CrossMark

On Optimal Representatives of Finite Coloured Linear Orders

F. Mwesigye[1,2] · J. K. Truss[1] (ID)

Abstract Two structures A and B are n-equivalent if player II has a winning strategy in the n-move Ehrenfeucht-Fraïssé game on A and B. We extend earlier results about n-equivalence classes for finite coloured linear orders, describing an algorithm for reducing to canonical form under 2-equivalence, and concentrating on the cases of 2 and 3 moves.

Keywords Coloured linear order · Ehrenfeucht-Fraïssé game · Optimality · Classification

1 Introduction

In [1] we studied the equivalence of finite coloured linear orders up to level n in an Ehrenfeucht-Fraïssé game, written as $\equiv_n$. This means that player II has a winning strategy in an n-move Ehrenfeucht-Fraïssé game. We also made some remarks about the infinite case. We gave some bounds for the optimal representatives in the finite case, meaning ones of minimum length, and the infinite case for up to 2 moves. These results were extended in [2] to all coloured ordinals, in the monochromatic case giving a precise list of optimal representatives, and in the coloured case giving bounds.

In this paper we return to the finite case, and extend the work of the first paper, by improving the bounds in some instances, and throwing further light on uniqueness of representatives. First we briefly recall the required definitions. A *coloured linear ordering*, also

✉ J. K. Truss
pmtjkt@leeds.ac.uk

F. Mwesigye
fmwesigye@must.ac.ug

1 Department of Pure Mathematics, University of Leeds, Leeds LS2 9JT, UK

2 Department of Mathematics, Mbarara University of Science and Technology, Mbarara, Uganda

called a *coloured string*, is a triple $(A, <, F)$ where $(A, <)$ is a linear order and F is a mapping from A onto a set C of 'colours'. We just write A instead of $(A, <, F)$ provided that the ordering and colouring are clear, and if $X \subseteq A$, we write $F(X)$ for $\{F(x) : x \in X\}$. In the n-move Ehrenfeucht-Fraïssé game on coloured linear orders A and B players I and II play alternately, I moving first. On each move I picks an element of either structure, and II responds by choosing an element of the other structure. After n moves, I and II between them have chosen elements $a_1, a_2, \ldots, a_n$ of A, and $b_1, b_2, \ldots, b_n$ of B, and player II *wins* if the map taking a_i to b_i for each i is an order and colour-isomorphism, and player I wins otherwise. We say that A and B are n-*equivalent* and write $A \equiv_n B$, if II has a winning strategy. Then $\equiv_n$ is an equivalence relation having just finitely many n-equivalence classes. An *optimal representative* is a member of an equivalence class of least possible length (which is chosen lexicographically least if there is more than one such, that is according to dictionary order assuming some fixed ordering of the set of colours), and a string or coloured string is called *optimal* if it is optimal in its equivalence class.

In [1] we gave upper bounds for the lengths of optimal representatives of $\equiv_n$-classes of m-coloured finite linear orders. Only for the case $n = 2$ were these bounds exact. We return to this case, describing explicitly the classification of finite m-coloured linear orders up to $\equiv_2$-equivalence (based on the idea of an 'm-configuration' introduced in [1]). From this we are able to read off which equivalence classes are finite or infinite, and provide an algorithm for determining an optimal representative corresponding to any given finite m-coloured linear order. We also show that a finite coloured linear order is $\equiv_2$-optimal if and only if each 1-character (see below for the definition) appears at most once.

The problem for more than 2 moves seems to be quite hard, so we concentrate on the case of 3 moves. The idea is that using a key inductive lemma from [1], we need to understand better how the 2-characters behave, and that is the reason for re-examining the case $n = 2$ in more detail.

Next we recall the notion of 'character' from [1], and the main result about characters. Let us fix an ordering of the set of colours. We write $\Gamma(m, n)$ for the family of all finite m-coloured linear orders A such that A has least length in its $\equiv_n$-class, and subject to this, A is lexicographically least (with respect to the ordering of the colours chosen). Since there are only finitely many $\equiv_n$-classes, $\Gamma(m, n)$ is finite. We may therefore define $g(m, n)$ to be the maximum of the lengths of members of $\Gamma(m, n)$. We write the representative for A as $[A]_n$. In a coloured linear order A, the n-*character* of $a \in A$ having colour c is the ordered triple $\langle [A^{<a}]_n, [A^{>a}]_n, c \rangle$ (where $A^{<a} = \{x \in A : x < a\}$ and $A^{>a} = \{x \in A : x > a\}$). We let $\rho_n(A) = \{\langle [A^{<a}]_n, [A^{>a}]_n, F(a) \rangle : a \in A\}$. Thus we shall always include the colour as part of the n-character of a, (unlike in [1]). This means for instance that the next theorem reads a little differently from there, but it has precisely the same content. We interchangeably say that a *realizes* this n-character, or *determines* it, or *exhibits* it.

Theorem 1.1 ([1]) *$A \equiv_{n+1} B$ if and only if $\rho_n(A) = \rho_n(B)$.*

We need the following result, which is Theorem 2.1 of [1] (which we call the 'Cutting lemma').

