A Logical Characterization of the Counting Hierarchy

JUHA KONTINEN
University of Helsinki

In this article we give a logical characterization of the counting hierarchy. The counting hierarchy is the analogue of the polynomial hierarchy, the building block being Probabilistic polynomial time PP instead of NP. We show that the extension of first-order logic by second-order majority quantifiers of all arities describes exactly the problems in the counting hierarchy. We also consider extending the characterization to general proportional quantifiers $\mathcal{Q}_r^k$ interpreted as "more than an r-fraction of k-ary relations". We show that the result holds for rational numbers of the form $s/2^m$ but for any other $0 < r < 1$ the corresponding logic satisfies the 0-1 law.

Categories and Subject Descriptors: F.1.2 [**Computation by Abstract Devices**]: Modes of Computation—*Probabilistic computation*; F.1.3 [**Computation by Abstract Devices**]: Complexity Measures and Classes—*Complexity hierarchies*; F.4.1 [**Mathematical Logic and Formal Languages**]: Mathematical Logic—*Computational Logic*

General Terms: Languages, Theory

Additional Key Words and Phrases: Counting hierarchy, majority quantifiers, quantifier elimination, the 0-1 law

ACM Reference Format:

Kontinen, J. 2009. A logical characterization of the counting hierarchy. ACM Trans. Comput. Logic. 10, 1, Article 7 (January 2009), 21 pages. DOI = 10.1145/1459010.1459017 http://doi.acm.org/10.1145/1459010.1459017

1. INTRODUCTION

The main goal of descriptive complexity theory is to give logical characterizations of central complexity classes. The seminal result in this field was Fagin's [1974] characterization of NP in terms of problems describable in existential second-order logic (∃SO). Since then, most of the central complexity classes have been given such logical characterization. Stockmeyer [1976] defined the polynomial hierarchy (PH) and observed that full second-order logic describes

The author was financially supported by the Academy of Finland, project No. 106300.

Author's address: Department of Mathematics and Statistics, P.O. Box 68, FIN-00014 University of Helsinki, Helsinki, Finland; email: juha.kontinen@helsinki.fi.

exactly the problems in the polynomial hierarchy. This result is a corollary of Fagin's characterization of NP and the fact that the levels Σ_k^{p} of the polynomial hierarchy can be characterized in terms of polynomial bounded existential and universal quantifiers. In other words, for a language L, $L \in \Sigma_k^{\mathrm{p}}$ if and only if there is a polynomial-time predicate R and a polynomial q such that

$$x \in L \Leftrightarrow \exists^{q(|x|)} x_1 \forall^{q(|x|)} x_2 \cdots Q x_k R(x, x_1, x_2, \ldots, x_k),$$

where the quantifiers range over words of length at most $q(|x|)$, and $Q = \forall^{q(|x|)}$ if k is even and $Q = \exists^{q(|x|)}$ otherwise.

The counting hierarchy (CH) was defined by Wagner [1986]. The counting hierarchy is the analogue of the polynomial hierarchy, the building block being Probabilistic polynomial time (PP):

—$C_0P = \mathrm{P}$,
—$C_{k+1}P = \mathrm{PP}^{C_kP}$,
—$\mathrm{CH} = \bigcup_{k \in \mathbb{N}} C_kP$.

The definition above in terms of oracles is due to Torán [1991]. Probabilistic polynomial time was defined by Gill [1977] in terms of probabilistic Turing machines, and independently by Simon [1975] in the context of threshold languages. The class PP can be also defined using ordinary nondeterministic Turing machines by altering the way a machine accepts its input. The class PP consists of languages L for which there is a polynomial time-bounded nondeterministic Turing machine N such that, for all inputs x, $x \in L$ iff more than half of the computations of N on input x end up accepting. The class PP has been studied extensively. Since Gill's paper it was known that NP $\subseteq$ PP and that PP is closed under complement. Russo [1985] showed that PP is closed under symmetric difference. Toda [1991] showed that surprisingly $\mathrm{PH} \subseteq \mathrm{P}^{\mathrm{PP}}$. It remained open for many years whether PP is closed under intersection until Beigel et al. [1995].

The counting hierarchy was originally defined in terms of the polynomial counting quantifier $\mathcal{C}$. Let K be a class of languages. Define $\mathcal{C}K$ to be the class of languages L for which there is some polynomial p and some $L' \in K$ such that $x \in L$ iff

$$|\{y \ : \ |y| = p(|x|) \text{ and } (x, y) \in L'\}| > 2^{p(|x|)-1}.$$

The counting hierarchy can be now defined using the quantifier $\mathcal{C}$ as follows:

—$C_0P = \mathrm{P}$,
—$C_{k+1}P = \mathcal{C}C_kP$.

Torán [1991] showed that the definition in terms of oracles is equivalent to the definition above, that is, he showed that

$$\mathcal{C}C_KP = \mathrm{PP}^{C_kP}$$

for all $k \in \mathbb{N}$. The original definition Wagner [1986] actually used a more general definition of the quantifier $\mathcal{C}$ but the definition given here is equivalent to the original one. The properties of the counting hierarchy are discussed in detail in Torán [1991] and Allender and Wagner [1990].

Generalized quantifiers have been studied extensively in the context of finite model theory. Mostowski [1957] was the first to consider quantifiers other than the familiar existential and universal ones. Lindström [1966] defined the concept of a first-order generalized quantifier in its most general form. We refer to Väänänen [1999] and Ebbinghaus and Flum [1999] for surveys of generalized quantifiers in finite model theory.

The study of second-order generalized quantifiers is a relatively new and unexplored area in finite model theory. It seems that second-order generalized quantifiers first appeared in the realm of finite model theory in the context of Leaf languages (see Burtschick and Vollmer [1998], and Vollmer [1999]). In this context acceptance of a word given as an input to a nondeterministic machine depends only on the values printed at the leaves of the computation tree. In other words, having fixed a leaf language B, an input is accepted if the values printed at the leaves of the computation tree, ordered by the natural order of the paths of the machine, constitute a word in B. Certain complexity classes have been characterized in this context on ordered structures in terms of logics Q^1_B FO, that is, sets of formulas of the form "second-order quantifier followed by a first-order formula". A crucial assumption, for these characterizations to hold, is that the leaf language B has a neutral element. In the logical side, having a neutral element gives a certain relativization property for the corresponding quantifier Q^1_B. It is worth noting that the leaf language

$$\{w \in \{0, 1\}^* \mid w \text{ has more "1"}s \text{ than "0"}s\},$$

corresponding to PP does not have a neutral element. In fact, we show by a definability argument that, for $k > 1$, the relativization of the k-ary majority quantifier Most^k is expressible in the logic $\mathrm{FO}(\mathrm{Most}^k)$.

Andersson [2002] has also studied second-order generalized quantifiers. Andersson [2002] showed that on finite structures with at most binary relations almost any countable logic is equivalent to a uniformly obtained sublogic of $\mathrm{FO}(\mathcal{Q})$, where $\mathcal{Q}$ is some second-order generalized quantifier of type ((1)), that is, a quantifier that applies to one formula and binds one unary second-order variable in it. He also showed that the result extends to all finite structures if $\mathcal{Q}$ is allowed to bind one binary second-order variable instead (type ((2))).

Kontinen [2004, 2006] has studied definability questions of second-order generalized quantifiers. Definability questions of first-order generalized quantifiers have been studied extensively in finite model theory. In the first-order case, definability of a quantifier Q in a logic $\mathcal{L}$ just means that the class of structures, used to interpret Q, is axiomatizable in $\mathcal{L}$. In the second-order case, things are not as direct, but a similar concept can be formulated.

The main result of this paper is a characterization of the counting hierarchy in terms of second-order majority quantifiers. Our proof is based on the observation that the k-ary second-order existential quantifier $\exists^2_k$ can be defined in terms of the quantifier Most^k and first-order logic. We also show that the k-ary second-order Rescher quantifier $\mathcal{R}^k$ can be defined in the logic $\mathrm{FO}(\mathrm{Most}^{k+1})$, and, for $k \geq 2$, already in $\mathrm{FO}(\mathrm{Most}^k)$. By using the fact that only one application of the quantifier Most^k is needed to express the quantifiers $\exists^2_k$ and $\mathcal{R}^{k-1}$, we also obtain characterizations of the levels C_nP of CH in terms of formulas having at most n nested applications of majority quantifiers. We also consider extending the characterization of CH to general proportional quantifiers $\mathcal{Q}^k_r$ interpreted as "more than an r-fraction of k-ary relations". We show that the result holds for rational numbers of the form $s/2^m$ but for any other $0 < r < 1$ the extension of FO by $\mathcal{Q}^k_r$, for $k \in \mathbb{N}^*$, satisfies the so-called 0-1 law for relational vocabularies. The 0-1 law implies, for example, that the class $\mathrm{EVEN}(\tau)$ of all τ-structures with an universe of even cardinality cannot be axiomatized for any relational vocabulary τ.

Many of the logical characterizations of complexity classes hold only on ordered structures. Ordering is needed in describing the computations of Turing machines using formulas. However, this assumption is not so crucial in the case of second-order logic, since an ordering can be created using second-order quantification. Therefore, the characterizations of NP and PH in terms of $\exists$SO and SO, respectively, hold also on unordered structures. For the same reason, the characterization of CH given in this article holds also on unordered structures.

