PROBABILITIES OF FIRST-ORDER SENTENCES
ABOUT UNARY FUNCTIONS

BY
JAMES F. LYNCH[1]

ABSTRACT. Let f be any fixed positive integer and σ a sentence in the first-order predicate calculus of f unary functions. For positive integers n, an n-structure is a model with universe $\{0, 1, \ldots, n-1\}$ and f unary functions, and $\mu(n, \sigma)$ is the ratio of the number of n-structures satisfying σ to n^{nf}, the number of n-structures. We show that $\lim_{n \to \infty} \mu(n, \sigma)$ exists for all such σ, and its value is given by an expression consisting of integer constants and the operators $+, -, \cdot, /,$ and e^x.

1. Introduction. Let f be a fixed positive integer and P a property that pertains to structures $\langle A, F_1, \ldots, F_f \rangle$, where A is a set and $F_1, \ldots, F_f$ are functions from A to A. If A is a finite set of cardinality n, we say that the probability of P on A is

$$\left| \left\{ \langle F_1, \ldots, F_f \rangle : P \text{ is true for } \langle A, F_1, \ldots, F_f \rangle \right\} \right| / n^{nf}.$$

That is, we regard $F_1, \ldots, F_f$ as random functions on A, where the n choices for $F_g(x)$ are equally likely for $x \in A$ and $1 \leqslant g \leqslant f$.

If P is isomorphism invariant, then this probability depends only on the cardinality of A, and we express it as $\mu(n, P)$. A number of papers have studied $\mu(n, P)$ for particular instances of P. N. Metropolis and S. Ulam [18] defined the notion of a random function and posed the problem of determining the expected number of components (for $f = 1$), conjecturing that it was $O(\log n)$. M. Kruskal [14] subsequently proved that it was $\log_e n + C + o(1)$, where $C = 0.5772\ldots$ is Euler's constant. L. Katz [13] found an expression for the probability that a random function is connected (i.e. has only one component), and showed that it was asymptotic to $(\pi/2n)^{1/2}$. B. Harris [12] computed the probability distributions of related quantities, such as component size and cycle size. Similar results have been found for some restricted classes of unary functions, for example random "hollow mappings", i.e. functions without fixed points, in [13] and random permutations in L. Shepp and S. Lloyd [21].

Some of the properties that these authors investigated are second-order. That is, they are expressible by a formal sentence which has variables that are interpreted as sets, as opposed to first-order sentences whose variables are all interpreted as individual elements. In fact, as H. Gaifman [9] and others have shown, the property

of connectedness for many kinds of finite binary relations, including unary functions, is not first-order expressible. Nevertheless, first-order sentences can express a wide variety of properties, for example, the properties of having a component or a cycle of fixed size. In this paper we study the asymptotic behavior of $\mu(n, \sigma)$, where σ is any sentence in the first-order predicate calculus of unary functions.

The first example of a limit law for probabilities of first-order sentences is the following, due to Y. Glebskii et al. [10]. Let σ be any first-order formula pertaining to structures $\langle A, R_1, \ldots, R_r \rangle$, where A is a set and each R_i is a d_i-ary relation on A. Let $d = \sum_{i=1}^{r} d_i$ and $x_1, \ldots, x_k$ be the free variables of σ. Then for $n \geq k$, $\nu(n, \sigma)$ is

$$\left| \{ \langle R_1, \ldots, R_r \rangle : \sigma \text{ is true for } \langle A, R_1, \ldots, R_r \rangle \} \right| / 2^{n^d},$$

where A is any fixed set of cardinality n and $x_1, \ldots, x_k$ are interpreted as any fixed distinct elements in A. Here we regard each R_i as a random relation on A, where the probability that $(y_1, \ldots, y_{d_i}) \in R_i$ is $1/2$ for $y_1, \ldots, y_{d_i} \in A$. Then

1.1 THEOREM (GLEBSKII ET AL.). (a) *For every first-order relational formula* σ, $\lim \nu(n, \sigma)$ *exists and is a dyadic rational, i.e. of the form* $u/2^v$ *for nonnegative integers* u *and* v.

(b) *If* σ *is a sentence, then* $\lim_{n \to \infty} \nu(n, \sigma)$ *is* 0 *or* 1.

It is not difficult to show that every dyadic rational in $[0, 1]$ is $\lim_{n \to \infty} \nu(n, \sigma)$ for some σ as in (a).

Part (b) of Theorem 1.1 was discovered independently by R. Fagin [6], who asked whether $\lim_{n \to \infty} \mu(n, \sigma)$ always exists if σ is a sentence in the first-order predicate calculus of one unary function. He gave an example showing that the limit need not be rational: let σ be $\forall x(F(x) \neq x)$, i.e. F has no fixed points. Then $\lim_{n \to \infty} \mu(n, \sigma)$ $= e^{-1}$. The main result of this paper is the following. Let Q_0 be the smallest set of real numbers containing 1 and satisfying the following. If $q \in Q_0$ and $j \in \omega$, then $q^j e^{-q}/j! \in Q_0$ and $1 - (\sum_{s \leq j} q^s/s!) e^{-q} \in Q_0$; if $q_0, q_1 \in Q_0$, then $q_0 q_1 \in Q_0$. Note that $Q_0 \subseteq (0, 1]$. Let Q consist of all reals in $[0, 1]$ which are sums of products of numbers of the form $(q/a)^j e^{-q/a}/j!$ or $1 - (\sum_{s \leq j} (q/a)^s/s!) e^{-q/a}$, where $q \in Q_0$, $1 \leq a \in \omega, j \in \omega$. We take the null sum to be 0 and the null product to be 1. Then

1.2 THEOREM. *Let* σ *be a sentence in the first-order predicate calculus of* f *unary functions. Then* $\lim_{n \to \infty} \mu(n, \sigma)$ *exists and is in* Q.

This extends a result announced in [16].

It is an open problem whether every element of Q is $\lim_{n \to \infty} \mu(n, \sigma)$ for some σ. Similar theorems apply to some closely related classes of finite structures.

1.3 THEOREM. *Let* σ *be a sentence with binary relation symbols* $F_1, \ldots, F_f$ *and*

$\mu_1(n, \sigma) = |\{ \langle F_1, \ldots, F_f \rangle : F_1, \ldots, F_f$ *are partial unary functions on* n *and* σ *is true for* $\langle n, F_1, \ldots, F_f \rangle \}| / (n+1)^{nf}$;

$\mu_2(n, \sigma) = |\{ \langle F_1, \ldots, F_f \rangle : F_1, \ldots, F_f$ *are unary functions on* n *with no fixed points and* σ *is true for* $\langle n, F_1, \ldots, F_f \rangle \}| / (n-1)^{nf}$;

$\mu_3(n, \sigma) = |\{ \langle F_1, \ldots, F_f \rangle : F_1, \ldots, F_f$ *are permutations on* n *and* σ *is true for* $\langle n, F_1, \ldots, F_f \rangle \}| / (n!)^f$.

Then for $i = 1, 2, 3$, $\lim_{n \to \infty} \mu_i(n, \sigma)$ *exists and is in* Q.

ADDED IN PROOF. A proof for the result on μ_1 can be derived from our proof of Theorem 1.2 by making minor changes to the counting arguments in §7. In fact, Theorem 1.2 and the result on μ_2 are immediate consequences of the result on μ_1. Also, Theorems 1.2 and 1.3 extend to the language with names for individual integers.

Recently, K. Compton (unpublished) proved

1.4 THEOREM. *Let σ be a sentence in the monadic second-order predicate calculus of one unary function. Then* $\lim_{n \to \infty} \sum_{m=0}^{n-1} \mu(m, \sigma)/n$ *exists.*

Theorem 1.4 does not characterize the possible values of the limit, and it is not known whether it can be extended to more than one unary function, nor is it known whether Theorem 1.2 can be extended to monadic second-order sentences.

While there are some similarities in the proofs of Theorems 1.2 and 1.4, in particular the use of the Ehrenfeucht game to characterize classes of indistinguishable structures, these theorems are representative of two different approaches to the study of asymptotic probabilities. The proof of Theorem 1.2 has the same general structure as the proofs of the limit laws in [15]. For $k \in \omega$ two technical conditions, k-richness and k-simplicity, are defined on the class of structures with f unary functions. k-richness corresponds roughly to k-extendibility of [15], and they are both somewhat involved. However, they are essentially modifications of an earlier notion used by Gaifman [8] to prove a 0-1 law for countable relational structures, by Reyes [20] to prove its topological analogue, and by Fagin [6] to prove the finite version (Theorem 1.1(b)).

An equivalence relation called k-agreeability is also defined. It corresponds to a simpler (unnamed) equivalence relation in [15], and is used in the same way: we show (Theorem 4.2) that there is a winning strategy for player II in the Ehrenfeucht game of length k on two k-simple and k-rich structures that are k-agreeable. Since there are only finitely many k-agreeability classes, the limit law follows when we show that almost all structures are k-simple (Theorem 4.3) and k-rich (Theorem 4.4), and that the probability of belonging to a given k-agreeability class approaches a limit for large structures (Theorem 4.5). However, the counting arguments used to calculate the probabilities in [15] are straightforward. Here, certain inductively defined formulas are used to characterize the k-agreeability classes. Then an extension of the principle of inclusion-exclusion is used to remove the quantifiers while inductively computing the probabilities of these formulas. This results in infinite series that converge to expressions involving the e^x operator. Infinite series also play a central role in Theorem 1.4 and related results of [2], but in a different way: they are generating series where the coefficient of x^n is determined by the number of structures of cardinality n. Also, the methods of [2] may be regarded as more general but less constructive than our methods, which give some idea of the values of the limits.

2. Model-theoretic definitions. For a more complete description of the notation and basic concepts of model theory, see Chang and Keisler [1]. Let A and B be sets. $^A B$ is the set of all functions from A to B. A partial function from A to B is a subset

of some member of ^AB. ω is the set of natural numbers, and we identify each $n \in \omega$ with the set $\{0, 1, \ldots, n-1\}$. Thus if $f, n \in \omega$, then $^f(^n n)$ is the set of all sequences $\langle F_1, \ldots, F_f \rangle$, where each F_g is a function from n to n.

Throughout the rest of this paper, $f \in \omega$ will be fixed. We shall consider models $\mathfrak{A}$ of the form $\langle A, F_1, \ldots, F_f, c_1, \ldots, c_k \rangle$, where A is a set (the universe of $\mathfrak{A}$), each F_g is a partial function from A to A, $k \in \omega$ and each c_j is an individual constant. Henceforth, *structure* will mean a model with f partial unary functions and finitely many constants. If all of the partial functions are actually functions, we will say that the structure is *functional*. We shall also consider substructures of $\mathfrak{A}$. If $B \subseteq A$ and $\{c_{j_1}, \ldots, c_{j_i}\} = \{c_1, \ldots, c_k\} \cap B$, then $\mathfrak{B} = \langle B, F_1 \cap {}^2B, \ldots, F_f \cap {}^2B, c_{j_1}, \ldots, c_{j_i} \rangle$ is the *substructure* of $\mathfrak{A}$ with universe B. We shall usually put F_g for $F_g \cap {}^2B$ in $\mathfrak{B}$ because the context will be clear.

The formal language we will use to express properties about structures is the first-order predicate calculus of f binary relations and arbitrarily many constants. Since the binary relations are always interpreted as partial functions, we will use functional instead of relational notation. That is, $x_0, x_1, \ldots$ are variables, $c_1, c_2, \ldots$ are constant symbols and the atomic formulas are $u = v$ and $F_g(u) = v$, where $1 \leqslant g \leqslant f$ and $u, v \in \{x_0, x_1, \ldots\} \cup \{c_1, c_2, \ldots\}$. Informally, we shall use x, y, z as variables, understanding that they are abbreviations for certain x_h, x_i, x_j. It will sometimes be convenient to write $F_g(u) = v$, where $-f \leqslant g \leqslant -1$. This will mean $F_{-g}(v) = u$. More complex formulas are constructed by applying the logical connectives $\neg, \vee, \wedge, \rightarrow, \leftrightarrow$ (not, or, and, implies, if and only if) and the quantifiers $\exists x_i$, $\forall x_i$ (there exists x_i, for all x_i) to the atomic formulas. The *depth* of a formula is defined inductively: atomic formulas have depth 0, if α and β have depths a and b respectively, then $\neg \alpha$ has depth a; $\alpha \vee \beta$, $\alpha \wedge \beta$, $\alpha \rightarrow \beta$ and $\alpha \leftrightarrow \beta$ have depth $\max(a, b)$; and $\exists x(\alpha)$ and $\forall x(\alpha)$ have depth $a + 1$. A sentence is a formula with no free variables. We put $\alpha(x_0, \ldots, x_m)$ for α a first-order formula with free variables $x_0, \ldots, x_m$. For $\mathfrak{A} = \langle A, F_1, \ldots, F_f, c_1, \ldots, c_k \rangle$ and $x_0, \ldots, x_m \in A$, $\mathfrak{A} \vDash \alpha(x_0, \ldots, x_m)$ means α is true in $\mathfrak{A}$ with each symbol x_i $(i \leqslant m)$, F_g $(1 \leqslant g \leqslant f)$, and c_j $(1 \leqslant j \leqslant k)$ interpreted as the corresponding object in $\mathfrak{A}$. If the interpretation of $x_0, \ldots, x_m$ is clear we often write $\mathfrak{A} \vDash \alpha$. $\alpha(x_0, \ldots, x_m)$ is *consistent* if there is a functional $\mathfrak{A}$ and $x_0, \ldots, x_m$ in $\mathfrak{A}$ such that $\mathfrak{A} \vDash \alpha(x_0, \ldots, x_m)$. $\alpha(x_0, \ldots, x_m)$ and $\beta(x_0, \ldots, x_m)$ are *equivalent* if for all functional $\mathfrak{A}$ and $x_0, \ldots, x_m$ in $\mathfrak{A}$, $\mathfrak{A} \vDash \alpha(x_0, \ldots, x_m) \leftrightarrow \beta(x_0, \ldots, x_m)$. In certain instances, the correspondence from symbol to object is ambiguous unless a distinction between them is made. Thus, if c_j is a constant symbol and $x \in A$, $c_j[x]$ is the constant c_j interpreted as x in $\mathfrak{A}$. We shall also use pairs of models $\mathfrak{A}^0$ and $\mathfrak{A}^1$, where for $p = 0, 1$ we put $\mathfrak{A}^p = \langle A^p, F_1^p, \ldots, F_f^p, c_1^p, \ldots, c_k^p \rangle$. In this case F_g $(1 \leqslant g \leqslant f)$ and c_j $(1 \leqslant j \leqslant k)$ are interpreted as F_g^p and c_j^p respectively.

