On Ramsey-type theorems and their applications*

Rod Downey
Mathematics Department
Victoria University
PO Box 600 Wellington
New Zealand

§1. Introduction

In [21], Frank Plumpton Ramsey proved what has turned out to be a remarkable and important theorem which is now known as Ramsey's theorem. This result is a generalization of the pigeonhole principle and can now be seen as part of a family of theorems of the same flavour. These Ramsey-type theorems all have the common feature that they assert, in some precise combinatorial way, that if we deal with large enough sets of numbers, there will be some well behaved fragment in the set. In Harrington's words, Ramsey-type theorems assert that complete chaos is impossible.

Ramsey-type theorems have turned out to be very important in a number of branches of mathematics. In this paper we shall survey a number of basic Ramsey-type theorems, and we will then look at a selection of applications of Ramsey-type theorems and Ramsey-type ideas.

In the applications we will concentrate on graph theory, logic and complexity theory. Proofs will mostly not be given in detail, but it is hoped that the reader will gain some appreciation of the usefulness and importance of the beautiful area of asymptotic combinatorics.

---

* The subject of this paper was presented as a seminar at the National University of Singapore in March 1989. The author wishes to thank Carl Jockusch for discussions concerning Shelah's proof of the Hales-Jewett Theorem.

§2. Basic Ramsey theory

Our starting point is the pigeonhole principle. This is the obvious statement that one cannot put more than n pigeons into n pigeonholes without at least two sharing. Whilst its statement is remarkably obvious, this principle has a number of very delicate and subtle applications. Here the reader is referred to, for example, [28, chapter III].

2.1 Example. Imagine numbers 1, ..., 36 are distributed randomly (but without repetition) in the 36 sectors of a roulette wheel. Let the contents of the sectors be $a_1, \ldots, a_{36}$. We claim that there are four consecutive sectors $a_i, a_{i+1}, a_{i+2}, a_{i+3}$ (where for example if $i + 1 = 36$ then a_{i+2} denotes a_1) such that $a_i + a_{i+1} + a_{i+2} + a_{i+3} \geq 74$.

To prove this first recall that $1 + 2 + \cdots + 36 = 18 \cdot 37 = 666$. Now consider the sums:

$$\begin{aligned}
a_1 + a_2 + a_3 + \cdots + a_{35} + a_{36} &= 666 \\
a_2 + a_3 + a_4 + \cdots + a_{36} + a_1 &= 666 \\
a_3 + a_4 + a_5 + \cdots + a_1 + a_2 &= 666 \\
a_4 + a_5 + a_6 + \cdots + a_2 + a_3 &= 666 \\
\hline
s_1 + s_2 + s_3 + \cdots + s_{35} + s_{36} &= 2664
\end{aligned}$$

Now there are 36 sums of consecutive sectors, whose sum is 2664. That is there are 2664 pigeons who must be packed into 36 pigeonholes. It follows that at least one hole must have $\geq 2664/36 = 74$ pigeons. Hence by the pigeonhole principle at least one sum s_k has $s_k \geq 74$. □

Another way to think of the pigeonhole principle is via *colouring*. The principle asserts that if I have n colours and more than n pigeons then two pigeons will be coloured the same.

Let A be a set. An *m-subset* of A is a subset with m-elements. The simplest form of Ramsey's theorem is concerned with colouring 2-sets rather than single elements. The simplest statement of Ramsey's theorem is then

2.2. Theorem (Ramsey [21]). *There is a number n so large that if I colour each of the 2-subsets of $\{1, \ldots, n\}$ in red or blue then there is a subset $B \subseteq A$ such that all the 2-subsets of B have the same colour.*

The statement above appears rather complex, but is best thought of as the "friend principle". Suppose in a party we will colour a pair of people red if they are friends and blue if they are non-friends. Then (2.2) asserts (2.3) below.

2.3. *Given k there is an n so large that in any party of at least n people there is either a set of k mutual friends or k mutual non-friends.*

Of course we cannot know which option will pertain since for some parties all people might be friends and for others all might be enemies!

The statements above are rather unwieldy, and so we shall introduce some notation to aid our discussion. We will write

$$n \to (k_1, k_2)$$

if n has the property that it is so large that in any party of n people there are either k_1 mutual friends or k_2 mutual non-friends. A more general statement is

2.4. $(\forall k_1,\ k_2)\ (\exists n)\ (n \to (k_1, k_2))$.

We will now describe how to prove (2.4) which implies (2.2) by setting $k = k_1 = k_2$. The proof of (2.4) uses the same idea as a much easier and familiar result:

2.5. *If A is any set with n elements then $P(A)$, the set of subsets of A, has 2^n elements.*

We will discuss the proof of (2.5) first. The proof we give uses an important combinatorial idea: a bijective proof. We will proceed by induction. In $n = 0$ then $A = \emptyset$ and $P(A) = \{\emptyset\}$, which has $2^0 = 1$ element.

For an induction suppose the result for any set with $n = k$ elements. Suppose A has $k + 1$ elements. We distinguish one element x of A. Then the subsets of A can be naturally divided into two types:

$$I_x = \{B\ :\ B \subseteq A \text{ and } x \in B\},$$

and

$$II_x = \{B\ :\ B \subseteq A \text{ and } x \notin B\}.$$

That is those that contain x and those that don't.

There is a bijection between I_x and II_x in the sense that to get a set in I_x we take a unique set in II_x and add x. Thus both I_x and II_x have the same size. Note that II_x is the collection of subsets of $A - \{x\}$. So, by the induction hypothesis, II_x has 2^k elements. Hence $P(A)$ has $2^k + 2^k = 2^{k+1}$ elements. □

To prove (2.4) we will again proceed by induction except that we will use a slightly trickier form of induction. Since we are dealing with a statement involving k_1 and k_2 we will proceed by induction on k_1 *and* k_2. Clearly $k_2 \to (2, k_2)$ and $k_1 \to (k_1, 2)$. For example, in any set of k_2 people they are either all friends or two of them are not, hence $k_2 \to (2, k_2)$.