This article is arranged as follows. After recalling the relevant definitions in Section 2, we present the definability results discussed above in Section 3. The characterization of CH in terms of second-order majority quantifiers is shown in Section 4, and in Section 5 we turn to general proportional quantifiers. The article ends with a result showing that the logic $\mathrm{FO}(\mathrm{Most}^1)$ is strictly stronger than monadic second-order logic.

2. PRELIMINARIES

Vocabularies τ are finite sets consisting of relation symbols and constant symbols. All structures are assumed to be finite. The universe of a structure $\mathbb{M}$ is denoted by M. The class of all τ-structures is denoted by $\mathrm{Str}(\tau)$. For a logic $\mathcal{L}$, the set of τ-formulas of $\mathcal{L}$ is denoted by $\mathcal{L}[\tau]$. If φ is a τ-sentence, then the class of τ-models of φ is denoted by $\mathrm{Mod}(\varphi)$. A class K of τ-models is said to be axiomatizable in a logic $\mathcal{L}$, if $K = \mathrm{Mod}(\varphi)$ for some sentence $\varphi \in \mathcal{L}[\tau]$. For logics $\mathcal{L}$ and $\mathcal{L}'$, we write $\mathcal{L} \leq \mathcal{L}'$, if for every τ and every sentence $\varphi \in \mathcal{L}[\tau]$ there is a sentence $\psi \in \mathcal{L}'[\tau]$ such that $\mathrm{Mod}(\varphi) = \mathrm{Mod}(\psi)$. The set of natural numbers is denoted by $\mathbb{N}$ and $\mathbb{N}^*$ denotes the set $\mathbb{N} \setminus \{0\}$.

In this article, we consider polynomial time-bounded nondeterministic Turing machines N over input alphabet $\Sigma = \{1, 0\}$. Every machine N is assumed to have two halting states: accepting and rejecting, and every computation path of N must end in one of these states. In particular, as in Beigel et al. [1995], we do not insist that all computation paths have the same length. We assume that the reader is familiar with the basic notions of complexity theory. For a language A, the class PP^A consists of languages L for which there is a polynomial time-bounded nondeterministic oracle Turing machine N with oracle A such that, for all inputs x, $x \in L$ iff more than half of the computations of N on input x end up accepting. For a set of languages K, $\mathrm{PP}^K = \bigcup\{\mathrm{PP}^A \mid A \in K\}$.

Structures, considered as inputs to Turing machines, are assumed to be ordered. Structures can be then encoded to binary strings by concatenating the bit strings coding the relations and constants in the following way. Let $\tau = \{<, R_1, \ldots , R_s, c_1, \ldots , c_m\}$ be a vocabulary and let $\mathbb{M}$ be a τ-structure. We may assume that $M = \{0, \ldots , n-1\}$ for some $n \in \mathbb{N}$. Now the interpretation $R_i^{\mathbb{M}}$ of each relation symbol R_i is encoded as a binary string $\text{bin}^{\mathbb{M}}(R_i)$ of length n^{r_i}, where r_i is the arity of R_i, such that "1" in a given position indicates that the corresponding tuple in the lexicographic ordering of M^k is in $R_i^{\mathbb{M}}$. Similarly, the interpretation of a constant c_i is encoded by the string $\text{bin}^{\mathbb{M}}(c_i)$ corresponding to number $c_i^{\mathbb{M}}$ in binary. The binary encoding $\text{bin}(\mathbb{M})$ of a structure $\mathbb{M}$ is defined as the concatenation of the bit strings coding its relations and constants:

$$\text{bin}(\mathbb{M}) = \text{bin}^{\mathbb{M}}(R_1) \cdots \text{bin}^{\mathbb{M}}(R_s)\,\text{bin}^{\mathbb{M}}(c_1) \cdots \text{bin}^{\mathbb{M}}(c_m).$$

On the other hand, any binary string can be viewed as a word structure over vocabulary $\tau = \{<, P\}$, where $<$ is a binary predicate, interpreted as an ordering, and P is unary. Given a binary word x, we sometimes denote the corresponding word structure by $\mathcal{W}_x$. The encoding of structures to binary word structures can be defined in a first-order way assuming all structures are equipped with an ordering and some numeric predicates such as $+$ and $\times$ (see Immerman [1999]). The predicates $+$ and $\times$ are defined as

$$+(i, j, k) \Leftrightarrow i + j = k,$$
$$\times(i, j, k) \Leftrightarrow i \times j = k.$$

Note that we do not include $<$ or any numeric predicates to $\text{bin}(\mathbb{M})$ since they can be easily recomputed.

Given a class K of ordered structures, we write

$$L_K = \{\text{bin}(\mathbb{M}) \mid \mathbb{M} \in K\}$$

for the language corresponding to K. We abbreviate $L_{\text{Mod}(\varphi)}$ to L_φ. Now that we have encoded classes of structures to languages over alphabet $\{0, 1\}$, we define what it means for a logic to capture a complexity class. We say that a logic $\mathcal{L}$ captures a complexity class C if for all τ of the form $\{<, R_1, \ldots , R_s, c_1, \ldots , c_m\}$, and all classes K of τ-structures,

$$L_K \in C \text{ iff } K = \text{Mod}(\varphi) \text{ for some } \varphi \in \mathcal{L}[\tau].$$

We shall next define the notion of a logic strongly capturing a complexity class. Given a class K of τ-structures (not necessarily ordered), the class $K_<$ of ordered representations of structures in K is defined as

$$K_< = \{(\mathbb{M}, <) \mid \mathbb{M} \in K, \; < \text{ an ordering of } M\}.$$

We say that a logic $\mathcal{L}$ strongly captures a complexity class C if for all vocabularies τ and all classes K of τ-structures,

$$L_{K_<} \in C \text{ iff } K = \text{Mod}(\varphi) \text{ for some } \varphi \in \mathcal{L}[\tau].$$

The prime example of a logic strongly capturing a complexity class is $\exists$SO. On the other hand, the characterizations of P and PSPACE in terms of fixed point logics LFP and PFP hold only on ordered structures since, for example, even cardinality cannot be expressed on unordered structures.

2.1 Generalized Quantifiers

First-order logic cannot express, for example, that a formula holds for an even number of elements. To acquire an extension of FO with this feature, we can extend it by a new quantifier Q_{even} with interpretation given by

$$\mathbb{M} \models Q_{\text{even}}\, x\, \varphi(x) \Leftrightarrow |\varphi^{\mathbb{M}}| \text{ is even,}$$

where $\varphi^{\mathbb{M}} = \{a \in M \mid \mathbb{M} \models \varphi(a)\}$. Let P be a unary predicate symbol. In general, the interpretation of a Lindström quantifier Q of type (1) is given by

$$\mathbb{M} \models Qx\, \varphi(x) \Leftrightarrow (M, \varphi^{\mathbb{M}}) \in K,$$

where M denotes the universe of $\mathbb{M}$ and K is a class of $\{P\}$-structures that is closed under isomorphisms. In fact, any class of relational structures gives rise to a quantifier.

Definition 2.1. Let $s = (l_1, \ldots, l_r)$ be a tuple of positive integers. A Lindström quantifier of type s is a class Q of structures of vocabulary $\tau_s = \{P_1, \ldots, P_r\}$ such that P_i is l_i-ary for $1 \leq i \leq r$, and Q is closed under isomorphisms.

The extension FO(Q) of first-order logic by a quantifier Q is defined as follows:

- The formula formation rules of FO are extended by the rule: if for $1 \leq i \leq r$, $\varphi_i(\bar{x}_i)$ is a formula and $\bar{x}_i$ is an l_i-tuple of pairwise distinct variables then $Q\bar{x}_1, \ldots, \bar{x}_r\, (\varphi_1(\bar{x}_1), \ldots, \varphi_r(\bar{x}_r))$ is a formula.
- The satisfaction relation of FO is extended by the rule:
$$\mathbb{M} \models Q\bar{x}_1, \ldots, \bar{x}_r\, (\varphi_1(\bar{x}_1), \ldots, \varphi_r(\bar{x}_r)) \text{ iff } \left(M, \varphi_1^{\mathbb{M}}, \ldots, \varphi_r^{\mathbb{M}}\right) \in Q,$$
where $\varphi_i^{\mathbb{M}} = \{\overline{a} \in M^{l_i} \mid \mathbb{M} \models \varphi_i(\overline{a})\}$.

Example 2.2. Let us look at some examples of Lindström quantifiers.

$$\begin{aligned}
\forall &= \{(M, P) \mid P \subseteq M \text{ and } P = M\} \\
\exists &= \{(M, P) \mid P \subseteq M \text{ and } P \neq \emptyset\} \\
Q_{\text{even}} &= \{(M, P) \mid P \subseteq M \text{ and } |P| \text{ is even}\} \\
\mathrm{R} &= \{(M, P, S) \mid P, S \subseteq M \text{ and } |P| > |S|\}
\end{aligned}$$

The first example is the familiar first-order universal quantifier. The quantifier Q_{even} says that a formula holds for an even number of elements. The last example R is the so-called Rescher quantifier. It allows us to compare the size of two definable sets.

We say that a quantifier Q is definable in a logic $\mathcal{L}$ if the class Q is axiomatizable in $\mathcal{L}$. If $\mathcal{L}$ has the substitution property and is closed under FO-operations, then definability of Q in $\mathcal{L}$ implies that FO(Q) $\leq \mathcal{L}$. So, among such logics, FO(Q) is the minimal logic in which Q is axiomatizable.

