Contents

1 Introduction

The theory of well-founded relations — recall that a relation R on a set X is well-founded if every subset $A \subseteq X$ has a minimal element with respect to R — is key to several areas of mathematics and computer science, since it allows us to perform *well-founded induction.* For example,

- in set theory, well-foundedness of set membership gives the Principle of ϵ-Induction, which can allow us define, for example, the ordinal rank of a set.
- in order theory, well-foundedness of the usual ordering on the ordinals gives us transfinite induction.
- In general, the fact that a given set of recursively-defined data structures is well-founded allows us to use structural induction to prove properties of all instances of the given data structure. Examples from computer science include lists and trees; in model theory it is used on the set of formulas in a given language to prove Łoś's theorem.

We are thus interested in studying well-founded relations in as much generality as is feasible. A very general class to consider are the quasiorders; reflexive and transitive relations. However, the property 'is a well-founded quasiorder on X' is often not preserved when we try to lift to a more complicated structure — in particular, to the power-set $P(X)$ where

$$A \leq B \iff \forall a \in A \exists b \in B : a \leq b.$$

To solve this problem, we restrict our attention to the class of well-quasiorders, which will turn out to be exactly those quasiorders $(X, \leq)$ for which $(P(X), \leq)$ is well-founded.

In 1930, Kurt Gödel proved his incompleteness theorems:

1. If a computable logical system, capable of describing the natural numbers, is consistent, then it not complete. That is, there is a true statement which is not proven by the system.
2. In particular, the statement "This system is consistent" cannot be proven inside the system.

The statements that Gödel builds in the course of proving these theorems are rather contrived; we must encode proofs and logical deduction in terms of natural numbers and arithmetic. The question of whether there is a 'natural' statement unproven by, say, the Peano Axioms was quickly raised. It took until 1977 for the first true example: the Paris-Harrington theorem states that the strengthened finite Ramsey theorem

> For every triple of positive integers n, k, m one can find N with the following property: if we colour each of the n-element subsets of $S = \{1, 2, 3, ..., N\}$ with one of k colours, then we can find a subset $Y \subseteq S$ with at least m elements, such that all n-element subsets of Y have the same colour, and the number of elements of Y is at least the smallest element of Y. [Wikipedia, accessed February 2020].

is unprovable in Peano Arithmetic (PA). In this essay I will develop the theory of well-quasiorders in order to prove Kruskal's tree theorem: the set of finite trees labelled with elements of a well-quasiorder is itself a well-quasiorder [3]. A suitably-chosen finite corollary of this (in much the same way that the strengthened finite Ramsey theorem is a corollary of the infinite Ramsey theorem) turns out also to be unprovable in PA.

Prerequisites

Some knowledge of order theory is assumed, including the basics of partial orders, total orders and well-orders, along with the theory of countable ordinals up to ε_0.

I also assume some knowledge of combinatorics, including the statement of Ramsey's theorem for k-partitions of $\mathbb{N}^{(2)}$. The proof that the infinite Ramsey theorem implies the finite Ramsey theorem may serve as a useful analogy to the finitisation proof in section 5 (Theorem 9).

Some familiarity with (directed) graph theory may help with visualisation, although I will be discussing the trees in question for the most part from an order-theoretic perspective.

2 Quasiorders and well-quasiorders

Definition 1. A quasiordering (or a preordering) $\leq$ on a set X is a reflexive and transitive relation on X. We call a set X equipped with such a relation a quasiorder (or a preorder).

We will write $a < b$ (and say 'a is strictly less than b') if $a \leq b \wedge b \not\leq a$. We will also write $a \not\lesseqgtr b$ (and say 'a and b are incomparable') if neither $a \leq b$ nor $b \leq a$.

Examples

- Quasiorders are very general objects: every partial order (and thus every total order) is a quasiorder with extra conditions:

  - a partial order also requires antisymmetry:

  $$\forall x, y \in X,\ x \leq y \wedge y \leq x \implies x = y$$

  - a total order further requires trichotomy:

  $$\forall x, y \in X,\ x \leq y \vee y \leq x.$$

  Indeed, the majority of quasiorders that come up naturally are partial orders, but the condition of being a partial order is not preserved under certain natural 'lifting' operations. For example, we might want to lift a quasiorder on a set X to the powerset $P(X)$ under the relation

  $$A \leq B \iff \forall x \in A\ \exists y \in B : a \leq b$$

  — that is, every element of A is 'dominated' by some element of B. Then this relation is reflexive and transitive (and thus the lift takes quasiorders to quasiorders), but the lift does not necessarily preserve antisymmetry. For example, taking $X = \mathbb{N}$, the sets $A = \mathbb{N}$, $B = \mathbb{N} \setminus \{0\}$ satisfy $A \leq B \leq A$ but $A \neq B$. This phenomenon, where the powerset lift does not preserve antisymmetry, can also be seen with similar constructions (the set of finite subsets of X, the set of trees labelled with elements of X, and so on). This gives some motivation for studying quasiorders, instead of the more familiar relations above.

- Note that every quasiorder can be turned into a partial order by taking a quotient by the equivalence relation $a \leq b \leq a$, but this is seldom a natural construction. For example, consider the set of subsets of $\mathbb{R}$ where $A \leq B$ if and only if there is an order-embedding $A \to B$; this relation is reflexive (the identity function is an order-embedding) and transitive (the composition of order-embeddings is an order-embedding). This quotient would identify, for instance $[0, 1]$ and $[0, 1)$, since we have the order-embeddings

  $$[0, 1) \to [0, 1] : x \mapsto x$$
  $$[0, 1] \to [0, 1) : x \mapsto \frac{x}{2}.$$

  But these sets differ in fundamental ways — the existence of a top element, for example! It is difficult to see the use of the quotient partial order in this case.

- Various contexts with a notion of 'embedding' admit a quasiorder interpretation. For example, we might take:
  - (some set of) groups, with $G \leq H$ exactly when there is an injective group homomorphism $G \to H$,
  - (some set of) topological spaces, with injective continuous maps,
  - (some set of) infinite graphs, with the subgraph relation, or the graph minor relation.

Well-foundedness and the good-pair condition

The principle of mathematical induction on the natural numbers $\mathbb{N}$ rests on the fact that, if a proposition fails for some natural number n, there must be a least natural number for which it fails. That is, it follows from the statement 'Every subset of the natural numbers has a least element'. We can generalise this property of the natural numbers to a general relation:

Definition 2. Given a set X, a relation $\leq$ on X is <u>well-founded</u> if every subset $A \subseteq X$ has a minimal element with respect to $\leq$. That is, for each $A \subseteq X$ there exists an $a \in A$ such that for every $b \in A$, $b \not\lneq a$.