Lemma 1.2 *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,*
(ii) *for every $x \in A$ with $a < x \leq b$, there is $y \leq a$ having the same n-character as x.*

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

Note that this may be applied in a trivial case, namely, that no two consecutive points of an $\equiv_{n+1}$-optimal finite string can have equal n-characters.

It is clear from Theorem 1.1 that if in an m-coloured linear order, no two points have the same n-character, then the ordering is $\equiv_{n+1}$-optimal, meaning that it is not $(n+1)$-equivalent to any shorter ordering. Based on this, we present a construction of a finite 2-coloured linear order of length 70 in which all points have distinct 2-characters, and which is therefore $\equiv_3$-optimal, and show that 70 is the greatest possible number in which all 2-characters are distinct. We also construct a finite coloured $\equiv_3$-optimal linear order of length 74, in which 2-characters must therefore repeat. It should be possible to find longer examples, but the details would be quite tedious, so giving one of this length is good enough to illustrate the idea. This casts some light on the hypothesis required for the 'cutting lemma' (that is, what it says is not that we can reduce the length just based on repetition of characters—more is required about what happens in between).

The typical case we have in mind is that when searching for optimal representatives, we start with a possibly long coloured order, and successively reduce it by removing pieces, retaining n-equivalence, till it becomes optimal. The proof of [1] is too indirect to guarantee immediately that the final ordering is a suborder of the one we start with. We therefore extend the material of [1] by showing that for 2-equivalence at any rate, we *can* guarantee that the optimal representative is contained in the original one; we present an algorithm for achieving this. We believe that this is false for $n = 3$, and in Section 3 explain why.

With regard to the general case, but particularly applied to $n = 3$, we use directed graphs to help analyze n-equivalence. One method would be to take $(n-1)$-characters themselves as vertices of the directed graph, with an arrow going from $\langle X_1, Y_1, c_1 \rangle$ to $\langle X_2, Y_2, c_2 \rangle$ if for some representatives x_1, x_2, y_1, y_2 of X_1, Y_1, X_2, Y_2, $x_1 c_1 \equiv_{n-1} x_2$ and $c_2 y_2 \equiv_{n-1} y_1$, where these are the strings obtained from x_1, y_2 by adding a c_1-coloured point on the right, a c_2-coloured point on the left respectively. The idea is that in scanning a (long) word from left to right, at each point we can view its $(n-1)$-character to left and right, and see how this varies. In practice in what we present here for $m = 2$, $n = 3$, we focus just on the 'middle' section of the given string, in which case a simplified directed graph gives all the information we require.

2 Classification of 2-Equivalence Classes

In this section we give a lot more detail about the 2-equivalence classes of finite coloured linear orders. In [1] we established the precise value $(m^2 + 2m)$ of the least upper bound of the lengths of the optimal representatives of $\equiv_2$-classes. Here we are able, using similar ideas, to give an explicit list of all the $\equiv_2$-classes, from which we can read off, for instance, the length of the optimal representative of each class, and also note which classes are finite or infinite. The key idea here is to use the notion of 'm-configuration' which was introduced in [1].

We fix m as the number of colours. An *m-configuration* is then defined to be a linear order of the form $T = \{x_i : 1 \leq i \leq m\} \cup \{y_i : 1 \leq i \leq m\}$ in which $x_1 < x_2 < \ldots < x_m$ and $y_1 > y_2 > \ldots > y_m$, and x_1 and y_1 are the least and greatest members of T respectively. Here all the x_i are therefore distinct, and so are the y_i, but it is not ruled out that $x_i = y_j$ for certain i and j. Each m-configuration therefore has size between m and $2m$. It is understood that the sequences (x_i) and (y_i) are part of the configuration. That is, to determine the configuration, we have to know not just which linear ordering it is, but also which of its entries are which x_i or y_j (as is clear from the examples given after Theorem 2.4).

Let $(A, \leq, F)$ be a finite coloured linear order having m colours. From this we can derive an associated m-configuration, which is defined to be the linear order induced on $\{x_i : 1 \leq i \leq m\} \cup \{y_i : 1 \leq i \leq m\}$ where for each i, x_i is the least point x of A such that $\{F(z) : z \in A, z \leq x\}$ has i elements, and y_i is the greatest point y of A such that $\{F(z) : z \in A, z \geq y\}$ has i elements. Under these circumstances, an m-configuration receives the colouring induced from that on A. However, the same m-configuration may be thereby coloured in several different ways, if it is viewed as a substructure of possibly different coloured linear orders. Furthermore, not all m-configurations are associated with a coloured linear order at all. The following lemma explains when this happens. An m-configuration together with a colouring that it receives in this way from some $(A, \leq, F)$ is called a *coloured m-configuration*.

Lemma 2.1 *The m-configuration* $T = \{x_i : 1 \leq i \leq m\} \cup \{y_i : 1 \leq i \leq m\}$ *is associated with some m-coloured linear order if and only if* $i + j \leq m + 1 \Rightarrow x_i \leq y_j$.

Proof First to check the necessity of the given condition, suppose that $\{x_i : 1 \leq i \leq m\} \cup \{y_i : 1 \leq i \leq m\}$ is the m-configuration arising from the coloured linear order $(A, \leq, F)$, and let $i + j \leq m + 1$. Let k be greatest such that $x_i \leq y_k$. Then $\{y_l : l > k\}$ are distinctly coloured points lying in $(-\infty, x_i)$, which exhibits $i - 1$ colours. Hence $m - k \leq i - 1$, so $m + 1 \leq i + k$. We deduce that $i + j \leq i + k$ and hence $j \leq k$, so that $x_i \leq y_j$.