Let us then turn to second-order generalized quantifiers. Assume $t = (s_1, \ldots, s_w)$, where $s_i = (l^i_1, \ldots, l^i_{r_i})$ is a tuple of positive integers for $1 \leq i \leq w$. A second-order structure of type t is a structure of the form $(M, P_1, \ldots, P_w)$, where $P_i \subseteq \mathcal{P}(M^{l^i_1}) \times \cdots \times \mathcal{P}(M^{l^i_{r_i}})$.

Definition 2.3. A second-order generalized quantifier $\mathcal{Q}$ of type t is a class of structures of type t such that $\mathcal{Q}$ is closed under isomorphisms.

Example 2.4.

$$\begin{aligned}
\exists^2_k &= \{(M, P) \mid P \subseteq \mathcal{P}(M^k) \text{ and } P \neq \emptyset\} \\
\mathrm{Most}^k &= \{(M, P) \mid P \subseteq \mathcal{P}(M^k) \text{ and } |P| > 2^{|M|^k - 1}\} \\
\mathrm{Most}^k_{\mathrm{r}} &= \{(M, P, S) \mid P, S \subseteq \mathcal{P}(M^k) \text{ and } |P \cap S| > |P \setminus S|\} \\
\mathcal{R}^k &= \{(M, P, S) \mid P, S \subseteq \mathcal{P}(M^k) \text{ and } |P| > |S|\}
\end{aligned}$$

The first example is the familiar k-ary second-order existential quantifier. The quantifier $\mathrm{Most}^k_{\mathrm{r}}$ is the relativization of the quantifier Most^k. The quantifier $\mathcal{R}^k$ is the k-ary second-order version of the Rescher quantifier.

The extension FO($\mathcal{Q}$) of FO by a quantifier $\mathcal{Q}$ is defined as follows:

—The formula formation rules of FO are extended by the rule: if for $1 \leq i \leq w$, $\varphi_i(\overline{X}_i)$ is a formula and $\overline{X}_i = (X_{1,i}, \ldots, X_{r_i,i})$ is a tuple of pairwise distinct predicate variables such that the arity of $X_{j,i}$ is l^i_j for $1 \leq j \leq r_i$, then

$$\mathcal{Q}\overline{X}_1, \ldots, \overline{X}_w\, (\varphi_1(\overline{X}_1), \ldots, \varphi_w(\overline{X}_w))$$

is a formula.

—Satisfaction relation of FO is extended by the rule:

$$\mathbb{M} \models \mathcal{Q}\overline{X}_1, \ldots, \overline{X}_w\, (\varphi_1, \ldots, \varphi_w) \text{ iff } \left(M, \varphi_1^{\mathbb{M}}, \ldots, \varphi_w^{\mathbb{M}}\right) \in \mathcal{Q},$$

where $\varphi_i^{\mathbb{M}} = \{\overline{R} \in \mathcal{P}(M^{l^i_1}) \times \cdots \times \mathcal{P}(M^{l^i_{r_i}}) \mid \mathbb{M} \models \varphi_i(\overline{R})\}$.

A notion of definability can also be formulated for second-order generalized quantifiers. Since second-order generalized quantifiers are interpreted using classes of second-order structures, there is no direct connection between quantifiers and classes of structures determined by sentences. Definability of a quantifier $\mathcal{Q}$ in a logic $\mathcal{L}$ can be formalized by considering axiomatizability in an extension of $\mathcal{L}$ by suitable second-order predicates (see Kontinen [2004, 2006]). For the purposes of this article it suffices to note that the definability results proved in the following section give us a uniform way to express the definable quantifier in our logic. In other words, definability of $\mathcal{Q}$ in $\mathcal{L}$ provides us with a compositional translation of $\mathcal{L}(\mathcal{Q})$-formulas into $\mathcal{L}$-formulas.

Example 2.5. The quantifier Most^k can be defined using the quantifier $\mathrm{Most}^k_{\mathrm{r}}$ as follows:

$$\models \mathrm{Most}^k\, X\, \psi \Leftrightarrow \mathrm{Most}^k_{\mathrm{r}}\, X, X\, (X = X, \psi).$$

The quantifier $\mathrm{Most}^k_{\mathrm{r}}$ in turn can be defined in terms of the quantifier $\mathcal{R}^k$:

$$\models \mathrm{Most}^k_{\mathrm{r}}\, X, Y\, (\psi, \phi) \Leftrightarrow \mathcal{R}^k\, X, Y\, (\psi \wedge \phi, \psi \wedge \neg\phi).$$

3. DEFINABILITY RESULTS FOR MAJORITY QUANTIFIERS

In this section, we show that the k-ary second-order existential quantifier is definable in the logic FO(Most^k). We also show that, for $k > 1$, the quantifier $\mathcal{R}^k$ can be defined in FO(Most^k).

THEOREM 3.1. *The quantifier $\exists^2_k$ is definable in the logic* FO(Most^k).

Note that the proof of Theorem 3.1 below is the logical analogue of the proof of the inclusion NP $\subseteq$ PP [Gill 1977].

PROOF OF THEOREM 3.1. Suppose $\psi(Y)$ is a formula with a free k-ary predicate variable Y. The idea is to construct a formula that is satisfied by more than half of the relations if and only if $\psi(Y)$ is satisfiable. The problem is to find a uniform way to do this, that is, a way that does not depend on the particular formula $\psi(Y)$. We claim that

$$\models \exists^2_k Y\ \psi(Y) \Leftrightarrow \phi_1 \vee \phi_2,$$

where $\phi_1 = \psi(Y(\bar{x})/\top)$, that is, the formula ϕ_1 is defined by substituting $Y(\bar{x})$ by some formula satisfied by all k-tuples in every model, and

$$\phi_2 = \exists x_1 \cdots \exists x_k(\mathrm{Most}^k\, Y\ (Y(\bar{x}) \vee \psi(Y))).$$

Suppose that $\mathbb{M}$ is a model and $\mathbb{M} \models \exists^2_k Y\ \psi(Y)$. We have two cases to consider. Assume first that $\mathbb{M} \models \psi(M^k)$. Since the formula ϕ_1 expresses exactly this, it holds that $\mathbb{M} \models \phi_1$. Suppose then that $\mathbb{M} \models \psi(A)$ for some $A \subsetneq M^k$. Let $\overline{a} \in M^k$ be such that $\overline{a} \notin A$. Now

$$\mathbb{M} \models (Y(\bar{x}) \vee \psi(Y))(B, \overline{a})$$

holds for more than half of the relations $B \subseteq M^k$, where Y and $\overline{x}$ are interpreted as B and $\overline{a}$. This holds because the first disjunct is satisfied by exactly half of all the relations $B \subseteq M^k$, that is, those relations for which $\overline{a} \in B$, and, by the assumption, the relation A satisfies the second disjunct but not the first one. Therefore,

$$\mathbb{M} \models \phi_2.$$

On the other hand, it is obvious that if $\mathbb{M} \models \phi_1 \vee \phi_2$, then $\mathbb{M} \models \exists^2_k Y\ \psi(Y)$. □

Remark 3.2. It is worth noting that the quantifier $\exists^2_k$ can be trivially defined using the relativized quantifier $\mathrm{Most}^k_{\mathrm{r}}$:

$$\models \exists^2_k Y\ \psi(Y) \Leftrightarrow \mathrm{Most}^k_{\mathrm{r}}\, Y, Y\ (\psi(Y), \psi(Y)).$$

We shall next show that the quantifier $\mathcal{R}^k$ can be expressed in terms of the quantifier Most^{k+1}.

PROPOSITION 3.3. *Let $k \geq 2$. Then the quantifier $\mathcal{R}^{k-1}$ is definable in the logic* FO(Most^k).

PROOF. Given formulas $\psi_1(X)$ and $\psi_2(Y)$ with free $(k-1)$-ary predicate variables X and Y, we claim that there is a uniform way to express

$$\mathcal{R}^{k-1}\, X, Y\ (\psi_1(X), \psi_2(Y)). \tag{1}$$

Let $\mathbb{M}$ be a model satisfying $|M| \geq 2$. For $A \subseteq M^{k-1}$ and $b \in M$, set $A_b = \{(b, \overline{a}) \in M^k \mid \overline{a} \in A\}$. As in the proof of Theorem 3.1, we first define a collection C containing exactly half of the k-ary relations using a k-tuple $\overline{a} = (a_1, \ldots, a_k) \in M^k$:

$$C = \{A \subseteq M^k \mid \overline{a} \in A\}.$$

Let $G_i = \{A \subseteq M^{k-1} \mid \mathbb{M} \models \psi_i(A)\}$ for $1 \leq i \leq 2$. The condition $|G_1| > |G_2|$ is clearly equivalent with the condition

$$|(C \cup G_1^*) \setminus G_2^*| > 2^{|M|^k - 1},$$

where $G_1^* = \{A_b \mid A \in G_1\}$ and $G_2^* = \{\{\overline{a}\} \cup A_b \mid A \in G_2\}$ for some $b \in M$ such that $b \neq a_1$.