Equivalently (given the Axiom of Dependent Choice, which I will assume for this essay), the relation is well-founded if it contains no countably-infinite descending chain $x_0 > x_1 > x_2 > ...$ in X.

Earlier, we saw that the property of being partial order is not necessarily preserved when taking the powerset lift, and so we study the weaker construction of a quasiorder. Now we will see that well-foundedness of a given quasiorders also need not be preserved under this lift.

For example, take $X = \mathbb{N}$, and consider it with the divisibility quasiorder $a \mid b \iff \exists m : am = b$. This is a well-founded relation: $a \mid b$ in X implies $a \leq b \vee b = 0$ in $\mathbb{N}$. Let $A = \{p_0, p_1, ...\}$ be the set of primes, and let

$$A_i := A \setminus \{p_0, ..., p_{i-1}\} \text{ for each } i \in \mathbb{N}.$$

Then in the lift $P(X)$ we have $A_0 > A_1 > A_2 > ...$, since for each $i \geq 1$,

$$\begin{aligned} & p_i \in A_{i-1} \setminus A_i \\ \implies & A_{i-1} \not\leq A_i. \end{aligned}$$

So $P(X)$ is not well-founded under this lift.

So what additional conditions can we place on a well-founded relation $\leq$ so that the lift to $P(X)$ (and ideally other similar lifts) is well-founded? In the above example, the thing which went wrong was that the primes formed an infinite set of incomparable elements (that is, an infinite antichain). Is this the only obstacle? (We shall see that the answer is 'yes'.)

Take a quasiorder X and consider sequences $\bar{a} : \mathbb{N} \to X$.

- A pair (a_i, a_j) is called good if $i < j$ as natural numbers and $a_i \leq a_j$ in X.
- The whole sequence is called good if it contains a good pair. Otherwise it is bad.

This allows us to define a stronger (as we shall see) notion than well-foundedness for our quasiorders.

Definition 3. A well-quasiorder X is a quasiorder for which every sequence $\bar{a} : \mathbb{N} \to X$ is good. (Henceforth we write 'wqo' for 'well-quasiorder'.)

Examples

- The natural numbers $(\mathbb{N}, \leq)$ with the usual order are a wqo — every well-order is wqo. The integers $(\mathbb{Z}, \leq)$ are not wqo, as the sequence of negative integers
$$0, -1, -2, -3, ...$$
is bad, and the naturals $(\mathbb{N}, |)$ under divisibility are not wqo, as the sequence of primes
$$2, 3, 5, 7, 11, ...$$
is bad. (These are in essence the only types of bad sequence; see Proposition 1).
- if $(X, \leq)$ is a wqo, then the finite product X^k with componentwise ordering is also wqo (See Proposition 4).
- If X is a finite set, the set X^* of finite strings of elements of X ordered by $a \leq b$ if and only if a is a subsequence of b (for example, $X = \{0, 1\}$, $a = 011$, $b = 01001$) is a wqo (this is called Higman's Lemma). This is a special case of Kruskal's Tree Theorem, which states that if Q is a wqo, then so is the set $T(Q)$ of finite trees labelled with elements of Q, under 'homeomorphic embedding' (See Section 3).

We cannot extend Higman's Lemma to the set of infinite sequences X^ω, or more generally to α-sequences for a given ordinal α, as this set is not a wqo in general. We can however restrict the class of well-quasiordering to so-called 'better-quasiorders' in order to generalise Higman's Lemma in this way.

The good-pair condition may seem somewhat contrived, but in fact it is equivalent to well-foundedness along with the 'no infinite antichains' condition.

Proposition 1. *Let A be a set with quasiorder $\leq$. Then the following are equivalent:*

(i) A is a well-quasiordering.

(ii) A contains no infinite strictly-decreasing sequence, nor an infinite sequence of pairwise-incomparable elements.

(iii) Every sequence $\bar{a} : \mathbb{N} \to A$ contains a non-decreasing subsequence $\bar{a}_u$.

Proof. We will show $(i) \implies (ii) \implies (iii) \implies (i)$.

$(i) \implies (ii)$ Let $\bar{a} : \mathbb{N} \to A$ be a sequence in A. By (i), $\bar{a}$ is good, so it contains a good pair $a_i \leq a_j$. Then because of this pair, $\bar{a}$ is neither an strictly-decreasing sequence, nor a sequence of pairwise-incomparable elements.

$(ii) \implies (iii)$ This implication may be shown with the following nice Ramsey-theoretic argument.

Given a sequence $\bar{a} : \mathbb{N} \to A$, partition the two-sets $\{i < j\} \in \mathbb{N}^{(2)}$ into three parts P_1, P_2, P_3, given respectively by the conditions $a_i \leq a_j$, $a_i > a_j$ and $a_i \not\lesseqgtr a_j$. Then Ramsey's theorem gives us a infinite monochromatic subset of $\mathbb{N}$.

But by (ii) this subset cannot be monochromatic in P_2 (giving a strictly-decreasing sequence), nor in P_3 (giving a sequence of pairwise-incomparable elements), and so it must be monochromatic in P_1. This is our non-decreasing subsequence $\bar{a}_u$.

$(iii) \implies (i)$ Let $\bar{a} : \mathbb{N} \to A$ be a sequence in A. By (iii), it contains a non-decreasing subsequence $\bar{a}_u$. In particular, $a_{u(0)} \leq a_{u(1)}$, and this is a good pair, so $\bar{a}$ is a good sequence.

□

This will allow is to show that the good-pair condition is precisely what we need to ensure the lift to $P(X)$ is well-founded.

Proposition 2. *Let X be a set with quasiorder $\leq$. Then X is a wqo if and only if the lift $P(X)$ is well-founded.*

Proof. In both directions we prove instead the contrapositive.

$\neg(i) \implies \neg(ii)$ Suppose X is not wqo, so we have a bad sequence $\bar{a} : \mathbb{N} \to X$. Define

$$A_i := \{a_j : j \geq i\}.$$

Then

$$A_0 > A_1 > A_2 > ...$$

is a strictly-decreasing sequence in $P(X)$ — if $A_i \leq A_j$ for some $i < j$, there is some $k \geq j > i$ such that $a_i \leq a_k$, contradicting the fact that $\bar{a}$ is bad.