Let α and β be sentences.

$$\mu(n, \alpha) = \left| \left\{ \langle F_1, \ldots, F_f \rangle \in {}^f(^n n) : \langle n, F_1, \ldots, F_f \rangle \vDash \alpha \right\} \right| / n^{nf}$$

and

$$\mu(n, \alpha | \beta) = \mu(n, \alpha) / \mu(n, \beta).$$

Of course, $\mu(n, \alpha)$ is also a conditional probability in a similar way: for any sentence γ with f binary relation symbols let

$$\xi(n, \gamma) = \left|\left\{\langle R_1,\ldots,R_f\rangle \in {}^f({}^{2^n}2): \langle n, R_1,\ldots,R_f\rangle \vDash \gamma\right\}\right|/(2^{n^2})^f.$$

Then if δ is the conjunction of the axioms of f unary functions, $\mu(n, \gamma) = \xi(n, \gamma|\delta)$.

For $p = 0, 1$ let $\mathfrak{A}^p = \langle A^p, F_1^p,\ldots,F_f^p\rangle$ be a structure, and let $k \in \omega$. The Ehrenfeucht game $\Gamma_k(\mathfrak{A}^0, \mathfrak{A}^1)$ is the following game of perfect information. It consists of k steps. For $1 \leqslant i \leqslant k$, step i consists of player I selecting an interpretation for c_i in either of the structures, say $\mathfrak{A}^{p_i}$, followed by player II selecting an interpretation for c_i in the other structure $\mathfrak{A}^{1-p_i}$. We shall also refer to step 0 as the initial condition of the game before step 1. Player II wins if

$$\left\langle \left\{c_1^0,\ldots,c_k^0\right\}, F_1^0,\ldots,F_f^0, \; c_1^0,\ldots,c_k^0\right\rangle$$

is isomorphic to

$$\left\langle \left\{c_1^1,\ldots,c_k^1\right\}, F_1^1,\ldots,F_f^1, c_1^1,\ldots,c_k^1\right\rangle.$$

The game and the following theorem can be extended to pairs of models of any finite type.

2.1 THEOREM (EHRENFEUCHT [4]). *Player II has a winning strategy for $\Gamma_k(\mathfrak{A}^0, \mathfrak{A}^1)$ if and only if $\mathfrak{A}^0$ and $\mathfrak{A}^1$ are indistinguishable by any sentence of depth k.*

3. Combinatorial definitions. Let $\mathfrak{A} = \langle A, F_1,\ldots,F_f, c_1,\ldots,c_k\rangle$ be a structure and $x, y \in A$. A *path* from x to y is a finite sequence $\langle x_0, g_0, x_1, g_1,\ldots,x_{m-1}, g_{m-1}, x_m\rangle$, where $x = x_0$, $y = x_m$, $x_i = x_j$ only if $i = j$ or $i, j \in \{0, m\}$, $g_i \in \{\pm 1,\ldots,\pm f\}$ and $\mathfrak{A} \vDash F_{g_i}(x_i) = x_{i+1}$ for $i < m$. Also, if $m = 2$ and $x = y$, then $g_0 \neq -g_1$.

The *length* of the path is m. If $m \geqslant 1$ and $x = y$, then the path is a *cycle* of size m and $\{x_0,\ldots,x_m\}$ is said to be a *cycle set*. $\delta(x, y)$ is the minimum path length from x to y. Putting $x \equiv y$ for $\delta(x, y) < \omega$, it is clear that $\equiv$ is an equivalence relation on A. Its equivalence classes are called *components* of $\mathfrak{A}$, and if $\mathfrak{A}$ has only one component it is said to be *connected*. It is also evident that δ is a metric on the components of $\mathfrak{A}$. We extend δ to subsets B, C of A: $\delta(x, B) = \delta(B, x) = \min\{\delta(x, y): y \in B\}$ and $\delta(B, C) = \min\{\delta(x, y): x \in B, y \in C\}$. We say that $\mathfrak{A}$ is a *cyclic* if it has no cycles.

The *center* of a component of $\mathfrak{A}$ is the minimal connected set containing the cycle sets and constants in the component, and the center of $\mathfrak{A}$ is the union of the centers of its components. It is easy to see that the center is unique. If Z is the center of $\mathfrak{A}$, the *radius* of $\mathfrak{A}$ is $\max\{\delta(Z, x): x \in A\}$. Therefore if $\mathfrak{A}$ has finite radius then every component has a nonempty center.

For $B \subseteq A$ and $r \in \omega$,

$$N(B, r) = \left\{x \in A: \delta(B, x) \leqslant r\right\}$$

and

$$\mathcal{N}(B, r) = \left\langle N(B, r), F_1,\ldots,F_f, c_{j_1},\ldots,c_{j_i}\right\rangle,$$

where $\{c_{j_1},\ldots,c_{j_i}\} = \{c_j: \delta(B, c_j) \leqslant r\}$.

Let $\mathfrak{B}$ be a structure with radius r. $\mathfrak{B}$ is said to be *realizable* if there is some functional structure $\mathfrak{A}$ with center Z such that $\mathfrak{B}$ is isomorphic to $\mathcal{N}(Z, r)$. Equivalently, for every x in $\mathfrak{B}$ whose distance from the center of $\mathfrak{B}$ is less than r and every $g = 1, \ldots, f$, $F_g(x)$ is in $\mathfrak{B}$.

4. Outline of proof of main theorem. The proof of Theorem 1.2 consists of three parts.

I. For each $j \in \omega$, an equivalence relation called *j-morphism* is defined on the class of structures with finite radius. In each j-morphism class, all members have isomorphic centers and equal radii. Thus, if P is an isomorphism invariant property and $r \in \omega$ we shall say that a j-morphism class has center with property P and radius r if each of its members do. Similarly, a j-morphism class is realizable if any (equivalently, all) of its members are.

It will be seen that there are only finitely many j-morphism classes with fixed (up to isomorphism) center and radius.

II. Let $1 \leqslant k \in \omega$ and $\mathfrak{A} = \langle A, F_1, \ldots, F_f \rangle$. $\mathfrak{A}$ is said to be *k-simple* if there are no cycle sets X_0, $X_1 \subseteq A$ such that $X_0 \neq X_1$, $|X_0|, |X_1| \leqslant 2 \cdot 3^{k-1} + 1$ and $\delta(X_0, X_1) \leqslant 2 \cdot 3^{k-1}$. $\mathfrak{A}$ is said to be *k-rich* if for every $i = 1, \ldots, k$, every realizable acyclic $(k - i)$-morphism class C with center $\{c_i\}$ and radius 3^{k-i} and every $x_1, \ldots, x_{i-1} \in A$, there is some $x \in A$ such that $\delta(x, \{x_1, \ldots, x_{i-1}\}) > 3^{k-i}$, $\delta(x, X) > 3^{k-i}$ for all cycle sets X of size $\leqslant 2 \cdot 3^{k-i} + 1$, and $\langle N(x, 3^{k-i}), F_1, \ldots, F_f, c_i[x] \rangle \in C$.

Let i, $j \in \omega$, and $D_0, \ldots, D_{d-1}$ be an enumeration of all j-morphism classes with a single cycle of size $\leqslant 2 \cdot 3^j + 1$, constants among $c_1, \ldots, c_i$ and radius $\leqslant 3^j$. For $p = 0, 1$ let $\mathfrak{A}^p = \langle A^p, F_1^p, \ldots, F_f^p, c_1^p, \ldots, c_i^p \rangle$. We put δ^p, N^p, $\mathcal{N}^p$ for δ, N, $\mathcal{N}$ in $\mathfrak{A}^p$. For $p = 0, 1$ and $a < d$ let

(4.1) $$S_a^p = \left\{ X : X \text{ is a cycle set in } \mathfrak{A}^p \text{ and } \mathcal{N}^p(X, 3^j) \in D_a \right\}.$$

Then $\mathfrak{A}^0$ is said to be *j-agreeable* to $\mathfrak{A}^1$ if and only if for all $a < d$,

$$|S_a^0| = |S_a^1| \quad \text{or} \quad |S_a^0|, |S_a^1| > j.$$

Clearly this is an equivalence relation with only finitely many (in fact $(j + 2)^d$) equivalence classes among structures with constants $c_1, \ldots, c_i$.

We will prove

4.2 Theorem. *Let $k \in \omega$ and $\mathfrak{A}^0$, $\mathfrak{A}^1$ be structures with no constants. If $\mathfrak{A}^0$ and $\mathfrak{A}^1$ are k-simple and k-rich and $\mathfrak{A}^0$ is k-agreeable to $\mathfrak{A}^1$, then $\Gamma_k(\mathfrak{A}^0, \mathfrak{A}^1)$ is a win for player II.*

III. Lastly, we prove the following for all $k \in \omega$.

4.3 Theorem. $\lim_{n \to \infty} \mu(n, \mathfrak{A} \text{ is } k\text{-simple}) = 1$.

4.4 Theorem. $\lim_{n \to \infty} \mu(n, \mathfrak{A} \text{ is } k\text{-rich}) = 1$.

4.5 Theorem. *For every k-agreeability class K on structures with no constants, there is $q \in Q$ such that $\lim_{n \to \infty} \mu(n, \mathfrak{A} \in K) = q$.*

Given these results the proof of Theorem 1.2 is straightforward. For fixed $k \in \omega$, let $R = \{\mathfrak{A} : \mathfrak{A} \text{ is } k\text{-simple and } k\text{-rich}\}$ and let $K_0, \ldots, K_{h-1}$ be the k-agreeability

classes on structures with no constants. Given a sentence σ of depth k, for any $i < h$ it follows by Theorems 4.2 and 2.1 that either $\mathfrak{A} \vDash \sigma$ for all $\mathfrak{A} \in R \cap K_i$ or $\mathfrak{A} \vDash \neg \sigma$ for all $\mathfrak{A} \in R \cap K_i$. Let $E = \{i < h: \mathfrak{A} \vDash \sigma$ for $\mathfrak{A} \in R \cap K_i\}$. Then

$$\mu(n, \sigma) = \mu(n, \sigma \wedge \mathfrak{A} \notin R) + \sum_{i \in E} \mu(n, \mathfrak{A} \in R \cap K_i).$$

Therefore

$$\lim_{n \to \infty} \mu(n, \sigma) = \sum_{i \in E} \lim_{n \to \infty} \mu(n, \mathfrak{A} \in K_i) \quad \text{by Theorems 4.3 and 4.4,}$$

$$\in Q \quad \text{by Theorem 4.5.}$$

5. Definition of j-morphism. The following notation will be used repeatedly in this and subsequent sections. Given $\mathfrak{A} = \langle A, F_1, \ldots, F_f, c_{j_1}, \ldots, c_{j_i} \rangle$ with center Z and $x \in A$,

(5.1) $$A_x = \{y \in A: \delta(Z, y) = \delta(Z, x) + \delta(x, y)\}$$

and

$$\mathfrak{A}_x = \langle A_x, F_1 - \{(x, x)\}, \ldots, F_f - \{(x, x)\}, c_1[x] \rangle.$$

Since $A_x \cap Z = \{x\} \cap Z$, $\mathfrak{A}_x$ is acyclic. Since $y \in A_x$ if and only if the unique path from Z to y includes x, $\mathfrak{A}_x$ is connected.

We first define j-morphism for pairs of acyclic connected structures with a single constant, say c_1. We use induction on r, the maximum of their radii. At each step of the induction it will be clear that among such structures with radius $\leqslant r$, isomorphism implies j-morphism, there are only finitely many j-morphism classes and in each class all structures have equal radii and either all of them are realizable or none of them are. To begin the induction, we define all pairs of such structures with $r = 0$ to be j-morphic.

Now assume $r > 0$ and that j-morphism has been defined for acyclic structures with center $\{c_1\}$ and radius $< r$. Let $C_0, \ldots, C_{c-1}$, where $c \in \omega$, be an enumeration of the j-morphism classes among such structures.'

For $p = 0, 1$ let $\mathfrak{A}^p = \langle A^p, F_1^p, \ldots, F_f^p, c_1^p \rangle$ be acyclic with radius $\leqslant r$. For each $x \in N^p(c_1^p, 1) - \{c_1^p\}$ let g_x^p be the unique $g \in \{\pm 1, \ldots, \pm f\}$ such that $F_g^p(c_1^p) = x$. Note that the radius of $\mathfrak{A}_x^p$ is less than r.

For $p = 0, 1$, $g \in \{\pm 1, \ldots, \pm f\}$ and $a < c$ let

(5.2) $$S_{ga}^p = \{x \in N^p(c_1^p, 1) - \{c_1^p\}: g_x^p = g \text{ and } \mathfrak{A}_x^p \in C_a\}.$$

Then we define $\mathfrak{A}^0$ to be j-morphic to $\mathfrak{A}^1$ if and only if for all such g and a,

$$\left|S_{ga}^0\right| = \left|S_{ga}^1\right| \quad \text{or} \quad \left|S_{ga}^0\right|, \left|S_{ga}^1\right| > j.$$

Clearly there are not more than $(2(j + 2))^{cf}$ j-morphism classes among acyclic structures with center $\{c_1\}$ and radius $\leqslant r$.