Suppose that $n_1 \to (k_1 - 1, k_2)$ and $n_2 \to (k_1, k_2 - 1)$ we claim that $n_1 + n_2 = n \to (k_1, k_2)$. The result will then follow by induction. Again we will use the "divide and conquer" technique. Take any element x of $\{1, \ldots, n\}$. Let

$$I_x = \{y : \{x, y\} \text{ is red}\},$$
$$\mathit{II}_x = \{y : \{x, y\} \text{ is not red}\}.$$

Now as $n_1 + n_2 = n$ and $|I_x| + |\mathit{II}_x| + 1 = n$, where $|B|$ denotes the number of elements in B. So by the pigeonhole principle it follows that either $|I_x| \geq n_1$ or $|\mathit{II}_x| \geq n_2$.

If $|I_x| \geq n_1$ then as $n_1 \to (k_1 - 1, k_2)$, there is a subset B of I_x such that either $|B| = k_2$ and all the 2-subsets of B are blue or $|B| = k_1 - 1$ and all the 2-subsets of B are red. In the first case we have the desired result. In the second $|B \cup \{x\}| = k_1$ and by *definition of* I_x, *all the 2-subsets of* $B \cup \{x\}$ are red. The case for II_x is essentially the same. □

There is nothing special here about *2-subsets* or *2-colourings*. There is a more general Ramsey's theorem concerning r-colourings of m-subsets. To facilitate discussion of this, we shall define $[A]^m$ to be the m-subsets of A. A subset B of A is called *homogeneous* for a *particular* r-colouring of A if $[B]^m$ is monochromatic.

The most general form of Ramsey's theorem asserts that given k, r, and m there is an n so large that for any r-colouring of $[\{1, \ldots, n\}]^m$ there is a homogeneous set of size k. This is written as

2.6. $$n \to (k)^m_r.$$

2.7. Ramsey's Theorem. $(\forall m, r, k)(\exists n)(n \to (k)^m_r)$.

There are many proofs of (2.7). Ramsey's original proof proceeds by an induction similar to the proof we gave of (2.4), but much more intricate. The reader should see [13] if he or she is interested in more details.

The original application of Ramsey's theorem (by Ramsey) was to a problem in formal logic. (For those readers who are familiar with logic, he used (2.7) to prove a sort of finite Lowenheim-Skolem Theorem.) This was essentially the state of Ramsey theory until the theorem (2.7) was rediscovered by P. Erdös and G. Szekeres [7].

Esther Szekeres observed that given 5 points in the plane with no three collinear, some 4 form a *convex* quadrilateral. Here a polygon is called convex if all of its internal angles are less than 180°. For example in Figure 1, polygon A is convex yet B is not as the angle at X is greater than 180°.

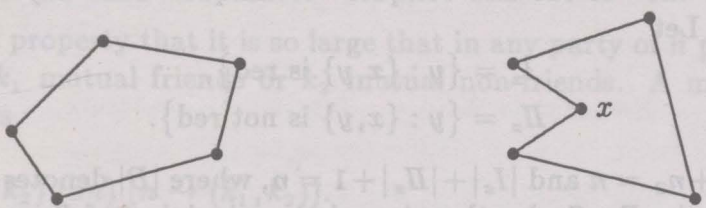

Figure 1

A natural question is to generalize this result to more than 4 points. Erdös and Szekeres rediscovered Ramsey's theorem to settle this question and showed that for any k there is an n so large that in any set of n points in the plane with no 3 collinear there is convex k-gon.

The original proof of this result used 2-colourings of 4-sets. We will not give this argument but give a simpler one due to Tarsy. At the time Tarsy was a student and apparently had failed to study this question for his exam. During the exam the question above was set. Tarsy got the question right, but with the following completely new proof. (I'm sure there is a moral here but ....) Let $n \to (k)^3_2$. Number any n points in the plane arbitrarily $1, \ldots, n$. Colour the 3-set $\{i,j,k\}$ red if $i < j < k$ and to get from i to j to k is clockwise, and blue if anticlockwise. Then the homogeneous set of size k must be convex.

A nice open question here is due to Erdös and Szekeres. If we let $n(k)$ denote the minimum number needed to guarantee a convex k-gon,

determine $n(k)$. It is known that

$$2^{k-2} < n(k) < \binom{2k-4}{k-2} + 1.$$

It is conjectured by Erdös, Szekeres and Klein that $n(k) = 2^{k-2} + 1$.

Another nice application of Ramsey's theorem is the following. Suppose $n \to (k)^2_2$. Then if I have a sequence of n distinct numbers $a_1, \ldots, a_n$, there is either an increasing subsequence of size k or a decreasing subsequence of size k. To see this, colour $\{a_i, a_j\}$ red if $i < j$ and $a_i < a_j$ and blue otherwise, and apply Ramsey's theorem. Again we have the theme that if n is large there is a well-behaved fragment. Here, however, optimal bounds are known. Using other pigeonhole type arguments, this result can be improved to say that in any sequence of $n^2 + 1$ distinct numbers, there is a monotone subsequence of size $n + 1$ (Erdös-Szekeres [7]). (The bounds given by Ramsey's theorem are more like 2^n rather than $n^2 + 1$). (This is also a consequence of Dilworth's theorem for readers familiar with that result).