$\neg(ii) \implies \neg(i)$ Conversely, suppose $P(X)$ is not well-founded. It then suffices by Proposition 1 to show that X contains either an infinite antichain or a strictly-decreasing sequence.

Since $P(X)$ is not well-founded, we have a strictly-decreasing chain of subsets

$$A_0 > A_1 > A_2 > ...;$$

take for each i some $a_i \in A_i$ such that $a_i \not\leq b$ for all $b \in A_{i+1}$. In particular, $a_i \not\leq a_{i+1}$, and in fact $a_i \not\leq a_j$ for all $j > i$. To see this, note that since $A_j \leq A_{i+1}$ there is some $b \in A_{i+1}$ with $a_j \leq b$. But then if $a_i \leq a_j$ we get $a_i \leq b$, which is impossible.

Now as we did earlier, we can apply Ramsey's theorem to the set $\{a_i : i \in \mathbb{N}\}$ to obtain either an infinite antichain or an infinite strictly-decreasing chain, as required (the conditions $a \leq b$, $a > b$, $a \nleqslant b$ form a trichotomy).

Hence X is not wqo.

□

3 The Minimal Bad Sequence

(The argument structure for this section is based on Forster [1].)

Every well-founded quasiorder X that is not a wqo must have at least one bad sequence (in fact, since all subsequences of a bad sequence are bad, it must have infinitely-many). Examining the notion of a minimal such sequence will allow us to write induction-style proofs that a given set is well-quasiordered.

Definition 4. Let X be a well-founded quasiorder which is not a wqo. A bad sequence $\bar{a} : \mathbb{N} \to X$ is a <u>minimal bad sequence</u> (an MBS) if for each $n \in \mathbb{N}$, a_n is minimal from the set

$$\{a \in X : \text{there is a bad sequence whose first } n \text{ terms are } a_0, ..., a_{n-1}, a\}.$$

That is, we first choose a_0 to be minimal from the set of elements which are first terms of bad sequences (this being non-empty since X is not wqo). Then we pick a_1 minimal from the set of elements for which (a_0, a_1) are the first two terms of bad sequences. We continue recursively; the resulting sequence $\bar{a} : \mathbb{N} \to X$ must be bad, since a good pair $a_i \leq a_j$ contradicts the definition of a_j.

For example, the naturals under division $(\mathbb{N}, |)$ are well-founded but not wqo, and the sequence of primes

$$2, 3, 5, 7, 11, ...$$

is an MBS. To see this, suppose the first i terms $(p_0, ..., p_{i-1})$ of the MBS have already been chosen. Then the set of elements which may be appended are

$$\{n \in \mathbb{N} : n \neq 1 \wedge p_j \nmid n \ \forall j < i\}.$$

The minimal elements in this set are those primes which have not already been chosen, and in particular we can choose the least such prime.

In the spirit of the ‘minimal counterexample’, we would like to say that every sequence which is in some sense ‘below’ an MBS must be a good sequence. And indeed we have the following lemma:

Lemma 3. *(The Minimal Bad Sequence Lemma) Let X be a well-founded quasiorder which is not wqo, and let $\bar{a} : \mathbb{N} \to X$ be an MBS as in Definition 4. Then the subset*

$$Y := \{y \in X : y < a_n \text{ for some } n \in \mathbb{N}\}$$

is wqo.

Proof. Let $\bar{b} : \mathbb{N} \to X$ be an arbitrary bad sequence in X. We show that there is some element of $\bar{b}$ which is not in Y.

Suppose for the sake of contradiction that every element of $\bar{b}$ is in Y; that is, suppose that for all i there is n such that $b_i < a_n$. Take a pair (i, n) with least possible n and consider the sequence

$$a_0, a_1, ..., a_{n-1}, b_i, b_{i+1}, b_{i+2}, ...$$

— it cannot be bad, or else a_n is not minimal among bad continuations of the initial segment $(a_0, a_1, ..., a_{n-1})$. Thus it contains a good pair, and this must be of the form $a_j \leq b_k$, since $\bar{a}$ and $\bar{b}$ are both bad.

But since $b_k \in Y$, there is some l with $b_k < a_l \implies a_j < a_l$, and by minimality of n we have $j < n \leq l$.

So in fact $a_j < a_l$ is a good pair, contradicting badness of $\bar{a}$. Thus $\bar{b}$ was not in Y, and so every sequence in Y is good.

Hence Y is wqo. $\square$

This lemma, as applied to the naturals under divisibility, is rather uninteresting — for all primes p the set of natural numbers below p is the singleton $\{1\}$, so in this case $Y = \{1\}$. This is indeed a wqo!

Nonetheless, in the general case this result is invaluable for proving that a given set $(X, \leq)$ is wqo. The general outline of such a proof proceeds by contradiction; we first show $\leq$ is reflexive, transitive and well-founded, so we can apply Lemma 3. Then, supposing we have an MBS $\bar{a}$, we show that the 'underneath' subset given by the Lemma is wqo, and use that to find a good pair in $\bar{a}$.

Well-quasiorders from well-quasiorders

The example of the power-set lift may prompt us to ask whether the condition of being wqo has implications for other constructions. As an example (which we will use later to prove Kruskal's Tree Theorem), we can consider the product of two well-quasiorders, or the set of finite subsets of a wqo.

Proposition 4. *Let A and B be wqo. Then the following are also wqo:*

(i) the product $A \times B$, given the ordering

$$(a, b) \leq (a', b') \iff a \leq b \wedge a' \leq b'.$$

(ii) the set $A^{(<\omega)}$ of finite subsets of A, given the ordering

$$B \leq C \iff \exists f : B \to C \text{ injective and non-decreasing.}$$

Proof. We will show (i), and use this result to prove (ii).

(i) Let $(\bar{a}, \bar{b}) : \mathbb{N} \to A \times B$ be a sequence in $A \times B$, with projections $\bar{a} : \mathbb{N} \to A$ and $\bar{b} : \mathbb{N} \to B$.

By Lemma 1, there is a non-decreasing subsequence $\bar{a}_u$ of $\bar{a}$, since A is a wqo. Since B is also a wqo, the corresponding subsequence $\bar{b}_u$ of $\bar{b}$ has a good pair $b_{u(i)} \leq b_{u(j)}$. Then $(a_{u(i)}, b_{u(i)}) \leq (a_{u(j)}, b_{u(j)})$ and so $(\bar{a}, \bar{b})$ is good.

So $A \times B$ is a wqo.