Formally, (1) can be expressed by the formula

$$\exists x_1 \cdots \exists x_k (x_1 \neq x_2 \wedge \varphi),$$

where $\varphi = \text{Most}^k\, Z\, ((Z(\overline{x}) \wedge \neg \chi_2) \vee \chi_1)$, and

$$\begin{aligned} \chi_1(Z) &= \forall \overline{z}(Z(\overline{z}) \rightarrow (z_1 = x_2)) \wedge \psi_1(X(\overline{y})/Z(x_2, \overline{y})), \\ \chi_2(Z) &= \forall \overline{z}(Z(\overline{z}) \rightarrow (z_1 = x_2 \vee \wedge_i z_i = x_i)) \wedge \psi_2(Y(\overline{y})/Z(x_2, \overline{y})). \end{aligned}$$

We assume that the variable x_2 does not appear in the formulas ψ_1 and ψ_2. □

The defining formulas above show that only one application of the quantifier Most^k is needed to express the quantifier $\mathcal{R}^{k-1}$. We need a bit more complicated argument to prove that the quantifier $\mathcal{R}^k$ can be defined in terms of the quantifier Most^k. The idea of the proof is adapted from Luosto [2004]. Note that we use more than one nested applications of the quantifier Most^k in expressing $\mathcal{R}^k$ in the proof of Theorem 3.4.

THEOREM 3.4. *Let $k \geq 2$. Then the quantifier $\mathcal{R}^k$ is definable in the logic* FO(Most^k).

PROOF. Suppose $\mathbb{M}$ is a model. We first existentially quantify a k-ary relation encoding an ordering $<$ over M, for example, a relation $A \subseteq M^k$ such that $\{(a, b) \mid \forall \overline{z} \in M^{k-2}((a, b, \overline{z}) \in A)\}$ is an ordering of M. Let $\delta(\overline{x}, \overline{y})$ be a formula defining the lexicographic order over M^k. It is now easy to construct a formula $\chi(X, Y)$ such that for $A, A' \subseteq M^k$, we have $\mathbb{M} \models \chi(A, A')$ iff $A < A'$ in the lexicographic ordering induced by $\delta(\overline{x}, \overline{y})$.

Now that we have a linear order of k-ary relations at our disposal, it is fairly straightforward to express

$$\mathcal{R}^k\, X, X\, (\psi_1(X), \psi_2(X)).$$

Let $G_i = \{A \subseteq M^k \mid \mathbb{M} \models \psi_i(A)\}$ for $1 \leq i \leq 2$. We may assume that $G_1 \cap G_2 = \emptyset$, since we can equivalently consider the formulas $\psi_1(X) \wedge \neg\psi_2(X)$ and $\psi_2(X) \wedge \neg\psi_1(X)$. Now, since only one of the sets G_i can satisfy $|G_i| > 2^{|M|^k - 1}$, the cases where either $|G_1| > 2^{|M|^k - 1}$ or $|G_2| > 2^{|M|^k - 1}$ can be directly taken care of using the quantifier Most^k. Hence, we may assume that $|G_1|, |G_2| \leq 2^{|M|^k - 1}$. In order to express $|G_1| > |G_2|$ using the quantifier Most^k, we consider collections $C_i(B)$ of relations of the form

$$G_i \cup \{A \subseteq M^k \mid A < B \text{ and } A \notin G_1 \cup G_2\}.$$

It is easy to construct a formula $\mu_i(X, Y)$ such that

$$C_i(B) = \{A \subseteq M^k \mid \mathbb{M} \models \mu_i(A, B)\}.$$

The condition $|G_1| > |G_2|$ can be now expressed by the formula φ

$$\varphi = \exists Y(\mathrm{Most}^k\, X\ \mu_1(X, Y) \wedge \neg\,\mathrm{Most}^k\, X\ \mu_2(X, Y)).$$

Finally, the defining formula is of the form

$$\gamma \vee \exists R(\text{“}R \text{ encodes an ordering”} \wedge \varphi),$$

where γ takes care of the situation in which either $|G_1| > 2^{|M|^k-1}$ or $|G_2| > 2^{|M|^k-1}$. By Theorem 3.1, we can express the above in the logic $\mathrm{FO}(\mathrm{Most}^k)$. □

4. THE CHARACTERIZATION OF CH IN TERMS OF MAJORITY QUANTIFIERS

In this section, we show that the extension of FO by the quantifiers Most^k, for $k \in \mathbb{N}^*$, strongly captures the counting hierarchy. We abbreviate $\{\mathrm{Most}^k \mid k \in \mathbb{N}^*\} = \mathrm{Most}$.

The following lemmas will be used in the proof. To be precise, we specify the pairing function used. We let (x, y) denote the string acquired by doubling the bits of x followed by the string $01y$. We shall next show that decoding the pair (x, y) can be done in a first-order way assuming we have $+$ and $\times$ available. Recall that $\mathcal{W}_x$ denotes the word structure determined by the binary word x.

Lemma 4.1. *Let $\tau = \{<, +, \times, P\}$, where P is unary, and let R be a r-ary predicate. Then, there is a* FO*-interpretation I of width $r+1$ mapping $\tau \cup \{R\}$ structures to τ-structures such that for all* $(\mathbb{M}, A)$

$$I((\mathbb{M}, A)) \cong \mathcal{W}_{(x,y_A)},$$

where y_A is the binary word corresponding to A and $x = \mathrm{bin}(\mathbb{M})$.

Proof. The interpretation I is defined by first-order $\tau \cup \{R\}$-formulas $\varphi_{dom}(\overline{x})$, $\varphi_P(\overline{x})$, $\varphi_<(\overline{x}_1, \overline{x}_2)$, $\varphi_+(\overline{x}_1, \overline{x}_2, \overline{x}_3)$, and $\varphi_\times(\overline{x}_1, \overline{x}_2, \overline{x}_3)$, where all the tuples $\overline{x}, \dots, \overline{x}_3$ are of length $r+1$, such that for all τ-structures $\mathbb{M}$ and all $A \subseteq M^r$ we have

$$\left(\varphi_{dom}^{(\mathbb{M},A)}, \varphi_P^{(\mathbb{M},A)}, \varphi_<^{(\mathbb{M},A)}, \varphi_+^{(\mathbb{M},A)}, \varphi_\times^{(\mathbb{M},A)}\right) \cong \mathcal{W}_{(x,y_A)}.$$

We assume that the domain of $\mathbb{M}$ is $\{0, \dots, n-1\}$ for some $n \in \mathbb{N}$. The domain of the structure on the left is defined as a set of $r+1$-tuples over $\{0, \dots, n-1\}$. We use the fact that the $(r+1)$-tuple versions of $+$ and $\times$ are first-order definable over $\mathbb{M}$ [Schweikardt 2005]. Let $<_{r+1}$ be the formula defining the lexicographic ordering over $\{0, \dots, n-1\}^{r+1}$. We write $\tilde{a}$ to denote a sequence $a \cdots a$, the length of which is clear from the context. We let $\varphi_{dom}(\overline{x})$ be the formula

$$(x_1, \dots, x_{r+1}) <_{r+1} (1, \tilde{0}, 2, 2).$$

Set $\varphi_P(\overline{x}) = \psi_1 \vee \psi_2 \vee \psi_3$, where

$$\begin{aligned}
\psi_1(\overline{x}) &= \exists y(P(y) \wedge (\overline{x} = (\tilde{0}, y) \times (\tilde{0}, 2) \vee \overline{x} = (\tilde{0}, y) \times (\tilde{0}, 2) + (\tilde{0}, 1))),\\
\psi_2(\overline{x}) &= \overline{x} = (\tilde{0}, 2, 1),\\
\psi_3(\overline{x}) &= R(\overline{x} - (\tilde{0}, 2, 2)).
\end{aligned}$$

Note that we have used definable constants in the formulas for readability. The other formulas can be simply defined by restricting the formulas over $\{0, \dots, n-1\}^{r+1}$ to tuples satisfying $\varphi_{dom}(\overline{x})$. □

The following proposition states the basic observation about interpretations.

PROPOSITION 4.2. *Let τ and R be as in Lemma* 4.1. *Then for any $\phi \in$* FO(Most)[τ] *there is $\phi^*(R) \in$* FO(Most)[τ], *where R is treated as a free second-order variable, such that for all $\mathbb{M}$ and $A \subseteq M^r$,*

$$\mathbb{M} \models \phi^*(A) \Leftrightarrow \mathcal{W}_{(x,y_A)} \models \phi.$$

PROOF. The formula $\phi^*(R)$ is defined from ϕ by replacing first-order variables by $r+1$-tuples and k-ary second-order variables by $k(r+1)$-ary variables, replacing the relation symbols of ϕ by the corresponding formulas, and restricting first-order quantifiers to $\varphi_{dom}(\overline{x})$. A formula of the form $\mathrm{Most}^k\, X\, \psi$ is translated simply as

$$\mathrm{Most}^{k(r+1)}\, X\, \psi^*.$$

Note that we do not need the relativized quantifier $\mathrm{Most}_{\mathrm{r}}^{k(r+1)}$ in the translation, since the formula ψ^* is already relativized to $\varphi_{dom}(\overline{x})$. In other words, the induction assumption is such that for all $B \subseteq M^{k(r+1)}$:

$$(\mathbb{M}, A) \models \psi^*(B) \Leftrightarrow I((\mathbb{M}, A)) \models \psi\left(B \cap \left(\varphi_{dom}^{(\mathbb{M},A)}\right)^k\right),$$

where B is interpreted as a k-ary relation on $r+1$-tuples on the right. □

Before going to the main result, we need to recall some concepts and properties of CH used in the proof. We say that a language L_1 is reducible to a language L_2 via a polynomial time disjunctive truth-table reduction if there exists a polynomial time computable function f mapping an input x to a polynomial number of inputs $y_1, \ldots, y_j$ such that $x \in L_1$ iff $y_i \in L_2$ for some $1 \leq i \leq j$.