§3. Classical Ramsey-type theorems

The last results of the last section suggest that Ramsey's theorem is not an isolated phenomenon. An easy generalization of Ramsey's theorem is to infinite sets. Thus if I colour the m-subsets of $\mathbb{N}$ in k colours, there is at least one infinite homogeneous set. A beautiful Ramsey-type theorem was given in 1927 by Van der Waerden solving a question of Schur (Van der Waerden [31]).

3.1. Theorem. *If I r-colour the positive integers then there are arbitrarily large monochromatic arithmetical progressions.*

In particular, if I colour the integers into 2 colours red and blue and give you a k, then either there is a blue arithmetical progression of size k or a red one of size k. There is a finite form of (3.1) which is the following.

3.2. Theorem. *Given k and r, there is an n so large that if we r-colour the integers $1, \ldots, n$ then there is a monochromatic arithmetical progression of size at least k.*

We denote the least such n by $W(k,r)$. To give the reader some idea as to how Van der Waerden proved (3.2) we will follow [13] and the case $k = 3$, $r = 2$. It is claimed that $W(3,2) \leq 325$. We would first break $[1, 325]$ into 65 blocks of length 5:

$$[1,325] = [1,5] \cup [6,10] \cup \cdots \cup [321,325]$$
$$= B_1 \cup B_2 \cup \cdots \cup B_{65}.$$

As these numbers are 2-coloured there are $2^5 = 32$ ways to 2-colour a block B_i. By the pigeonhole principle there are at least 2 blocks B_i and B_j in the first 33 blocks of the same colour, say B_{11} and B_{26}.

Consider $B_{11} = [51,55]$. Of the *first* 3 elements $\{51,52,53\}$ at least *two* must have the same colour, say j and $j+d$. Note that $j+2d$ belongs to B_{11} since $|B_{11}| = 5$. If j, $j+d$, $j+2d$ all have the same colour, done. Otherwise, without loss of generality, j, $j+d$ are red and $j+2d$ is blue. For example, we would have 51, 53, 54 red and 52, 55 blue. This means that in B_{26}, 126, 128, 129 are red and 127, 130 are blue. But then we are done: consider B_{41}. If $205(\in B_{41})$ is blue then 55, 130, 205 form a blue arithmetic progression and if 205 is red then 51, 128, 205 form a red arithmetic progression. We win no matter what the colour.

The basic idea here is to focus on B_i and B_j with $i < j \leq 33$ and B_i, B_j with the same colouring. We then argue that amongst B_i, B_j, $B_{i+2(j-i)}$ we get the arithmetic progression by a process of elimination as above. □

The ideas above can be generalized to give a proof of the full theorem. A nice account of how this theorem was discovered and how the general argument was found can be found in [32]. This is an excellent study for students of psychology of mathematical discovery. The reader might wonder what sorts of numbers the proof above gives for $W(k,r)$. For example it gives

$$W(3,3) \leq 7(2 \cdot 3^7 + 1)(2 \cdot 3^{7(2 \cdot 3^7 + 1)} + 1).$$

In general it gives astronomical bounds for $W(k,r)$. How astronomical? Is it important if the bounds are astronomical? We will return to these questions in the next section.

The reader may wonder if (3.1) can be improved to saying that there must be *infinitely long* arithmetical progressions of the same colour, rather than *arbitrarily* long (but finite) ones. The answer is no (exercise).

A powerful generalization of (3.1) was proved by Szemerédi. Szemerédi's theorem is called the *density version* of Van der Waerden's theorem and asserts that for any k there is an n so large that for any r-colouring of $\{1,\ldots,n\}$ there is a monochromatic arithmetic progression of size at least k in the *most common colour*. More precisely we say a set of natural numbers A has *positive upper density* if

$$\limsup \frac{|A \cap \{1,\ldots,n\}|}{n} > 0.$$

3.3. Theorem (Szemerédi [30]). *If A has positive upper density, then A contains arbitrarily long arithmetic progressions.*

This result had quite a long history. It was conjectured in 1936 by Erdös and Turan. In 1952, using analytic number theory, Roth [23] proved that if A has positive upper density then A contains a 3-element arithmetic progression and later in 1969 E. Szemerédi [29] improved this to 4-element arithmetic progressions. It was not until 1974 that Szemerédi proved 3.3 using very complex but elementary combinatorial arguments. Another proof of 3.3 was given in 1977 by Furstenberg using probabilistic methods – or rather, ergodic theory.

Erdös has conjectured that if A is a set of positive integers and

$$\sum_{a \in A} \frac{1}{a} = \infty,$$

then A contains arbitrarily long arithmetic progressions. This generalization of Szemerédi's result appears very difficult, since, in particular it would solve the very old question of whether the primes contain arbitrarily long arithmetic progressions. Incidentally, Erdös has offered US$3,000 for solving the conjecture above.

Van der Waerden's theorem was generalized in several ways. A central generalization of this type is the Hales-Jewett theorem. This result is very important as it is a purely combinatorial one about finite sets of integers and has many applications, only one of which is Van der Waerden's theorem. In many ways the most difficult aspect of the Hales-Jewett theorem is understanding its statement. The following method was suggested to the author by Carl Jockusch and is quite nice.

We let ${}^k n$ denote the set of all vectors $(x_1,\ldots,x_k)$ with $0 \leq x_i \leq n$. This is sometimes called the *k-dimensional n-cube*. In some sense it is

rather like a vector space except that there are no operations and we only have integers from 0 to n. What then would be the analogues of a line, plane, etc., in this setting. For vector spaces a d-dimensional subspace is the set of solutions to a set of equations with d degrees of freedom. Exploring this analogue in the primitive situation here, what can equations look like? As there are no operations for variables $x_1, \ldots, x_k$ equations can only say that

3.4. $\qquad x_i = x_j$ or $x_i = a$ for some constant $a \leq n$.