(ii) Note that the relation $\leq$ on $A^{(<\omega)}$ is reflexive (take $f = 1_B : B \to B$) and transitive (since the composition of non-decreasing functions is itself non-decreasing).

Moreover, it is well-founded: take a subset $\mathcal{A} \subseteq A^{(<\omega)}$, and let $n := \min\{|B| : B \in \mathcal{A}\}$. Since $B \leq C \implies |B| \leq |C|$, a minimal element among the finitely-many elements of size n is minimal in $\mathcal{A}$.

Hence we can take an MBS $\bar{B} : \mathbb{N} \to A^{(<\omega)}$. As the empty set is the minimum element in $A^{(<\omega)}$, none of the B_i is empty; pick $b_i \in B_i$ for each i, and write $C_i := B_i \setminus \{b_i\}$.

Note that $C_i < B_i$ (the inclusion is injective and non-decreasing). Then by the MBS Lemma (Lemma 3), the set

$$X := \{C_i \mid i \in \mathbb{N}\} \subseteq A^{(<\omega)}$$

is wqo.

Now, we know by (i) that $A \times X$ is a wqo, and thus that the sequence $(\bar{b}, \bar{C})$ is good. But a good pair $(b_i, C_i) \leq (b_j, C_j)$ yields a good pair $B_i \leq B_j$ in $\bar{B}$, contradicting the fact that $\bar{B}$ is a bad sequence.

Hence $A^{(<\omega)}$ is a wqo.

□

Corollary 5. *If $n \in \mathbb{N}$ and $(X, \leq)$ is a wqo, then $(X^n, \leq)$ with componentwise ordering is also wqo.*

Proof. Use induction on n and the proof of Proposition 4 (i). □

The proof of Higman's Lemma (the set of finite lists of elements from a wqo, under the relation

> $(a_0, ..., a_m) \leq (b_0, ..., b_n)$ if and only if there is a strictly increasing function $f : \{0, ..., m\} \to \{0, ..., n\}$ such that for all $0 \leq i \leq m$,
>
> $$a_i \leq b_{f(i)}.$$

is wqo) is almost identical to the proof of Proposition 4 (ii). For a contradiction, we pick an MBS and take a subsequence for which the heads of the lists form a non-decreasing sequence. Then the set of tails of sequences are wqo by the MBS Lemma, and so there is a good pair of tails, which together with the heads give a good pair in the MBS.

4 Trees and homeomorphic embedding

One structure to which we can lift a quasiorder is the finite (rooted) tree, which here we can consider as a generalisation of the finite list.

Definition 5. A finite (unlabelled) <u>tree</u> is a finite partially-ordered set t, whose elements are called <u>vertices</u>, such that

- t has a minimum vertex $r = \mathrm{root}(t)$, called the root of t, and
- for every $b \in t$, the set of vertices below b, $\{a : a < b\}$ (the under-set of b), is linearly-ordered.

In this way, we might say that trees 'look like lists when looking down'.

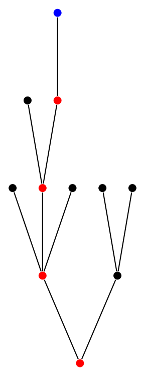

Figure 1: A tree, in which $a \leq b$ if there is a path upwards from a to b. Here the blue vertex has its under-set highlighted in red.

As Figure 1 suggests, there is a natural correspondence between finite trees as defined and 'rooted cycle-free graphs'; that is, a graph G with no cycles and a designated vertex r, called the root. We shall not need this description in this essay, although it is useful for visualisation.

We say 'a is the parent of b' if $a = \max\{x : x < b\}$ (which exists because the set is a finite linear order), and we say that 'b is a child of a' if a is the parent of b (see Figure 2). Note that a vertex can have multiple children.

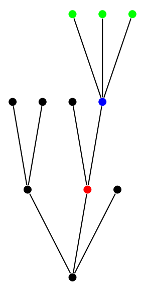

Figure 2: A vertex in blue: its children are in green, and its parent is in red.

Some authors choose to require that the children of each vertex be themselves linearly-ordered. The diagrams used here force a choice of ordering on each vertex's

Figure 3: Two drawings of the same tree.

children, in drawing them from left to right. Nevertheless, in this essay I will treat, for example, the two trees in Figure 3 as identical.

For a vertex $b \in t$, the branch at b is the subset $\{a : a \geq b\}$ of t with the induced partial ordering. This is itself a finite tree with root b. In fact, this allows for an inductive definition of trees:

> A tree is either a single vertex or a finite set of trees with a single vertex below them all.

A labelled tree (with labels in the quasiorder Q) is function $\tau : t \to Q$, where t is an unlabelled tree. We say 'a is a vertex of τ with label q' if $a \in t$, $q \in Q$ and $\tau(a) = q$.

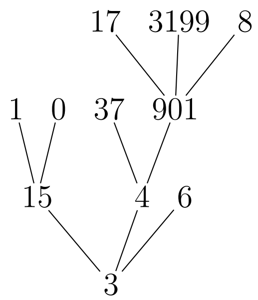

Figure 4: A tree labelled with elements from the quasiorder $Q = \mathbb{N}$.

Definition 6. A homeomorphic embedding (henceforth a map) $f : t \to u$ between finite trees is an injective function f satisfying, for all $a, b \in t$,

$$f(a \wedge b) = f(a) \wedge f(b),$$

where $a \wedge b$ is the infimum of a and b — that is, the greatest element in both their under-sets. If there is a map $t \to u$ write $t \leq u$; since the composition of maps is again a map, and the identity function is a map, the resulting relation $\leq$ is a quasiorder.

Figure 5: A tree homeomorphically embeds into another; vertices in the range are colored blue.

For labelled trees a <u>non-decreasing homeomorphic embedding</u> (henceforth also called a <u>map</u>) $f : \tau \to \upsilon$ is the corresponding notion: we require that f be a map, considered a a function $t \to u$ (ignoring labels), and that for every vertex a of τ, $\tau(a) \leq \upsilon(f(a))$.

This relation of (non-decreasing) homeomorphic embedding generalises the relation on lists we saw earlier, treating a list as a tree in which every vertex has at most one child.

Notice that a map f of unlabelled trees is an order-embedding:

$$\begin{aligned} a \leq b &\iff a \wedge b = a \\ &\iff f(a \wedge b) = f(a) \text{ since } f \text{ is injective} \\ &\iff f(a) \wedge f(b) = f(a) \\ &\iff f(a) \leq f(b). \end{aligned}$$

In particular, this means that if f is a surjective map, it is in fact an order-isomorphism.