Conversely, assuming the given condition holds, let the m-configuration $T = \{x_i : 1 \leq i \leq m\} \cup \{y_i : 1 \leq i \leq m\}$ be given, and we have to find a coloured linear order $(A, \leq, F)$ such that T is the associated m-configuration. We take $A = T$, and have to show how the points can be coloured so that x_i is the least point x of A such that $\{F(z) : z \in A, z \leq x\}$ has i elements, and y_i is the greatest point y of A such that $\{F(z) : z \in A, z \geq y\}$ has i elements. Let us start by colouring the x_i by distinct colours. Clearly this ensures that x_i is the least point such that $\{F(x_k) : x_k \in A, x_k \leq x_i\}$ has i elements. We have to colour the y_j so that no member of $\{F(z) : z \in A, z \leq x_i\}$ has a 'new' colour. We assign colours successively to $y_m, y_{m-1}, \ldots, y_1$ according to which of the sets $\{x_1\}$, (x_1, x_2), $\{x_2\}$, (x_2, x_3), $\ldots$, $\{x_m\}$, (x_m, ∞) they lie in. Given i, let j be the least such that $y_j < x_{i+1}$ (if any). Then by hypothesis, $i + 1 + j \not\leq m + 1$, so $j > m - i$. Hence there are at most i values of j such that $y_j < x_{i+1}$.

The colouring is now given as follows. If $y_j = x_i$ then we let $F(y_j) = F(x_i)$. Otherwise consider colouring all the y_js which lie in (x_i, x_{i+1}). By the remark just made, there are at most i values of j such that $y_j < x_{i+1}$, and there are i colours available for $\{y_j : y_j < x_{i+1}\}$. We have so far used $|\{j : y_j \leq x_i\}|$ of these colours, so the number remaining is $i - |\{j : y_j \leq x_i\}| \geq |\{j : y_j < x_{i+1}\}| - |\{j : y_j \leq x_i\}| = |\{j : x_i < y_j < x_{i+1}\}|$, and these points are coloured in any way using the available colours.

The construction has explicitly ensured that for each i, x_i is the least point such that $(-\infty, x_i]$ is coloured by i colours. To verify the corresponding condition for y_i, note that there are certainly exactly i values of $k \leq i$ such that $y_k \geq y_i$, and these points are all coloured by distinct colours. Suppose that $x_j \geq y_i$. Then as there are m colours, and all y_k points are distinctly coloured, there is k such that $F(x_j) = F(y_k)$. If $k > i$ then $y_k < y_i \leq x_j$, contrary to x_j the *least* point coloured $F(x_j)$. We deduce that $k \leq i$, and so $F(x_j) \in \{F(y_k) : k \leq i\}$ as required.   □

To specify a finite coloured order $(A, \leq, F)$ up to 2-equivalence, we need to know first of all what its associated m-configuration T is, and in addition what colouring T receives (so we need to know what its associated *coloured m*-configuration is). We also need to

know which colours arise as the colours of points (of A) lying between any two consecutive members of T, and we write the set of colours between u and v as $\chi(u, v)$. We write $\mathcal{C}_{T,\chi}$ for the set of all finite coloured orders such that T is the associated coloured m-configuration and colours between the points are given by χ. Note that not all sets of colours are possible for $\chi(u, v)$ and they will be constrained by the x_i and y_j. If for ease we write $x_{m+1} = \infty$ and $y_{m+1} = -\infty$ (not coloured) then a point with colour c can be inserted (without altering the m-configuration realized) in $(x_i, x_{i+1}) \cap (y_{j+1}, y_j)$ if and only if $c \in F(-\infty, x_i] \cap F[y_j, \infty)$.

Theorem 2.2 *Two finite C-coloured linear orders are 2-equivalent if and only if for some T and χ, where T is a coloured m-configuration, they both lie in $\mathcal{C}_{T,\chi}$. In other words, the $\mathcal{C}_{T,\chi}$ classify the $\equiv_2$-classes of m-coloured finite linear orders.*