LEMMA 4.3. *Let $k \in \mathbb{N}$. Then the following holds:*

(1) *The class C_kP is closed under complement.*
(2) *The class C_kP is closed under intersection.*
(3) *The class C_kP is closed under polynomial time disjunctive truth-table reductions.*

PROOF. Claim 4.3 is proved in Torán [1991]. Claim 2 is proved in Gupta [1990] and it follows from a relativization of Beigel, Reinhold, and Spielman's closure of PP under intersection [Beigel et al. 1995]. Since, for all oracles A, PP^A is closed under polynomial time disjunctive truth-table reductions [Beigel et al. 1995], Claim 3 follows. □

THEOREM 4.4. *The logic* FO(Most) *strongly captures* CH.

PROOF. We first show that FO(Most) $\subseteq$ CH, that is, we show that for all τ and for all $\varphi \in$ FO(Most)[τ], the language $L_{\mathrm{Mod}(\varphi)_<} = L_\varphi$, corresponding to the class $\mathrm{Mod}(\varphi)_<$, is contained in CH. We prove the claim using induction on φ. We treat formulas with free variables as sentences in an enlarged vocabulary. If φ is atomic, then $L_\varphi \in \mathrm{P} = C_0P$. Also, if $\varphi = \neg\psi$ or $\varphi = \psi \wedge \phi$, then the claim holds since C_kP is closed under complement and intersection by Lemma 4.3. Assume then that $\varphi = \exists x \psi$ and that $L_{\psi(c)} \in C_kP$. It is easy to see that L_φ is reducible to $L_{\psi(c)}$ via a polynomial time disjunctive truth-table reduction. Therefore, by Lemma 4.3, we have that $L_\varphi \in C_k P$. Let us then assume that

$$\varphi = \mathrm{Most}^k\, R\, \psi(R).$$

By induction hypothesis, $L_{\psi(R)} \in C_k P$ for some $k \in \mathbb{N}$. We show that $L_\varphi \in C_{k+1}P = \mathrm{PP}^{C_k P}$. In particular, the machine we shall describe uses $L_{\psi(R)}$ as an oracle. Let $\mathbb{M}$ be a structure. The machine N_φ, started with bin($\mathbb{M}$), guesses a word of length n^k, intended as the code of the interpretation A of R, and then consults the oracle whether bin$((\mathbb{M}, A)) \in L_{\psi(R)}$, that is, whether $\mathbb{M} \models \psi(A)$ holds. Then, the machine halts and accepts iff the oracle answered "yes". Now, by the definition of PP, the string bin($\mathbb{M}$) is accepted by N_φ iff more than half of the computations accept, that is, iff for more than half of the relations A the oracle answered positively. This is clearly equivalent with

$$\mathbb{M} \models \mathrm{Most}^k\, R\, \psi(R).$$

Let us then show that CH $\subseteq$ FO(Most). It suffices to prove the claim for binary word structures. We also expand our language by the numeric predicates $+$ and $\times$. Note that the predicates $+$ and $\times$ can be finally existentially quantified out, see Proposition 4.7. Let $\tau = \{<, +, \times, P\}$. We show that for all $L \subseteq \{1, 0\}^* \setminus \{\lambda\}$: $L \in$ CH iff $L = L_\varphi$ for some $\varphi \in$ FO(Most)$[\tau]$.

We prove using induction on k that $C_k P \subseteq$ FO(Most). The case $k = 1$ ($C_1 P =$ PP) is analogous to Fagin's Theorem on $\exists$SO and NP. The computation of a nondeterministic machine N, using time n^k for inputs of length n, can be coded using first-order formulas. In particular, let $\phi(X)$ say that "Relation X codes a n^k time-bounded run of N" and let $\psi(Y)$ say that "Relation Y codes a n^k time-bounded run which accepts", where the arity of X and Y is l. Since the formulas $\phi(X)$ and $\psi(Y)$ can be constructed so that each computation path on N corresponds to a unique l-ary relation on M, it is immediate that

$$\mathbb{M} \models \mathrm{Most}^l_{\mathrm{r}} X, Y\, (\phi(X), \psi(Y)) \Leftrightarrow N \text{ accepts bin}(\mathbb{M}).$$

By Theorem 3.4 and Example 2.5, the formula above can be expressed in FO(Most).

Assume then that $L \in C_{k+1}P = \mathcal{C}C_k P$. Then there is some language $L' \in C_k P$ and a polynomial p such that $x \in L$ iff

$$|\{y \;:\; |y| = p(|x|) \text{ and } (x, y) \in L'\}| > 2^{p(|x|)-1}.$$

By a simple padding argument, we may assume that $p(x) = x^d$ for some $d \in \mathbb{N}$. By induction hypothesis, there is $\phi \in$ FO(Most)$[\tau]$ such that $L_\phi = L'$. By Proposition 4.2, there is a formula $\phi^*(R) \in$ FO(Most)$[\tau]$, having a free d-ary predicate variable R, such that for all τ-structures $\mathbb{M}$ and $A \subseteq M^d$ we have

$$\mathbb{M} \models \phi^*(A) \Leftrightarrow \mathcal{W}_{(x, y_A)} \models \phi,$$

where $x =$ bin($\mathbb{M}$) and y_A is the word corresponding to A. It then follows that $L = L_\chi$, where $\chi = \mathrm{Most}^d\, R\, \phi^*(R)$. □

We denote by $qr(\varphi)$ the maximal nesting depth of the quantifiers Most^k in a formula $\varphi \in$ FO(Most). In particular, we are not taking account of first-order quantifiers. The proof of Theorem 4.4 shows that the level C_kP of a language L_φ in CH is determined entirely by the value $qr(\varphi)$. The first part of the proof shows that if $qr(\varphi) \leq k$, then $L_\varphi \in C_kP$. On the other hand, by Proposition 3.3, the quantifier $\mathrm{Most}^k_{\mathrm{r}}$ can be expressed using just one application of Most^{k+1}.

PROPOSITION 4.5. *Let $k \geq 1$ and $\tau = \{<, +, \times, P\}$. Then*

(1) $\mathrm{PP} = \{L_\varphi \mid \varphi \in \mathrm{FO(Most)}[\tau],\ qr(\varphi) \leq 1\}$,
(2) $C_kP = \{L_\varphi \mid \varphi \in \mathrm{FO(Most)}[\tau],\ qr(\varphi) \leq k\}$,
(3) $\mathrm{PH} \subseteq \{L_\varphi \mid \varphi \in \mathrm{FO(Most)}[\tau],\ qr(\varphi) \leq 2\}$.

PROOF. Case 1 follows directly from Proposition 3.3 and the proof of Theorem 4.4. Case 2 follows by induction on k: given ϕ defining the language $L' \in C_kP$, the formula $\phi^*(R)$ satisfies $qr(\phi) = qr(\phi^*)$ by Proposition 4.2, and hence the formula χ defining L satisfies $qr(\chi) = qr(\phi) + 1$. Case 3 follows by Toda's Theorem [Toda 1991]. □

Remark 4.6. By (3) of Proposition 4.5, every sentence $\varphi \in \mathrm{SO}[\tau]$ is equivalent to some $\psi \in \mathrm{FO(Most)}[\tau]$ having $qr(\psi) \leq 2$.

Proposition 4.5 holds also for arbitrary vocabularies τ assuming $\{<, +, \times\} \subseteq \tau$. Without ordering or the numeric predicates the following holds.

PROPOSITION 4.7. *Let $k \geq 1$ and τ a vocabulary. Then*

$$\{K \subseteq \mathrm{Str}(\tau) \mid L_{K_<} \in C_kP\} \subseteq \{\mathrm{Mod}(\varphi) \mid \varphi \in \mathrm{FO(Most)}[\tau],\ qr(\varphi) \leq k+3\}.$$

PROOF. The proof of this proposition is analogous to the proof of Theorem 7.5.14 in Ebbinghaus and Flum [1999] showing that ∃SO strongly captures NP. Suppose that K is class of τ-structures such that $L_{K_<} \in C_kP$. Then, by Proposition 4.5, there is φ over vocabulary $\{<, +, \times, P\}$ having $qr(\varphi) \leq k$ such that $L_\varphi = L_{K_<}$. Let $\mathbb{M}$ be a τ-structure. We can now first existentially quantify relations $<$, $+$, and $\times$ over $\mathbb{M}$ and then, using an FO-interpretation of $\mathcal{W}_{\mathrm{bin}(\mathbb{M})}$ in $\mathbb{M}$, write a formula ψ which evaluates φ in $\mathcal{W}_{\mathrm{bin}(\mathbb{M})}$. It is now easy to verify that the sentence

$$\chi = \exists < \exists + \exists \times \psi,$$

satisfies $\mathrm{Mod}(\chi) = K$ and that $qr(\chi) = k + 3$, since $\exists^2_k$ can be expressed using one application of Most^k and, as in Proposition 4.2, ψ can be constructed so that $qr(\psi) = qr(\varphi)$. □

5. GENERAL PROPORTIONAL QUANTIFIERS

On the complexity-theoretic side, it holds that, for any rational $0 < r < 1$,

$$\mathrm{PP}_r = \mathrm{PP},$$

where PP_r is defined by changing the input acceptance condition to "more than an r-fraction of accepting computations" (see Papadimitriou [1994]). It turns out that in the logical side this is not the case.