5 Kruskal's Tree Theorem

We now have all the tools we need to prove the main theorem of this essay. First, so that we can use the MBS Lemma (Lemma 3), we must show that finite trees labelled by a wqo form a well-founded quasiorder under homeomorphic embedding. The identity function is a map, and the composition of two maps is again a map: suppose $f : \tau \to \upsilon$, $g : \upsilon \to \phi$ are maps. Then for $a, b \in \tau$,

$$\begin{aligned} g \circ f(a \wedge b) &= g(f(a \wedge b)) = g(f(a) \wedge f(b)) \\ &= g(f(a)) \wedge g(f(b)) \\ &= g \circ f(a) \wedge g \circ f(b). \end{aligned}$$

Also,

$$\begin{aligned} &\tau(a) \leq \upsilon(f(a)) \leq \phi(g(f(a))) \\ \implies &\tau(a) \leq \phi(g \circ f(a)). \end{aligned}$$

Thus it remains to show that the relation is well-founded.

Lemma 6. *Let Q be wqo. Then the set of finite trees labelled by Q, $T(Q)$, is well-founded under homeomorphic embedding.*

Proof. For a contradiction, suppose not. Then we have a strictly-decreasing chain in $T(Q)$

$$\bar{\tau} := (\tau_1, \tau_2, \tau_3, ...), \ \tau_1 > \tau_2 > \tau_3 > ...$$

Consider the underlying chain of unlabelled trees $t_i := \text{dom}(\tau_i)$. Then since $\mathbb{N}$ is well-founded and $t_i \geq t_j \implies |t_i| \geq |t_j|$, we have a subsequence of trees of equal size. But then, in this subsequence, the maps $t_i \to t_j$ are surjective, and thus order-isomorphisms.

Hence we may restrict to the case where $\text{dom}(\tau_i) = \text{dom}(\tau_j) := t$ for all $i, j \in \mathbb{N}$.

Let the vertices of t be $a_1, ..., a_n$, and consider for $i = 1, ..., n$ the sequence

$$\bar{a}_i : \mathbb{N} \to Q : k \mapsto \tau_k(a_i)$$

— that is to say, $\bar{a}_i$ is the sequence of labels at the vertex a_i. Since Q is wqo, by Lemma 1 there is a subsequence $\bar{\tau}_1 \subseteq \bar{\tau}$ such that the corresponding subsequence of $\bar{a}_1$ is non-decreasing. Inductively, if $\bar{\tau}_i \subseteq \bar{\tau}$ is such that the corresponding subsequence of $\bar{a}_j$ is non-decreasing for all $j \leq i$, by Lemma 1 there is a subsequence $\bar{\tau}_{i+1} \subseteq \bar{\tau}_i$ such that the corresponding subsequence of $\bar{a}_{i+1}$ is also non-decreasing.

Then if υ, υ' are the first two elements of the subsequence $\bar{\tau}_n$, we have $\upsilon(a) \leq \upsilon'(a)$ for all $a \in t$, and so $\upsilon \leq \upsilon'$. But these trees are in $\bar{\tau}$ so $\upsilon > \upsilon'$, which is a contradiction.

Hence in fact $T(Q)$ is well-founded under homeomorphic embedding. □

Theorem 7. *(Kruskal's tree theorem) The set of finite trees labelled by elements of a wqo Q, $T(Q)$, is itself wqo under homeomorphic embedding.*

Proof. The proof proceeds very similarly to part (ii) of Proposition 4. The use of Lemma 4 (ii) inspired by Nash-Williams' proof [4].

For a contradiction, suppose $T(Q)$ is not wqo. Then since $T(Q)$ is a well-founded quasiorder we can take an MBS $\bar{t} : \mathbb{N} \to T(Q)$. As Q is quasiordered, the sequence $\mathrm{root}(\bar{\tau}) : \mathbb{N} \to Q$ has a non-decreasing subsequence $\mathrm{root}(\bar{\tau})_u$ by Proposition 1 (iii).

Consider the corresponding sequence $\bar{\tau}_u$ in $T(Q)$, and define for each i the set A_i of branches at the children of the root of τ_{ui}. Define also

$$A := \bigcup_{i \in \mathbb{N}} A_i;$$

then for all $r \in A$, $r \in A_i$ for some $i \implies r < \tau_{ui}$. Thus by the MBS Lemma A is wqo.

Moreover, by Proposition 4 (ii) $A^{(<\omega)}$ is also wqo. So we have a good pair $A_i \leq A_j$, which is to say a non-decreasing function

$$f : A_i \to A_j.$$

Since $r \leq f(r)$ for all $r \in A_i$, we have maps $h_r : r \to f(r)$. This lets us define a map $h : \tau_{ui} \to \tau_{uj}$ as follows:

- $h(\mathrm{root}(\tau_{ui})) := \mathrm{root}(\tau_{uj})$,
- $h|_r := h_r$ for each branch $r \in A_i$.

But this means $\tau_{ui} \leq \tau_{uj}$, contradicting the fact that $\bar{\tau}$ is bad.

Hence $T(Q)$ is wqo. □

Finitisations of the tree theorem

The statement of Kruskal's tree theorem as we have just proven it is not entirely finitary in nature, as it involves quantifying over all infinite sequences of finite trees:

> For every infinite sequence of finite trees labelled by elements of a well-quasiorder Q, there are indices $i < j$ for which $t_i \leq t_j$.

We seek to replace this statement with a finitisation, which should be not much more complex than the original formulation, and yet should still be strong (in fact, it will still not be provable in Peano Arithmetic).

To prove this finitisation from Kruskal's tree theorem, we will need the following lemma, which is a consequence of the Axiom of Dependent Choice:

Lemma 8. *(Kőnig's Lemma) Every infinite tree t (where an infinite tree is defined as in Definition 5, replacing 'finite partial order' with 'infinite partial order' and requiring that each under-set be finite), in which every vertex has only finitely-many children, contains an infinite path*

$$(a_0, a_1, a_2, ...),$$

where $a_0 = \mathrm{root}(t)$ *and* a_{i+1} *is a child of* a_i *for every* $i \in \mathbb{N}$.

Proof. We define recursively a sequence $\bar{a}$ of vertices in t. First, set $a_0 := \mathrm{root}(t)$. The finite set of branches at the children of a_0 must contain at least one infinite tree, or else the underlying set of t is a finite union of finite sets — but t is infinite. So there is at least one child whose branch is infinite; let this child be a_1.