What then is the solution space for the equations of the form (3.4)? They can only be of the form $x_{i_1} = x_{i_2} = x_{i_3}$, $x_{j_1} = x_{j_2} = x_{j_3} = \cdots = x_{j_p}$ or $x_m = a$, etc. with "a" constant. Thus here "dimension" will denote *the number of blocks of variables*. A d-parameter subset of ${}^k n$ is a d-dimensional subspace in the sense above.

A 1-parameter set is called a *line*. For example if $n = 4$ and $k = 5$, consider

$$x_1 = x_3 = x_5, \qquad x_2 = 3, \qquad x_4 = 1.$$

The line generated by the equations above is $\{(0,3,0,1,0), (1,3,1,1,1), (2,3,2,1,2), (3,3,3,1,3), (4,3,4,1,4)\}$.

3.5. Theorem (Hales-Jewett [14]). *Given n and c there exists k so large that for any c-colouring of ${}^k n$ there is a monochromatic line.*

The Hales-Jewett theorem also has a natural extension to asserting for a given p there is a k with ${}^k n$ having a p-parameter subset. The case above is for $p = 1$. The Hales-Jewett theorem implies Van der Waerden's theorem as follows.

3.6. Theorem (Hales-Jewett [14]). *(3.5) implies (3.2).*

Proof. Given n and c where c is the number of colours and n is the desired length of the arithmetic progression, take k as in (3.5). Let $R = nk + 1$. Consider a c-colouring $\hat{d}$ of $\{1, \ldots, R\}$. We define a colouring d on ${}^k n$ by

$$d(x_1, \ldots, x_k) = \hat{d}(x_1 + \cdots + x_k + 1).$$

By (3.5) we have a monochromatic line in ${}^k n$. There are n points in such a line. We claim the image of this line under the induced map

above is the relevant arithmetic progression. Why? The line will count of a block of variables which vary from $0,\dots,n$ and all other coordinates will be constant. (For example, for the line $\{(0,1,0,3,0), (1,1,1,3,1), (2,1,2,3,2), \dots\}$ we have

$$\begin{aligned} -3 &= (0+1+0+3+0)-(1+1+1+3+1) \\ &= (1+1+1+3+1)-(2+1+2+3+2) \\ &\dots \end{aligned}$$

□

The Hales-Jewett theorem has a number of important and interesting consequences in, for example, complexity theory, as well as asymptotic combinatorics. One example in combinatorics is Gallai's theorem [see 13] which is a sort of generalization of Van der Waerden's theorem to higher dimensions. An extension of (3.5) was conjectured by Rota for vector spaces over finite fields. This conjecture was finally verified by Graham, Leeb and Rothschild.

3.7. Theorem (Graham, Leeb, Rothschild [12]). *Let F be a fixed finite field. For all $r, t, k \geq 1$ there exists an n so large that if the t-dimensional subspaces of F^n are r-coloured then there exists a k-dimensional subspace all of whose t-dimensional subspaces have the same colour.*

A simplified proof of (3.7) was discovered by Spencer [27].

To finish this section, we remark that it is unknown if there is density version of the Hales-Jewett theorem as follows:

Conjecture (L. Moser). *For all $t \geq 2$ and $\epsilon > 0$ there exists $N = N(t,\epsilon)$ such that if $n \geq N$ and $B \subseteq {}^t n$ has at least ϵt^n elements, then B contains a line.*

§4. Bounds and applications to logic and graph theory: unprovable theorems

In this section we will look at the bounds we get for the earlier results and some further Ramsey-type results.

A nice place to start here is with Gödel's incompleteness theorem. This is one of the great theorems of mathematics and proves that mathematics can never be mechanized. Without the language of logic this means

the following. By a system we mean a collection of axioms generated by a machine, together with a set of deduction rules. A proof is then a finite ordered set of statements each of which is either an axiom or a consequence of earlier lines by the deduction rules. Gödel's incompleteness theorem states roughly that for any system rich enough to capture a reasonable fragment of arithmetic, there is always a statement *true of the system* (and statable within the system) but not *provable within the system*.

This beautiful and powerful theorem destroyed forever the hopes of Hilbert's programme, which attempted to reduce all of mathematics to the production of a machine that would eventually calculate all the theorems of mathematics. The one problem with Gödel's theorem was that true but unprovable theorems were very uninteresting "mathematically" in the sense they were artificial statements mathematicians would not wish to prove in "ordinary" mathematics. In some sense the question became: Does Gödel's theorem matter to "ordinary" mathematics? This has obvious implications to the aspirations of the designers of automated theorem provers, etc. In recent years, various Ramsey-type theorems have shown that the answer is "yes" and in turn this rich intersection of logic and combinatorics yielded the Robinson-Seymour theorem [22] in graph theory, as we will now see.

The breakthrough came from the work of Paris and Harrington [19]. The system PA (*Peano Arithmetic*) is a very primitive one that can be described as "finite combinatorics". The axioms are not all that important, but express remarkable facts such as "0 exists", "if x exists so does $x+1$", "$x+y=y+x$, $xy=yx$", etc., together with simple logical rules of deduction. Roughly speaking if φ can be proved in "finite combinatorics" then φ can be proved in PA. Gödel's theorem applies to PA and hence there are φ in PA statements true of PA but not provable in PA.

Paris-Harrington[19] gave a mathematical example of the incompleteness of PA. To state the Paris-Harrington result, we will need a variation of the original Ramsey theorem. Define a set S to be *relatively large* if $|S| > \min(S)$. For example $S = \{31, 62, 78\}$ is not relatively large as $3 = |S| \not> \min(S) = 31$. The notation

$$n \underset{*}{\longrightarrow} (k)^m_r$$

means that for any r-colouring of $[\{1, \ldots, n\}]^m$ there is a *relatively large* homogeneous set of size k.