To extend j-morphism to all structures with finite radius, for $p = 0, 1$ let $\mathfrak{A}^p = \langle A^p, F_1^p, \ldots, F_f^p, c_{j_1}^p, \ldots, c_{j_i}^p \rangle$ with center Z^p. If there is an isomorphism θ from $\langle Z^0, F_1^0, \ldots, F_f^0, c_{j_1}^0, \ldots, c_{j_i}^0 \rangle$ onto $\langle Z^1, F_1^1, \ldots, F_f^1, c_{j_1}^1, \ldots, c_{j_i}^1 \rangle$ such that for each $x \in Z^0$, $\mathfrak{A}_x^0$ is j-morphic to $\mathfrak{A}_{\theta(x)}^1$, then we say that $\mathfrak{A}^0$ is j-morphic to $\mathfrak{A}^1$ via θ. It is

clear that for all $r \in \omega$, there are not more than $c^{|Z^0|}$ j-morphism classes among structures with center isomorphic to Z^0 and radius $\leqslant r$, where c is the number of j-morphism classes among acyclic structures with center $\{c_1\}$ and radius $\leqslant r$. Also, isomorphism implies j-morphism, and in each class all structures have equal radii and either all of them are realizable or none of them are.

6. Proof of Theorem 4.2.

6.1 LEMMA. *Let $\mathfrak{A}^0$ and $\mathfrak{A}^1$ be j-morphic with centers Z^0 and Z^1 respectively. Then for every $k \leqslant j$ and $s \in \omega$, $\mathcal{N}^0(Z^0, s)$ is k-morphic to $\mathcal{N}^1(Z^1, s)$.*

PROOF. We first prove the lemma for the case when $\mathfrak{A}^p = \left\langle A^p, F_1^p, \ldots, F_f^p, c_1^p \right\rangle$ is acyclic ($p = 0, 1$). Thus $Z^p = \{c_1^p\}$. We use induction on s. For $s = 0$, the result is immediate because $\mathcal{N}^0(Z^0, 0)$ and $\mathcal{N}^1(Z^1, 0)$ are isomorphic.

Now assume that $s > 0$ and the result holds for $0, \ldots, s - 1$. Let r be the radius of $\mathfrak{A}^p$ ($p = 0, 1$), let A_x^p and $\mathfrak{A}_x^p$ be as in (5.1) for $x \in A^p$, and let

$$B_x^p = A_x^p \cap N^p(Z^p, s)$$

and

$$\mathfrak{B}_x^p = \left\langle B_x^p, F_1^p - \{(x, x)\}, \ldots, F_f^p - \{(x, x)\}, c_1[x] \right\rangle$$

for $x \in N^p(Z^p, s)$. Note that the radius of $\mathfrak{B}_x^p$ is less than or equal to s, and is equal to s if and only if $x = c_1^p$ and $s \leqslant r$.

Let $C_0, \ldots, C_{c-1}$ (respectively $D_0, \ldots, D_{d-1}$) be an enumeration of all acyclic j-morphism (respectively k-morphism) classes with center $\{c_1\}$ and radius $< r$ (respectively s). Let S_{ga}^p be as in (5.2) for $p = 0, 1$, $g \in \{\pm 1, \ldots, \pm f\}$ and $a < c$, and let

$$T_{gb}^p = \left\{ x \in N^p(c_1^p, 1) - \{c_1^p\} : g_x^p = g \text{ and } \mathfrak{B}_x^p \in D_b \right\}$$

for $b < d$. Fixing g and b, it is clear that

$$T_{gb}^p \subseteq \bigcup_{a < c} S_{ga}^p.$$

By the induction assumption, for $a < c$,

$$S_{ga}^p \subseteq T_{gb}^p \quad \text{for } p = 0, 1 \quad \text{or} \quad S_{ga}^p \cap T_{gb}^p = \varnothing \quad \text{for } p = 0, 1.$$

Letting $E = \{a : S_{ga}^p \subseteq T_{gb}^p\}$,

$$\left| T_{gb}^p \right| = \sum_{a \in E} \left| S_{ga}^p \right| \quad \text{for } p = 0, 1.$$

Since $\mathfrak{A}^0$ and $\mathfrak{A}^1$ are j-morphic, for $a < c$,

$$\left| S_{ga}^0 \right| = \left| S_{ga}^1 \right| \quad \text{or} \quad \left| S_{ga}^0 \right|, \left| S_{ga}^1 \right| > j.$$

Therefore

$$\left| T_{gb}^0 \right| = \left| T_{gb}^1 \right| \quad \text{or} \quad \left| T_{gb}^0 \right|, \left| T_{gb}^1 \right| > k,$$

and it follows that $\mathcal{N}^0(Z^0, s)$ is k-morphic to $\mathcal{N}^1(Z^1, s)$.

To extend the proof to all structures, let $\mathfrak{A}^0$ and $\mathfrak{A}^1$ be j-morphic via θ. Then for each $x \in Z^0$, $\mathfrak{A}_x^0$ is j-morphic to $\mathfrak{A}_{\theta(x)}^1$. By what was shown above, $\mathfrak{B}_x^0$ is k-morphic to $\mathfrak{B}_{\theta(x)}^1$. Therefore $\mathcal{N}^0(Z^0, s)$ is k-morphic to $\mathcal{N}^1(Z^1, s)$ via θ. $\square$

6.2 LEMMA. *Let $\mathfrak{A}^0$ and $\mathfrak{A}^1$ be j-morphic via θ with centers Z^0 and Z^1 respectively. Then for every $s \in \omega$ and $X^0 \subseteq Z^0$ such that $\mathcal{N}^0(X^0, s)$ has finite radius, $\mathcal{N}^0(X^0, s)$ is j-morphic to $\mathcal{N}^1(\theta(X^0), s)$ via θ.*

PROOF. Let $X^1 = \theta(X^0)$ and let Y^p be the center of $\mathcal{N}^p(X^p, s)$ for $p = 0, 1$. Then $Y^p \subseteq Z^p$ and $Y^1 = \theta(Y^0)$. For $x, y \in N^p(X^p, s)$ let $\zeta^p(x, y)$ be the minimum path length in $\mathcal{N}^p(X^p, s)$ from x to y. For $x \in A^p$ let A_x^p and $\mathfrak{A}_x^p$ be as in (5.1), and for $x \in N^p(X^p, s)$ let

$$B_x^p = \{ y \in N^p(X^p, s) \colon \zeta^p(Y^p, y) = \zeta^p(Y^p, x) + \zeta^p(x, y) \},$$
$$\mathfrak{B}_x^p = \langle B_x^p, F_1^p - \{(x, x)\}, \ldots, F_f^p - \{(x, x)\}, c_1[x] \rangle,$$
$$\rho^p(x) = \max\{ \zeta^p(x, y) \colon y \in B_x^p \cap Z^p \}.$$

If we prove the following, we will be done.

CLAIM. For all $x \in N^0(X^0, s) \cap Z^0$, $\mathfrak{B}_x^0$ is j-morphic to $\mathfrak{B}_{\theta(x)}^1$.

We prove the Claim by induction on $\rho^0(x)$.

Fix $x \in N^0(X^0, s) \cap Z^0$. We will sometimes put x^0 for x and x^1 for $\theta(x)$. If $\rho^0(x) = 0$, then since θ is an isomorphism from Z^0 onto Z^1, $\rho^1(\theta(x)) = 0$. Therefore, for $p = 0, 1$,

$$B_{x^p}^p = A_{x^p}^p \cap N^p(x^p, s - \delta^p(X^p, x^p)).$$

Since $\mathfrak{A}_x^0$ is j-morphic to $\mathfrak{A}_{\theta(x)}^1$, it follows by Lemma 6.1 that $\mathfrak{B}_x^0$ is j-morphic to $\mathfrak{B}_{\theta(x)}^1$.

Now assume that $\rho^0(x) > 0$ and the Claim holds for all $y \in N^0(X^0, s) \cap Z^0$ such that $\rho^0(y) < \rho^0(x)$. If $\delta^0(X^0, x) = s$, then $B_{x^p}^p \subseteq Z^p$ for $p = 0, 1$, implying $\mathfrak{B}_x^0$ is isomorphic to $\mathfrak{B}_{\theta(x)}^1$. Thus let us assume $\delta^0(X^0, x) < s$. Let r be the maximum of the radii of $\mathfrak{A}_{x^p}^p$ and $\mathfrak{B}_{x^p}^p$ ($p = 0, 1$), and let $C_0, \ldots, C_{c-1}$ be an enumeration of all acyclic j-morphism classes with center $\{c_1\}$ and radius $< r$. For $p = 0, 1$, $g \in \{ \pm 1, \ldots, \pm f \}$ and $a < c$, let

$$S_{ga}^p = \{ y \in N^p(x^p, 1) \cap A_{x^p}^p - \{x^p\} \colon g_y^p = g \text{ and } \mathfrak{A}_y^p \in C_a \},$$
$$T_{ga}^p = \{ y \in N^p(x^p, 1) \cap B_{x^p}^p - \{x^p\} \colon g_y^p = g \text{ and } \mathfrak{B}_y^p \in C_a \},$$
$$U_{ga}^p = T_{ga}^p - Z^p, \qquad V_{ga}^p = T_{ga}^p \cap Z^p.$$

Since $\delta^p(X^p, x^p) < s$, it is easily seen that $N^p(x^p, 1) \cap B_{x^p}^p - Z^p = N^p(x^p, 1) \cap A_{x^p}^p - \{x^p\}$. Therefore, fixing g and $b < c$,

$$U_{gb}^p \subseteq \bigcup_{a < c} S_{ga}^p,$$

and by Lemma 6.1, for $a < c$,

$$S_{ga}^p \subseteq U_{gb}^p \quad \text{for } p = 0, 1 \quad \text{or} \quad S_{ga}^p \cap U_{gb}^p = \varnothing \quad \text{for } p = 0, 1.$$

Since $\mathfrak{A}_x^0$ is j-morphic to $\mathfrak{A}_{\theta(x)}^1$, for $a < c$,

$$\left| S_{ga}^0 \right| = \left| S_{ga}^1 \right| \quad \text{or} \quad \left| S_{ga}^0 \right|, \left| S_{ga}^1 \right| > j.$$

Therefore

$$\left| U_{gb}^0 \right| = \left| U_{gb}^1 \right| \quad \text{or} \quad \left| U_{gb}^0 \right|, \left| U_{gb}^1 \right| > j.$$

For every $y \in N^0(x, 1) \cap B_x^0 \cap Z^0 - \{x\}$, $\rho^0(y) \leqslant \rho^0(x) - 1$. Therefore by the induction assumption, $\mathfrak{B}_y^0$ is j-morphic to $\mathfrak{B}_{\theta(y)}^1$. Since θ is an isomorphism from Z^0 to Z^1, $g_y^0 = g_{\theta(y)}^1$. Therefore $\theta(V_{gb}^0) = V_{gb}^1$, and

$$\left| T_{gb}^0 \right| = \left| T_{gb}^1 \right| \quad \text{or} \quad \left| T_{gb}^0 \right|, \left| T_{gb}^1 \right| > j,$$

implying $\mathfrak{B}_x^0$ is j-morphic to $\mathfrak{B}_{\theta(x)}^1$. $\square$

6.3 LEMMA. *For $p = 0, 1$ let $\mathfrak{A}^p = \left\langle A^p, F_1^p, \ldots, F_f^p, c_1^p, \ldots, c_{i-1}^p \right\rangle$, where $1 \leqslant i \in \omega$. If $j > 0$ and $\mathfrak{A}^0$ is j-morphic to $\mathfrak{A}^1$, then for every $c_i^0 \in A^0$, there is $c_i^1 \in A^1$ such that $\left\langle A^0, F_1^0, \ldots, F_f^0, c_1^0, \ldots, c_i^0 \right\rangle$ is $(j - 1)$-morphic to $\left\langle A^1, F_1^1, \ldots, F_f^1, c_1^1, \ldots, c_i^1 \right\rangle$.*

PROOF. Let $\mathfrak{A}^p$ have center Z^p ($p = 0, 1$) and say $\mathfrak{A}^0$ is j-morphic to $\mathfrak{A}^1$ via θ. We use induction on $s = \delta^0(Z^0, c_i^0)$. If $s = 0$, let $c_i^1 = \theta(c_i^0)$. The lemma then follows immediately from Lemma 6.1.

Now assume that $s > 0$ and the lemma holds for $s - 1$. For $p = 0, 1$ and $x \in A^p$ let A_x^p and $\mathfrak{A}_x^p$ be as in (5.1). There exist unique $x^0 \in Z^0$, $y^0 \in A^0 - Z^0$ and $h \in \{\pm 1, \ldots, \pm f\}$ such that $c_i^0 \in A_{y^0}^0 \subseteq A_{x^0}^0$ and $F_h^0(x^0) = y^0$. Let $x^1 = \theta(x^0)$. Since $\mathfrak{A}_{x^0}^0$ is j-morphic to $\mathfrak{A}_{x^1}^1$, there is $y^1 \in A_{x^1}^1$ such that $F_h^1(x^1) = y^1$ and $\mathfrak{A}_{y^0}^0$ is j-morphic to $\mathfrak{A}_{y^1}^1$. Since $\delta^0(y^0, c_i^0) = s - 1$, by the induction assumption there is $c_i^1 \in A_{y^1}^1$ such that $\left\langle A_{y^0}^0, F_1^0, \ldots, F_f^0, c_1[y^0], c_i^0 \right\rangle$ and $\left\langle A_{y^1}^1, F_1^1, \ldots, F_f^1, c_1[y^1], c_i^1 \right\rangle$ are $(j - 1)$-morphic. (If $i = 1$, just use $c_2[y^p]$ for $c_1[y^p]$.)

For $p = 0, 1$ let $\mathfrak{B}^p = \left\langle A^p, F_1^p, \ldots, F_f^p, c_1^p, \ldots, c_i^p \right\rangle$. To prove that $\mathfrak{B}^0$ is $(j - 1)$-morphic to $\mathfrak{B}^1$, first note that the center X^p of $\left\langle A_{y^p}^p, F_1^p, \ldots, F_f^p, c_1[y^p], c_i^p \right\rangle$ is the set of points in the unique path from y^p to c_i^p, $Z^p \cap X^p = \varnothing$, and the center Y^p of $\mathfrak{B}^p$ is $Z^p \cup X^p$. Since X^0 is isomorphic to X^1 and $F_h^p(x^p) = y^p$ for $p = 0, 1$, θ can be extended to an isomorphism from Y^0 onto Y^1.