Proof This relies on Theorem 1.1, which tells us that $A \equiv_2 A'$ if and only if they exhibit the same 1-characters. Let $\{x_i : 1 \le i \le m\} \cup \{y_i : 1 \le i \le m\}$ and $\{x'_i : 1 \le i \le m\} \cup \{y'_i : 1 \le i \le m\}$ be the coloured m-configurations associated with A and A', and suppose first that $A \equiv_2 A'$. Thus A and A' exhibit the same 1-characters. Now by definition of x_i, $|\{F(z) : z < x_i\}| = i - 1$ and $|\{F(z) : z \le x_i\}| = i$. Let $x' \in A'$ realize the same 1-character in A' as x_i does in A. Then $F'(x') = F(x_i)$, $|\{F'(z) : z < x'\}| = i - 1$ and $|\{F'(z) : z \le x'\}| = i$. This implies that $x' = x'_i$. Similarly for y_j and y'_j. Next we have to see that $x_i \le y_j \Leftrightarrow x'_i \le y'_j$, and similarly for $<$. Suppose that $x_i \le y_j$ ($x_i < y_j$ respectively). Then x_i has at least $j - 1$ colours to the right (at least j respectively), and as it realizes the same character as x'_i, this is also true of x'_i in A', and so it follows that $x'_i \le y'_j$ ($x'_i < y'_j$ respectively). We deduce that the same coloured m-configurations arise from A and A'. To see that the corresponding functions χ and χ' are equal, let $u < v$ be consecutive members of $\{x_i : 1 \le i \le m\} \cup \{y_i : 1 \le i \le m\}$, and $u' < v'$ be the corresponding consecutive members of $\{x'_i : 1 \le i \le m\} \cup \{y'_i : 1 \le i \le m\}$. Then the left and right 1-characters of each member of (u, v) are $\{F(z) : z \le u\}$ and $\{F(z) : z \ge v\}$ respectively, and furthermore, these characters are not realized by any other members of A. Precisely these same left and right characters are realized in (u', v'), and since the only extra ingredient required to specify the character is the colour of the point, it follows that exactly the same set of colours arises in (u, v) and (u', v'). In other words, $\chi(u, v) = \chi'(u', v')$.

Conversely, supposing that A and A' both lie in the same $\mathcal{C}_{T,\chi}$, we see that they both realize the same 1-characters, so are 2-equivalent.   □

Corollary 2.3 *A $\equiv_2$-class of finite linear coloured orders is finite if and only if it is a singleton, which holds if and only if $\chi(u, v) = \emptyset$ for each u, v.*

From Theorem 2.2 we derive an algorithm for determining an optimal member of the 2-equivalence class of a finite coloured linear order A. From A we first evaluate x_i and y_j. Then we replace each interval (u, v) such that u and v are consecutive points of the resulting m-configuration by one in which each of its colours only arises once. This leads to the following result.

Theorem 2.4 *Any finite coloured string has a 2-equivalent $\equiv_2$-optimal substring.*

Let us see how the algorithm works out for small values of m:
If $m = 0$ there is only one order, namely $\emptyset$.

If $m = 1$ with colour r, then there are two possible m-configurations, with $x_1 = y_1$ or $x_1 < y_1$. The former gives us just a singleton r (since there is no interval of consecutive points into which new elements can be inserted), and the latter rr which is a singleton $\equiv_2$-class, and rrr, which lies in the infinite $\equiv_2$-class $\{r^n : 3 \leq n\}$ (where $\chi(x_1, y_1) = \{r\}$).

If $m = 2$, these are the possible m-configurations, with the corresponding singleton $\equiv_2$-classes given:

$x_1 < x_2 < y_2 < y_1, rbrb, rbbr, brbr, brrb,$
$x_1 < x_2 = y_2 < y_1, rbr, brb,$
$x_1 < y_2 < x_2 < y_1, rrbb, bbrr,$
$x_1 = y_2 < x_2 < y_1, rbb, brr,$
$x_1 < y_2 < x_2 = y_1, rrb, bbr,$
$x_1 = y_2 < x_2 = y_1, rb, br.$

Including the allowed insertions, where we write r^k for an arbitrary sequence of k r's ($k \geq 0$), similarly b^l, and $w(r, b)$ an arbitrary string of r's and b's, this gives rise to the following list

for $rbrb$: $rr^k bw(r, b)rb^l$, 16 $\equiv_2$-classes (two options for each of k and l, and four for $w(r, b)$), similarly for $brbr$, $rbbr$ and $brrb$,
for rbr: $rr^k br^l r$, 4 $\equiv_2$-classes, similarly for brb,
for $rrbb$: $rr^k rbb^l b$, 4 $\equiv_2$-classes, similarly for $bbrr$,
for rbb: $rbb^l b$, 2 $\equiv_2$-classes, similarly for brr,
rrb, bbr are similar to rbb,
rb, just one $\equiv_2$-class, and br is similar.

This gives a total of 90 $\equiv_2$-classes in which two colours appear. Note that the optimal representative of each class is unique, except when there is a 'middle' section in which both colours appear. For instance, $rrbrbrbb \equiv_2 rrbbrrbb$, though each is of optimal length.

If $m = 3$, there are 26 possible m-configurations, of which all but four fulfil the stipulations of Lemma 2.1 (the four which do not are given by $x_1 \leq y_3 < y_2 < x_2 < x_3 \leq y_1$). To list even these is quite laborious, and when their possible colourings are taken into account, as well as the possible insertions, it is seen that the list increases dramatically over the case $m = 2$. For instance, for the m-configuration $x_1 < x_2 < x_3 < y_3 < y_2 < y_1$ there are 36 ways of colouring the points, and for the rbg (red/blue/green) colouring of x_i and y_i points, the $\equiv_2$-classes are of the forms $rr^{i_1} br^{i_2} b^{j_2} gr^{i_3} b^{j_3} g^{k_3} rb^{j_4} g^{k_4} bg^{k_5} g$ where the indices are all 0 or 1, giving 2^9 possibilities, so even for this case there are $36 \times 2^9 = 18432 \equiv_2$-classes.

