Order (2011) 28:387–397
DOI 10.1007/s11083-010-9178-9

# Classification of Finite Coloured Linear Orderings

**Feresiano Mwesigye · John Kenneth Truss**

Received: 30 March 2010 / Accepted: 20 August 2010 / Published online: 7 September 2010


**Abstract** This paper concerns the classification of finite coloured linear orders up to $n$-equivalence. Ehrenfeucht–Fraïssé games are used to define what this means, and also to help analyze such structures. We give an explicit bound for the least number $g(m, n)$ such that any finite $m$-coloured linear order is $n$-equivalent to some ordering of size $\leq g(m, n)$, from which it follows that $g$ is computable. We give exact values for $g(m, 1)$ and $g(m, 2)$. The method of *characters* is developed and used.

**Keywords** Colours · Classification · Coloured linear orderings

**Mathematics Subject Classifications (2010)** 06A05 · 03C64

## 1 Introduction

We say that relational structures $A$ and $B$ are *n-equivalent*, written $A \equiv_n B$, if player $II$ has a winning strategy in the $n$-move Ehrenfeucht–Fraïssé game $G_n(A, B)$ on $A$ and $B$. In this game, players $I$ and $II$ move alternately, starting with $I$. On each move, $I$ chooses a point of $A$ or $B$ (not necessarily the same one each time), and $II$ replies by choosing a point of the other structure (which is as 'similar' to $I$'s choice as possible). After $n$ moves the players have between them chosen $a_1, a_2, \ldots, a_n \in A$ and $b_1, b_2, \ldots, b_n \in B$, and player $II$ wins if the map sending $a_i$ to $b_i$ for each $i$ is an isomorphism. Player $I$ wins otherwise. Note that $\equiv_n$ is automatically an equivalence relation on the class of structures of any finite relational language.

This paper is based on part of the first author's Ph.D. thesis at the University of Leeds [2], which was supported by a Commonwealth Scholarship.

F. Mwesigye · J. K. Truss (✉)
Department of Pure Mathematics, University of Leeds, Leeds LS2 9JT, UK
e-mail: pmtjkt@leeds.ac.uk

F. Mwesigye
e-mail: feresiano@yahoo.com

A coloured linear ordering is a triple $(A, <, F)$ where $(A, <)$ is a linear order and $F$ is a mapping from $A$ onto a set $C$ which we think of as a set of colours. In order that we can use the theory of Ehrenfeucht–Fraïssé games, we need to think of a coloured linear order as a relational structure, which is done by using (finitely many) unary predicates to stand for the colours (which must therefore pick out pairwise disjoint sets whose union is the whole set), though in practice this is more easily expressed by using the colouring function $F$. In this paper we consider the classification of finite $m$-coloured linear orders up to $n$-equivalence. The isomorphisms used in the corresponding Ehrenfeucht–Fraïssé games must therefore preserve colours as well as the ordering. We may think of finite coloured linear orders as 'words' (or finite strings) over an alphabet consisting of the colour set. If we (linearly) order the colours in some way, then the words of any fixed length may be ordered by the first point of difference, giving the 'lexicographic', or dictionary order. As an illustration of this, we shall remark in Lemma 3.1 that two coloured linear orders $A$ and $B$ are 1-equivalent if and only if the sets of colours appearing in $A$ and $B$ are same. Therefore a shortest representative of an $\equiv_1$-class will have each colour appearing exactly once, but this will not be unique unless we make an arbitrary choice, for instance by taking the lexicographically least.

By general considerations (see [5] for instance) it is known that for any finite relational language there are only finitely many $\equiv_n$-classes (which follows from the fact that up to logical equivalence there are only finitely many sentences of given quantifier depth). Hence there is a least number $g(m, n)$ such that for any finite $m$-coloured linear order $A$, $A \equiv_n B$ for some $m$-coloured linear order $B$ with $|B| \leq g(m, n)$. (We note that there are infinite linear orders (even without colours) which are 2-equivalent to no finite linear orders; see Lemma 3.2 for example.) It seems quite difficult however to give any precise information about the values of $g(m, n)$. It is our main aim to give an explicit (recursively defined) bound $f(m, n)$ for $g(m, n)$, and from this it follows easily that $g(m, n)$ is computable. The bound is almost certainly much larger than necessary, but getting a more accurate value would be a lot harder (though in Section 3 we give exact values for $g(m, 1)$ and $g(m, 2)$).

We remark on some related work. A detailed analysis of Ehrenfeucht–Fraïssé games on linear orderings was done by Ryan Bissell–Siders in [1], but his results are different in character and do not apply directly to coloured linear orderings. The decidability of the theory of finite $m$-coloured linear orders follows from the decidability of the weak second order theory of linear orders established in [4]. We also remark that if we define a (related) function $h(m, n)$ as the least number such that that every finite $m$-coloured linear order $A$ has a subset $B$ such that $A$ is $n$-equivalent to $B$ and $|B| \leq h(m, n)$, then our proof will also show that $g(m, n) \leq h(m, n) \leq f(m, n)$, though we do not know at present whether $g(m, n) = h(m, n)$ for all values of $n$ (it holds for $n = 1$ and 2 as is shown in Section 3).