Going back to our party analogy, the idea is that each person has a unique number $\leq n$. A person thinks a set is big if his number is less

than the size of the set. The Paris-Harrington (PH) variant of Ramsey's theorem is

4.1. Theorem (Paris-Harrington). $(\forall m,r,k)(\exists n)(n \longrightarrow (k)^M_r)$.

That is for $m = r = 2$, in any party there is a set of k mutual friends or non-friends and *furthermore somebody in this set thinks the set is big*. The remarkable fact is that although Ramsey's theorem is easily proven in PA, Paris and Harrington showed that (4.1) cannot be proven in PA. In other words, any proof of (4.1) *essentially involves the use of infinite sets* (this can be made precise). Indeed, the proof that (4.1) is true uses in an essential way the *infinite* Ramsey theorem (which, although not provable in PA is not statable in PA either).

How should we *prove* that (4.1) is *not provable* in PA? Originally this was done using proof theoretic techniques from logic. Later it was realized that easier and more revealing proofs were possible by analyzing the lower bounds for the $n = n(m,r,k)$'s of (4.1).

We will now try to give a flavour of the proofs here. The Ackermann Hierarchy of functions is defined as follows. For any positive integer x,

$$f_1(x) = 2x,$$
$$f_{n+1}(x) = f_n^{(x)}(x)$$

where $f^{(x)}$ denotes the x^{th} iterate of f (under composition). For example $f_1(x) = 2x$, $f_2(x) = x2^x$,

$$f_3(x) \geq 2^{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}, \quad \text{a tower of height } x.$$

One can see the function grows explosively. This defines $f_n(x)$ for all $n = 1,2,3,\ldots$. With the use of set theory, it is possible to extend the finite integers by artificial constructs called adding *ordinals* beyond the finite ordinals (numbers). We add a symbol $\omega = \{0,1,2,\ldots\}$, together with its natural extensions $\omega+1, \omega+2, \ldots, \omega+\omega, \omega+\omega+1, \ldots$. If an ordinal γ is of the form $\beta+1$ we say γ is a *successor ordinal*, otherwise $\gamma = \cup_{\eta<\gamma}\eta$ and we say γ is a *limit ordinal*. We can extend multiplication, addition, etc., to the ordinals. The ordinals we are interested in are those that can be put into a form like, for example,

$$\omega^{\omega^{\omega^{\omega+1}}} + \omega^{\omega^{\omega+3}} + \omega.$$

where the exponentiation is finite.

These ordinals β are called *the ordinals below* ϵ_0. They have the property that there is countable set of ordinals $\beta[0], \beta[1], \ldots$ such that

$$\beta = \lim_{i \in \omega} \beta[i].$$

This allows one to extend the definition of the Ackerman Hierarchy to the *Grzegorczyk Hierarchy* of ordinals below ϵ_0. We do this as follows:

$$f_1(x) = x$$
$$f_{\alpha+1}(x) = f_\alpha^{(x)}(x) \quad \text{and}$$
$$f_\beta(x) = f_{\beta[x]}(x) \quad \text{for } \beta \text{ a limit ordinal.}$$

We then need only *one* result from logic. The one classical result from proof theory (due to Kreisel) we need is that if f is a function whose construction can be carried out in PA, then for some $\eta < \epsilon_0$, for all x, $f(x) \leq f_\eta(x)$.

Using a direct analysis of the size of growth of (the least) $n(k,m,r)$ of (4.1), Ketonen and Solovay showed that if $\hat{n}(2^k 3^m 5^r) = n(k,m,r)$, then

4.2. Theorem (Ketonen and Solovay [15]). *$\hat{n}$ grows faster than all of f_α for $\alpha < \epsilon_0$.*

In fact they showed that $\hat{n}(x)$ grows at exactly the same speed as $f_{\epsilon_0}(x)$. As a point of comparison the functions for the original Ramsey theorem is dominated by $f_4(x)$. At this stage, the Ketonen-Solovay and Paris-Harrington results revealed an avalanche of mathematical results that were independent of PA, or indeed systems much stronger than PA. Indeed, it was recently discovered that a theorem from the 40's was *already an example of the Gödel phenomenon for PA.*

Let b be a positive integer ≥ 2. Then any nonnegative integer n can be written uniquely as

$$n = b^{n_1} c_1 + \cdots + b^{n_k} c_k$$

where $k \geq 0$, $n_1 > n_2 > \cdots > n_k \geq 0$ and $0 < c_i < b$.

We can extend this by similarly writing the exponents $n_1, \ldots, n_k$, etc., yielding the *complete base b representation* of n. For example, for $b = 2$, following the presentation of Simpson [26], we have

$$266 = 2^{2^{2+1}} + 2^{2+1} + 2.$$

Let $R_b(n)$ be the integer that results from taking the complete base b representation of n and replacing each b by $b+1$. Thus

$$R_2(266) = 3^{3^{3+1}} + 3^{3+1} + 3.$$

We define the *Goodstein sequence* of n by

$$(n)_0 = n$$

$$(n)_{k+1} = \begin{cases} R_{k+2}\big((n)_k\big) - 1 & \text{if } (n)_k > 0 \\ 0 & \text{otherwise.} \end{cases}$$

For example

$$(266)_0 = 266$$

$$(266)_1 = 3^{3^{3+1}} + 3^{3+1} + 2$$

$$(266)_2 = 4^{4^{4+1}} + 4^{4+1} + 1$$

$$(266)_3 = 5^{5^{5+1}} + 5^{5+1}$$

$$(266)_4 = 6^{6^{6+1}} + 6^6 \cdot 5 + 6^5 \cdot 5 + \cdots + 6 \cdot 5 + 5$$

etc.