Definition 5.1. Let $0 < r < 1$ be a real number. The k-ary proportional quantifier $\mathcal{Q}_r^k$ is defined by the class

$$\mathcal{Q}_r^k = \{(M, P) : \ P \subseteq \mathcal{P}(M^k) \ \text{ and } |P| > r2^{|M|^k}\}.$$

Denote by FO($\mathcal{Q}_r$) the extension of FO by the quantifiers $\mathcal{Q}_r^k$ for $k \in \mathbb{N}^*$.

We shall show the following:

THEOREM 5.2. *Let $0 < r < 1$ be a real number. Then the following holds:*

(1) *If $r = s/2^m$ for some $s, m \in \mathbb{N}^*$, then the logic* FO($\mathcal{Q}_r$) *strongly captures the counting hierarchy.*

(2) *If r is not of the form $s/2^m$, then the logic* FO($\mathcal{Q}_r$) *satisfies the 0-1 law.*

Note that (2) implies that, for example, the class EVEN($\emptyset$) of all sets of even cardinality, which is computable in P $= C_0P$, cannot be axiomatized in the logic FO($\mathcal{Q}_r$). It also shows that definability results analogous to Theorem 3.1 and Proposition 3.3 do not hold in this case.

5.1 Claim 1 of Theorem 5.2

In this section, we show that Theorem 4.4 remains valid if the majority quantifiers are replaced by proportional quantifiers with threshold r of the form $s/2^m$.

Theorem 3.1 and Proposition 3.3 are based on the observation that we can easily define a set of relations containing exactly one half of the k-ary relations over any M. By fixing $\overline{a}_1, \ldots, \overline{a}_m \in M^k$, instead of just one tuple, we can divide $\mathcal{P}(M^k)$ into 2^m many disjoint sets S all having cardinality $2^{|M|^k - m}$. These sets can be indexed by binary words of length m, "1" in position j indicating that $\overline{a}_j \in A$ for all $A \in S$. The following is now easily obtained.

LEMMA 5.3. *Let $k, s, m \in \mathbb{N}^*$ and $s < 2^m$. Then*

(1) *The quantifier $\exists_k^2$ is definable in the logic* FO($\mathcal{Q}_{s/2^m}^k$).

(2) *The quantifier $\mathcal{R}^k$ is definable in the logic* FO($\mathcal{Q}_{s/2^m}^{k+1}$).

PROOF. The proof of Claim 1 is analogous to the proof of Theorem 3.1. In particular, the formula ϕ_1 in the proof of Theorem 3.1 is replaced by a formula with meaning

$$\psi(M^k) \vee \exists \overline{a}_1 \cdots \exists \overline{a}_m \psi(M^k \setminus \{\overline{a}_1, \ldots, \overline{a}_m\}).$$

Let us then turn to the proof of Claim 2. Given formulas $\psi_1(X)$ and $\psi_2(Y)$ with free k-ary predicate variables X and Y, we show how to express

$$\mathcal{R}^k X, Y\ (\psi_1(X), \psi_2(Y)), \qquad (2)$$

in terms of the quantifier $\mathcal{Q}_{s/2^m}^{k+1}$. Let $\overline{x}^1, \ldots, \overline{x}^m$ be $(k+1)$-tuples of pairwise distinct variables and Z a $(k+1)$-ary second-order variable. Denote by $\chi_j(Z)$, for $j \in \{0, 1\}^m$, the formula $\gamma_1 \wedge \cdots \wedge \gamma_m$, where $\gamma_i = Z(\overline{x}^i)$ if $j(i) = 1$ and $\gamma_i = \neg Z(\overline{x}^i)$ if $j(i) = 0$. Let $n(j)$ be the integer whose length m binary representation is given by $j(m)\cdots j(1)$. For all models $\mathbb{M}$ such that $|M| \geq m+1$, we can now express (2) by the formula

$$\exists w \exists \overline{x}_1 \cdots \exists \overline{x}_m \left(\varphi \wedge \bigwedge_{1 \leq i < j \leq m} x_1^i \neq x_1^j \wedge \bigwedge_{1 \leq i \leq m} x_1^i \neq w \right),$$

where $\varphi = \mathcal{Q}_{s/2^m}^{k+1} Z \left(\left(\left(\bigvee_{1 \leq n(j) \leq s} \chi_j(Z) \right) \wedge \neg \theta_2 \right) \vee \theta_1 \right)$, and

$$\begin{aligned} \theta_1(Z) &= \forall \overline{z}(Z(\overline{z}) \to z_1 = w) \wedge \psi_1(X(\overline{y})/Z(w, \overline{y})), \\ \theta_2(Z) &= \forall \overline{z}(Z(\overline{z}) \to (z_1 = w \vee \overline{z} = \overline{x}_1)) \wedge \psi_2(Y(\overline{y})/Z(w, \overline{y})). \end{aligned}$$

We assume that the variable w does not appear in the formulas ψ_1 and ψ_2. The case $|M| \leq m$ can be already expressed in a first-order way. □

We are now ready to prove Claim 1 of Theorem 5.2:

CLAIM 1 OF THEOREM 5.2. *Let* $s, m \in \mathbb{N}^*$ *and* $s < 2^m$. *Then* FO(Most) $\equiv$ FO($\mathcal{Q}_{s/2^m}$).

PROOF. By (2) of Lemma 5.3 and Example 2.5, the quantifiers Most^k can be expressed in the logic FO($\mathcal{Q}_{s/2^m}$). Therefore, we have that FO($\mathcal{Q}_{s/2^m}$) $\geq$ FO(Most). On the other hand, since $\text{PP}^A = \text{PP}_{s/2^m}^A$ for all A, the same argument as in the case of FO(Most) shows that FO($\mathcal{Q}_{s/2^m}$) $\subseteq$ CH. □

5.2 Claim 2 of Theorem 5.2

In this section, we show that the logic FO($\mathcal{Q}_r$) satisfies the 0-1 law if r is not of the form $s/2^m$. Glebskiĭ et al. [1969] and Fagin [1976] independently showed that first-order logic satisfies the 0-1 law. Since then, many extensions of FO have been shown to satisfy the 0-1 law. Our argument is based on a "almost sure" quantifier elimination result in the lines of Knyazev [1990] and Dawar and Grädel [1995]. In Knyazev [1990], the 0-1 law is shown to hold for a certain fragment of the extension of FO by first-order proportional quantifiers. In Dawar and Grädel [1995], the 0-1 law is established for the extension of first-order logic by the quantifier expressing rigidity.

We begin with some definitions and notation. In this section, we restrict attention to relational vocabularies.

For a class K of τ-structures, we write $\mu_n(K)$ for the fraction of τ-structures in K with universe $\{1, \ldots, n\}$. Define

$$\mu(K) = \lim_{n \to \infty} \mu_n(K),$$

if this limit exists. If $\mu(K) = a$, we say that K has asymptotic probability a. In the case $K = \text{Mod}(\varphi)$ for some sentence φ, we abbreviate $\mu(K)$ to $\mu(\varphi)$. We say that a logic $\mathcal{L}$ satisfies the 0-1 law if for every relational τ and every sentence $\varphi \in \mathcal{L}[\tau]$ we have that $\mu(\varphi)$ exists and

$$\mu(\varphi) \in \{0, 1\}.$$

Since we also consider formulas with free variables, we say that $\varphi(\overline{x})$ and $\theta(\overline{x})$ are almost everywhere equivalent if

$$\mu(\forall\overline{x}(\varphi(\overline{x}) \leftrightarrow \theta(\overline{x}))) = 1.$$

In this section, we treat free second-order variables as predicate symbols.

Example 5.4. Let $0 < r < 1$, and suppose that $\tau = \{R\}$, where R is k-ary, and $\varphi \in \mathrm{FO}[\tau]$ is a sentence. Then, the formula $\mathcal{Q}^k_r R\, \varphi$ is equivalent to a FO-sentence, since, by the 0-1 law of first-order logic, there is $n \in \mathbb{N}$ such that $\mu_m(\varphi) > r$ or $\mu_m(\varphi) < r$ for $m > n$. Therefore, in models of cardinality greater than n, the sentence $\mathcal{Q}^k_r R\, \varphi$ is equivalent with $\top$ if $\mu(\varphi) = 1$ and $\bot$ if $\mu(\varphi) = 0$.

The following lemma is essential for the result of this section.

LEMMA 5.5 ([GLEBSKIĬ ET AL. 1969]). *For every formula $\psi(\overline{x})$ of first-order logic, there is a quantifier-free formula $\theta(\overline{x})$ such that the sentence*

$$\forall\overline{x}(\psi(\overline{x}) \leftrightarrow \theta(\overline{x}))$$

has asymptotic probability 1.

Definition 5.6. An atomic type in variables $x_1, \ldots, x_k$ over τ is a maximal consistent set of atomic and negated atomic τ-formulas in the variables $x_1, \ldots, x_k$. We denote atomic types by t, s, or $t(\overline{x})$ to display the variables.

We do not distinguish between an atomic type and the conjunction over all formulas in it.