We remark that it would also be possible to find an algorithm to reduce to an optimal form inductively on the number of colours. If we define the subsets L, R, and M of A, for 'left', 'right', and 'middle', by $L = (-\infty, x_m)$, $R = (y_m, \infty)$, and $M = A \setminus (L \cup R)$ in the above notation, then the induction would be based on the fact that each of L and R exhibit only $m - 1$ colours. There are however some (minor) complications in the case where L and R overlap, so the method presented just before the statement of Theorem 2.4 is preferable.

Now the easiest way to demonstrate that a finite string is optimal in its $\equiv_2$-class is to show that all its points have distinct 1-characters (then appeal to Theorem 1.1), and in fact this suffices for all 90 strings for $m = n = 2$, as one sees by inspection. The same holds for any number of colours (though not with greater values of n, as we see in the next section).

Theorem 2.5 *For any m, no $\equiv_2$-optimal m-coloured string realizes the same 1-character more than once.*

Proof Suppose on the contrary that $(A, <, F)$ is $\equiv_2$-optimal but $a < b$ realize the same 1-character. We show that $A \equiv_2 A \setminus \{b\}$, contradicting optimality of A. This is similar to the Cutting Lemma, Lemma 1.2. We just need to show that A and $A' = A \setminus \{b\}$ realize the same 1-characters, and the result then follows by appeal to Theorem 1.1. First let $x \neq b$, and we show that x realizes the same 1-character in A and A'. The 1-character is determined by the colour of the point, and by which colours occur on its left and right, so we have to show that $A^{<x}$ and $A'^{<x}$ exhibit the same colours, and so do $A^{>x}$ and $A'^{>x}$. If $x < b$ then $A^{<x} = A'^{<x}$, and the colours in $A^{>x}$ and $A'^{>x}$ could only possibly differ on $F(b)$, but as $A^{>a} \equiv_1 A^{>b}$, and $b \in A^{>a}$, there is a point $> a$ coloured $F(b)$, and therefore also a point $> b$ (and hence $> x$) coloured $F(b)$. A similar argument applies if $x > b$ (using $A^{<a} \equiv_1 A^{<b}$).

If $x = b$, we show that it realizes the same 1-character in A as a does in A'. Since a and b have the same colour, we just have to show that $A^{<b}$ and $A'^{<a}$ exhibit the same colours, and so do $A^{>b}$ and $A'^{>a}$. Now $A^{<b} \equiv_1 A^{<a} = A'^{<a}$ and $A^{>b} \equiv_1 A^{>a} \equiv_1 A'^{>a}$. In all cases, $\equiv_1$ means that the two sets exhibit the same sets of colours, and this holds for $A^{>a}$ and $A'^{>a}$ because the only point of difference between these sets is b, but because $A^{>a} \equiv_1 A^{>b}$, and $b \in A^{>a}$, there must be a point $> b$ coloured by $F(b)$. □

3 3-Equivalence Classes

To help analyze the behaviour of strings up to 3-equivalence, we introduce various labelled directed graphs to keep track of the transitions between 2-characters as we pass through the string. The basic idea is that if $a_1 a_2 a_3 \ldots a_k$ is a string over an alphabet of m colours, then a node of the digraph will be taken to be a 2-character of the form $\langle x, y, c \rangle$ and we include a directed edge from $\langle x_1, y_1, c_1 \rangle$ to $\langle x_2, y_2, c_2 \rangle$ provided that for some strings u_1, v_1, u_2, v_2, $x_1 = [u_1]_2, x_2 = [u_2]_2, y_1 = [v_1]_2, y_2 = [v_2]_2$ and $u_2 = u_1 c_1, v_1 = c_2 v_2$. This corresponds to the fact that the string $a_1 a_2 a_3 \ldots a_k$ gives rise to a path

$$\langle [\emptyset]_2, [a_2 \ldots a_k]_2, F(a_1) \rangle \longrightarrow \langle [a_1]_2, [a_3 \ldots a_k]_2, F(a_2) \rangle \longrightarrow$$
$$\langle [a_1 a_2]_2, [a_4 \ldots a_k]_2, F(a_3) \rangle \longrightarrow \ldots \longrightarrow \langle [a_1 a_2 a_3 \ldots a_{k-1}]_2, [\emptyset]_2, F(a_k) \rangle.$$

In practice, retaining all of both co-ordinates is too cumbersome, and we use an abbreviated string which at any rate for points in the 'middle', suffices to describe the 2-character. In [1] Theorem 2.3, a very crude upper bound for $g(m, n)$, the maximum of the lengths of optimal representatives of finite m-coloured strings under $\equiv_n$, is given. The object here is to obtain some *lower* bounds on $g(m, 3)$, by producing as long optimal strings as possible. The easiest way in which optimality can be assured is to arrange that all points have distinct 2-characters. That this is not *necessary* for optimality is later remarked (by contrast with Theorem 2.5 for $\equiv_2$).