4.3. Theorem (Goodstein[11]). *For all n there is a k such that $(n)_k = 0$.*

To prove (4.3), Goodstein associated with the sequence $(n)_i$ a corresponding sequence $m(i)$ of infinite ordinals below ϵ_0 such that $m(i+1) < m(i)$ for all i. The result then easily follows from the set-theoretic axiom that the (countable) ordinals are well-ordered (that is in particular there are no infinite descending sequences of ordinals). Infinitary methods are necessary as (4.4) below shows.

4.4. Theorem (Kirby and Paris [16]). *Goodstein's theorem is not provable in PA. In fact $k(n)$ grows as fast as $f_{\epsilon_0}(n)$.*

Until recently, all known proofs of Van der Waerden's theorems or the Hales-Jewett theorem also seemed to give very large bounds. Indeed careful analysis of the *proofs* showed that the witnessing functions eventually exceeded f_η for all $\eta < \epsilon_0$. Therefore it seemed possible that perhaps these theorems too were examples of mathematical incompleteness in PA and the bounds really did grow that astronomically. Further work by Girard on Szemerédi's theorem also suggested this. It was therefore somewhat surprising that in 1988, Shelah proved:

4.5. Theorem (Shelah [24]). *The Hales-Jewett, Van der Waerden and vector space Ramsey theorems are provable in PA.*

The proof is elementary, but a little involved to include here. In some ways it is much easier than the original proofs. It is not known if the finite version of Szemerédi's theorem is provable in PA.

To finish this section we will examine one further outcome of this remarkable interaction of logic and combinatorics.

A finite tree will be taken to be a partially ordered set with a single root (the minimum) that forms a tree. The meet of nodes x and y on a tree T is denoted by $x \wedge y$. We say that a tree T_1 is *embeddable* into a tree T_2 (written $T_1 \to T_2$) if there is a 1–1 function f taking nodes to nodes such that $x \leq y \to f(x) \leq f(y)$ and $f(x \wedge y) = f(x) \wedge f(y)$. For example, in the trees below $T_1 \to T_2$ with f as indicated.

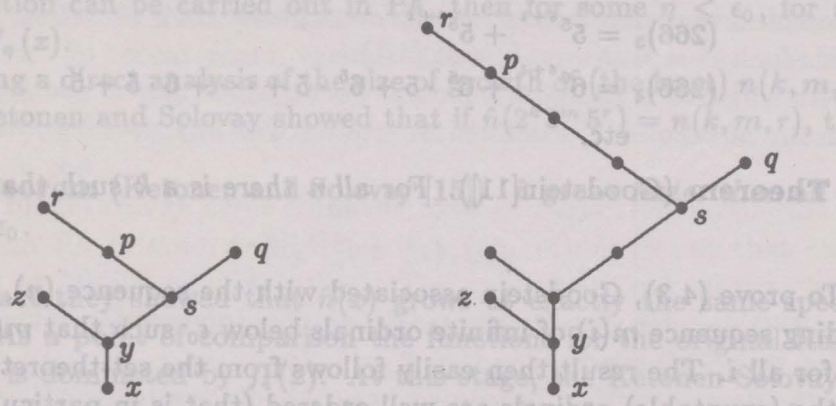

Figure 2

In [17] Kruskal showed that if $T_1, T_2, \ldots$ is any *infinite sequence of finite trees, there exist T_i and T_j with $T_i \to T_j$*. Harvey Friedman proved a finite form of Kruskal's theorem:

4.6. Theorem (Friedman see [26]). *For any c there exists $n = n(c)$ such that if $T_1, \ldots, T_n$ is any set of finite trees with $|T_i| \leq c \cdot i$, for all i, then there exists $k < j$ with $T_k \to T_j$.*

Again (4.6) is *provable equivalent to* (the infinite form of) Kruskal's theorem, and in particular cannot be proven in PA. In fact, Friedman however showed that any proof of (4.6) needs *uncountable sets*, (not just

infinite sets) a much stronger result. The growth rate of the $n(c)$ of (4.6) totally dwarfs f_{ϵ_0}!

Actually, a variant of (4.6) we shall call (FG) (Friedman's finite form of Kruskal's theorem with the gap condition) is not provable even in systems where (4.6) is provable and any system strong enough to prove virtually all theorems of classical analysis. The rate of growth of the witness function $n(c)$ for (FG) is startling. Friedman also has various combinatorial principles statable in PA whose truth or falsity is actually equivalent to exotic set-theoretic assumptions such as the existence of "n-Mahlo cardinals".

This remarkable interaction turned out to have spinoffs in combinatorics too. One is the following.

Let G be a finite graph. A *minor* of G is any graph obtained by deleting and contracting edges of G. In a long series of papers, Robertson and Seymour have shown that given any infinite set $G_1, G_2, \ldots$ of finite graphs, then there is $i < j$ with G_i a minor of G_j. A beautiful consequence of this result is a solution to Wagner's conjecture:

> *For any 2-manifold M there are only finitely many finite graphs not embeddable into M and are minimal with this property.*

This result is a generalization of Kuratowski's theorem that a graph is planar if and only if it does not have K_5 or $K_{3,3}$ as a minor. A crucial ingredient of the proof of the Robertson-Seymour theorem is the use of (FG). Friedman, Robertson and Seymour [9] have shown in fact that the Robertson-Seymour theorem is equivalent to (FG).

We refer the reader to Simpson [25] for an excellent account of this exciting area.

