorem (see [Papadimitriou, 1994]).

8 Conclusion

The question of whether one can establish a hierarchy on the number of first-order variables needed to express properties of finite ordered graphs has been known for many years within a community of researchers working in finite model theory. Little has been written about this question, largely because there are not many concrete partial results indicating progress towards solving it. The present paper aims to bring the question to a wider audience by demonstrating the deep links this apparently simple problem has with a number of areas of logic and complexity. At the same time, the problem illustrates the fundamental difficulties in proving inexpressibility results for first-order logic on finite ordered structures. We have two basic techniques with which we can establish such results. On the one hand there are Ehrenfeucht-Fraïssé games, which establish fine-grained results on unordered structures but which we do not know how to wield in the presence of order. On the other hand, we have methods from circuit complexity which work not only on ordered structures but on structures where the order is enriched with arbitrary numerical predicates. However, in the latter case the results they yield are not sufficiently fine-grained for our purposes. In the intermediate ground, of ordered structures with a high degree of uniformity, it seems that entirely new methods would be required. It is to be hoped that the development of such methods would yield new insights into the relation between logic and computational complexity. The problem of establishing the L^k hierarchy on ordered graphs appears to be an important testing ground for new techniques.

BIBLIOGRAPHY

[Abiteboul and Vianu, 1991] S. Abiteboul and V. Vianu. Datalog extensions for database queries and updates. *Journal of Computer and System Sciences*, 43:62–124, 1991.

[Abiteboul and Vianu, 1995] S. Abiteboul and V. Vianu. Computing with first-order logic. *Journal of Computer and System Sciences*, 50(2):309–335, 1995.

[Barrington *et al.*, 1990] D.M. Barrington, N. Immerman, and H. Straubing. On uniformity within NC_1. *Journal of Computer and System Sciences*, 41:274–306, 1990.

[Barwise, 1977] J. Barwise. On Moschovakis closure ordinals. *Journal of Symbolic Logic*, 42:292–296, 1977.

[Dawar and Gurevich, 2002] A. Dawar and Y. Gurevich. Fixed-point logics. *Bulletin of Symbolic Logic*, 8:65–88, 2002.

[Dawar and Hella, 1995] A. Dawar and L. Hella. The expressive power of finitely many generalized quantifiers. *Information and Computation*, 123(2):172–184, 1995.

[Dawar *et al.*, 1995] A. Dawar, S. Lindell, and S. Weinstein. Infinitary logic and inductive definability over finite structures. *Information and Computation*, 119(2):160–175, 1995.

[Dawar *et al.*, 1996] A. Dawar, S. Lindell, and S. Weinstein. First order logic, fixed point logic and linear order. In H. Kleine-Büning, editor, *Computer Science Logic '95*, volume 1092 of *LNCS*, pages 161–177. Springer-Verlag, 1996.

[Denenberg *et al.*, 1986] L. Denenberg, Y. Gurevich, and S. Shelah. Definability by constant-depth polynomial-size circuits. *Information and Control*, 70:216–240, 1986.

[Downey and Fellows, 1999] R.G. Downey and M.R. Fellows. *Parametrized Complexity.* Springer-Verlag, 1999.

[Ebbinghaus and Flum, 1999] H-D. Ebbinghaus and J. Flum. *Finite Model Theory.* Springer, 2 edition, 1999.

[Fagin, 1974] R. Fagin. Generalized first-order spectra and polynomial-time recognizable sets. In R. M. Karp, editor, *Complexity of Computation, SIAM-AMS Proceedings, Vol 7*, pages 43–73, 1974.

[Fagin, 1990] R. Fagin. Finite model theory – a personal perspective. In S. Abiteboul and P. C. Kanellakis, editors, *Third International Conference on Database Theory (LNCS 470)*, pages 3–24. Springer-Verlag, 1990.

[Furst *et al.*, 1984] M. Furst, J. Saxe, and M. Sipser. Parity, circuits, and the polynomial time hierarchy. *Math Systems Theory*, 17:13–27, 1984.

[Gabbay *et al.*, 1994] D. Gabbay, I. Hodkinson, and M. Reynolds. *Temporal Logic, vol. 1.* Oxford University Press, 1994.