For $x \in A^p$ let

$$B_x^p = \left\{ y \in A^p \colon \delta^p(Y^p, y) = \delta^p(Y^p, x) + \delta^p(x, y) \right\},$$
$$\mathfrak{B}_x^p = \left\langle B_x^p, F_1^p - \{(x, x)\}, \ldots, F_f^p - \{(x, x)\}, c_1[x] \right\rangle.$$

We complete the proof by showing that for $x \in Y^0$, $\mathfrak{B}_x^0$ is $(j - 1)$-morphic to $\mathfrak{B}_{\theta(x)}^1$. There are three cases: $x \in Z^0 - \{x^0\}$, $x \in X^0$, and $x = x^0$. If $x \in Z^0 - \{x^0\}$, then $\mathfrak{B}_x^0 = \mathfrak{A}_x^0$ and $\mathfrak{B}_{\theta(x)}^1 = \mathfrak{A}_{\theta(x)}^1$. Now $\mathfrak{A}_x^0$ and $\mathfrak{A}_{\theta(x)}^1$ are $(j - 1)$-morphic by Lemma 6.1, and therefore so are $\mathfrak{B}_x^0$ and $\mathfrak{B}_{\theta(x)}^1$. If $x \in X^0$, then $\mathfrak{B}_x^0$ and $\mathfrak{B}_{\theta(x)}^1$ are $(j - 1)$-morphic because

$$\left\langle A_{y^0}^0, F_1^0, \ldots, F_f^0, c_1[y^0], c_i^0 \right\rangle \quad \text{and} \quad \left\langle A_{y^1}^1, F_1^1, \ldots, F_f^1, c_1[y^1], c_i^1 \right\rangle$$

are $(j - 1)$-morphic and $B_x^0 \subseteq A_{y^0}^0$, $B_{\theta(x)}^1 \subseteq A_{y^1}^1$.

If $x = x^0$, let r be the radius of $\mathfrak{A}_{x^p}^p$ ($p = 0, 1$), and let $C_0, \ldots, C_{c-1}$ (respectively $D_0, \ldots, D_{d-1}$) be an enumeration of all acyclic j-morphism (respectively $(j - 1)$-morphism) classes with center $\{c_1\}$ and radius $< r$. For $p = 0, 1$, $g \in \{\pm 1, \ldots, \pm f\}$ and $a < c$ let

$$S_{ga}^p = \left\{ y \in N^p(x^p, 1) \cap A_{x^p}^p - \{x^p\} \colon g_y^p = g \text{ and } \mathfrak{A}_y^p \in C_a \right\},$$

and for $b < d$ let

$$T_{gb}^p = \left\{ y \in N^p(x^p, 1) \cap B_{x^p}^p - \{x^p\} : g_y^p = g \text{ and } \mathfrak{B}_y^p \in D_b \right\}.$$

Note that $N^p(x^p, 1) \cap B_{x^p}^p = N^p(x^p, 1) \cap A_{x^p}^p - \{y^p\}$, and for $y \in N^p(x^p, 1) \cap B_{x^p}^p$, $\mathfrak{B}_y^p = \mathfrak{A}_y^p$ so the radius of $\mathfrak{B}_y^p$ is less than r.

Therefore, fixing g and b,

$$T_{gb}^p \subseteq \bigcup_{a < c} S_{ga}^p - \{y^p\},$$

and by Lemma 6.1, for $a < c$,

$$S_{ga}^p - \{y^p\} \subseteq T_{gb}^p \quad \text{for } p = 0, 1 \qquad \text{or} \qquad \left(S_{ga}^p - \{y^p\}\right) \cap T_{gb}^p = \varnothing \quad \text{for } p = 0, 1.$$

Let $k < c$ such that $y^p \in S_{hk}^p$ for $p = 0, 1$, and let $E = \{a : S_{ga}^p - \{y^p\} \subseteq T_{gb}^p\}$. If $g \neq h$ or $k \notin E$,

$$\left|T_{gb}^p\right| = \sum_{a \in E} \left|S_{ga}^p\right| \quad \text{for } p = 0, 1,$$

and if $g = h$ and $k \in E$,

$$\left|T_{gb}^p\right| = \sum_{a \in E} \left|S_{ga}^p\right| - 1.$$

Since $\mathfrak{A}_{x^0}^0$ and $\mathfrak{A}_{x^1}^1$ are j-morphic, for $a < c$,

$$\left|S_{ga}^0\right| = \left|S_{ga}^1\right| \quad \text{or} \quad \left|S_{ga}^0\right|, \left|S_{ga}^1\right| > j.$$

Therefore

$$\left|T_{gb}^0\right| = \left|T_{gb}^1\right| \quad \text{or} \quad \left|T_{gb}^0\right|, \left|T_{gb}^1\right| > j - 1,$$

and it follows that $\mathfrak{B}_x^0$ is $(j - 1)$-morphic to $\mathfrak{B}_{\theta(x)}^1$. $\square$

We now proceed with the proof of Theorem 4.2. Let $k \in \omega$ and $\mathfrak{A}^p = \left\langle A^p, F_1^p, \ldots, F_f^p \right\rangle$ for $p = 0, 1$. If $i \leqslant k$ and $c_1^p, \ldots, c_i^p \in A^p$ are constants chosen in the Ehrenfeucht game, we put $\mathfrak{A}^p\left\langle c_1^p, \ldots, c_i^p \right\rangle$ for $\left\langle A^p, F_1^p, \ldots, F_f^p, c_1^p, \ldots, c_i^p \right\rangle$, and for $x \in A^p$ and $r \in \omega$, we put $\mathcal{N}^p(x, r)\left\langle c_1^p, \ldots, c_i^p \right\rangle$ for $\mathcal{N}^p(x, r)$ in $\mathfrak{A}^p\left\langle c_1^p, \ldots, c_i^p \right\rangle$. Player II's strategy will be to choose so that at each step $i \leqslant k$, the following conditions hold:

(6.4)   For $1 \leqslant h \leqslant i$, $\mathcal{N}^0\left(c_h^0, 3^{k-i}\right)\left\langle c_1^0, \ldots, c_i^0 \right\rangle$ and $\mathcal{N}^1\left(c_h^1, 3^{k-i}\right)\left\langle c_1^1, \ldots, c_i^1 \right\rangle$ are $(k - i)$-morphic.

(6.5) $\mathfrak{A}^0\left\langle c_1^0, \ldots, c_i^0 \right\rangle$ and $\mathfrak{A}^1\left\langle c_1^1, \ldots, c_i^1 \right\rangle$ are $(k - i)$-agreeable.

If (6.4) holds for $i = k$, then player II has won. The proof that player II can choose so that (6.4) and (6.5) hold at each step i is by induction on i. For $i = 0$, (6.4) is vacuously true, and (6.5) holds by assumption.

Now let us assume that $i > 0$ and (6.4) and (6.5) hold for $i - 1$. By symmetry, say player I chooses $c_i^0 \in A^0$. There are three cases.

CASE I. $\delta^0(c_i^0, \{c_1^0, \ldots, c_{i-1}^0\}) > 3^{k-i}$ and $\delta^0(c_i^0, X) > 3^{k-i}$ for all cycle sets $X \subseteq A^0$ of size $\leqslant 2 \cdot 3^{k-i} + 1$. Then $\mathcal{N}^0(c_i^0, 3^{k-i})\left\langle c_1^0, \ldots, c_i^0 \right\rangle$ is acyclic with center $\{c_i^0\}$

and radius 3^{k-i}. Since $\mathfrak{A}^1$ is k-rich, there is $c_i^1 \in A^1$ such that $\delta^1(c_i^1, \{c_1^1, \ldots, c_{i-1}^1\}) > 3^{k-i}$, $\delta^1(c_i^1, X) > 3^{k-i}$ for all cycle sets $X \subseteq A^1$ of size $\leqslant 2 \cdot 3^{k-i} + 1$, and $\mathcal{N}^0(c_i^0, 3^{k-i})\langle c_1^0, \ldots, c_i^0 \rangle$ is $(k-i)$-morphic to $\mathcal{N}^1(c_i^1, 3^{k-i})\langle c_1^1, \ldots, c_i^1 \rangle$.

To show (6.4) holds, let $h < i$. By the induction hypothesis for (6.4) and Lemma 6.1 and Lemma 6.2, $\mathcal{N}^0(c_h^0, 3^{k-i})\langle c_1^0, \ldots, c_{i-1}^0 \rangle$ is $(k-i)$-morphic to $\mathcal{N}^1(c_h^1, 3^{k-i})\langle c_1^1, \ldots, c_{i-1}^1 \rangle$. Since $\delta^p(c_h^p, c_i^p) > 3^{k-i}$ for $p = 0, 1$,

$$\mathcal{N}^p\big(c_h^p, 3^{k-i}\big)\langle c_1^p, \ldots, c_i^p \rangle = \mathcal{N}^p\big(c_h^p, 3^{k-i}\big)\langle c_1^p, \ldots, c_{i-1}^p \rangle,$$

therefore $\mathcal{N}^0(c_h^0, 3^{k-i})\langle c_1^0, \ldots, c_i^0 \rangle$ is $(k-i)$-morphic to $\mathcal{N}^1(c_h^1, 3^{k-i})\langle c_1^1, \ldots, c_i^1 \rangle$.

To show (6.5) holds, let D be a $(k-i)$-morphism class with a single cycle of size $\leqslant 2 \cdot 3^{k-i} + 1$, constants among $c_1, \ldots, c_i$ and radius $\leqslant 3^{k-i}$, and for $p = 0, 1$

$$T^p = \big\{ X \colon X \text{ is a cycle set in } \mathfrak{A}^p \text{ and } \mathcal{N}^0(X, 3^{k-i})\langle c_1^p, \ldots, c_i^p \rangle \in D \big\}.$$

For all cycle sets X in $\mathfrak{A}^p$ of size $\leqslant 2 \cdot 3^{k-i} + 1$, $\mathcal{N}^p(X, 3^{k-i})\langle c_1^p, \ldots, c_i^p \rangle = \mathcal{N}^p(X, 3^{k-i})\langle c_1^p, \ldots, c_{i-1}^p \rangle$ because $\delta^p(X, c_i^p) > 3^{k-i}$. Therefore $T^p = \varnothing$ if c_i is a constant in D. If c_i is not a constant in D, let $C_0, \ldots, C_{c-1}$ be an enumeration of all $(k-i+1)$-morphism classes with a single cycle of size $\leqslant 2 \cdot 3^{k-i} + 1$, constants among $c_1, \ldots, c_{i-1}$ and radius $\leqslant 3^{k-i+1}$, and for $p = 0, 1$ and $a < c$,

$$S_a^p = \big\{ X \colon X \text{ is a cycle set in } \mathfrak{A}^p \text{ and } \mathcal{N}^p(X, 3^{k-i+1})\langle c_1^p, \ldots, c_{i-1}^p \rangle \in C_a \big\}.$$

Then

$$T^p \subseteq \bigcup_{a < c} S_a^p,$$

and by Lemmas 6.1 and 6.2, for $a < c$,

$$S_a^p \subseteq T^p \quad \text{for } p = 0, 1, \qquad \text{or} \qquad S_a^p \cap T^p = \varnothing \quad \text{for } p = 0, 1.$$

Letting $E = \{a \colon S_a^p \subseteq T^p\}$,

$$|T^p| = \sum_{a \in E} |S_a^p| \quad \text{for } p = 0, 1.$$

By the induction assumption for (6.5), for $a < c$,

$$|S_a^0| = |S_a^1| \quad \text{or} \quad |S_a^0|, |S_a^1| > k - i + 1.$$

Therefore

$$|T^0| = |T^1| \quad \text{or} \quad |T^0|, |T^1| > k - i,$$

which implies (6.5).

CASE II. $\delta^0(c_i^0, X^0) \leqslant 3^{k-i}$ for some cycle set $X^0 \subseteq A^0$ of size $\leqslant 2 \cdot 3^{k-i} + 1$. By the induction assumption for (6.5), there is $X^1 \subseteq A^1$ such that

$$\mathcal{N}^0\big(X^0, 3^{k-i+1}\big)\langle c_1^0, \ldots, c_{i-1}^0 \rangle$$

is $(k-i+1)$-morphic to $\mathcal{N}^1(X^1, 3^{k-i+1})\langle c_1^1, \ldots, c_{i-1}^1 \rangle$. By Lemma 6.3, there is $c_i^1 \in N^1(X^1, 3^{k-i+1})$ such that

$$\mathcal{N}^0\big(X^0, 3^{k-i+1}\big)\langle c_1^0, \ldots, c_i^0 \rangle$$

is $(k-i)$-morphic to $\mathcal{N}^1(X^1, 3^{k-i+1})\langle c_1^1, \ldots, c_i^1 \rangle$.

To show (6.4) holds, take $h \leqslant i$. First, suppose $\delta^0(X^0, c_h^0) > 2 \cdot 3^{k-i}$. Then $\delta^1(X^1, c_h^1) > 2 \cdot 3^{k-i}$. Therefore by the triangle inequality for δ^p ($p = 0, 1$), $\delta^p(c_h^p, c_i^p) > 3^{k-i}$. As in Case I, it then follows that $\mathcal{N}^0(c_h^0, 3^{k-i})\langle c_1^0, \ldots, c_i^0 \rangle$ is $(k-i)$-morphic to $\mathcal{N}^1(c_h^1, 3^{k-i})\langle c_1^1, \ldots, c_i^1 \rangle$.

Now suppose $\delta^0(X^0, c_h^0) \leqslant 2 \cdot 3^{k-i}$. Then $\delta^1(X^1, c_h^1) \leqslant 2 \cdot 3^{k-i}$ and, by the triangle inequality, $N^p(c_h^p, 3^{k-i}) \subseteq N^p(X^p, 3^{k-i+1})$ for $p = 0, 1$. Therefore by Lemma 6.2, $\mathcal{N}^0(c_h^0, 3^{k-i})\langle c_1^0, \ldots, c_i^0 \rangle$ is $(k-i)$-morphic to $\mathcal{N}^1(c_h^1, 3^{k-i})\langle c_1^1, \ldots, c_i^1 \rangle$.