To illustrate this, we show how for two colours we can construct an $\equiv_3$-optimal string A of length 70. Certain features of this seem to depend heavily on the specifics of this case, and we are unsure of how to generalize. The string is composed of three sections, L, M, and R (of lengths 19, 32, and 19) with $L < M < R$. The idea is that, by the time we get to the middle section M, the 2-character has sufficiently 'settled down', to enable us to handle substrings rather uniformly. To describe what L, M, and R are in this case, we emphasize that this subdivision applies just to the 3-move case. The subdivision used in Section 2 for the 2-move case is also used, but on the left and right subsets, where inductively, and using Theorem 1.1 we need to look at 2-characters. To avoid confusion, we use L, M, and R to

stand only for the subdivision of the whole string, and if we need to refer to the subdivision of a left or right segment, we use the terms 'left', 'middle', and 'right'. The definition here is that M comprises all those points whose left and right 2-characters both themselves have non-empty middle sections. Since it is clear that M so defined is convex, we can then let L and R be the subsets of its complement which are to its left, right respectively.

Note that we adopt the notation from the previous section, usually with respect to the initial segment L, or more generally, initial segments of the form $A^{<a}$ for $a \in M \cup R$, (and this will be applied 'similarly' to R, but without corresponding details). That is, x_i will stand for the least point of this set such that exactly i colours are used up to (and including) it, and y_j is the greatest point of the same set such that exactly j colours are used to its right. This perspective enables us to give a more explicit description of what M is. It comprises exactly those points a such that in $A^{<a}$, $x_m < y_m$, and $(x_m, y_m) \neq \emptyset$, and similarly for $A^{>a}$.

Since we shall take $L = rrrrrrbbbbbbrbbbbbr$, the discussion given in the previous section shows that for any a in M, $[A^{<a}]_2$ begins rrb, and it ends with rb, rbb, br, or brr (since we must have $x_1 < x_2 < y_2 < y_1$), and as (x_2, y_2) contains points of both colours, the middle may be taken as rb. This means that we can essentially describe the left 2-character of a point a in M by the ending of $A^{<a}$ (and its colour). Although the ending will actually have length 2 or 3, we can tell what it is just from its last two points. Taking $[A^{>a}]_2$ into account in a similar way, a point is entirely characterized by just 5 entries, two on the left, two on the right, and the colour of a in the middle. The following general lemma is invoked here just for $m = 2$ and $k = 5$, but may be more widely applicable.

Lemma 3.1 *If m, k are integers ≥ 2, then there is a string of length m^k with entries in $\{0, 1, \ldots, m - 1\}$ such that every string of length k in $\{0, 1, \ldots, m - 1\}$ arises exactly once as a substring of k consecutive entries (counting cyclically).*

Proof This method, using an eulerian circuit, was pointed out by P J Cameron.

We form a digraph having as vertices all strings over $\{0, 1, \ldots, m - 1\}$ of length $k - 1$, and including a directed edge from $(x_1, x_2, \ldots, x_{k-1})$ to $(y_1, y_2, \ldots, y_{k-1})$ provided that $(x_2, x_3, \ldots, x_{k-1}) = (y_1, y_2, \ldots, y_{k-2})$. Then the directed edges starting at $(x_1, x_2, \ldots, x_{k-1})$ are those of the form $(x_1, x_2, \ldots, x_{k-1}) \to (x_2, \ldots, x_{k-1}, y)$ so there are exactly m of them, and similarly the in-degree of each vertex is also m. Furthermore, the digraph is strongly connected, since there is a path from $(x_1, x_2, \ldots, x_{k-1})$ to $(y_1, y_2, \ldots, y_{k-1})$ passing by way of $(x_2, x_3, \ldots, x_{k-1}, y_1)$, $(x_3, x_4, \ldots, x_{k-1}, y_1, y_2)$, ..., $(x_{k-1}, y_1, y_2, \ldots, y_{k-2})$. Hence the digraph has an eulerian circuit, and this provides the desired string of length m^k.   □

Given this lemma, we can form a binary string of length 32, such that cyclically ordered, every 5-element string arises exactly once, and this may be taken explicitly as

$$M = rbrbrrbbbrbrbbrbbbbbrrrrbrrrbbr.$$

To form our sequence of length 70, we precede M by $L = rrrrrrbbbbbbrbbbbbr$ and succeed it by $R = rbbbbbrbbbbbbrrrrrr$ (which is L in reverse, easing some verifications). Let us write this string as $a_1 a_2 a_3 \ldots a_{70}$. We verify that all 70 points have distinct 2-characters.

First we can see that for every point a of $M \cup R$, in $A^{<a}$, $x_1 = a_1$, $x_2 = a_7$, $y_1 = a_i$ and $y_2 = a_j$ where $i \geq 19$ and $i > j \geq 18$, so its left 2-character begins $rrbrb$, and ends rb, rbb, br, or brr. However, if $a \in L$, $A^{<a}$ has the form $r^i b^j r b^k$, $r^i b^j r$, $r^i b^j$, or r^i for some i, j, k, so its left 2-character is *not* of this form. By symmetry, we can see that no

point of $L \cup M$ shares a right 2-character with a point of R. We now treat each of L and M individually (and R is similar to L).

For L, the left 2-characters of the 19 points are $\emptyset, r, rr, rrr, rrr, rrr, rrr, rrrb, rrrbb,$ $rrrbbb, rrrbbb, rrrbbb, rrrbbb, rrbbbr, rrbbrb, rrbbrbb, rrbbrbb, rrbbrbb, rrbbrbb$ respectively. The fourth point for each of the repeated left 2-characters has a different colour from the others, and the three remaining points are distinguished by their right 2-characters, which in each case are distinct members of $\{rrbrbbrr, rbrbbrr, bbrrbbrr, brrbbrr\}$.