LEMMA 5.7. *Every quantifier-free formula is equivalent to a formula of the form*

$$\bigvee_i s_i(\overline{x}),$$

where $s_i(\overline{x})$ is an atomic type in the variables $x_1, \ldots, x_k$.

Definition 5.8. Let τ be a vocabulary, $R \notin \tau$ k-ary, $\varphi(\overline{x})$ a $\tau \cup \{R\}$-formula, $\mathbb{M}$ a τ-model over $\{1, \ldots, n\}$, and $\overline{a} \in M$. Denote by $F(\mathbb{M}, \overline{a}, \varphi(R))$ the fraction

$$|\{A \subseteq M^k \mid (\mathbb{M}, A) \models \varphi(\overline{a})\}|/2^{n^k}.$$

If φ does not have free first-order variables, we write $F(\mathbb{M}, \varphi(R))$.

The argument in Example 5.4 fails if τ is not of the form $\{R\}$. However, the following lemma can be used to show that in the general case $\mu(\mathcal{Q}^k_r R\, \varphi) = a$ if $\mu(\varphi) = a$, where $a \in \{0, 1\}$.

LEMMA 5.9. *Let τ be a vocabulary, $R \notin \tau$ k-ary, and let φ be a $\tau \cup \{R\}$-sentence such that $\mu(\varphi) = 1$ ($\mu(\varphi) = 0$). Then, for every $\epsilon > 0$, there is n_ϵ such that, for $n > n_\epsilon$, the fraction of τ-models $\mathbb{M}$ over domain $\{1, \ldots, n\}$ satisfying*

$$F(\mathbb{M}, \varphi(R)) > 1 - \epsilon \ (< \epsilon) \tag{3}$$

is greater than $1 - \epsilon$.

Proof. Suppose that there is $0 < \epsilon < 1$ for which there is no such n_ϵ. Let n' be such that $\mu_m(\varphi) > 1 - \epsilon^2$ for $m \geq n'$. By the assumption, the fraction of models $\mathbb{M}$ of size m for which (3) fails is at least ϵ. Therefore, we must have $\mu_m(\varphi) < 1 - \epsilon^2$, which is a contradiction. □

The following lemma gives us the inductive step for the quantifier elimination.

Lemma 5.10. *Let $0 < r < 1$ be a real number not of the form $s/2^m$ and let τ and R be as above. Suppose $\varphi(\overline{x})$ is a $\mathrm{FO}(\mathcal{Q}_r)[\tau \cup \{R\}]$-formula that is almost everywhere equivalent to a quantifier-free formula. Then, the formula $\mathcal{Q}_r^k R\, \varphi$ is almost everywhere equivalent to a quantifier-free formula.*

Proof. Suppose first that φ is a sentence, that is, it does not have free first-order variables. Then, we have that $\mu(\varphi) \in \{0, 1\}$. Without loss of generality, we may assume that $\mu(\varphi) = 1$. Let $\mathbb{M}$ be a τ-model. Now

$$\mathbb{M} \models \mathcal{Q}_r^k R\, \varphi \;\text{ iff }\; F(\mathbb{M}, \varphi(R)) > r.$$

Therefore, assuming $1 - \epsilon > r$, Lemma 5.9 implies that

$$\mu_m\big(\mathcal{Q}_r^k R\, \varphi\big) > 1 - \epsilon,$$

for $m > n_\epsilon$, and thus $\mu(\mathcal{Q}_r^k R\, \varphi) = 1$. Note that the above holds for every $0 < r < 1$.

Assume then that $\varphi(\overline{x})$ has free variables $x_1, \ldots, x_m$. Now, by Lemma 5.7, we can find a disjunction of atomic types

$$\bigvee_{1 \leq i \leq w} s_i(\overline{x}), \tag{4}$$

which is equivalent to φ almost everywhere. We assume that $s_i \neq s_j$ for $i \neq j$. Let $t_i(\overline{x})$ denote the reduct of $s_i(\overline{x})$ to a type in $\overline{x}$ over τ. The idea of the proof goes as follows. We shall first show that the formula

$$\mathcal{Q}_r^k R \bigvee_{1 \leq i \leq w} s_i(\overline{x})$$

is equivalent to the disjunction ψ of those types $t_i(\overline{x})$ for which

$$F\left(\mathbb{M}, \overline{a}, \bigvee_{1 \leq i \leq w} s_i(R)\right) > r,$$

where $\mathbb{M}$ and $\overline{a}$ satisfy $\mathbb{M} \models t_i(\overline{a})$, but are otherwise arbitrary. Then, using the assumption that φ and the formula in (4) are almost everywhere equivalent, it follows that $\mathcal{Q}_r^k R\, \varphi$ is almost everywhere equivalent with ψ.

Let $l_i \leq m$ denote the number of variables such that $s_i(\overline{x})$ forces their interpretations to be distinct, that is, l_i is the cardinality of a maximal set

$$\{y_1, \ldots, y_{l_i}\} \subseteq \{x_1, \ldots, x_m\}$$

such that $s_i \models y_k \neq y_j$ for $1 \leq k < j \leq l_i$. Since already $t_i(\overline{x})$ determines the identity formulas of $s_i(\overline{x})$, we have that $l_i = l_j$ if $t_i = t_j$. Since either $R(\overline{y}) \in s_i(\overline{x})$ or $\neg R(\overline{y}) \in s_i(\overline{x})$ for every $(y_1, \ldots, y_k) \in \{x_1, \ldots, x_m\}^k$, all interpretations of

R satisfying s_i agree with respect to l_i^k many tuples. On the other hand, the interpretation of R can be chosen arbitrarily outside the set of parameters, hence,

$$|\{A \subseteq M^k \mid (\mathbb{M}, A) \models s_i(\overline{a})\}| = 2^{n^k - l_i^k},$$

assuming $\mathbb{M} \models t_i(\overline{a})$ and n is the cardinality of $\mathbb{M}$. Over any τ-model $\mathbb{M}$ and $\overline{a} \in M$, the set

$$\left\{A \subseteq M^k \mid (\mathbb{M}, A) \models \bigvee_{1\leq i\leq w} s_i(\overline{a})\right\}$$

can be written as

$$\bigcup_{t_i = t} \{A \subseteq M^k \mid (\mathbb{M}, A) \models s_i(\overline{a})\}, \tag{5}$$

where t is the atomic type of $\overline{a}$ in $\mathbb{M}$. Let l denote the number of distinct parameters in $\overline{a}$. The sets of relations in (5) are pairwise disjoint and each of them has cardinality $2^{n^k - l^k}$. Therefore,

$$F\left(\mathbb{M}, \overline{a}, \bigvee_{1\leq i\leq w} s_i(R)\right) = z/2^{l^k},$$

where z is the number of types s_i such that $t_i = t$.

Consequently, the formula

$$\mathcal{Q}_r^k R \bigvee_{1\leq i\leq w} s_i(\overline{x})$$

is logically equivalent to the disjunction of those types $t_i(\overline{x})$ for which $z_i/2^{m_i} > r$, where $m_i = l_i^k$, and z_i is the number of extensions of t_i among $s_1, \ldots, s_w$. We let $t_{i_1}(\overline{x}), \ldots, t_{i_v}(\overline{x})$ enumerate these types without repetitions. We shall now show that $\mathcal{Q}_r^k R\, \varphi$ is almost everywhere equivalent with

$$\bigvee_{1\leq j\leq v} t_{i_j}(\overline{x}).$$

Let $\epsilon > 0$ be such that $\epsilon < r$ and $\epsilon < |z_i/2^{m_i} - r|$ for $1 \leq i \leq w$. By Lemma 5.9, there is n_ϵ such that the fraction of models $\mathbb{M}$ over $\{1, \ldots, m\}$ satisfying

$$F\left(\mathbb{M}, \forall\overline{x}\left(\varphi \leftrightarrow \bigvee_{1\leq i\leq w} s_i\right)(R)\right) > 1 - \epsilon \tag{6}$$

is greater than $1 - \epsilon$ for $m > n_\epsilon$. Suppose $\mathbb{M}$ is a τ-model satisfying (6) and $\overline{a} \in M$. Since, by (6),

$$|F(\mathbb{M}, \overline{a}, \varphi(R)) - F\left(\mathbb{M}, \overline{a}, \bigvee_{1\leq i\leq w} s_i(R)\right)| < \epsilon,$$

and

$$F\left(\mathbb{M}, \overline{a}, \bigvee_{1\leq i\leq w} s_i(R)\right) \in \{0, z_1/2^{m_1}, \ldots, z_w/2^{m_w}\},$$

we have, by the choice of ϵ,

$$\mathbb{M} \models \left(\mathcal{Q}_r^k R\, \varphi \leftrightarrow \mathcal{Q}_r^k R \bigvee_{1 \le i \le w} s_i \right) (\overline{a}).$$

Since the right-hand side is equivalent with $\bigvee_{1 \le j \le v} t_{i_j}(\overline{x})$ and $\overline{a}$ was arbitrary, we get

$$\mathbb{M} \models \forall \overline{x} \left(\mathcal{Q}_r^k R\, \varphi \leftrightarrow \bigvee_{1 \le j \le v} t_{i_j}(\overline{x}) \right). \tag{7}$$

We have shown that for $m > n_\epsilon$

$$\mu_m \left(\forall \overline{x} \left(\mathcal{Q}_r^k R\, \varphi \leftrightarrow \bigvee_{1 \le j \le v} t_{i_j}(\overline{x}) \right) \right) > 1 - \epsilon. \quad \square$$