Likewise, the set of branches at the children of a_1 must contain at least one infinite tree, so pick a_2 to be a child whose branch is infinite. Continuing, we obtain an infinite path $(a_0, a_1, a_2, ...)$ as required. □

In order that our finitisation reference only finite objects, let us restrict our attention to the case of unlabelled trees (equivalently, to the case where Q is the singleton wqo). Kruskal's theorem states that every infinite sequence of trees contains a good pair $t_i \leq t_j$. Now, we cannot simply claim that all sufficiently-long finite sequences of finite trees contain a good pair — for every $n \in \mathbb{N}$, we can write down a sequence

$$t_0, ..., t_n$$

where $|t_i| = n - i$ for each i, and this sequence has no good pair. So we must also consider only certain finite sequences of trees. It turns out that constraining the size of the trees in our sequences is sufficient:

Theorem 9. *(Friedman's finite form of Kruskal's tree theorem) Let* $c, k \in \mathbb{N}$. *Then there is a sufficiently-large* $n \in \mathbb{N}$ *such that whenever* $t_0, ..., t_n$ *is a sequence of trees satisfying*

$$|t_l| \leq c \cdot (k + l) \text{ for all } 0 \leq l < k,$$

there are $i < j \leq n$ *such that* $t_i \leq t_j$.

Proof. For a contradiction, suppose not. That is, for every $n \in \mathbb{N}$ there is a sequence with no good pair. We construct a tree whose vertices are given by finite sequences $(t_0, ..., t_l)$ which have no good pair, ordered by the prefix relation

$$(t_0, ..., t_l) \leq (u_0, ..., u_m) \iff l \leq m \text{ and } t_i = u_i \text{ for all } 0 \leq i \leq l.$$

Since there are bad sequences of all finite lengths, this tree is infinite. Suppose $(t_0, ..., t_{l-1})$ is a vertex of this tree; there are only finitely-many trees of order at most $c \cdot (k+l)$, and so this vertex has finitely-many children. Hence by Lemma 8, there is an infinite path

$$\emptyset, (t_0), (t_0, t_1), (t_0, t_1, t_2), \ldots$$

— then the infinite sequence $\bar{t}$ given by $\bar{t}|_{[n]} := t_n$ has no good pair, contradicting Kruskal's tree theorem. $\square$

6 Independence proofs

It was shown by Gentzen in 1936 [2] that the Peano axioms are proven consistent by primitive recursive arithmetic along with the statement

$$WO(\varepsilon_0) := \text{the ordinal } \varepsilon_0 \text{ is well-ordered.}$$

(In fact, he showed that it suffices to assume the statement 'for all quantifier-free formulas $p(x)$, if there is an ordinal $\alpha < \varepsilon_0$ for which $p(\alpha)$ is false, there is a least such ordinal'.)

Here ε_0 is the first fixed point of the normal function $\alpha \mapsto \omega^\alpha$; it is the supremum of the set

$$\omega, \omega^\omega, \omega^{\omega^\omega}, \omega^{\omega^{\omega^\omega}}, \ldots$$

In this way we know that (if PA is consistent) PA cannot prove $WO(\varepsilon_0)$. Indeed, since PA interprets primitive recursive arithmetic, such a proof would imply that PA proves its own consistency, which is false by Gödel's second incompleteness theorem.

Thus, to show that a statement is independent from PA, it suffices to show that it implies $WO(\varepsilon_0)$, which appears much more tractable if we are to start from Kruskal's tree theorem.

[Aside: Since Peano Arithmetic and ZFfin (that is, the ZF axioms with the Axiom of Infinity replaced with its negation, and the Axiom of Foundation stated as the Principle of ϵ-Induction) are bi-interpretable — which is to say that they are essentially relabellings of each other — I will 'work in PA' by doing normal mathematics while only using finite objects.]

Tree representation of ordinals less than ε_0

Every ordinal less than ε_0 may be represented uniquely in its Cantor Normal Form:

$$\alpha = \omega^{\alpha_0} + \omega^{\alpha_1} + \ldots + \omega^{\alpha_n},$$

where $\alpha_0 \geq \alpha_1 \geq \ldots \geq \alpha_n$ are finitely-many ordinals, each strictly less than α.

Recursively expanding out the α_i in Cantor Normal form until nothing remains but 0 and ω^x yields a very tree-like structure:

$$\omega^{\omega\cdot 2+1} + 3 = \omega^{\omega^{\omega^0}+\omega^{\omega^0}+\omega^0} + \omega^0 + \omega^0 + \omega^0$$

and indeed this is the essence of how we will encode ordinals up to ε_0 as finite trees. If T is the set of finite trees, we define $F : \varepsilon_0 \to T$ as follows:

- we define $F(0)$ to be the singleton tree (call it 1_T), and
- given trees $F(\alpha_i)$ for $0 \leq i \leq n$ and $\alpha = \omega^{\alpha_0} + \omega^{\alpha_1} + ... + \omega^{\alpha_n}$, $F(\alpha)$ is the tree with branches $F(\alpha_0), ..., F(\alpha_n)$ joined to a single root.

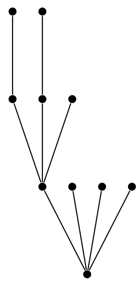

Figure 6: The tree corresponding to $\omega^{\omega\cdot 2+1} + 3$.

Now this encoding F is particularly nice — it is a bijection, and if $t \leq u$ under homeomorphic embedding, $F^{-1}(t) \leq F^{-1}(u)$ as ordinals.

To show that F is bijective, we can define its inverse. First, let the height of a tree, $\mathrm{ht}(t)$, be defined recursively by

- letting $\mathrm{ht}(1_T) = 0$, and otherwise
- the height of a tree is one more than the maximum of the heights of the branches at the children of the root.

Note that this defines height for all finite trees — for every finite tree, the size of its branches is strictly less than the size of the whole tree, and so the recursive definition terminates.