To prove (6.5), let D and T^p ($p = 0, 1$) be as in Case I. If c_i is a constant in D, then since $\mathfrak{A}^p$ is k-simple, by Lemma 6.2

$$T^p = \varnothing \quad \text{for } p = 0, 1 \qquad \text{or} \qquad T^p = \{ X^p \} \quad \text{for } p = 0, 1.$$

If c_i is not a constant in D, let $C_0, \ldots, C_{c-1}$ and S_a^p ($p = 0, 1$, $a < c$) be as in Case I. Then

$$T^p \subseteq \bigcup_{a < c} S_a^p - \{ X^p \},$$

and by Lemmas 6.1 and 6.2, for $a < c$,

$$S_a^p - \{ X^p \} \subseteq T^p \quad \text{for } p = 0, 1 \qquad \text{or} \qquad \left(S_a^p - \{ X^p \} \right) \cap T^p = \varnothing \quad \text{for } p = 0, 1.$$

Let $h < c$ such that $X^p \in S_h^p$ for $p = 0, 1$, and $E = \{ a: S_a^p - \{ X^p \} \subseteq T^p \}$. If $h \notin E$,

$$|T^p| = \sum_{a \in E} |S_a^p| \quad \text{for } p = 0, 1$$

and if $h \in E$,

$$|T^p| = \sum_{a \in E} |S_a^p| - 1.$$

In either case,

$$|T^0| = |T^1| \quad \text{or} \quad |T^0|, |T^1| > k - i,$$

and (6.5) holds.

CASE III. $\delta^0(c_i^0, \{ c_1^0, \ldots, c_{i-1}^0 \}) \leqslant 3^{k-i}$ but $\delta^0(c_i^0, X) > 3^{k-i}$ for all cycle sets $X \subseteq A^0$ of size $\leqslant 2 \cdot 3^{k-i} + 1$. Let $1 \leqslant h < i$ such that $\delta^0(c_i^0, c_h^0) \leqslant 3^{k-i}$. By the induction assumption for (6.4), $\mathcal{N}^0(c_h^0, 3^{k-i+1})\langle c_1^0, \ldots, c_{i-1}^0 \rangle$ is $(k-i+1)$-morphic to $\mathcal{N}^1(c_h^1, 3^{k-i+1})\langle c_1^1, \ldots, c_{i-1}^1 \rangle$. By Lemma 6.3, there is $c_i^1 \in N^1(c_h^1, 3^{k-i+1})$ such that $\mathcal{N}^0(c_h^0, 3^{k-i+1})\langle c_1^0, \ldots, c_i^0 \rangle$ is $(k-i)$-morphic to $\mathcal{N}^1(c_h^1, 3^{k-i+1})\langle c_1^1, \ldots, c_i^1 \rangle$.

The proof of (6.4) is similar to the proof of (6.4) in Case II, and the proof of (6.5) is the same as the proof of (6.5) in Case I.

This completes the proof of Theorem 4.2.

7. Proof of Theorems 4.3, 4.4, 4.5.

7.1 PROOF OF THEOREM 4.3. We estimate $\mu(n, \mathfrak{A}$ is not k-simple) from above. Let K be the class of all connected structures of cardinality $\leqslant 2 \cdot 3^k + 1$ with at least two distinct cycles. Clearly there are only finitely many, say m, isomorphism classes in K. For $n \in \omega$ let $q(n) = \max\{ \mu(n, \mathfrak{A}$ has a substructure isomorphic to $\mathfrak{B})$: $\mathfrak{B} \in K \}$. Then $\mu(n, \mathfrak{A}$ is not k-simple$) \leqslant mq(n)$, and we need only show $\lim_{n \to \infty} q(n) = 0$.

For any $\mathfrak{B} = \langle B, G_1, \ldots, G_f \rangle \in K$, let $b = |B|$ and $c = \sum_{1 \leqslant g \leqslant f} |G_g|$. By a graph-theoretic argument (see Harary [11, p. 39]), $b + 1 \leqslant c$. Therefore

$$\mu(n, \mathfrak{A} \text{ has a substructure isomorphic to } \mathfrak{B}) \leqslant n(n-1) \cdots (n-b+1)/n^c < n^{-1},$$

and it follows that $\lim_{n \to \infty} q(n) = 0$. $\square$

The main tool for proving Theorems 4.4 and 4.5 is Lemma 7.10 below. It in turn depends on the next sequence of definitions and lemmas.

The first two lemmas are generalizations of the principle of inclusion-exclusion and Bonferroni's inequalities (see Feller [7, pp. 106–111]). Their proofs were suggested by the referee. Let G, H, I be finite sets. For each $(h, i) \in H \times I$ let $P(h, i) \subseteq G$. In the usual terminology G is said to be a collection of objects and each $P(h, i)$ is a property of some of the objects in G. For $S \subseteq H \times I$ and $i \in I$ let

$$S(i) = S \cap H \times \{i\}, \quad E^{\geqslant}(S) = \bigcap_{(h,i) \in S} P(h, i),$$

$$E^{=}(S) = E^{\geqslant}(S) - \bigcup_{(h,i) \in H \times I - S} P(h, i).$$

That is, $E^{\geqslant}(S)$ is the set of objects that have all the properties in S, and $E^{=}(S)$ is the set of objects that have all the properties in S and none of the properties in $H \times I - S$.

For $s \in {}^I(|H| + 1)$ let $L(s) = \sum_S |E^{\geqslant}(S)|$, where the sum is taken over all S such that $|S(i)| = s(i)$, $i \in I$, and for $J \subseteq I$ let $M(J, s) = \bigcup_S E^{=}(S)$, where the union is taken over all S such that $|S(i)| = s(i)$ for $i \in J$ and $|S(i)| \geqslant s(i)$ for $i \in I - J$. Thus $M(J, s)$ is the set of objects with exactly $s(i)$ properties in $H \times \{i\}$ for $i \in J$ and at least $s(i)$ properties in $H \times \{i\}$ for $i \in I - J$. Lastly, for $s, t \in {}^I(|H| + 1)$ we put $s \leqslant t$ if $s(i) \leqslant t(i)$ for $i \in I$.

7.2 LEMMA. *Let $s \in {}^I(|H| + 1)$ and $J \subseteq I$ such that $s(i) > 0$ for $i \in I - J$. Then*

$$|M(J, s)| = \sum_{\substack{t \in {}^I(|H|+1) \\ t \geqslant s}} \prod_{i \in J} (-1)^{t(i)-s(i)} \binom{t(i)}{s(i)}$$

$$\times \prod_{i \in I - J} (-1)^{t(i)-s(i)} \binom{t(i)-1}{s(i)-1} \times L(t).$$

PROOF. For all $S \subseteq H \times I$, $E^{\geqslant}(S) = \bigcup_{T \supseteq S} E^{=}(T)$. Since the sets $E^{=}(T)$ are disjoint,

$$|E^{\geqslant}(S)| = \sum_{T \supseteq S} |E^{=}(T)|,$$

and so by Möbius inversion

$$|E^{=}(S)| = \sum_{T \supseteq S} (-1)^{|T|-|S|} |E^{\geqslant}(T)|.$$

Now

$$|M(J,s)| = \sum_S |E^=(S)| = \sum_S \sum_{T \supseteq S} (-1)^{|T|-|S|}|E^{\geqslant}(T)|,$$

where the sums are taken over all S such that $|S(i)| = s(i)$ for $i \in J$ and $|S(i)| \geqslant s(i)$ for $i \in I - J$. Reversing the order of summation gives

$$|M(J,s)| = \sum_T |E^{\geqslant}(T)| \sum_{S \subseteq T} (-1)^{|T|-|S|},$$

where T ranges over all sets such that $|T(i)| \geqslant s(i)$, $i \in I$, and S is restricted as above.

Thus it remains to evaluate

$$\sum_{S \subseteq T} (-1)^{|T|-|S|} = \left(\prod_{i \in J} \sum_{\substack{U \subseteq T(i) \\ |U| = s(i)}} (-1)^{|T(i)|-|U|} \right)\left(\prod_{i \in I-J} \sum_{\substack{U \subseteq T(i) \\ |U| \geqslant s(i)}} (-1)^{|T(i)|-|U|} \right)$$

$$= \prod_{i \in J} (-1)^{t(i)-s(i)}\binom{t(i)}{s(i)} \prod_{i \in I-J} (-1)^{t(i)-s(i)}\binom{t(i)-1}{s(i)-1},$$

where $|T(i)| = t(i)$, $i \in I$. This proves the lemma. $\square$

7.3 LEMMA. *Let s and J be as in Lemma 7.2. For $t \in {}^I(|H|+1)$ we put Σt for $\sum_{i \in I} t(i)$. Then for every $v \geqslant \Sigma s$*

$$\sum_{\substack{t \in {}^I(|H|+1) \\ t \geqslant s \\ \Sigma t \geqslant v}} (-1)^{\Sigma t - v} \prod_{i \in J} \binom{t(i)}{s(i)} \times \prod_{i \in I-J} \binom{t(i)-1}{s(i)-1} \times L(t) \geqslant 0.$$

PROOF. Note that for $t \in {}^I(|H|+1)$ such that $t(i) > 0$ for $i \in I - J$,

$$L(t) = \sum_{\substack{u \in {}^I(|H|+1) \\ u \geqslant t}} \prod_{i \in J} \binom{u(i)}{t(i)} \prod_{i \in I-J} \binom{u(i)-1}{t(i)-1} \times |M(J,u)|.$$

To prove this, we regard the equations in the statement of Lemma 7.2, where s varies and J is fixed, as a system of $(|H|+1)^{|J|}|H_i|^{|I-J|}$ linear equations with the same number of unknowns $L(t)$, where t varies. Thus by showing that the above formula is a solution to the system of equations, it follows that it is the unique solution.

Substituting the above formula for $L(t)$ in the equation of Lemma 7.2 and reversing the order of summation gives

$$|M(J,s)| = \sum_{\substack{u \in {}^I(|H|+1) \\ u \geqslant s}} |M(J,u)| \sum_{\substack{t \in {}^I(|H|+1) \\ s \leqslant t \leqslant u}} \prod_{i \in J} (-1)^{t(i)-s(i)}\binom{t(i)}{s(i)}\binom{u(i)}{t(i)}$$

$$\times \prod_{i \in I-J} (-1)^{t(i)-s(i)}\binom{t(i)-1}{s(i)-1}\binom{u(i)-1}{t(i)-1}.$$

The second sum on the right equals

$$\prod_{i \in J}\left[\sum_{s(i) \leqslant w \leqslant u(i)}(-1)^{w-s(i)}\binom{w}{s(i)}\binom{u(i)}{w}\right]$$

$$\times \prod_{i \in I-J}\left[\sum_{s(i) \leqslant w \leqslant u(i)}(-1)^{w-s(i)}\binom{w-1}{s(i)-1}\binom{w(i)-1}{w-1}\right]$$

$$=\prod_{i \in J}\left[\binom{u(i)}{s(i)}\sum_{s(i) \leqslant w \leqslant u(i)}(-1)^{w-s(i)}\binom{u(i)-s(i)}{w-s(i)}\right]$$

$$\times \prod_{i \in I-J}\left[\binom{u(i)-1}{s(i)-1}\sum_{s(i) \leqslant w \leqslant u(i)}(-1)^{w-s(i)}\binom{u(i)-s(i)}{w-s(i)}\right]$$

$$=\begin{cases}1 & \text{if } u=s, \\ 0 & \text{if } u>s.\end{cases}$$

Thus we get the identity $|M(J, s)| = |M(J, s)|$, proving that the formula for $L(t)$ is a solution.

To complete the proof of the lemma, by substitution of the formula for $L(t)$ and reversing the order of summation

$$\sum_{\substack{t \in {}^I(|H|+1) \\ t \geqslant s \\ \Sigma t \geqslant v}}(-1)^{\Sigma t-v}\prod_{i \in J}\binom{t(i)}{s(i)} \times \prod_{i \in I-J}\binom{t(i)-1}{s(i)-1} \times L(t)$$

$$=\sum_{\substack{u \in {}^I(|H|+1) \\ u \geqslant s \\ \Sigma u \geqslant v}}|M(J, u)|\sum_{\substack{t \in {}^I(|H|+1) \\ s \leqslant t \leqslant u \\ \Sigma t \geqslant v}}(-1)^{\Sigma t-v}\prod_{i \in J}\binom{t(i)}{s(i)}\binom{u(i)}{t(i)}$$

$$\times \prod_{i \in I-J}\binom{t(i)-1}{s(i)-1}\binom{u(i)-1}{t(i)-1}.$$

The second sum on the right equals

$$\prod_{i \in J}\binom{u(i)}{s(i)}\prod_{i \in I-J}\binom{u(i)-1}{s(i)-1}\sum_{v \leqslant w \leqslant \Sigma u}(-1)^{w-v}\binom{\Sigma u-\Sigma s}{w-\Sigma s},$$

and this last sum equals 1 if $u = s$ and equals $\binom{\Sigma u-\Sigma s-1}{v-\Sigma s-1} \geqslant 0$ if $u > s$. $\square$

7.4 LEMMA. *Let* $u \in {}^{\omega \times \omega}[0, \infty)$ *and* $\langle u_m: m \in \omega \rangle$ *be a sequence in* $[0, \infty)$ *satisfying the following*:

(7.5) $\displaystyle\sum_{m \in \omega}(-1)^m u(n, m)$ *converges for all* $n \in \omega$.

(7.6) $\displaystyle\sum_{m \in \omega}(-1)^m u_m$ *converges*.

(7.7) $\displaystyle\lim_{n \to \infty}u(n, m) = u_m$ *for all* $m \in \omega$.

(7.8) $\displaystyle\sum_{m \geqslant v}(-1)^{m-v}u(n, m) \geqslant 0$ *for all* $n, v \in \omega$.