[Gabbay, 1981] D. Gabbay. Expressive functional completeness in tense logic. In U. Mönnich, editor, *Aspects of Philosophical Logic*, pages 91–117. Reidel, 1981.

[Gurevich and Shelah, 1986] Y. Gurevich and S. Shelah. Fixed-point extensions of first-order logic. *Annals of Pure and Applied Logic*, 32:265–280, 1986.

[Gurevich *et al.*, 1994] Y. Gurevich, N. Immerman, and S. Shelah. McColm's conjecture. In *Proc. 9th IEEE Symp. on Logic in Computer Science*, 1994.

[Hodges, 1993] W. Hodges. *Model Theory.* Cambridge University Press, 1993.

[Hodkinson and Simon, 1997] I. Hodkinson and A. Simon. The k-variable property is stronger than h-dimension k. *J. Philosophical Logic*, 26:81–101, 1997.

[Immerman, 1982] N. Immerman. Upper and lower bounds for first-order expressibility. *Journal of Computer and System Sciences*, 25:76–98, 1982.

[Immerman, 1986] N. Immerman. Relational queries computable in polynomial time. *Information and Control*, 68:86–104, 1986.

[Immerman, 1991] N. Immerman. DSPACE[n^k]=VAR[$k+1$]. In *Proc. 6th IEEE Symp. on Structure in Complexity Theory*, pages 334–340, 1991.

[Immerman, 1999] N. Immerman. *Descriptive Complexity.* Springer, 1999.

[Kamp, 1968] H. Kamp. *Tense Logic and the Theory of Linear Order.* PhD thesis, University of California, Los Angeles, 1968.

[Kolaitis and Vardi, 1992a] Ph. G. Kolaitis and M. Y. Vardi. Fixpoint logic vs. infinitary logic in finite-model theory. In *Proc. 7th IEEE Symp. on Logic in Computer Science*, pages 46–57, 1992.

[Kolaitis and Vardi, 1992b] Ph. G. Kolaitis and M. Y. Vardi. Infinitary logics and 0-1 laws. *Information and Computation*, 98(2):258–294, 1992.

[Kolaitis and Vardi, 1995] Ph. G. Kolaitis and M. Y Vardi. On the expressive power of Datalog: Tools and a case study. *Journal of Computer and System Sciences*, 51:110–134, 1995.

[Kolaitis and Vardi, 2000] Ph. G. Kolaitis and M. Y Vardi. A game-theoretic approach to constraint satisfaction. In *Proc. 17th National Conference on Artificial Intelligence*, pages 175–181, 2000.

[Kreutzer, 2002] S. Kreutzer. Expressive equivalence of least and inflationary fixed-point logic. In *Proc. of the 17th IEEE Symp. on Logic in Computer Science (LICS).*, pages 403–410, 2002.

[Lynch, 1986] J.F. Lynch. A depth-size tradeoff for boolean circuits with unbounded fan-in. In *Proc. of the Conf. on Structure in Complexity Theory*, pages 234–248, 1986.

[McColm, 1990a] G. L. McColm. Parametrization over inductive relations of a bounded number of variables. *Annals of Pure and Applied Logic*, 48:103–134, 1990.

[McColm, 1990b] G. L. McColm. When is arithmetic possible? *Annals of Pure and Applied Logic*, 50:29–51, 1990.

[Papadimitriou, 1994] Ch. H. Papadimitriou. *Computational Complexity.* Addison-Wesley, 1994.

[Poizat, 1982] B. Poizat. Deux ou trois choses que je sais de L_n. *Journal of Symbolic Logic*, 47(3):641–658, 1982.

[Stolbouskin and Taitslin, 1995] A. Stolbouskin and M. Taitslin. Is first order contained in an initial segment of PTIME? In *Computer Science Logic 94*, volume 933 of *LNCS*. Springer-Verlag, 1995.

[Vardi, 1982] M. Y. Vardi. The complexity of relational query languages. In *Proc. of the 14th ACM Symp. on the Theory of Computing*, pages 137–146, 1982.

[Venema, 1990] Y. Venema. Expressiveness and completeness of an interval tense logic. *Notre Dame Journal of Formal Logic*, 31:529–547, 1990.