We now define $G : T \to \varepsilon_0$ similarly:

- $G(1_T) = 0$, and
- if $\mathrm{ht}(t) = k > 0$, let $S := \{s_0, ..., s_n\}$ be the set of the branches at the children of the root. Order them so that

$$G(s_0) \geq G(s_1) \geq ... \geq G(s_n);$$

note that G is already defined on the s_i since they each have height at most $k-1$. Then set

$$G(t) = \omega^{G(s_0)} + ... + \omega^{G(s_n)}.$$

Proposition 10. *F and G as defined above are inverses. Furthermore, if $t \leq u$, then $G(t) \leq G(u)$.*

Proof. Consider $GF : \varepsilon_0 \to \varepsilon_0$. We have $GF(0) = G(1_T) = 0$; let $\alpha = \omega^{\alpha_0} + \omega^{\alpha_1} + ... + \omega^{\alpha_n}$, and suppose $GF(\alpha_i) = \alpha_i,\ 0 \leq i \leq n$.

Then by definition $F(\alpha)$ is the tree with branches $F(\alpha_0), ..., F(\alpha_n)$ joined to a single root, and so

$$\begin{aligned} GF(\alpha) &= \omega^{GF(\alpha_0)} + ... + \omega^{GF(\alpha_n)} \\ &= \omega^{\alpha_0} + ... + \omega^{\alpha_n} \\ &= \alpha. \end{aligned}$$

So $GF = \mathrm{id} : \varepsilon_0 \to \varepsilon_0$.

Moreover, G is injective, by uniqueness and well-definition of Cantor Normal Form. So $GFG = G \implies FG = \mathrm{id} : T \to T$.

So $G = F^{-1}$. Now suppose $t \leq u$; that is, there is a map $f : t \to u$. We proceed by induction on $\mathrm{ht}(t)$.

If $\mathrm{ht}(t) = 0$ then $t = 1_T$ and $G(t) = 0 \implies G(t) \leq G(u)$. Otherwise, let $\mathrm{ht}(t) = k \leq 1$, let $s_0, ..., s_n$ be the set of branches at the children of $\mathrm{root}(t)$, and let $v_0, ..., v_m$ be the set of branches at the children of $\mathrm{root}(u)$ — both in G-decreasing order, as above. Without loss of generality, we may assume that $f(\mathrm{root}(t)) = \mathrm{root}(u)$; writing b for the branch of u at $f(\mathrm{root}(t))$ we have $G(b) \leq G(u)$, so $G(t) \leq G(b) \implies G(t) \leq G(u)$.

Given this, each s_i is mapped by f into some v_{j_i}, and since f maps $\mathrm{root}(t)$ to $\mathrm{root}(u)$, we get for every $i \neq j$ that

$$\begin{aligned} f(\mathrm{root}(s_i)) \wedge f(\mathrm{root}(s_j)) &= f(\mathrm{root}(s_i) \wedge \mathrm{root}(s_j)) \\ &= f(\mathrm{root}(t)) = \mathrm{root}(u). \end{aligned}$$

Hence f maps the s_i to distinct v_{j_i} (in particular, we have that $n \leq m$). By the induction hypothesis, $G(s_i) \leq G(v_{j_i})$ for each $0 \leq i \leq n$. Note that $G(s_0) \leq G(v_{j_0}) \leq G(v_0)$ by the chosen ordering on the v_j.

Now we have two cases: if $G(s_0) < G(v_0)$, the Cantor Normal Forms

$$G(t) = \omega^{G(s_0)} + ...,\ G(u) = \omega^{G(v_0)} + ...$$

imply that $G(t) < G(u)$. Otherwise $G(s_0) = G(v_0)$ and so $G(v_{j_0}) = G(v_0) \implies v_{j_0} = v_0$ by injectivity of G. Hence by relabelling we can arrange for f to map s_0 into v_0.

Now $j_1 \neq 0$ as $j_0 = 0$. Thus, in continuing, we find that $G(s_1) < G(v_{j_1}) \leq G(v_1)$. Comparing Cantor Normal Forms again, we find either that $G(t) < G(u)$ or that we can relabel so that f maps s_1 into v_1.

We can continue this procedure; either we show that $G(t) < G(u)$ or we find maps $s_i \to v_i$ for every $0 \leq i \leq n$. But then

$$\begin{aligned} G(t) &= \omega^{G(s_0)} + ... + \omega^{G(s_n)} \\ &\leq \omega^{G(v_0)} + ... + \omega^{G(v_n)} \\ &\leq \omega^{G(v_0)} + ... + \omega^{G(v_n)} + ... + \omega^{G(v_m)} \\ &= G(u). \end{aligned}$$

Hence $t \leq u \implies G(t) \leq G(u)$ as required. □

Now that we have this nice encoding, proving $WO(\varepsilon_0)$ from Kruskal's tree theorem is straightforward.

Theorem 11. *The implication 'Kruskal $\implies$ $WO(\varepsilon_0)$' is provable in PA.*

Proof. Take an arbitrary sequence of (notations for) ordinals below ε_0

$$\bar{\alpha} = (\alpha_0, \alpha_1, \alpha_2, ...).$$

By Proposition 10, for each of these ordinals there is a unique tree t_i with $G(t_i) = \alpha_i$, giving a corresponding sequence $\bar{t} = (t_0, t_1, t_2, ...)$. But then by Kruskal's tree theorem, there is a good pair $t_i \leq t_j$, which yields a pair $\alpha_i \leq \alpha_j$. Hence $\bar{\alpha}$ is not a strictly-decreasing sequence. So ε_0 is well-founded. □

Corollary 12. *Kruskal's theorem is not provable in PA.*

Moreover, it was shown by Simpson in 1985 [5] that the finitisation Theorem 9 also proves $WO(\varepsilon_0)$, although the proof is more involved. This is because one must show that, given a strictly-decreasing sequence in ε_0, we can build a strictly-decreasing sequence for which the size of the corresponding trees grows no faster than linearly.

Definition 7. Given an ordinal $\alpha < \varepsilon_0$, let its <u>size</u> $|\alpha|$ be the size of the corresponding finite tree $F(\alpha)$.

Call a strictly-decreasing sequence of ordinals $\bar{\alpha} = (\alpha_0, \alpha_1, ...)$ <u>slow</u> if there are $c, k \in \mathbb{N}$ such that

$$|\alpha_i| \leq c \cdot (k + i) \text{ for all } i \in \mathbb{N}.$$

We will sketch a proof of Simpson's result. The results I will assume are:

- To show $WO(\varepsilon_0)$, it is enough to show that there is no primitive recursive (PR) strictly-decreasing sequence in ε_0 (that is, a sequence definable by a primitive recursive function).
- For every primitive recursive function $f : \mathbb{N} \to \mathbb{N}$, there is a primitive recursive function $h : \mathbb{N}^2 \to \omega^\omega$ satisfying
  - $h(i,j) > h(i,j+1)$ for all $j < f(i)$,
  - $|h(i,j)| \leq c \cdot (i+j+1)$ for some constant c.