Now the values of $g(m, n)$ are given inductively, and so we concentrate on the step from $g(m, n)$ to $g(m, n + 1)$. Assuming therefore that we know that $g(m, n)$ exists, we shall write $\Gamma(m, n)$ for the set of all finite $m$-coloured linear orders $A$ of size at most $g(m, n)$ that satisfy the following conditions:

(i) if $B$ is a finite $m$-coloured linear order such that $A \equiv_n B$, then $|A| \leq |B|$.
(ii) $A$ is the lexicographically least member of its $n$-equivalence class (subject to clause (i)).

Thus the members of $\Gamma(m,n)$ are representatives of the $\equiv_n$-classes of finite $m$-coloured linear orders of least size. As remarked above, $g(m,n)$ exists for each $m$ and $n$ (and is finite), and it follows that $\Gamma(m,n)$ is finite, though we can also deduce this from the proofs given below. We let $[A]_n$ stand for the member of $\Gamma(m,n)$ which is $n$-equivalent to $A$. For $a \in A$ we write $A^{<a}$, $A^{\leq a}$, $A^{>a}$, $A^{\geq a}$ for $\{x \in A : x < a\}$, $\{x \in A : x \leq a\}$, $\{x \in A : x > a\}$, and $\{x \in A : x \geq a\}$ respectively, and these subsets of $A$ (or indeed any others) are viewed as coloured linear orders under the structure inherited from $A$.

**Definition 1.1** Let $A$ be a finite $m$-coloured linear ordering, $n \geq 2$. Then the *$n$-character* of $a \in A$ is the ordered triple $\langle F(a), [A^{<a}]_n, [A^{>a}]_n \rangle$.

For any $A$, we let $\rho_n(A) = \{\langle F(a), [A^{<a}]_n, [A^{>a}]_n \rangle : a \in A\}$ and for $c \in C$, $\rho_n^c(A)$ be the set of members of $\rho_n(A)$ whose first co-ordinate is $c$. Thus $\rho_n(A)$ is equal to the disjoint union $\bigcup_{c \in C} \rho_n^c(A)$. If $A$ is a finite $m$-coloured linear ordering and $a \in A$, then we shall also write $\rho_n(a)$ for $\langle F(a), [A^{<a}]_n, [A^{>a}]_n \rangle$, and sometimes refer to *character* instead of $n$-character. Let $\langle [X]_n, [Y]_n \rangle \in \Gamma(m,n)^2$. Then $\langle c, [X]_n, [Y]_n \rangle$ is said to be an *$n$-character* of $A$ if $A^{<a} \equiv_n X$ and $A^{>a} \equiv_n Y$ for some $c$-coloured element $a \in A$.

The following result (which is a version of a theorem proved in [5] but here stated in terms of characters) gives a necessary and sufficient condition for two $m$-coloured linear orders $A$ and $B$ to be $n$-equivalent, which helps us reduce consideration of games with $n+1$ moves to games with $n$ moves, so enabling us to deal with $\Gamma(m,n)$ and $g(m,n)$ inductively. It applies whether the coloured orderings are finite or infinite, and will be used again in Section 3.

**Theorem 1.2** *$A \equiv_{n+1} B$ if and only if $\rho_n(A) = \rho_n(B)$.*

*Proof* Suppose that $A \equiv_{n+1} B$. Thus player $II$ has a winning strategy $\sigma$ in $G_{n+1}(A, B)$. Let $\langle X, Y \rangle \in \rho_n^c(A)$. This means there is a $c$-coloured element $a \in A$ such that $X \equiv_n A^{<a}$ and $Y \equiv_n A^{>a}$. Let player $I$ choose $a$ on his first move, and let $b \in B$ be $II$'s response using $\sigma$. Because $\sigma$ wins for $II$, $b$ must be a $c$-coloured element of $B$ such that $A^{<a} \equiv_n B^{<b}$ and $A^{>a} \equiv_n B^{>b}$. For as player $II$ wins in the remaining $n$ moves using $\sigma$, if either of these failed, $I$ could defeat $II$ by playing only on the relevant side of $a$ and $b$. We therefore find that $B^{<b} \equiv_n X$ and $B^{>b} \equiv_n Y$. Hence $\langle X, Y \rangle \in \rho_n^c(B)$. This shows that $\rho_n^c(A) \subseteq \rho_n^c(B)$ and the reverse inclusion is established similarly.