Then

$$\lim_{n \to \infty}\sum_{m \in \omega}(-1)^m u(n, m) = \sum_{m \in \omega}(-1)^m u_m.$$

PROOF. Let $\varepsilon > 0$. By (7.6) there is $M \in \omega$ such that

$$\left| \sum_{m < M} (-1)^m u_m - \sum_{m \in \omega} (-1)^m u_m \right| < \varepsilon/3 \quad \text{and} \quad u_M < \varepsilon/3.$$

By (7.7) there is $N \in \omega$ such that $|u(n, m) - u_m| < \varepsilon/3(M + 1)$ for $m \leqslant M$ and $n \geqslant N$. Then for all $n \geqslant N$,

$$\left| \sum_{m \in \omega} (-1)^m u(n, m) - \sum_{m \in \omega} (-1)^m u_m \right|$$

$$\leqslant \left| \sum_{m \in \omega} (-1)^m u(n, m) - \sum_{m < M} (-1)^m u(n, m) \right|$$

$$+ \left| \sum_{m < M} (-1)^m u(n, m) - \sum_{m < M} (-1)^m u_m \right|$$

$$+ \left| \sum_{m < M} (-1)^m u_m - \sum_{m \in \omega} (-1)^m u_m \right|$$

$$\leqslant u(n, M) \, [\text{by (7.8)}] + \sum_{m < M} |u(n, m) - u_m| + \varepsilon/3$$

$$\leqslant \varepsilon/3 + \varepsilon/3(M + 1) + \varepsilon M/3(M + 1) + \varepsilon/3 = \varepsilon. \quad \square$$

The following notation and definitions will be used in Lemmas 7.9 and 7.10 below.

Let $j \in \omega$ be fixed and $C_0, C_1, \ldots$ be an enumeration of all acyclic j-morphism classes with center $\{c_1\}$ such that the radius of $C_d \leqslant$ the radius of C_i for all $d \leqslant i$. For each $d \in \omega$, we define a certain $q_d \in Q_0$, using induction on the radius of C_d. $q_0 = 1$. Let $r > 0$, $C_0, \ldots, C_{c-1}$ be all the classes with radius $< r$, and let C_d have radius r. Assume $q_0, \ldots, q_{c-1}$ have been defined. Now C_d is characterized by a sequence $\langle s_{dga} : g \in \{\pm 1, \ldots, \pm f\}, a < c \rangle$ of elements in $j + 2$ in the following way. For any acyclic structure $\mathfrak{A} = \langle A, F_1, \ldots, F_f, c_1 \rangle$ and $g \in \{\pm 1, \ldots, \pm f\}$, $a < c$, let S_{ga} be as in (5.2). Then $\mathfrak{A} \in C_d$ if and only if for all such g and a,

$$|S_{ga}| = s_{dga} \quad \text{or} \quad |S_{ga}|, \quad s_{dga} > j.$$

Since C_d is nonempty, if $g > 0$ then $s_{dga} \leqslant 1$ for all $a < c$, and there is at most one $a < c$ such that $s_{dga} = 1$. Let

$$I_d = \left\{ (g, a) \in \{\pm 1, \ldots, \pm f\} \times c : g < 0 \text{ or } s_{dga} > 0 \right\},$$

$$v_{dga} = q_a^s e^{-q_a}/s! \quad \text{if } g < 0 \text{ and } s_{dga} = s \leqslant j,$$

$$= 1 - \left(\sum_{s \leqslant j} q_a^s/s! \right) e^{-q_a} \quad \text{if } g < 0 \text{ and } s_{dga} = j + 1,$$

$$= q_a \quad \text{if } g > 0 \text{ for } (g, a) \in I_d,$$

and

$$q_d = \prod_{(g, a) \in I_d} v_{dga}.$$

Let $p, r \in \omega$. A formula $\alpha(x_0, \ldots, x_{p-1})$ is a *diagram* of $x_0, \ldots, x_{p-1}$ if it is of the form $\bigwedge_{1 \leqslant g \leqslant f, h, i < p} \gamma_{ghi}$, where each γ_{ghi} is $F_g(x_h) = x_i$ or $F_g(x_h) \neq x_i$. For any

structure $\mathfrak{A} = \langle A, F_1, \ldots, F_f \rangle$, $x_0, \ldots, x_{p-1} \in A$ and $x \in N(\{x_0, \ldots, x_{p-1}\}, r)$ let

$$B_x(x_0, \ldots, x_{p-1}, r) = \left\{ y \in N(\{x_0, \ldots, x_{p-1}\}, r) : \delta(\{x_0, \ldots, x_{p-1}\}, y) \right.$$

$$= \left. \delta(\{x_0, \ldots, x_{p-1}\}, x) + \delta(x, y) \right\},$$

$$\mathfrak{B}_x(x_0, \ldots, x_{p-1}, r)$$

$$= \left\langle B_x(x_0, \ldots, x_{p-1}, r), F_1 - \{(x, x)\}, \ldots, F_f - \{(x, x)\}, c_1[x] \right\rangle.$$

When $x_0, \ldots, x_{p-1}, r$ are understood, we write B_x and $\mathfrak{B}_x$. Let C be an acyclic j-morphism class with center c_1 and radius $\leqslant r$. For notational convenience, we will treat "$\mathfrak{B}_x \in C$" as a first-order formula. Although it is irrelevant to the combinatorial arguments that follow, it can be shown that it is equivalent to a first-order formula with free variables $x, x_0, \ldots, x_{p-1}$. This follows easily by induction on r (see §5).

For $i < p$ and $n \geqslant p$, $\mu(n, \mathfrak{B}_{x_i} \in C | \alpha)$ is independent of the choices of $x_0, \ldots, x_{p-1} \in n$ provided they are distinct, because j-morphism is isomorphism invariant. Thus $\mu(n, \mathfrak{B}_{x_i} \in C | \alpha)$ will be taken to mean the probability for any distinct $x_0, \ldots, x_{p-1} \in n$. We will use this convention implicitly for other formulas whose probability is constant for all assignments of their free variables to distinct elements.

Let $\beta_r(x_0, \ldots, x_{p-1})$ be a formula such that for all structures $\mathfrak{A}$ and $x_0, \ldots, x_{p-1}$ in $\mathfrak{A}$, $\mathfrak{A} \vDash \beta_r(x_0, \ldots, x_{p-1})$ if and only if, for every $x \in N(\{x_0, \ldots, x_{p-1}\}, r)$, there is a unique path in $N(\{x_0, \ldots, x_{p-1}\}, r)$ from $\{x_0, \ldots, x_{p-1}\}$ to x.

7.9 LEMMA. *If $\alpha \wedge \bigwedge_{h < i < p} x_h \neq x_i$ is consistent, where α is a diagram, then $\lim_{n \to \infty} \mu(n, \beta_r | \alpha) = 1$.*

PROOF. We show $\lim_{n \to \infty} \mu(n, \neg \beta_r | \alpha) = 0$. If $\mathfrak{A} \vDash \neg \beta_r(x_0, \ldots, x_{p-1})$, then there is a sequence $\langle y_0, g_0, y_1, \ldots, g_{m-1}, y_m \rangle$, where $2 \leqslant m \leqslant 2r + 1$, $\{x_0, \ldots, x_{p-1}\} \cap \{y_0, \ldots, y_{m-1}\} = \{y_0\}, y_m \in \{x_0, \ldots, x_{p-1}\} \cup \{y_0, \ldots, y_{m-1}\}, y_h \neq y_i$ for $h < i < m$, $g_i \in \{\pm 1, \ldots, \pm f\}$ and $\mathfrak{A} \vDash F_{g_i}(y_i) = y_{i+1}$ for $i < m$, and $g_{m-2} \neq -g_{m-1}$.

For $2 \leqslant m \leqslant 2r + 1$ and $n \in \omega$ let $q(n, m) = \mu(n, \mathfrak{A}$ has such a sequence of length $m | \alpha)$. Then $\mu(n, \neg \beta_r | \alpha) \leqslant \sum_{2 \leqslant m \leqslant 2r+1} q(n, m)$, and we need only show $\lim_{n \to \infty} q(n, m) = 0$ for $2 \leqslant m \leqslant 2r + 1$.

Since $\alpha \wedge \bigwedge_{h < i < p} x_h \neq x_i$ is consistent, for $n \geqslant p + m$,

$$q(n, m) \leqslant p(p + m)(2f)^m(n - p) \cdots (n - p - m + 2)/n^m.$$

Therefore $\lim_{n \to \infty} q(n, m) = 0$. $\quad\square$

7.10 LEMMA. *For all $p, r \in \omega$, $\alpha(x_0, \ldots, x_{p-1})$ a diagram, $E \subseteq p$ and $b \in {}^E\omega$, where the radius of C_{b_i} is $\leqslant r$ for each $i \in E$, if $\alpha \wedge \bigwedge_{i \in E} \mathfrak{B}_{x_i} \in C_{b_i} \wedge \bigwedge_{h < i < p} x_h \neq x_i$ is consistent, then*

$$\lim_{n \to \infty} \mu\left(n, \bigwedge_{i \in E} \mathfrak{B}_{x_i} \in C_{b_i} | \alpha\right) = \prod_{i \in E} q_{b_i}.$$

PROOF. We use induction on r. If $r = 0$, then $b_i = 0$ for $i \in E$, and for all $\mathfrak{A}$ and distinct $x_0, \ldots, x_{p-1}$ in $\mathfrak{A}$, $\mathfrak{A} \vDash \bigwedge_{i \in E} \mathfrak{B}_{x_i} \in C_0$. Since $q_0 = 1$, the lemma obviously holds.

Now assume that $r > 0$ and the lemma holds for $r - 1$. As before, $C_0, \ldots, C_{c-1}$ are the classes with radius $< r$. For $g \in \{\pm 1, \ldots, \pm f\}$ and $t \in \omega$ let $\alpha_g^t(x, y_0, \ldots, y_{t-1})$ be $\bigwedge_{i < t} F_g(x) = y_i$, let $\sigma_a^t(x_0, \ldots, x_{p-1}, y_0, \ldots, y_{t-1})$ be $\bigwedge_{i < t} \mathfrak{B}_{y_i} \in C_a$ for $a < c$, and for d such that C_d has radius r let $\tau_{dga}(x, x_0, \ldots, x_{p-1})$ be

$$\exists y_0 \cdots \exists y_{s-1} \left[\bigwedge_{\substack{h < p \\ i < s}} x_h \neq y_i \wedge \bigwedge_{h < i < s} y_h \neq y_i \wedge \alpha_g^s(x) \wedge \sigma_a^s \right.$$

$$\left. \wedge \forall y_s \left(F_g(x) = y_s \wedge \mathfrak{B}_{y_s} \in C_a \to \bigvee_{i < s} y_s = y_i \right) \right] \quad \text{if } s_{dga} = s \leqslant j,$$

or

$$\exists y_0 \cdots \exists y_j \left[\bigwedge_{\substack{h < p \\ i \leqslant j}} x_h \neq y_i \wedge \bigwedge_{h < i \leqslant j} y_h \neq y_i \wedge \alpha_g^{j+1}(x) \wedge \sigma_a^{j+1} \right] \quad \text{if } s_{dga} = j + 1.$$

For every $i \in E$, $\alpha \wedge \mathfrak{B}_{x_i} \in C_{b_i}$ is consistent by assumption. Therefore for $1 \leqslant g \leqslant f$, if $s_{b_i ga} = 0$ for all $a < c$, then $F_g(x_i) = x_h$ is a conjunct in α for some $h < p$. Then for every functional $\mathfrak{A}$ and distinct $x_0, \ldots, x_{p-1}$ in $\mathfrak{A}$,

$$\mathfrak{A} \vDash \alpha \wedge \beta_r \to \left(\mathfrak{B}_{x_i} \in C_{b_i} \leftrightarrow \bigwedge_{(g, a) \in I_{b_i}} \tau_{b_i ga}(x_i) \right).$$

Let $D = \{ i \in E: b_i < c \}$ and $I = \{ (i, g, a): i \in E - D \text{ and } (g, a) \in I_{b_i} \}$. Then by Lemma 7.9, we will be done if we prove

$$(7.11) \quad \lim_{n \to \infty} \mu \left(n, \bigwedge_{i \in D} \mathfrak{B}_{x_i} \in C_{b_i} \wedge \bigwedge_{(i, g, a) \in I} \tau_{b_i ga}(x_i) | \alpha \right) = \prod_{i \in D} q_{b_i} \times \prod_{(i, g, a) \in I} v_{b_i ga}.$$

Let $J = \{ (i, g, a) \in I: s_{b_i ga} \leqslant j \}$, and for $m, n \in \omega$ let

$$u(n, m) = \sum_{\langle t_{iga} \rangle} \prod_{(i, g, a) \in J} \binom{t_{iga}}{s_{b_i ga}} \times \prod_{(i, g, a) \in I - J} \binom{t_{iga} - 1}{j}$$

$$\times \left[(n - p)(n - p - 1) \cdots (n - p - m + 1) \middle/ \prod_{(i, g, a) \in I} t_{iga}! \right]$$

$$\times \mu \left(n, \bigwedge_{i \in D} \mathfrak{B}_{x_i} \in C_{b_i} \wedge \bigwedge_{(i, g, a) \in I} \alpha_g^{t_{iga}}(x_i) \wedge \sigma_a^{t_{iga}} | \alpha \right)$$

and

$$u_m = \prod_{i \in D} q_{b_i} \times \sum_{\langle t_{iga} \rangle} \prod_{(i, g, a) \in J} \left[s_{b_i ga}! \left(t_{iga} - s_{b_i ga} \right)! \right]^{-1} q_a^{t_{iga}}$$

$$\times \prod_{(i, g, a) \in I - J} \left[j! (t_{iga} - j - 1)! t_{iga} \right]^{-1} q_a^{t_{iga}}.$$

Both sums are taken over all sequences $\langle t_{iga}: (i, g, a) \in I \rangle$ of elements in ω such that for all $(i, g, a) \in I$, $t_{iga} \geqslant s_{b_i ga}$ and $t_{iga} = 1$ if $g > 0$, and $\sum_{(i, g, a) \in I} t_{iga} = m$. In

the first sum, we rename the free variables $y_0, \ldots, y_{t_{iga}-1}$ in the conjuncts $\alpha_g^{t_{iga}}(x_i) \wedge \sigma_a^{t_{iga}}$ so that every variable $y_0, \ldots, y_{m-1}$ occurs in exactly one conjunct.