Reference for these can be found in [5]. Armed with these facts, we have the following lemma.

Lemma 13. *Let $\bar{\alpha} = (\alpha_0, \alpha_1, ...)$ be a strictly-decreasing PR sequence in ε_0. Then there is a slow strictly-descending PR sequence $\bar{\beta} = (\beta_0, \beta_1, ...)$ in ε_0.*

Proof. Since $\bar{\alpha}$ is PR, so is the function $f : n \mapsto |\alpha_n|$. Thus we have a PR function $h : \mathbb{N}^2 \to \omega^\omega$ by satisfying

- $h(i,j) > h(i,j+1)$ for all $j < f(i) = |\alpha_i|$,
- $|h(i,j)| \leq c \cdot (i+j+1)$ for some constant c.

With this h we can define the sequence $\bar{\beta}$ by

$$\beta_k := \omega^\omega \cdot \alpha_i + h(i,j),$$

where i, j are given by

$$k = \sum_{r=0}^{i-1} |\alpha_r| + j, \;\; j < |\alpha_i| \, .$$

Intuitively, $\bar{\beta}$ 'spreads out' the terms of $\bar{\alpha}$ so that the sizes of the ordinals in the sequence doesn't grow too quickly. Indeed, we find that

$$\begin{aligned} |\beta_k| &= |\omega^\omega \cdot \alpha_i + h(i,j)| \\ &\leq 4\,|\alpha_i| + c \cdot (i+j+1) \\ &\leq (c+4) \cdot (1+k). \end{aligned}$$

So $\bar{\beta}$ is a slow strictly-descending PR sequence in ε_0. $\square$

From here, the proof that our finitisation implies $WO(\varepsilon_0)$ is much the same as it was for Kruskal's tree theorem itself.

Theorem 14. *The implication 'Friedman's finite form of Kruskal $\implies WO(\varepsilon_0)$' is provable in PA.*

Proof. We work in PA. For a contradiction, suppose that in fact we have a PR strictly-decreasing sequence of (notations for) ordinals below ε_0

$$\bar{\alpha} = (\alpha_0, \alpha_1, \alpha_2, ...);$$

by Lemma 13 we can build a slow PR strictly-decreasing sequence

$$\bar{\beta} = (\beta_0, \beta_1, \beta_2, ...).$$

Now the corresponding sequence of finite trees $\bar{t}$ with $t_i := G(\beta_i)$ satisfies the conditions of Theorem 9, and so there is a good pair $t_i \leq t_j$. But then $\beta_i \leq \beta_j$, and hence $\bar{\beta}$ is not strictly-descending, which is a contradiction.

So in fact, Theorem 9 implies $WO(\varepsilon_0)$ as required. □

Corollary 15. *PA cannot prove Friedman's finite form of Kruskal's tree theorem.*

So finally we have a statement entirely about finite sets, but which is not provable in PA. Interestingly, although PA is not strong enough to prove Friedman's finite form of Kruskal (here given with $c = 1$ for brevity):

> $\forall k \ p(k) \equiv \forall k \exists m$ such that if $t_0, ..., t_m$ is a sequence of trees with $|t_r| \leq k + r$, then $t_i \leq t_j$ for some $i < j \leq m$,

it can prove $p(k)$, given a fixed k [6].

7 Conclusion

We arrived at the definition of well-quasiorder in refining the class of well-founded quasi-orders so that their well-foundedness is preserved in lifting the quasi-order to more complicated structures. By providing several equivalent definitions, we were able to show that this class is closed under several common constructions, including

- finite products under componentwise ordering,
- set of finite subsets or finite lists ordered by non-decreasing embeddings, and eventually
- the set of finite trees labelled by our original set and ordered by homeomorphic embedding.

This last result was strong enough to prove the consistency of Peano Arithmetic, even when reduced to a statement referencing only finite objects. In this way we found a comparatively-simple statement in finite combinatorics which is unprovable in Peano Arithmetic. Thus we see vividly that Gödel's incompleteness theorems do not only apply to unnatural and concocted statements, such that their implication may be ignored in the course of 'working' mathematics, but that they apply also to propositions encountered in other fields of mathematics.

I close by considering two directions in which the results I have presented here have been extended.

1. The technique of encoding ordinals by finite trees can be used to encode ordinals far larger than ε_0, and indeed almost any system for encoding countable ordinals uses trees. Simpson [5] describes an encoding up to the Feferman-Schütte ordinal Γ_0, defined to be the least ordinal not reachable by the Veblen hierarchy of normal functions ϕ_α:

   - $\phi_0(\beta) := \omega^\beta$,
   - $\phi_{\alpha+1}(\beta) := \beta^{\text{th}}$ fixed point of ϕ_α.

   That is, Γ_0 is the the least fixed point of the map $\beta \mapsto \phi_\beta(0)$. Note that in this notation $\varepsilon_0 = \phi_1(0)$.

2. The class of well-quasiorders $(X, \leq)$ is not closed under taking infinitary structures such as the set of α-sequences of elements of X for ordinals $\alpha \geq \omega$ (This would be a generalisation of Higman's Lemma). Attempting to refine the definition to obtain a class closed under such operations yield the <u>better-quasiorders</u>. Several quasi-orders have been shown to be wqo by proving them to be better-quasiorders: for example, the class of scattered linear orders (proven by Richard Laver, 1971).

References

[1] T. E. Forster, N. Bowler and M. Seisenberger *An Introduction to WQO and BQO Theory* `http://dpmms.cam.ac.uk/~tf/BQObok.pdf/` [accessed January 2020].

[2] G. Gentzen *Die Widerspruchsfreiheit der reinen Zahlentheorie* Mathematische Annalen, 112, 1936.

[3] J. B. Kruskal *Well-quasiordering, the tree theorem, and Vazsonyi's Conjecture.* Transactions of the American Mathematical Society, American Mathematical Society, 1960.

[4] C. St.J. A. Nash-Williams *On well-quasiordering finite trees* Proceedings of the Cambridge Philosophical Society, 1963.

[5] S. G. Simpson *Non-provability of certain combinatorial properties of finite trees* Harvey Friedman's Research on the Foundations of Mathematics, Studies in Logic and the Foundations of Mathematics, 1985.

[6] R. L. Smith *The consistency strengths of some finite forms of the Higman and Kruskal theorems* Harvey Friedman's Research on the Foundations of Mathematics, 1985.