Conversely suppose that $\rho_n^c(A) = \rho_n^c(B)$ for all $c \in C$. We give a winning strategy for player $II$ in $G_{n+1}(A, B)$. If player $I$ chooses a $c$-coloured element $a \in A$ for some $c \in C$, then $a$ determines a character $\langle c, [A^{<a}]_n, [A^{>a}]_n \rangle \in \rho_n^c(A)$. Since $\rho_n^c(A) = \rho_n^c(B)$, also $\langle c, [A^{<a}]_n, [A^{>a}]_n \rangle \in \rho_n^c(B)$. There is therefore a $c$-coloured element $b \in B$ such that $A^{<a} \equiv_n B^{<b}$ and $A^{>a} \equiv_n B^{>b}$. Player $II$ plays $b$ on his first move, and thereafter uses his winning strategies in $G_n(A^{<a}, B^{<b})$ or $G_n(A^{>a}, B^{>b})$ to respond to whatever player $I$ plays, depending on whether $I$ plays points to the left or right of $a$ or $b$. This gives a winning strategy for player $II$, so $A \equiv_{n+1} B$. □

As illustration of the use of characters, we remark that we can easily use them to derive the classification of (monochromatic) finite linear orders up to $n$-equivalence (see [5]), which says that finite linear orders $A$ and $B$ are $n$-equivalent if and only if they either have equal size, which is less than $2^n - 1$, or they both have sizes $\geq 2^n - 1$ (which need not be equal). We omit the straightforward details.

This idea is extended to infinite ordinals in [3], and also to the coloured case. In this paper we concentrate however on the extension to finite linear orders (and general linear orders for $n \leq 2$), but now with finitely many ($m$) colours allowed. Unfortunately the classification is not (at present) as explicit as in the monochromatic case, and the best we can do is to give suitable bounds. These are sufficient however to establish the computability of the least bound $g(m, n)$ on the lengths of representatives for the $\equiv_n$-classes (of finite $m$-coloured linear orders), and also to show that the problem of determining an optimal $\equiv_n$ representative corresponding to any given finite $m$-coloured linear order is decidable. In Section 3 we give precise values for $g(1, m)$ and $g(2, m)$ for all $m$.

The following lemma will be needed at various points.

**Lemma 1.3** *Let $X$, $Y$, $A$ and $B$ be coloured linear orderings. If $A \equiv_n B$, then $X + A + Y \equiv_n X + B + Y$.*

*Proof* (i) By assumption, player $II$ has a winning strategy $\sigma$ in $G_n(A, B)$. He plays in $G_n(X + A + Y, X + B + Y)$ as follows. Whenever player $I$ plays in $X$ or $Y$, he just copies the move and plays the same point in the other structure. If player $I$ plays in $A$ or $B$, he uses $\sigma$ to provide his next move. This describes a strategy in $G_n(X + A + Y, X + B + Y)$ for player $II$, which wins since at most $n$ of the moves are in $A$ or $B$. The moves in $X$ and $Y$ precisely match, so as $\sigma$ wins in $G_n(A, B)$, the map produced must be a partial isomorphism. Hence $II$ has a winning strategy for $G_n(X + A + Y, X + B + Y)$. □

Note that it follows from this lemma that if $A \equiv B$, then $X + A + Y \equiv X + B + Y$ (though this is not needed in this paper).

## 2 Classification of Finite Coloured Linear Orderings

Let $a < b$ be elements of a coloured linear order $A$ such that $\langle F(a), [A^{<a}]_n, [A^{>a}]_n\rangle = \langle F(b), [A^{<b}]_n, [A^{>b}]_n\rangle$. Suppose that, for every $x \in A$ with $a < x \leq b$, there is $y \leq a$ such that $\langle F(x), [A^{<x}]_n, [A^{>x}]_n\rangle = \langle F(y), [A^{<y}]_n, [A^{>y}]_n\rangle$. Then we call $(a, b]$ an *unnecessary interval*. Otherwise it is a *necessary interval*. If $(a, b] \subseteq A$ is a necessary interval, we refer to any element $x \in (a, b]$ for which there is no corresponding element $y \leq a$ such that $\langle F(x), [A^{<x}]_n, [A^{>x}]_n\rangle = \langle F(y), [A^{<y}]_n, [A^{>y}]_n\rangle$ as a *new element* of $(a, b]$ and its character is called a *new character*. We note that $(a, b]$ is unnecessary if and only if it has no new character.

The following theorem is the key result in our inductive determination of $g(m, n)$.

**Theorem 2.1** *Let $A$ be a finite $m$-coloured linear order and let $a$ and $b$ be elements of $A$ such that $a < b$ satisfying the following conditions:*

(i) *$a$ and $b$ determine the same $n$-character, that is, $\langle F(a), [A^{<a}]_n, [A^{>a}]_n\rangle = \langle F(b), [A^{<b}]_n, [A^{>b}]_n\rangle$.*

(ii) $(a, b]$ *is unnecessary.*

*Then* $A$ *is* $n+1$*-equivalent to* $B = A - (a, b]$.