Finally, we can see that all points of M have distinct 2-characters since they are midpoints of distinct 5-element strings—notice that we have arranged things so that $a_{18}a_{19}a_{20}a_{21} = a_{50}a_{51}a_{52}a_{53}$, which means that distinctness of the 5-element strings persists even at the 'ends'.

This shows that $g(2, 3)$, which is the maximum of the lengths of optimal representatives of finite 2-coloured strings under $\equiv_3$ is at least 70. The upper bound given in [1] is clearly absurdly high, but even so, 70 is a big increase on the optimal length for $m = n = 2$ which is 8.

Let us see that this is the best we can do by these methods, in which optimality is guaranteed by distinctness of the 2-characters. We can always subdivide a given 2-coloured finite linear order into 3 sections, L, M, and R, where in M, both left and right 2-characters have rb as 'middle'. Clearly M is convex, so we may take L and R to be the sets of points to the left, right of M respectively. If $a \in M$, then the left and right 2-characters of a must have at least 6 entries, and as in the discussion above, the 2-characters of the points of M are entirely determined by the 5-element strings of which they are mid-points. Hence $|M| \leq 32$.

Now consider what L can be. Without loss of generality, suppose it begins with r. If it has an initial segment of the form $r^i b^j r^k b^l r^p b^q$ with positive exponents, then the next point does not lie in L, and similarly, $q = 1$ (since otherwise the final point does not lie in L), and by similar arguments, $j = l = 1$, giving $L = r^i br^k br^p b$. If $i \geq 7$ then the fourth and fifth points of A realize the same 2-characters, contrary to assumption. Hence $i \leq 6$. Similarly, $k, p \leq 6$. If $k = 6$, then 2-characters are repeated for the middle two entries in that block. Hence $k \leq 5$, and similarly $p \leq 5$. Hence $|L| \leq 6 + 5 + 5 + 3 = 19$ (and one can check that $rrrrrbrrrrrbrrrrrb$ is possible).

If $L = r^i b^j r^k b^l r^p$ it again follows that $j = l = 1, i \leq 6, k, p \leq 5$, so $|L| \leq 18$.
If $L = r^i b^j r^k b^l$ then $j = 1$ or $k = 1$, and again, $i \leq 6, l \leq 5$, and also $j, k \leq 5$, so $|L| \leq 6 + 1 + 5 + 5 = 17$.
If $L = r^i b^j r^k$ then $i \leq 6, j \leq 7, k \leq 6$ so $|L| \leq 19$.
If $L = r^i b^j$ then $|L| \leq 14$ and if $L = r^i$ then $|L| \leq 7$.

It follows similarly that $|R| \leq 19$, and hence $|A| \leq 19 + 32 + 19 = 70$.

Finally we remark that in $\equiv_3$-optimal strings, 2-characters may be repeated, and using this we are able to construct a longer $\equiv_3$-optimal 2-coloured string. We first give a small example. Consider $A = rbrbrbrbrbrbrbr$, which has length 15, is a palindrome (reading the same forwards and backwards), and whose 7th and 9th entries realize the same 2-character (though apart from this, all 2-characters are distinct). To see that A is $\equiv_3$-optimal, suppose that $B \equiv_3 A$, and we show that B has length at least 15. Since A realizes 14 2-characters, so does B, and hence it has length at least 14. Now A realizes the 2-character $\langle rbrb, brrbbr, r \rangle$, so B must realize this as well, and as $rbrb$ lies in a singleton $\equiv_2$-class, B begins $rbrbr$. Similarly, B realizes $\langle rbrbr, rbrbbr, b \rangle$, so as the $\equiv_2$-class of $rbrbr$ is $\{rbr^p br : p \geq 1\}$, B begins with $rbr^p brb$ for some $p \geq 1$. Since B begins with $rbrbr$ it follows that $p = 1$, and that B begins $rbrbrb$. Similarly B ends with $brbrbr$. The other two 2-characters realized by B are $\chi_1 = \langle rbrbrb, brrbbr, r \rangle$ and $\chi_2 = \langle rbrbbr, rbrbr, b \rangle$.

Since $B^{<b_7} = rbrbrb$, b_7 must realize χ_1, and so $b_7 = r$. Similarly, the 7th point from the right realizes χ_1 and is r. Since $|B| \geq 14$, these two points are distinct, and as B also realizes χ_2, there must be another point between them, so B has length at least 15.

We now present a 3-optimal string of length 74 having the same L and R as in the example given of length 70, but with longer M:

$A = rrrrrbbbbbbr\,bbbbbr\,|rbr\,bbbbbbr\,bbrr\,br\,brrrrr\,brrr\,bbr\,brrr\,bbbr\,|r\,bbbbbr\,bbbbbbrrrrr.$