§5. Complexity Theory

In this section we will very briefly discuss some fascinating recent developments which have arisen from the use of Ramsey-type theorems in complexity theory. A fundamental problem which confronts us in complexity is to try to give tight lower bounds for algorithms for various processes. In general, it is often easy to work out a lower bound for a *specific* algorithm (e.g. via generating functions and recurrence relations). The real difficulty is to give a lower bound for a *problem* and hence *any possible algorithm* to solve it.

A classic example of this is the Hamilton circuit problem: Given a graph G decide if there is a circuit through the vertices that passes through each vertex exactly once. This is an example of what is called an NP-complete problem. There are many important "real life" examples of such problems. It can be shown that if we could solve one of these problems *fast*, for example in polynomial time (relative to the size of the graph), then we could solve *all of them fast*.

Unfortunately, we strongly suspect that we cannot solve any of this list of problems fast, and any known algorithm to solve them takes exponential time. In fact the best algorithms are not much better than trial and error. However, it may be that we are not smart enough. Maybe there is a fast algorithm and we have not yet found it. We would therefore like to *prove* that we cannot solve, for example, the Hamilton circuit problem in polynomial time. Our state of knowledge of such separation results (i.e., separating one complexity class from another) is appalling. It is only recently that we have been able to prove *nonlinear* bounds for various problems like the Hamilton circuit problem.

Many of these recent weak separation results uses delicate Ramsey-type arguments. The reason Ramsey theory seems to apply is as follows. Suppose we have a process P we feel takes a long time. Suppose further we feel that for P to be performed, any algorithm must at some stage implement a procedure Q (or one of a set of procedures) that we *know* takes a long time. The use of Ramsey-type arguments will be to show that to perform P faithfully we must perform Q somewhere.

We will content ourselves with one example and refer the reader to [2, 4, 6, 18, 21] for further examples. The example is due to Alon and Maass[1].

For a sequence $M = x_1 \dots x_m$ of length m with $x_i \in \{1, 2, \dots, n\} = N$ we shall say that an interval $x_i x_{i+1} \dots x_{i+j}$ is a *link between* disjoint sets S and $T \subseteq N$ if

(i) $x_{i+1}, \dots, x_{i+j-1} \notin S \cup T$ but

(ii) ($x_i \in S$ and $x_{i+j} \in T$) or ($x_i \in T$ and $x_{i+j} \in S$).

We shall say that M is a *meander* if for *any* two disjoint sets $S, T \subseteq N$ with $|S| = |T|$ there are in M at least $|S|$ links between S and T. More generally, if $g : N \to \mathbb{R}^+$ we call a sequence M a *g-meander* if between any two disjoint $S, T \subseteq N$ with $|S| = |T|$, there are at least $g(|S|)$ links between S and T. If g is the identity, a g-meander is a meander.

Intuitively, a meander is a sequence of integers that is rather random and wanders back and forth a great deal.

Let $X = x_1 \ldots x_r$ be a sequence of integers from a set N. For an ordered pair of distinct elements (a, b) of N define the *order type vector* $V(a,b) = V_x(a,b)$ as the vector obtained from x by replacing each occurrence of a by 0 and each occurrence of b by 1 and omitting all other numbers in x. Alon and Maass established the following Ramsey-theoretic result which is in the spirit of Hales-Jewett theorem.

5.1. Theorem (Alon and Maass[1]). *Let $X = x_1 \ldots x_r$ be a sequence in which each $a \in N$ appear exactly k times (so $r = nk$). Suppose that $N_1 \cup N_2$ is a partition of N. Then there exists $S \subseteq N_1$ and $T \subseteq N_2$ with $|S| \geq |N_1|/2^{2k-1}$ and $|T| \geq |N_2|/2^{2k-1}$ such that the set of all order type vectors $\{V_x(s,t) : s \in S, t \in T\}$ contains only one element.*

The proof of Theorem 5.1 is not difficult, and is by an induction similar in spirit to our proof of Ramsey's theorem in §3.

Using 5.1 it is not too difficult to obtain estimates on the sizes of g-meanders. For example Alon and Maass show that if $g(x)$ dominates $x \log x$ asymptotically, then the minimum length of a g-meander $(L_g(n))$ on $\{1, \ldots, n\}$ dominates $n \log n$. Furthermore if $g(x) \to \infty$ then $L_g(n)$ is superlinear in n.

Alon and Maass apply this work on g-meanders to what are called "superconcentrators" [20] and the related notion of *lower bounds for branching programmes* which we now examine.

A *branching programme* is a directed acyclic graph with a special vertex S (start) that has no ingoing edges and other special vertices called terminal vertices or *sinks*. A branching programme will compute a Boolean function on $x_1, \ldots, x_n$ as follows. All nonsink vertices are labeled by a variable x_i and all sinks by 0 or 1. Each nonsink vertex has two children ("fan-out 2"). These are labeled 0 or 1. Each assignment b of values to $x_1, \ldots, x_n$ defines a unique computation path through the programme. A programme computes f if for all assignments $b_1, \ldots, b_n$, $f(b_1, \ldots, b_n)$ is the label of the sink at the end of the path.

The *width* of programme is the maximum number of nodes on a level (= nodes of the same distance from S) and the *length* of a programme is the length of the longest computation path.

Clearly branching programmes provide a good and natural model of computation, and have attracted a great deal of attention. It is of great interest to obtain lower bounds for the time needed for various types of branching programmes to compute functions.

In particular, one open question from Borodin and Cook [2] is to find polynomial time computable functions f that cannot be computed in a linear length, polynomial width (= linear time, logarithmic space) branching programme.