*Proof* By Theorem 1.2 it suffices to show that $\rho_n(A) = \rho_n(B)$.

If $x \leq a$, then $A^{<x} = B^{<x}$ and

$$\begin{aligned} A^{>x} &= (x, a] \cup A^{>a} \\ &\equiv_n (x, a] \cup A^{>b} \qquad \text{(by Lemma 1.3)} \\ &= B^{>x}. \end{aligned}$$

If $x > b$, then $A^{>x} = B^{>x}$ and

$$\begin{aligned} A^{<x} &= A^{<b} \cup [b, x) \\ &\equiv_n A^{<a} \cup [b, x) \\ &\cong A^{\leq a} \cup (b, x) \qquad \text{(since } F(a) = F(b)\text{)} \\ &= B^{<x}. \end{aligned}$$

Otherwise suppose that $a < x \leq b$, and let $y$ be given by the assumption that $(a, b]$ is unnecessary, so that $y \leq a$ with $F(x) = F(y)$. Then $A^{<x} \equiv_n A^{<y} = B^{<y}$, and $A^{>x} \equiv_n A^{>y} = (y, a] \cup A^{>a} \equiv_n (y, a] \cup A^{>b} = B^{>y}$. So this shows that $\rho_n(A) = \rho_n(B)$. Therefore $A \equiv_{n+1} B$. □

**Corollary 2.2** *If* $A$ *is a finite* $m$*-coloured linear ordering and* $P = |\rho_n(A)|$*, then the smallest* $m$*-coloured linear ordering that is* $(n+1)$*-equivalent to it has size at most* $P(P+1)$.

*Proof* Let $B$ be a finite $m$-coloured linear ordering of least size in the $\equiv_n$-class of $A$, and suppose for a contradiction that $B$ contains at least $P+2$ points having the same character. Then there are $P+1$ 'gaps' and one gap must fulfil the condition that no new character has appeared in that gap for the first time, since $P$ is equal to the total number of possible characters. By Theorem 2.1, such a gap can be cut out. We therefore deduce that $B$ contains at most $P+1$ such points. Hence $|B| \leq P(P+1)$. □

Now we can deduce from this result an explicit (though very large) bound $f(m, n)$ for $g(m, n)$ (and at the same time remark that it is also a bound for $h(m, n)$ defined in the introduction). Observe that since by definition, $g(m, n)$ is the greatest length of a member of $\Gamma(m, n)$, every member of $\Gamma(m, n)$ is an ordering of length at most $g(m, n)$, and so $|\Gamma(m, n)| \leq \sum_{q=0}^{g(m,n)} m^q$. In view of Corollary 2.2 this leads us to define $f(m, n)$ by:

$$f(m, 1) = m, \; f(m, n+1) = m \cdot \left( \sum_{q=0}^{f(m,n)} m^q \right)^2 \cdot \left( m \cdot \left( \sum_{q=0}^{f(m,n)} m^q \right)^2 + 1 \right).$$

**Theorem 2.3** *For every* $m, n \geq 1$*,* $g(m, n) \leq h(m, n) \leq f(m, n)$.

*Proof* We use induction on $n$. We can easily deduce from consideration of a 1-move game that all members of $\Gamma(m, 1)$ are $m$-coloured linear orders in which no colour is repeated (see Lemma 3.1). Hence $g(m, 1) = m = f(m, 1)$.

For the induction step, we observe that for any finite $m$-coloured linear ordering $A$, $P = |\rho_n(A)| \le m \times |\Gamma(m, n)|^2$. As remarked before the statement of the theorem, this is at most $m.(\sum_{q=0}^{g(m,n)} m^q)^2 \le m.(\sum_{q=0}^{f(m,n)} m^q)^2$ by induction hypothesis. By Corollary 2.2, $g(m, n+1) \le m \cdot (\sum_{q=0}^{f(m,n)} m^q)^2 \cdot (m \cdot (\sum_{q=0}^{f(m,n)} m^q)^2 + 1) = f(m, n+1)$.

The fact that $h(m, n) \le f(m, n)$ follows from the proof since the reduction of the ordering to length at most $f(m, n)$ is carried out by cutting out unnecessary intervals; and $g(m, n) \le h(m, n)$ is immediate. The reason that we cannot in general deduce from our proof that $g(m, n) = h(m, n)$ is that there may be a shorter coloured linear order $n$-equivalent to $A$ which is not actually a substructure of $A$. □

**Corollary 2.4**

(i) *$g(m, n)$ is a computable function.*
(ii) *There is an effective procedure for determining, for any finite $m$-coloured linear order $A$, the (unique) member of $\Gamma(m, n)$ which is $n$-equivalent to $A$.*

*Proof* We do the two parts simultaneously by induction on $n$.