Let $\Sigma s = \sum_{(i,g,a)\in I} s_{b_i ga}$. By Lemma 7.2,

$$(7.12) \quad \mu\left(n, \bigwedge_{i\in D} \mathfrak{B}_{x_i} \in C_{b_i} \wedge \bigwedge_{(i,g,a)\in I} \tau_{b_i ga}(x_i)|\alpha\right) = \sum_{m \geqslant \Sigma s} (-1)^{m - \Sigma s} u(n, m).$$

(We take G to be the set of functional structures with universe n satisfying $\alpha \wedge \bigwedge_{i\in D} \mathfrak{B}_{x_i} \in C_{b_i}$, $H = n - \{x_0, \ldots, x_{p-1}\}$, and for $y \in H$, $(i, g, a) \in I$, $P(y, (i, g, a)) = \{\mathfrak{A} \in G : \mathfrak{A} \vDash F_g(x_i) = y \wedge \mathfrak{B}_y \in C_a\}$.)We will apply Lemma 7.4 to the sum in (7.12). Clearly (7.5) holds because $u(n, m) = 0$ for $m > n$. For $-f \leqslant g \leqslant -1$, if $(i, g, a) \in J$ then letting $s = s_{b_i ga}$,

$$v_{b_i ga} = \sum_{t \geqslant s} (-1)^{t-s} [s!(t-s)!]^{-1} q_a^t,$$

and if $(i, g, a) \in I - J$, then

$$v_{b_i ga} = 1 - \sum_{s \leqslant j} \left[\sum_{t \geqslant s} (-1)^{t-s} \binom{t}{s} q_a^t/t!\right]$$

$$= 1 - \sum_{t \leqslant j} \left[\sum_{s \leqslant t} (-1)^{t-s} \binom{t}{s}\right] q_a^t/t! - \sum_{t > j} \left[\sum_{s \leqslant j} (-1)^{t-s} \binom{t}{s}\right] q_a^t/t!$$

$$= \sum_{t > j} (-1)^{t-j-1} \left[\sum_{s \leqslant j} (-1)^{j-s} \binom{t}{s}\right] q_a^t/t!$$

$$= \sum_{t > j} (-1)^{t-j-1} [j!(t-j-1)!t]^{-1} q_a^t.$$

Similarly,

$$(7.13) \qquad \prod_{i\in D} q_{b_i} \times \prod_{(i,g,a)\in I} v_{b_i ga} = \sum_{m \geqslant \Sigma s} (-1)^{m - \Sigma s} u_m,$$

and (7.6) holds. Take any $m \in \omega$ and $\langle t_{iga} : (i, g, a) \in I\rangle$ such that $t_{iga} = 1$ for $g > 0$ and $\sum_{(i,g,a)\in I} t_{iga} = m$. Letting $\alpha' = \bigwedge_{(i,g,a)\in I} \alpha_g^{t_{iga}}(x_i)$ and $\sigma' = \bigwedge_{(i,g,a)\in I} \sigma_a^{t_{iga}}$, for $n \in \omega$,

$$\mu\left(n, \bigwedge_{i\in D} \mathfrak{B}_{x_i} \in C_{b_i} \wedge \alpha' \wedge \sigma'|\alpha\right) = \mu\left(n, \bigwedge_{i\in D} \mathfrak{B}_{x_i} \in C_{b_i} \wedge \sigma'|\alpha \wedge \alpha'\right) \times \mu(n, \alpha'|\alpha).$$

Let $\alpha''(x_0, \ldots, x_{p-1}, y_0, \ldots, y_{m-1})$ be the conjunction of all formulas $F_g(z_0) \neq z_1$, where $1 \leqslant g \leqslant f$, $z_0, z_1 \in \{x_0, \ldots, x_{p-1}\} \cup \{y_0, \ldots, y_{n-1}\}$ and $F_g(z_0) = z_1$ is not a conjunct of $\alpha \wedge \alpha'$.

Then $\alpha \wedge \alpha' \wedge \alpha''$ is a diagram and $\lim_{n\to\infty} \mu(n, \alpha''|\alpha \wedge \alpha') = 1$. Also

$$\alpha \wedge \alpha' \wedge \alpha'' \wedge \bigwedge_{i\in D} \mathfrak{B}_{x_i} \in C_{b_i} \wedge \sigma' \wedge \bigwedge_{h<i<p} x_h \neq x_i \wedge \bigwedge_{h<i<m} y_h \neq y_i \wedge \bigwedge_{\substack{h<p \\ i<m}} x_h \neq y_i$$

is consistent because $t_{iga} = 1$ for $g > 0$, and

$$\alpha \wedge \bigwedge_{i\in E} \mathfrak{B}_{x_i} \in C_{b_i} \wedge \bigwedge_{h<i<p} x_h \neq x_i$$

is consistent. For all functional $\mathfrak{A}$, distinct $x_0,\ldots,x_{p-1}, y_0,\ldots,y_{m-1}$ in $\mathfrak{A}$ and $i < m$,

$$\beta_{r-1}(x_0,\ldots,x_{p-1}, y_0,\ldots,y_{m-1}) \wedge \alpha' \to B_{y_i}(x_0,\ldots,x_{p-1}, r)$$

$$= B_{y_i}(x_0,\ldots,x_{p-1}, y_0,\ldots,y_{m-1}, r-1).$$

Therefore by Lemma 7.9 and the induction assumption,

$$\lim_{n \to \infty} \mu\left(n, \bigwedge_{i \in D} \mathfrak{B}_{x_i} \in C_{b_i} \wedge \sigma' | \alpha \wedge \alpha'\right) = \prod_{i \in D} q_{b_i} \times \prod_{(i,g,a) \in I} q_a^{t_{iga}}.$$

It is easily seen that

$$\mu(n, \alpha' | \alpha) = n^{-m}.$$

Therefore $\lim_{n \to \infty} u(n, m) = u_m$ and (7.7) holds.

Lastly, (7.8) holds by Lemma 7.3, and by Lemma 7.4,

$$\lim_{n \to \infty} \sum_{m \geqslant \Sigma s} (-1)^{m-\Sigma s} u(n, m) = \sum_{m \geqslant \Sigma s} (-1)^{m-\Sigma s} u_m,$$

which together with (7.12) and (7.13) completes the proof of (7.11) and the lemma.
$\square$

7.14 PROOF OF THEOREM 4.4. Let $1 \leqslant i \leqslant k$ and C be a realizable acyclic $(k-i)$-morphism class with center $\{c_1\}$ and radius 3^{k-i}. Taking $j = k - i$ and $r = 3^{k-i}$, let B_x, $\mathfrak{B}_x$ and β_{2r} be as defined prior to Lemma 7.9, and let $q \in Q_0$ be associated with C. For any $\mathfrak{A} = \langle A, F_1,\ldots,F_f \rangle$, if there are $y_0,\ldots,y_{i-1} \in A$ such that $\mathfrak{A} \vDash \beta_{2r}(y_0,\ldots,y_{i-1})$, then

$$N(y_d, r) \cap N(y_h, r) = \varnothing \quad \text{for } d < h < i$$

and

$$\delta(y_h, X) > r \quad \text{for } h < i \text{ and all cycle sets } X \text{ of size} \leqslant 2r + 1.$$

Therefore for every $x_1,\ldots,x_{i-1} \in A$, there is some $h < i$ such that $\delta(y_h, \{x_1,\ldots,x_{i-1}\}) > r$, and for every $h < i$,

$$\mathfrak{B}_{y_h}(y_0,\ldots,y_{i-1}, r) = \langle N(y_h, r), F_1,\ldots,F_f, c_1[y_h] \rangle.$$

For $t \in \omega$ let $\zeta^t(y_0,\ldots,y_{t-1})$ be

$$\beta_{2r}(y_0,\ldots,y_{t-1}) \wedge \bigwedge_{h < t} \mathfrak{B}_{y_h}(y_0,\ldots,y_{t-1}, r) \in C.$$

By showing that, for every such i and C,

$$\lim_{n \to \infty} \mu\left(n, \exists y_0 \cdots \exists y_{i-1} \zeta^i(y_0,\ldots,y_{i-1})\right) = 1,$$

we will be done.

For $p \geqslant i$ let $\gamma_p(z_0,\ldots,z_{p-1})$ be

$$\exists y_0 \cdots \exists y_{i-1} \left[\{y_0,\ldots,y_{i-1}\} \subseteq \{z_0,\ldots,z_{p-1}\} \wedge \zeta^i(y_0,\ldots,y_{i-1})\right].$$

Then $\mu(n, \gamma_p) \leqslant \mu(n, \exists y_0 \cdots \exists y_{i-1} \zeta^i)$, and we complete the proof by showing $\lim_{n \to \infty} \mu(n, \gamma_p) = u_p$ for some $u_p \in [0, 1]$, and $\lim_{p \to \infty} = 1$. By the principle of inclusion-exclusion,

$$\mu(n, \gamma_p) = 1 - \sum_{s < i} \sum_{s \leqslant t \leqslant p} (-1)^{t-s} \binom{t}{s} \binom{p}{t} \mu(n, \zeta^t).$$

For $t \in \omega$ let α^t be $\bigwedge_{1 \leqslant g \leqslant f, h, i < t} F_g(y_h) \neq y_i$. Then

$$\lim_{n \to \infty} \mu(n, \zeta^t) = \lim_{n \to \infty} \mu(n, \zeta^t | \alpha^t)$$

$$= \lim_{n \to \infty} \mu\left(n, \bigwedge_{h < t} \mathfrak{B}_{y_h}(y_0, \ldots, y_{t-1}, r) \in C | \alpha^t\right)$$

by Lemma 7.9. Since C is realizable, $\alpha^t \wedge \bigwedge_{h < t} \mathfrak{B}_{y_h} \in C \wedge \bigwedge_{d < h < t} y_d \neq y_h$ is consistent. Therefore by Lemma 7.10, $\lim_{n \to \infty} \mu(n, \zeta^t) = q^t$ for all $t \in \omega$, and

$$\lim_{n \to \infty} \mu(n, \gamma_p) = 1 - \sum_{s < i} \sum_{s \leqslant t \leqslant p} (-1)^{t-s} \binom{t}{s} \binom{p}{t} q^t = 1 - \sum_{s < i} \binom{p}{s} q^s (1 - q)^{p-s}$$

$$\to 1 \quad \text{as } p \to \infty \text{ because } q \neq 0. \quad \square$$

7.15 **PROOF OF THEOREM 4.5.** Let $C_0, \ldots, C_{c-1}$ be an enumeration of all acyclic k-morphism classes with center $\{c_1\}$ and radius $\leqslant 3^k$ and $q_0, \ldots, q_{c-1}$ be the associated probabilities as described above. For any k-morphism class D with a single cycle of size $p \leqslant 2 \cdot 3^k + 1$, no constants, and radius $\leqslant 3^k$, a *characteristic* of D is a formula γ of the form

$$\alpha(x_0, \ldots, x_{p-1}) \wedge \sigma(x_0, \ldots, x_{p-1}),$$

where α is a diagram, σ is $\bigwedge_{i < p} \mathfrak{B}_{x_i} \in C_{b_i}$, $b_i < c$ for $i < p$, and for all structures $\mathfrak{A}$ and distinct $x_0, \ldots, x_{p-1}$ in $\mathfrak{A}$, if $\mathfrak{A} \models \gamma(x_0, \ldots, x_{p-1})$ then $\mathcal{N}(\{x_0, \ldots, x_{p-1}\}, 3^k) \in D$ and it has center $\{x_0, \ldots, x_{p-1}\}$. Let $\text{aut}(D)$ be the set of all permutations θ on p such that $\gamma(x_{\theta(0)}, \ldots, x_{\theta(p-1)})$ is equivalent to $\gamma(x_0, \ldots, x_{p-1})$.

It is easily seen that $|\text{aut}(D)|$ does not depend on the choice of characteristic, and $p!/|\text{aut}(D)|$ is the number of nonequivalent characteristics of D. Further, for every $n \in \omega$ and $X \subseteq n$ such that $|X| = p$,

$$(7.16) \qquad \mu(n, \mathcal{N}(X, 3^k) \in D) = (p!/|\text{aut}(D)|)\mu(n, \gamma).$$

Let $D_0, \ldots, D_{d-1}$ be an enumeration of all k-morphism classes with a single cycle of size $\leqslant 2 \cdot 3^k + 1$, no constants, and radius $\leqslant 3^k$. For each $a < d$, let p_a be the size of D_a's cycle and γ_a be any characteristic of D_a with α_a, σ_a and b_{ai} ($i < p_a$) as above. The k-agreeability class K is characterized by a sequence $\langle s_a : a < d \rangle$ of elements in $k + 2$ as follows. For any structure $\mathfrak{A} = \langle A, F_1, \ldots, F_f \rangle$ and $a < d$, let S_a be as in (4.1) with $j = k$. Then $\mathfrak{A} \in K$ if and only if for all such a,

$$|S_a| = s_a \quad \text{or} \quad |S_a|, \quad s_a > k.$$

In the following formulas, $X_0, X_1, \cdots$ will represent subsets of the universe. This is for convenience; at the cost of more cumbersome notation, we could use first-order formulas.