This question is still open, but Alon and Maass solve the Borodin-Cook problem for *input oblivious programmes*. Here a programme is input oblivious if all non-sink vertices of the same distance from S have the same label. Alon and Maass use g-meanders to construct a number of functions that cannot be so computed. The detailed results are a little complicated to state but the basic idea is clear enough. If the width of the programme is constrained we would like to argue that the length m of the programme is great. To do this we show that constrained width allows us to construct a *meander-like sequence of length* m. Once we show that this can be done, Ramsey-theoretic results like 5.1 will show that m is large. As an example, by this technique Alon and Maass show

5.2. Theorem (Alon and Maass[1]). *Let* $T_k = T(x_1, \ldots, x_r)$ *be a Boolean function of* n *variables such that* $T_k(x_1, \ldots, x_n) = 1$ *if and only if* $x_i \geq k$. *Suppose* $\frac{1}{2} > \delta > 0$. *Then any input oblivious branching programme of width* w *that computes* T_k *for some* k *with* $n^\delta \leq k \leq n - n^\delta$ *has length at least of the order of* $\delta n \log n / \log w$.

We believe lower bound arguments await the development of the appropriate asymptotic combinatorics.

References

[1] N. Alon and W. Maass: Meanders and their applications in lower bound arguments, *J. Comput. and Systems Sci.* **37**(1988) 118-129.

[2] D. A. Barrington: Bounded width polynomial size branching programmes recognize exactly those languages in NC^1, in *Proceedings 18th ACM STOC 1986* 1-5.

[3] B. Bollobas: *Extremal Graph Theory*, Academic Press, New York 1976.

[4] A. Borodin and S. Cook: A space-time trade-off for sorting on a general sequential model of computation, *SIAM J. Comp.* **11**(1982) 287-297.

[5] W. Buchholz and S. Wainer: Provably computable functions and the fast growing hierarchy, in Simpson [25] 179-198.

[6] M. Dietzfelbinger and W. Maass: Lower bound arguments via "inaccessible" numbers, *J. Comput. System. Sci.* **36** (1988) 313-335.

[7] P. Erdős and G. Szekeres: A combinatorial problem in geometry, *Compositio Math.* **2**(1935) 464-470.

[8] P. Erdős and P. Turan: On some sequences of integers, *J. London Math. Soc.* **11**(1936) 261-264.

[9] H. Friedman, N. Robertson and P. Seymour: The metamathematics of the graph minor theorem, in Simpson [25] 229-262.

[10] H. Furstenberg: Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Anal. Math.* **31**(1977) 204-256.

[11] R. Goodstein: On the restricted ordinal theorem, *J. Symbolic Logic*, **9** (1944) 33-41.

[12] R. Graham, K. Leeb and B. Rothschild: Ramsey's theorem for a class of categories, *Adv. Math.* **8**(1972) 417-433.

[13] R. Graham, B. Rothschild and J. Spencer: *Ramsey Theory* , Wiley 1980.

[14] A. Hales and R. Jewett: Regularity and positional games, *Trans. Amer. Math. Soc.* **106**(1963) 222-229.

[15] J. Ketonen and R. Solovay: Rapidly growing Ramsey functions, *Ann. Math.* **113**(1981) 267-314.

[16] L. Kirby and J. Paris: Accessible independence results for Peano arithmetic, *Bull. London Math. Soc.* **14** (1982) 285-293.

[17] J. Kruskal: Well-quasi-ordering, the tree theorem and Vazsonyi's conjecture, *Trans. Amer. Math. Soc.* **95** (1960) 210-225.

[18] S. Moran, M. Snir and V. Manber: Applications of Ramsey's theorem to decision tree complexity, *J. Assoc. Comput. Mach.* **32** (1985) 938-949.

[19] J. Paris and L. Harrington: A mathematical incompleteness in Peano arithmetic, in *Handbook of Mathematical Logic* (J. Barwise, ed.), North Holland(1977) 1133-1142.

[20] N. Pippenger: Superconcentrators of depth 2, *J. Comput. System Sci.* **24**(1982) 82-90.

[21] F. Ramsey: On a problem of formal logic, *Proc. London Math. Soc.* **30** (1930) 264-286.

[22] N. Robertson and P. Seymour: Graph minors III: Planar tree width, *J. Comb. Theory (B)* **36** (1984) 49-64.

[23] K. Roth: Sur quelques ensembles d'entiers, *C. R. Acad. Sci. Paris*, **234** (1952) 388-390.

[24] S. Shelah: Primitive recursive bounds for Van der Waerden numbers, *J. Amer. Math. Soc.* **1** (1988) 638-697.

[25] S. Simpson (editor): *Logic and Combinatorics*, Amer. Math. Soc. Publ. Contemporary Mathematics vol 65, Providence, Rhode Island, 1985.

[26] S. Simpson: Unprovable theorems and fast growing functions, in Simspon [25] 259-394.

[27] J. Spencer: Ramsey's theorem for spaces, *Trans. Amer. Math. Soc.* **249** (1979) 363-371.

[28] A. Street and W. Wallis: *Combinatorics: a first course*, Charles Babbage, Winnipeg (1982).

[29] E. Szemerédi: On sets of integers containing no four elements in arithmetic progression, *Acta. Math. Acad. Sci. Hung.* **20** (1969) 89-104.

[30] E. Szemerédi: On sets of integers containing k elements in arithmetic progression, *Acta. Arith.* **27** (1975) 199-245.

[31] B. Van der Waerden: Beweis einer Baudetschen Vermutung, *Nieuw Arch. Wisk.* **15** (1927) 212-216.

[32] B. Van der Waerden: How the proof of Baudet's conjecture was found, in *Studies in Pure Mathematics* (L. Mirsky, ed.), Academic Press, (1971) 251-260.