If $n = 1$ then $g(m, n) = m$. Given $A$ we have to find the lexicographically least finite $m$-coloured linear order $[A]_1$ with the same colours appearing as for $A$, but only once each. For this we search through $A$ from the left, and delete any point having a colour which has already appeared. This produces an ordering of the correct length, and since it only has finitely many rearrangements, we can effectively find which is lexicographically least.

For the induction step, assume that we have computed $g(m, n)$, and that we have an effective procedure for finding the representative in $\Gamma(m, n)$ corresponding to any finite coloured linear order. We know the value of $f(m, n+1)$, and there are only finitely many $m$-coloured linear orders of length at most $f(m, n+1)$. List them all, and for each such $A$, using the effective procedure assumed inductively to exist, calculate the value of $\rho_n(A)$. Using Theorem 1.2, we can now tell which of these are $n+1$-equivalent, and hence select the one of minimal length and subject to that lexicographically least for each $\equiv_{n+1}$-class. The greatest length of such a choice gives the value of $g(m, n+1)$.

Finally, suppose we have an arbitrary finite $m$-coloured linear order $A$. Again using the procedure assumed inductively to exist, we can calculate $\rho_n(A)$, and hence find which of the members of $\Gamma(m, n+1)$ just determined it is $n+1$-equivalent to. □

## 3 Finite *m*-coloured Linear Orderings up to 2-equivalence

We now deal with the special case $n \le 2$, as here all the calculations are a lot easier, and we can obtain stronger conclusions, in fact for *all* coloured linear orders, not just the finite ones. Furthermore, in the finite case, we are able to give precise values for $g(m, 1)$ and $g(m, 2)$. We first justify the remark we have used in the previous section on 1-equivalence classes, applied to this more general case.

**Lemma 3.1** *For any finite colour set C, the* 1*-equivalence classes of all C-coloured linear orders correspond to the subsets of C. In fact, two C-coloured linear orders A and B are* 1*-equivalent if and only if they are coloured by the same subset of C.*

*Proof* If a colour appears in one of $A$ and $B$ but not the other, then $I$ can immediately win by playing a point of that colour, so that $A$ and $B$ are not 1-equivalent. If however, $A$ and $B$ exhibit precisely the same colours, then whatever $I$ plays in one of $A$ and $B$ can be matched by $II$. □

We remark that as a special case of this lemma, for $A$ 1-equivalent to $B$, if one of $A$ and $B$ is non-empty, then so is the other (and in the monochromatic case, this is also sufficient for 1-equivalence). We note that the order in which the colours appear is also of no consequence.

The point of stating this lemma is that it is immediate that even for infinite coloured linear orders, we have a fixed finite family of (short) finite representatives. This is no longer true for $n = 2$, but at least here we may use Theorem 1.2 to deduce what the possibilities are. They are still quite limited, and a key remark is that one cannot detect *density* of the ordering in just 2 moves. A method which works well in getting (reasonably) good bounds for $g(m, 2)$ (with similar and related results for general coloured linear orders) is to use induction on $m$ rather than $n$, though in fact in the finite case we argue directly to get the optimal value. We begin therefore by looking at the monochromatic case (also given in [5]), partly because this may also be done quite neatly by the use of characters.

**Lemma 3.2** *Any linear order is* 2*-equivalent to a unique order in the list* $0, 1, 2, 3, \omega, \omega^*, \omega^* + \omega$ *(where* $\omega$ *is the least infinite ordinal, and* $\omega^*$ *is the same set under the reverse ordering).*

*Proof* By Theorem 1.2 we just need to see what combinations of 1-characters can arise. Since we are in the monochromatic case, Lemma 3.1 tells us that the only possible characters are 0 and 1. Hence there are just four ordered pairs which can arise, and hence 16 sets of ordered pairs. However, several of these are impossible. In particular, there is a unique linear order in which $\langle 0, 0\rangle$ arises, namely 1, since this character tells us precisely that there is a point with nothing to left or right of it. In addition we cannot have just either $\langle 0, 1\rangle$ or $\langle 1, 0\rangle$ on its own, for in the first case this would say that there is a first but not last element, but any following element would exhibit a different character, and similarly for the second. This leaves just 7 possible sets, which correspond as shown to the 7 orders listed in the statement of the lemma:

$\emptyset$: 0,
$\{\langle 0, 0\rangle\}$: 1,
$\{\langle 0, 1\rangle, \langle 1, 0\rangle\}$: 2,
$\{\langle 0, 1\rangle, \langle 1, 1\rangle\}$: $\omega$,
$\{\langle 0, 1\rangle, \langle 1, 0\rangle, \langle 1, 1\rangle\}$: 3,
$\{\langle 1, 0\rangle, \langle 1, 1\rangle\}$: $\omega^*$,
$\{\langle 1, 1\rangle\}$: $\omega^* + \omega$. □

Before we derive the explicit value for $g(m, 2)$ we need a fairly easy combinatorial lemma. To formulate this, let us say that $T$ is an *m-configuration* if it is a finite linear order of the form $\{x_r : 1 \leq r \leq m\} \cup \{y_r : 1 \leq r \leq m\}$ where $x_1 < x_2 < \ldots < x_m$ and $y_1 > y_2 > \ldots > y_m$ and $x_1$ and $y_1$ are the least and greatest members of $T$ respectively. (We note that it is *not* required that $x_i \neq y_j$.) We associate with any $m$-configuration its 'weight' as follows. For each pair $(u, v)$ of consecutive elements of $T$, its *weight* is the greatest number $r$ such that $x_r \leq u$ and $y_r \geq v$. The *weight* of $T$ is then just the sum of the weights of all its consecutive pairs.