For $a < d$ and $t \in \omega$ let α_a^t be $\bigwedge_{i < t} \alpha_a(x_{i \cdot p_a}, \ldots, x_{i \cdot p_a + p_a - 1})$ and similarly for σ_a^t. Let $\Phi_a^t(X_0, \ldots, X_{t-1})$ be $\bigwedge_{i < t} \mathcal{N}(X_i, 3^k) \in D_a$ and Ψ_a be

$$\exists X_0 \cdots \exists X_{s-1} \left[\bigwedge_{h < i < s} X_h \neq X_i \wedge \Phi_a^s(X_0, \ldots, X_{s-1}) \right.$$

$$\left. \wedge \forall X_s \left(\mathcal{N}(X_s, 3^k) \in D_a \to \bigvee_{i < s} X_s = X_i \right) \right] \quad \text{if } s_a = s \leqslant k,$$

or

$$\exists X_0 \cdots \exists X_k \left[\bigwedge_{h < i \leqslant k} X_h \neq X_i \wedge \Phi_a^{k+1}(X_0, \ldots, X_k) \right] \quad \text{if } s_a = k + 1.$$

Then, for every $\mathfrak{A}$, $\mathfrak{A} \in K$ if and only if $\mathfrak{A} \vDash \bigwedge_{a<d} \Psi_a$. Using methods similar to those in the proof of Lemma 7.10, we will complete the proof by showing

$$(7.17) \qquad \lim_{n \to \infty} \mu\left(n, \bigwedge_{a<d} \Psi_a\right) = q \in Q.$$

Let $J = \{a < d : s_a \leqslant k\}$, for $a < d$ let

$$v_a = \left(\prod_{i < p_a} q_{b_{ai}} \right) / |\mathrm{aut}(D_a)|,$$

and for $m, n \in \omega$ let

$$u(n, m) = \sum_{\langle t_a \rangle} \prod_{a \in J} \binom{t_a}{s_a} \times \prod_{a \in d-J} \binom{t_a - 1}{k}$$

$$\times \left[n(n-1) \cdots \left(n + 1 - \sum_{a<d} p_a t_a \right) \Big/ \prod_{a<d} (p_a!)^{t_a} t_a! \right]$$

$$\times \mu\left(n, \bigwedge_{a<d} \Phi_a^{t_a} \right)$$

and

$$u_m = \sum_{\langle t_a \rangle} \prod_{a \in J} \left[s_a!(t_a - s_a)! \right]^{-1} v_a^{t_a} \times \prod_{a \in d-J} \left[k!(t_a - k - 1)! t_a \right]^{-1} v_a^{t_a}.$$

Both sums are over all sequences $\langle t_a : a < d \rangle$ in ω such that $t_a \geqslant s_a$ for all $a < d$ and $\sum_{a<d} t_a = m$. In the first sum, we rename the free variables $X_0, \ldots, X_{t_a - 1}$ in each conjunct $\Phi_a^{t_a}$ so that every variable $X_0, \ldots, X_{m-1}$ occurs in exactly one conjunct, and $\mu(n, \bigwedge_{a<d} \Phi_a^{t_a})$ means the probability for distinct $X_0, \ldots, X_{m-1}$.

Let $\Sigma s = \sum_{a<d} s_a$. By Lemma 7.2,

$$(7.18) \qquad \mu\left(n, \bigwedge_{a<d} \Psi_a \right) = \sum_{m \geqslant \Sigma s} (-1)^{m - \Sigma s} u(n, m).$$

(We take G to be the set of functional structures with universe n, $H = \{X : X \subseteq n\}$, $I = d$, and for $X \in H$, $a < d$, $P(X, a) = \{\mathfrak{A} \in G : \mathfrak{A} \vDash \mathcal{N}(X, 3^k) \in D_a\}$.) As in the proof of Lemma 7.10, (7.5) and (7.8) hold. Letting

$$q = \prod_{a \in J} v_a^{s_a} e^{-v_a} / s_a! \times \prod_{a \in d-J} \left(1 - \left(\sum_{s \leqslant k} v_a^s / s! \right) e^{-v_a} \right),$$

$\sum_{m \geqslant \Sigma s} (-1)^{m - \Sigma s} u_m = q$ and (7.6) holds. By the same reasoning used in (7.16),

$$\mu\left(n, \bigwedge_{a<d} \Phi_a^{t_a} \right) = \left[\prod_{a<d} \left(p_a! / |\mathrm{aut}(D_a)| \right) \right]^{t_a}$$

$$\times \mu\left(n, \bigwedge_{a<d} \sigma_a^{t_a} \,\Big|\, \bigwedge_{a<d} \alpha_a^{t_a} \right) \times \mu\left(n, \bigwedge_{a<d} \alpha_a^{t_a} \right).$$

Again we rename the free variables in $\bigwedge_{a<d} \alpha_a^{t_a}$ and $\bigwedge_{a<d} \sigma_a^{t_a}$ so that each free variable occurs in exactly one conjunct. By Lemma 7.10,

$$\lim_{n\to\infty} \mu\left(n, \bigwedge_{a<d}\sigma_a^{t_a}\middle| \bigwedge_{a<d}\alpha_a^{t_a}\right)\middle/ \prod_{a<d}|\mathrm{aut}(D_a)|^{t_a} = \prod_{a<d}v_a^{t_a}.$$

It is clear that

$$\mu\left(n, \bigwedge_{a<d}\alpha_a^{t_a}\right) = \prod_{a<d}\left[(n-p_a)^{f-1}/n^f\right]^{p_a t_a} = \prod_{a<d}n^{-p_a t_a}[1+o(1)].$$

Therefore $\lim_{n\to\infty} u(n,m) = u_m$, and (7.7) holds. Applying Lemma 7.4 to (7.18) gives us (7.17). $\square$

8. Related problems. An obvious question is whether there is an effective procedure for calculating $\lim_{n\to\infty}\mu(n,\sigma)$, given any sentence σ. By this we mean generating an expression in closed form whose value is the limit. We sketch a method for doing this. Let S be the set of all formulas σ of the form $\bigwedge_{i\in E}\mathfrak{B}_{x_i}(x_0,\ldots,x_{p-1},r)\in C_i$, where $p, r \in \omega$, $E\subseteq p$, and for some $j\in\omega$ each C_i is an acyclic j-morphism class with center $\{c_1\}$ and radius $\leqslant r$. As previously mentioned, $\mathfrak{B}_{x_i}\in C_i$ is an abbreviation for some canonical first-order formula. Also, it is not difficult to define this set of canonical formulas so that S is recursive. Let $\alpha(x_0,\ldots,x_{p-1})$ be a diagram. The proof of Lemma 7.10 implicitly gives an algorithm for calculating $\lim_{n\to\infty}\mu(n,\sigma|\alpha)$. First, if $\alpha\wedge\sigma\wedge\bigwedge_{h<i<p}x_h\neq x_i$ is inconsistent, then $\lim_{n\to\infty}\mu(n,\sigma|\alpha)=0$. By using the definition of j-morphism in §5, it is easy to see that there is a recursive procedure for determining the consistency of such a formula. If it is consistent, then by Lemma 7.10, the problem of calculating $\lim_{n\to\infty}\mu(n,\sigma|\alpha)$ reduces to calculating the probabilities q_i for $i\in E$, which in turn can be calculated recursively from the definition prior to Lemma 7.9. Thus

8.1 LEMMA. *There is a recursive procedure such that, given a partial diagram* $\alpha(x_0,\ldots,x_{p-1})$ *and* $\sigma(x_0,\ldots,x_{p-1})\in S$, *it calculates* $\lim_{n\to\infty}\mu(n,\sigma|\alpha)$.

Similarly, there is a recursive set T consisting of all sentences $\bigvee_{i<p}\mathfrak{A}\in K_i$, where $p\in\omega$ and for some $k\in\omega$ each K_i is a k-agreeability class. Of course, we mean $\mathfrak{A}\in K_i$ is an abbreviation for some canonical first-order sentence, similar in form to $\bigwedge_{a<d}\Psi_a$ in the proof of Theorem 4.5. Clearly, if $K_h\neq K_i$ for $h<i<p$, then

$$\lim_{n\to\infty}\mu\left(n, \bigvee_{i<p}\mathfrak{A}\in K_i\right) = \sum_{i<p}\lim_{n\to\infty}\mu(n,\mathfrak{A}\in K_i).$$

The proof of Theorem 4.5 shows that calculating $\lim_{n\to\infty}\mu(n,\mathfrak{A}\in K_i)$ reduces to the problem of calculating $\lim_{n\to\infty}\mu(n,\sigma|\alpha)$ for $\sigma\in S$ and α a partial diagram. Thus

8.2 COROLLARY. *There is a recursive procedure such that, given* $\tau\in T$, *it calculates* $\lim_{n\to\infty}\mu(n,\tau)$.

Given an arbitrary sentence σ of depth k, by Theorems 4.3 and 4.4, $\lim_{n\to\infty}\mu(n,\sigma) = \lim_{n\to\infty}\mu(n,\sigma\wedge\gamma)$, where γ is a first-order sentence true for

exactly those structures which are k-simple and k-rich. By Theorem 4.2, $\sigma \wedge \gamma$ is equivalent to $\tau \wedge \gamma$ for some $\tau \in T$. Let Th_f be the first-order theory of f unary functions. Since T is recursive and Th_f is recursively enumerable, we have a procedure for calculating $\lim_{n \to \infty} \mu(n, \sigma)$. We simply generate sentences in Th_f until one of the form $\sigma \wedge \gamma \leftrightarrow \tau \wedge \gamma$, $\tau \in T$, occurs, and then calculate $\lim_{n \to \infty} \mu(n, \tau) = \lim_{n \to \infty} \mu(n, \tau \wedge \gamma)$. Thus

8.3 THEOREM. *The problem of calculating* $\lim_{n \to \infty} \mu(n, \sigma)$ *is recursive.*

As is well known, Th_1 is decidable (Ehrenfeucht, unpublished), but Th_f is undecidable for $f > 1$ (B. Trachtenbrot; see Monk [19, pp. 295–296]). This shows that the problem of calculating $\lim_{n \to \infty} \mu(n, \sigma)$ is in a sense easier than determining the validity of σ. However, this is not always the case. Compton [3] has shown that there are classes of finite structures with 0-1 limit laws for which the problem of computing asymptotic probabilities is unsolvable. Note also that we have not considered the computational efficiency of calculating the limits.

A more general category of problems is finding limit laws for other classes of finite structures, such as binary functions or structures with relations and functions. When one attempts to use the methods of this paper, difficulties arise very quickly. In particular, it is not obvious how to generalize the metric δ even to a single binary function. For a unary function, the intuitive idea of $\delta(x, y)$ is the smallest number of backward or forward edges needed to get from x to y, but there does not seem to be any corresponding notion for binary functions. For other classes of structures, there are obvious metrics, but they do not seem to satisfy any condition similar to k-richness. For example, let σ be a first-order sentence with the binary relation symbol $\leqslant$ interpreted in the usual way and the r-ary relation symbol R. It is not known whether

$$\lim_{n \to \infty} \left| \left\{ R \subseteq {}^r n \colon \langle n, \leqslant, R \rangle \vDash \sigma \right\} \right| / 2^{n^r}$$

always exists when $r > 1$. (See [15] for a proof by Ehrenfeucht that it always exists when $r = 1$, and an example that the limit can be irrational when $r = 2$.) A related problem has been posed by J. Mycielski. For $n \in \omega$ let $T_n = \{(x, y, z) \in {}^3 n \colon x \leqslant y \leqslant z$ or $z \leqslant x \leqslant y$ or $y \leqslant z \leqslant x\}$. (Intuitively, if one thinks of an n-gon whose vertices are labeled with $0, 1, \ldots, n - 1$ in clockwise order, then $T_n(x, y, z)$ holds if and only if the clockwise path from x to z includes y.) It seems likely that

$$\lim_{n \to \infty} \left| \left\{ R \subseteq {}^r n \colon \langle n, T_n, R \rangle \vDash \sigma \right\} \right| / 2^{n^r}$$

is 0 or 1 for all $r \geqslant 1$ and all σ. However, it is not even known if the limit always exists.

As Ulam [22] has remarked, there is a need for a general theory of combinatorial metrics. We would add that such a theory could have important applications to finite model theory. Metrics defined on the universes of finite models have been central to the proofs of the undefinability results in [5, 9, 17] and the probabilistic limit laws of [15] and this paper. A more comprehensive theory of such metrics could possibly unify and extend these results.

REFERENCES

1. C. C. Chang and H. J. Keisler, *Model theory*, North-Holland, Amsterdam, 1977.

2. K. J. Compton, *A logical approach to asymptotic combinatorics*, Adv. in Math. (submitted).

3. _____, *An undecidable problem in finite combinatorics*, J. Symbolic Logic (submitted).

4. A. Ehrenfeucht, *An application of games to the completeness problem for formalized theories*, Fund. Math. **49** (1961), 129–141.

5. R. Fagin, *Monadic generalized spectra*, Z. Math. Logik Grundlag. Math. **21** (1975), 89–96.

6. _____, *Probabilities on finite models*, J. Symbolic Logic **41** (1976), 50–58.

7. W. Feller, *An introduction to probability theory*. Vol. I, 3rd ed., Wiley, New York, 1968.

8. H. Gaifman, *Concerning measures in first-order calculi*, Israel J. Math. **2** (1964), 1–18.

9. _____, *On local and non-local properties* (to appear).

10. Y. V. Glebskii, D. I. Kogan, M. I. Liogon'kii and V. A. Talanov, *Range and degree of realizability of formulas in the restricted predicate calculus*, Cybernetics **5** (1969), 142–154.

11. F. Harary, *Graph theory*, Addison-Wesley, Reading, Mass., 1969.

12. B. Harris, *Probability distributions related to random mappings*, Ann. Math. Statist. **31** (1960), 1045–1062.

13. L. Katz, *Probability of indecomposability of a random mapping function*, Ann. Math. Statist. **26** (1955), 512–517.

14. M. D. Kruskal, *The expected number of components under a random mapping function*, Amer. Math. Monthly **61** (1954), 392–397.

15. J. F. Lynch, *Almost sure theories*, Ann. Math. Logic **18** (1980), 91–135.

16. _____, *A limit law for sentences about unary functions*, Abstracts Amer. Math. Soc. **1** (1980), 494.

17. _____, *On sets of relations definable by addition*, J. Symbolic Logic **47** (1982), 659–668.

18. N. Metropolis and S. Ulam, *A property of randomness of an arithmetical function*, Amer. Math. Monthly **60** (1953), 252–253.

19. J. D. Monk, *Mathematical logic*, Springer-Verlag, New York, 1976.

20. G. E. Reyes, *Local definability theory*, Ann. Math. Logic **1** (1970), 95–137.

21. L. A. Shepp and S. P. Lloyd, *Ordered cycle lengths in a random permutation*, Trans. Amer. Math. Soc. **121** (1966), 340–357.

22. S. M. Ulam, *On the notion of analogy and complexity in some constructive mathematical schemata* (to appear).

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, CLARKSON UNIVERSITY, POTSDAM, NEW YORK 13676