 The subdivision into L, M, R is indicated. By the previous discussion, this must have repeated 2-characters in M. We let $L' = br$ (the end of L) and $R' = rb$ (the beginning of R). To verify that A is $\equiv_3$-optimal we first note that by the previous arguments, the 2-characters of all elements of $L \cup R$ are uniquely determined and different from all those occurring in M, so any 3-equivalent string must begin and end in this way. Now A determines a path through the digraph D having as vertices strings of length 4 over $\{r, b\}$ arising as convex subsets of $L' \cup M \cup R'$, and the edges of D which may be viewed as quintuples $(x_1, x_2, x_3, x_4, x_5)$ (though they are 'officially' pairs of overlapping quadruples $(x_1, x_2, x_3, x_4) \rightarrow (x_2, x_3, x_4, x_5)$ as in the proof of Lemma 3.1) tell us precisely which 2-characters are realized in the path. So although some will now be repeated, we have to check that no shorter path through D can realize precisely the same 2-characters. One checks that D has 28 edges. Suppose therefore that P is a path through D traversing precisely the same edges (though not necessarily the same number of times). We shall show that P has length at least 36. We may also view P as a 'multi-digraph' in which the multiplicities with which the edges of D arise in P are also recognized, and in this sense we may talk of the 'in-degree' $in(x)$ and 'out-degree' $out(x)$ of a vertex x. By the usual theory of eulerian paths, if i and f are the initial and final vertices of P, then $i \neq f$ implies that $out(i) = in(i) + 1$ and $in(f) = out(f) + 1$; all other vertices x, and also i and f if they are equal, satisfy $in(x) = out(x)$. Furthermore, since $L' = br$, i must equal $brrr$, $brrb$, $brbr$, or $brbb$, and similarly f must equal $rrrb$, $rbrb$, $brrb$, or $bbrb$. The digraph D is shown in Fig. 1.

Now we note that the vertex $rrbb$ is therefore an internal vertex of P, so has equal in- and out-degrees in P. Since its out-neighbours (in P) $rbbr$ and $rbbb$ are distinct, it follows that its in-degree is at least 2. Similarly, $in(rbrb)$, $out(rbrr)$, $out(brbr)$, $out(bbrr) \geq 2$. In D, each of $rrbb$ and $rbrb$ has only one in-neighbour, so the corresponding edges in P must each appear at least twice. Similarly for the out-neighbours of $rbrr$, $brbr$, and $bbrr$. This already assures us that P has length at least 33. But now we know that $in(rbrr) \geq 3$, and as $rbrr$ is internal, also $out(rbrr) \geq 3$. Similarly, $in(rrbr)$, $in(rrrb) \geq 3$. Since the extra edges thereby assured and contributing to $out(rbrr)$, $in(rrbr)$, and $in(rrrb)$ must be $rbrr \rightarrow brrr$, $rrrb \rightarrow rbr$ or $brrb \rightarrow rrbr$, and $rrrr \rightarrow rrrb$ or $brrr \rightarrow rrrb$, this gives at least 3 extra edges in P, showing that it has length at least 36, as desired.
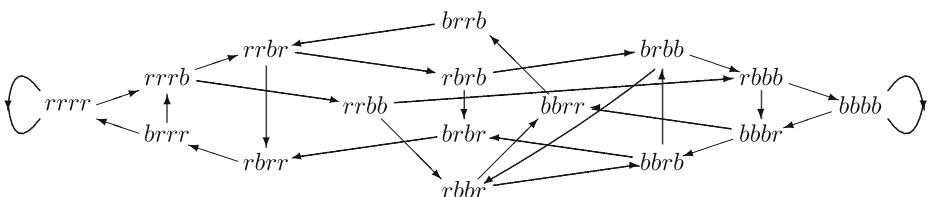

Fig. 1 Digraph helping verify optimality of the string A

Future Work We have really only scratched the surface of this topic, in [1] and here, and a great deal more effort would be required to understand fully the structure of $\equiv_n$-optimal strings for all n, and for all colour set sizes. The method in the final example just given seems very laborious, merely to increase the length by 4, and that is only for two colours and $n = 3$. Undoubtedly there will be longer examples, requiring more careful checking. We remark that we also believe that one can construct finite strings having no $\equiv_3$-equivalent optimal substring. The idea would be to find a string as above obtained by modifying the length 70 example, but such that this time, the path is *not* optimal, but that any optimal path traversing the same edges as D would have to have them in a different order, so that the optimal string would not actually be a substring of the original one.

We conclude by illustrating the specific problem which applies even to increase the number of colours by 1. We would like to apply the same kind of analysis as for the case of 2 colours, which relied on the subdivision of the string A into L, M, R. We can still do this, and for long enough strings there will be a non-trivial M, comprising those points such that the left and right 2-characters both themselves have a 'middle' in which all 3 colours, red, blue, and green, appear. Since this time 2-characters have length up to 15 (see [1]) the lengths of L and R will usually be a lot longer. The main problem comes about in M however. Last time we were able to pin down the 2-character of a point in M by a sequence of length 5. This time though, the right end of the left 2-character may be rbg, $rbgg$, $rbbgg$, $rbgbg$, $rbggb$, as well as others obtained by permuting the three colours, and we seem to need the final 4 entries at least to tell which is which, and so we'd have to have sequences of length 7, 8, or 9, in place of 5, that is, not constant. Thus the trick of using an Euler circuit does not seem to apply.

References

1. Mwesigye, F., Truss, J.K.: Classification of finite coloured linear orderings. Order **28**, 387–397 (2011)
2. Mwesigye, F., Truss, J.K.: Ehrenfeucht-Fraïssé games on ordinals, to appear in the Annals of Pure and Applied Logic