**Lemma 3.3** *The weight of any m-configuration is at most* $m^2$.

*Proof* Let us say that a pair $(r, s)$ is *out of order* if $x_r \geq y_s$. The proof goes by induction on the number of pairs that are out of order. If no pairs are out of order then $x_r < y_s$ for all $r$ and $s$, so

$$x_1 < x_2 < \ldots < x_m < y_m < \ldots < y_2 < y_1$$

The weights of the pairs, of which there are $2m - 1$, are $1, 2, \ldots, m - 1, m, m - 1, \ldots, 2, 1$ respectively, and their sum is $2\sum_{r=1}^{m-1} r + m = m^2$.

Now for the induction step, we suppose that there is at least one pair out of order in $T$, and find a revised $m$-configuration $T'$ having one less pair out of order. First suppose that $x_r = y_s$ for some $r$ and $s$. Let $T'$ be obtained from $T$ by letting $x_r < y_s$ (all other relationships being unchanged). Let $u$ and $v$ be the greatest element of $T$ less than $x_r$ and the least element of $T$ greater than $x_r$ respectively. (One of these is not defined if we are at an endpoint.) Then the contribution to the weight of $T'$ from $(u, x_r)$ and $(y_s, v)$ is unchanged from what it was in $T$, and the contribution from $(x_r, y_s)$ is $\min(r, s)$. By induction hypothesis, the weight of $T'$ is at most $m^2$, and it follows that the same is true for $T$. If there is no pair such that $x_r = y_s$ then all the $x_r$ and $y_s$ are distinct, and there is a pair such that $x_r > y_s$. For this value of $r$, choose the smallest possible value of $s$, and this means that $y_s < x_r < y_{s-1}$. For this value of $s$, choose the least value of $r$, and this means that $x_{r-1} < y_s < x_r < y_{s-1}$. Hence $(y_s, x_r)$ is a consecutive pair, and if $(u, y_s)$ and $(x_r, v)$ are consecutive, then $x_{r-1} \leq u$ and $v \leq y_{s-1}$. Let $T'$ be obtained from $T$ by swapping $x_r$ and $y_s$, and retaining all other relationships. Then we see that the sum of the weights for the pairs $(u, y_s)$, $(y_s, x_r)$ and $(x_r, v)$ in $T$ is $\min(r - 1, s) + \min(r - 1, s - 1) + \min(r, s - 1)$, but in $T'$ this is replaced by the pairs $(u, x_r)$, $(x_r, y_s)$ and $(y_s, v)$, for which the sum of the weights is $\min(r - 1, s) + \min(r, s) + \min(r, s - 1)$. Since by induction hypothesis, the weight of $T'$ is at most $m^2$, and the weight of $T$ is less than that for $T'$, it follows that the weight of $T$ is also at most $m^2$. □

**Theorem 3.4** *For each m,* $g(m, 1) = h(m, 1) = m$ *and* $g(m, 2) = h(m, 2) = m^2 + 2m$.

*Proof* First note that the value of $g(m, 1)$ follows at once from Lemma 3.1, and this is also equal to $h(m, 1)$ since any finite $m$-coloured linear order has a substructure in which the same colours all appear exactly once.

We first show that $g(m, 2) \leq m^2 + 2m$ and afterwards verify equality. So we let $A$ be a finite $m$-coloured linear order, which is of minimal length in its $\equiv_2$-class, and we suppose that all the $m$ colours actually arise. For $1 \leq r \leq m$ let $x_r$ and $y_r$ be the least and greatest points in $A$ such that $A^{\leq x_r}$ ($A^{\geq y_r}$ respectively) is coloured by exactly $r$

colours. Thus for instance $x_1$ is the least point of $A$, $y_1$ is the greatest, and it is clear that $x_1 < x_2 < \ldots < x_m$ and $y_1 > y_2 > \ldots > y_m$.

To illustrate the idea, we first suppose that $x_m < y_m$. Then we notice that all points of $(x_r, x_{r+1})$ have the same character, since on the left, precisely the same set of colours (of size $r$) arises for each of the points, and on the right, all colours arise. Similarly, all points of $(x_m, y_m)$ have the same character, and so do all points of $(y_{r+1}, y_r)$. So if we replace each of these intervals by a set of points in which precisely the same set of colours arises, but only once each, then by Theorem 1.2, the result is 2-equivalent to $A$. Since $A$ is minimal in its $\equiv_2$-class, this replacement actually results in no change, and we deduce that $|A| \leq 2m + 2\sum_{r=1}^{m-1} r + m = m^2 + 2m$ as desired.