By a repeated application of Lemma 5.10, we obtain the proof of Claim 2 of Theorem 5.2:

Claim 2 of Theorem 5.2. *Let $0 < r < 1$ not of the form $s/2^m$. Then every formula of the logic* FO($\mathcal{Q}_r$) *is almost everywhere equivalent to a quantifier-free formula. In particular, the logic* FO($\mathcal{Q}_r$) *satisfies the* 0-1 *law.*

Proof. We prove the claim using induction on φ. If φ is atomic, then the claim holds trivially. Also, assuming that the claim holds for ψ and χ, it follows easily for $\neg\psi$, $\psi \wedge \chi$, and for $\exists x \psi$ by Lemma 5.5. Suppose then that $\varphi = \mathcal{Q}_r^k Y\ \psi$. Now, the claim follows by Lemma 5.10 and the induction assumption. $\square$

Recall that every formula $\varphi \in$ SO is equivalent to a formula in prenex normal form, that is, to a formula of the form

$$Q_1 \alpha_1 \cdots Q_k \alpha_k \psi, \tag{8}$$

where $Q_1, \ldots, Q_k \in \{\forall, \exists\}$, and where $\alpha_1, \ldots, \alpha_k$ are first-order or second-order variables, and ψ is quantifier-free. In addition, we may assume that in (8) each second-order quantifier precedes all first-order quantifiers. Corollary 5.11 below implies that there is no such normal form in the case of FO(Most). For example, by Corollary 5.11 the class EVEN($\emptyset$) cannot be defined with a sentence of the form

$$\mathrm{Most}^{i_1} X_1 \cdots \mathrm{Most}^{i_j} X_j\, \varphi,$$

where $\varphi \in$ FO, but it can be defined in FO(Most) by Proposition 4.7.

Corollary 5.11. *Every sentence φ acquired by applying the connectives $\wedge$, $\neg$, and any proportional quantifiers $\mathcal{Q}_r^k$ ($k \in \mathbb{N}^*$ and $0 < r < 1$) to arbitrary first-order sentences has asymptotic probability* 0 *or* 1.

Proof. The claim is proved using induction on φ. It suffices to note that by Lemma 5.10, for any $\mathcal{Q}_r^k$, it holds that if $\mu(\varphi) = a$, where $a \in \{1, 0\}$, then $\mu(\mathcal{Q}_r^k R\, \varphi) = a$. $\square$

6. ON FO(Most^1)

In this section, we study the monadic fragment of the logic FO(Most). We show that monadic second-order logic (MSO) is strictly contained in FO(Most^1).

Theorem 3.1 implies that MSO $\leq$ FO(Most^1). Theorem 3.4 does not apply to FO(Most^1), since there is no obvious way to create an ordering over a model. On the other hand, the idea of the proof of Proposition 3.3 can be used to show the following.

PROPOSITION 6.1. *The first-order Rescher quantifier* R *is definable in the logic* FO(Most^1).

PROOF. We show that there is a sentence $\psi \in$ FO(Most^1) over vocabulary $\{P_1, P_2\}$, P_i unary, such that for all (M, P_1, P_2)

$$(M, P_1, P_2) \models \psi \Leftrightarrow |P_1| > |P_2|.$$

We abbreviate the formulas $P_1(x) \wedge \neg P_2(x)$ and $P_2(x) \wedge \neg P_1(x)$ by $Q_1(x)$ and $Q_2(x)$, respectively. We define ψ as follows:

$$\psi = \chi_1 \vee \chi_2,$$

where $\chi_1 = \exists x Q_1(x) \wedge \neg \exists x Q_2(x)$, and

$$\begin{aligned}
\chi_2 &= \exists x_1 \exists x_2 (Q_1(x_1) \wedge Q_2(x_2) \wedge \mathrm{Most}^1\, Y\, ((Y(x_1) \wedge \neg\chi_3) \vee \chi_4)),\\
\chi_3 &= \forall y (Y(y) \rightarrow (y = x_1 \vee (y \neq x_2 \wedge Q_2(y)))),\\
\chi_4 &= \forall y (Y(y) \rightarrow (y \neq x_1 \wedge Q_1(y))). \quad \square
\end{aligned}$$

Proposition 6.1 can be easily generalized to quantifiers $\mathcal{Q}_r^1$, where r is of the form $s/2^m$.

It is well known that MSO collapses to FO over vocabularies consisting of unary predicates only and that the quantifier R is not definable in FO.

COROLLARY 6.2. MSO $<$ FO(Most^1).

ACKNOWLEDGMENTS

I would like to thank Lauri Hella, Kerkko Luosto, and Hannu Niemistö for valuable comments and suggestions. I would also like to thank Hannu Niemistö for pointing out to me the proof of Theorem 3.4.

REFERENCES

ALLENDER, E., AND WAGNER, K. W. 1990. Counting hierarchies: Polynomial time and constant depth circuits. *Bull. EATCS 40*, 182–194.

ANDERSSON, A. 2002. On second-order generalized quantifiers and finite structures. *Ann. Pure Appl. Logic 115,* 1-3, 1–32.

BEIGEL, R., REINGOLD, N., AND SPIELMAN, D. 1995. PP is closed under intersection. *J. Comput. Syst. Sci. 50,* 2, 191–202.

BURTSCHICK, H.-J., AND VOLLMER, H. 1998. Lindström quantifiers and leaf language definability. *Int. J. Found. Comput. Sci. 9,* 3, 277–294.

DAWAR, A. AND GRÄDEL, E. 1995. Generalized quantifiers and 0-1 laws. In *Proceedings of the 10th Annual IEEE Symposium on Logic in Computer Science*. IEEE Computer Society Press, Los Alamitos, CA, 54–64.

Ebbinghaus, H.-D., and Flum, J. 1999. *Finite model theory,* 2nd edition. Perspectives in Mathematical Logic. Springer-Verlag, Berlin, Germany.

Fagin, R. 1974. Generalized first-order spectra and polynomial-time recognizable sets. In *Complexity of Computation*, American Mathematical Society, Providence, RI, 43–73.

Fagin, R. 1976. Probabilities on finite models. *J. Symb. Logic 41*, 1, 50–58.

Gill, J. 1977. Computational complexity of probabilistic Turing machines. *SIAM J. Comput. 6,* 4, 675–695.

Glebskiĭ, J. V., Kogan, D. I., Liogon′kiĭ, M. I., and Talanov, V. A. 1969. Range and degree of realizability of formulas in the restricted predicate calculus. *Cybernetics* 5, 142–154.

Gupta, S. 1990. A note on the counting hierarchy. Tech. Rep. OSU-CISRC-8/90-TR24, Computer and Information Research Center, Ohio State University.

Immerman, N. 1999. *Descriptive complexity*. Graduate Texts in Computer Science. Springer-Verlag, Berlin, Germany.

Knyazev, V. V. 1990. A zero-one law for an extension of the first-order predicate language. *Kibernetika (Kiev) 2*, 110–113.

Kontinen, J. 2004. Definability of second order generalized quantifiers. http://www.helsinki.fi/~jkontine/. (*Arch. Math. Logic*, to appear).

Kontinen, J. 2006. The hierarchy theorem for second order generalized quantifiers. *J. Symb. Logic 71,* 1, 188–202.

Lindström, P. 1966. First order predicate logic with generalized quantifiers. *Theoria 32*, 186–195.

Luosto, K. 2004. Equicardinality on linear orders. In *Proceedings of the 19th IEEE Symposium on Logic in Computer Science*. IEEE Computer Society Press, Los Alamitos, CA, 458–465.

Mostowski, A. 1957. On a generalization of quantifiers. *Fund. Math. 44*, 12–36.

Papadimitriou, C. H. 1994. *Computational complexity*. Addison-Wesley, Reading, MA.

Russo, D. A. 1985. Structural properties of complexity classes. Ph.D. dissertation. Univ. California at Santa Barbara, Santa Barbara, CA.

Schweikardt, N. 2005. Arithmetic, first order logic, and counting quantifiers. *ACM Trans. Comp. Logic 6,* 3, 634–671.

Simon, J. 1975. On some central problems in computational complexity. Ph.D. dissertation. Cornell Univ., Ithaca, NY.

Stockmeyer, L. J. 1976. The polynomial-time hierarchy. *Theoret. Comput. Sci. 3,* 1, 1–22.

Toda, S. 1991. PP is as hard as the polynomial-time hierarchy. *SIAM J. Comput. 20,* 5, 865–877.

Torán, J. 1991. Complexity classes defined by counting quantifiers. *J. ACM 38,* 3, 753–774.

Väänänen, J. 1999. Generalized quantifiers, an introduction. In *Generalized Quantifiers and Computation (Aix-en-Provence, 1997)*. Lecture Notes in Computer Science, vol. 1754. Springer-Verlag, Berlin, Germany, 1–17.

Vollmer, H. 1999. A generalized quantifier concept in computational complexity theory. In *Generalized Quantifiers and Computation (Aix-en-Provence, 1997)*. Lecture Notes in Computer Science, vol. 1754. Springer-Verlag, Berlin, Germany, 99–123.

Wagner, K. 1986. The complexity of combinatorial problems with succint input representation. *Acta Inf. 23*, 325–356.

Received June 2006; accepted June 2007