The main complication in the proof comes about if $y_m \leq x_m$. But here we see that the subset $T = \{x_r : 1 \leq r \leq m\} \cup \{y_r : 1 \leq r \leq m\}$ of $A$ is an $m$-configuration, and furthermore, the optimality of $A$ in its $\equiv_2$-class tells us that on any interval of consecutive points of $T$, the greatest number of points that $A$ can have is equal to its weight (since the weight tells us the greatest number of colours available for this interval, and we can only have one occurrence of each). Therefore the number of points in $A$ is at most equal to $2m$ (the greatest size of $T$) added to its weight. By Lemma 3.3 this is at most $m^2 + 2m$. (This was done explicitly in the previous paragraph in the easiest case.)

Finally we establish optimality. Let $C = \{c_r : 1 \leq r \leq m\}$, and let the coloured linear order $A$ be the union of 'blocks' $L_1 < L_2 < \ldots < L_{m-1} < L_m < M < R_m < R_{m-1} < \ldots < R_2 < R_1$, where $|L_r| = r$ having points coloured $c_1, c_2, \ldots, c_r$ (in that order), $|M| = m$ with points coloured $c_1, c_2, \ldots, c_m$, and $|R_r| = r$ with points coloured $c_{m-r+1}, \ldots, c_m$. Then $|A| = \frac{1}{2}m(m+1) + m + \frac{1}{2}m(m+1) = m^2 + 2m$. So we just have to see that all points of $A$ realize distinct characters which will mean by Theorem 1.2 that it is minimal in its $\equiv_2$-class.

Now the points of $M$ have distinct colours, and they have all colours both to left and right. The points of $L_r$ have all colours to the right, and those of $R_r$ have all colours to the left. However, the points of $L_r$ do not have any $c_r$-coloured point to the left, and so do not have the same character as any point of $M$ or $R_s$. Now let $x \in L_r$ and $y \in L_s$ where $r < s$. Then $y$ has a $c_r$-coloured point to the left (the maximum element of $L_r$) but $x$ does not. So $x$ and $y$ have different characters. Distinct members of the same $L_r$ have different colours. Similar arguments apply to $M$ and $R_r$, concluding the proof.

Once again, the fact that $g(m, 2) = h(m, 2)$ follows from the fact that we can cut down any given $A$ to its optimal length by passing to a substructure. □

Now we move on to the general (not necessarily finite) case.

**Theorem 3.5** *For any linear order $A$, coloured by $m$ colours, there is a 2-equivalent $m$-coloured linear order having order-type of the form $A_0 + A_1 + \ldots + A_{k-1}$ for some finite $k$, where at most $m$ of the $A_i$ are $\omega^*$, at most $m$ of them are $\omega$, and the rest are finite. Conversely, there is an $m$-coloured linear order of order-type $m \times \omega^* + m \times \omega$ which is not 2-equivalent to any coloured order of smaller order-type.*

*Proof* We remark that the notation used here for lexicographic products is that $X \times Y$ stands for '$X$ copies of $Y$', which is contrary to that used for ordinals (which is the reversed lexicographic order, as in [2]).

We begin by proving by induction on $m$ that $A$ is 2-equivalent to a coloured linear order which is a subset of $(2^m - 1) \times \mathbb{Z}$. For $m = 1$, this follows by Lemma 3.2 (remembering that $\omega^* + \omega \cong \mathbb{Z}$).

Now we assume the result for $m$, and let $A$ be $(m+1)$-coloured. Let $L$ and $R$ be the sets of elements $a$ of $A$ such that $A^{\leq a}$, $A^{\geq a}$ respectively are coloured by $\leq m$ colours. By induction hypothesis, we may suppose that each of $L$ and $R$ has order-type at most $(2^m - 1) \times \mathbb{Z}$. If $L \cup R = A$ then the order-type of $A$ is at most $2 \cdot (2^m - 1) \times \mathbb{Z} \leq (2^{m+1} - 1) \times \mathbb{Z}$. Otherwise, let $M = A - (L \cup R)$. We replace $M$ by $M_1$ which is given by four possible cases:

If $M$ has no greatest or least, then $M_1$ is a copy of $\mathbb{Z}$, coloured periodically by the colours which appear as colours of elements of $M$ (so that if the colours are $\{c_i : 0 \leq i < k\}$ then for each $n \in \mathbb{Z}$, $kn + i$ is coloured by $c_i$, for instance).

If $M$ has a least but no greatest, then $M_1$ is a copy of $\omega$, whose first point is coloured as the first point of $M$, and after that it is coloured periodically by the colours which appear as colours of elements of $M$.

If $M$ has a greatest but no least, then $M_1$ is a copy of $\omega^*$, whose last point is coloured as the last point of $M$, and before that it is coloured periodically by the colours which appear as colours of elements of $M$.

If $M$ has a least and greatest, then $M_1$ is finite, with least and greatest coloured as the corresponding elements of $M$, and otherwise, just one point of each possible colour (as in the proof of Theorem 3.4).

We can now appeal to Theorem 1.2 to see that the resulting coloured ordering is 2-equivalent to $A$. This is because for all points $a$ of $M$ except the greatest and least (if they exist) the value of $\langle [A^{<a}]_1, [A^{>a}]_1 \rangle$ is the same, as all colours appear to their left and right, and so we just have to retain either of the endpoints if present, and ensure that all colours appear in between. If neither endpoint exists then we have to insert an ordering of type $\mathbb{Z}$ (as we must avoid inserting an endpoint) and similarly if there is just one endpoint (with $\omega^*$ or $\omega$). Since the order-type of $M_1$ is at most $\mathbb{Z}$, it also follows that we have found a coloured ordering as required of order-type at most $(2^{m+1} - 1) \times \mathbb{Z}$.

Now therefore assume that $A \subseteq (2^m - 1) \times \mathbb{Z}$, and write $A$ as a sum of the form $A_0 + A_1 + \ldots + A_{k-1}$ where each $A_i$ is either finite, or ordered in type $\omega$ or $\omega^*$. Using similar arguments as above, we can further reduce $A$ by cutting each $A_i$ down so that at most $m$ of the $A_i$ are of type $\omega$, and at most $m$ of them are of type $\omega^*$. If $A_i$ is finite, then we make no change. We concentrate on showing how to handle a typical $A_i$ which is of type $\omega$ (and $\omega^*$ is done similarly). Consider the set $C_i$ of colours which occur cofinally in $A_i$. If some member of $C_i$ does not occur as the colour of any point of $A$ greater than $A_i$, then we again leave $A_i$ unchanged. Otherwise, we form $A'_i$ by removing a final segment of $A_i$ coloured only by members of $C_i$, and in such a way that there are $x < y < z$ in $A'_i$ where $z$ is the greatest member of $A'_i$ such that $C_i$ is exactly equal to the set of colours appearing in each of $(x, y]$ and $(y, z]$. The main point is to show that if $A'$ is the result of replacing $A_i$ by $A'_i$ (and leaving all other $A_j$s unchanged), then $A \equiv_2 A'$. For this we appeal to Theorem 1.2, and it suffices to remark that for each point $a$ of $A$, there is some $a' \in A'$ having the same colour as $a$ such that $A^{<a}$ and $A'^{<a'}$ exhibit the same colours, as do $A^{>a}$ and $A'^{>a'}$, and similarly for the converse statement. If $a \in A'$, then we may take $a' = a$ (and this also works for the converse), and if $a \in A - A'$, then we may take a point lying in $(y, z]$.

By repeating the argument just given, we may suppose that for each $i$, if $A_i$ has order-type $\omega$, then some colour which occurs cofinally in $A_i$ does not occur to the right of $A_i$, and similarly, if $A_i$ has order-type $\omega^*$, then some colour which occurs coinitially in $A_i$ does not occur to the left of $A_i$. Now for each colour $c$ there is at most one value of $i$ such that $c$ occurs cofinally in $A_i$ and not to the right of $A_i$, and it follows that there are at most $m$ such values of $i$ such that $A_i$ has order-type $\omega$, and similarly for $A_i$ of order-type $\omega^*$.

To establish the final sentence, we can just consider each of the $m$ copies of each of $\omega^*$ and $\omega$ to be coloured by distinct colours. □

We remark that where we have replaced $M$ by an ordering coloured periodically, in fact any colouring in which all the colours of $M$ (except of the least and greatest points if they exist) appear coterminally (or cofinally/coinitially) would do just as well. The reason for colouring periodically is to give an explicit solution, and so that we have a finite and 'easily described' set of representatives. (One could formalize this by saying that the function producing the representative is 'computable' in an appropriate sense.)

## References

1. Bissell-Siders, R.: Ehrenfeucht–Fraïssé games on linear orders, logic, language, information and computation. In: Lecture Notes in Computer Science, vol. 4576, pp. 72–82. Springer, Berlin (2007)
2. Mwesigye, F.: Elementary equivalence of linear orders and coloured linear orders, Ph.D. thesis, University of Leeds (2009)
3. Mwesigye, F., Truss, J.K.: Finitely coloured ordinals (in preparation)
4. Rabin, M.O.: Decidability of second-order theories and automata on infinite trees. Trans. Am. Math. Soc. **141**, 1–15 (1969)
5. Rosenstein, J.G.: Linear Orderings. Academic, New York (1982)