Skiing Is Easy, Gymnastics Is Hard: Complexity of Routine Construction in Olympic Sports

James Koppel
Electrical Engineering and Computer Science, MIT, Cambridge, MA, USA

Yun William Yu
Computer and Mathematical Sciences, University of Toronto at Scarborough, Canada
Department of Mathematics, University of Toronto, Canada

Abstract

Some Olympic sports, like the marathon, are purely feats of human athleticism. But in others such as gymnastics, athletes channel their athleticism into a routine of skills. In these disciplines, designing the highest-scoring routine can be a challenging problem, because the routines are judged via a combination of artistic merit, which is largely subjective, and technical difficulty, which comes with complicated but objective scoring rules. Notably, since the 2006 Code of Points, FIG (International Gymnastics Federation) has sought to make gymnastics scoring more objective by encoding more of the score in those objective technical side of scoring, and in this paper, we show how that push is reflected in the computational complexity of routine optimization.

Here, we analyze the purely-technical component of the scoring rules of routines in 17 different events across 5 Olympic sports. We identify four attributes that classify the common rules found in scoring functions, and, for each combination of attributes, prove hardness results or provide algorithms for designing the highest-scoring routine according to the objective technical component of the scoring functions. Ultimately, we discover that optimal routine construction for events in artistic, rhythmic, and trampoline gymnastics is NP-hard, while optimal routine construction for all other sports is in P.

2012 ACM Subject Classification Theory of computation; Theory of computation → Problems, reductions and completeness

Keywords and phrases complexity, games, sports

Digital Object Identifier 10.4230/LIPIcs.FUN.2022.17

Funding *Yun William Yu*: This work is supported by faculty startup funds from UTSC.

Acknowledgements The authors thank Aviv Adler and Quanquan Liu for comments on earlier drafts of this paper, and thank the members of the MIT Gymnastics team for fostering our love of gymnastics.

1 Introduction: Olympic Sports and Complexity

There has been much work on determining the complexity classes of various logic puzzles and video games [13, 17, 11, 12], as players must strategize and solve constraints to win. Sensibly, there has been little work on the complexity of physical sports, where the challenges are purely athletic. The winning strategy of the 100-meter dash is very simple: run fast.

But in other sports, winning is more complicated. We turn our attention to judged routine-based sports such as figure skating, freestyle skiing, and gymnastics, where athletes demonstrate a string of moves (or "skills") called a "routine," where the routine must meet prescribed constraints (e.g.: at most 7 jumps in figure skating), and the routine's score is different from the sum of its parts (e.g.: a "connection bonus" for chaining together two difficult moves in some gymnastics events). Many an athlete has sat by the sidelines trying to think of the highest-scoring routine they can do. For it is hard. NP-hard?

11th International Conference on Fun with Algorithms (FUN 2022).
Editors: Pierre Fraigniaud and Yushi Uno; Article No. 17; pp. 17:1–17:20
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

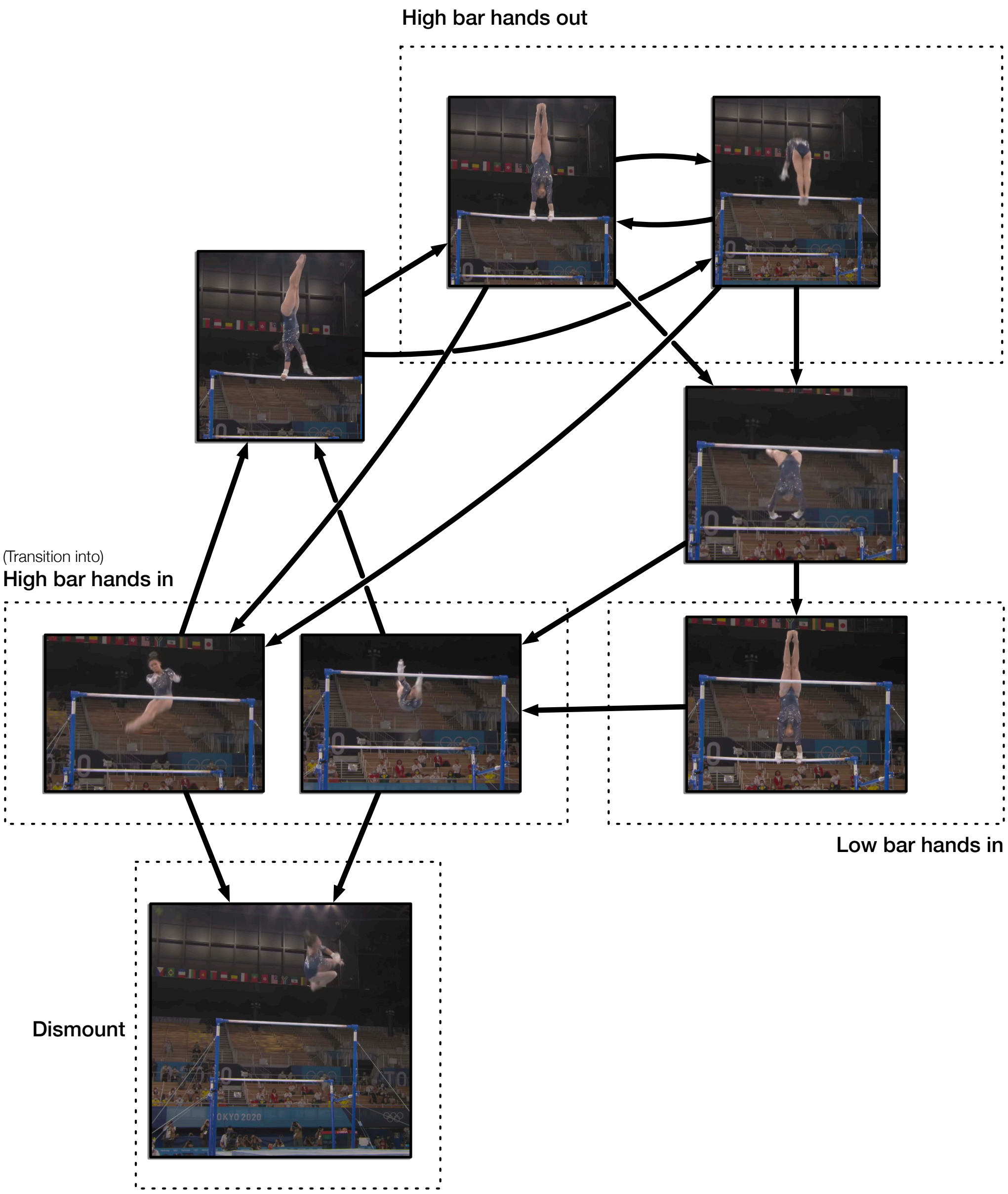


Figure 1 Graph of possible transitions between skills, constructed from skills seen in Suni Lee's uneven bars routine at the Tokyo Olympics. Source: [19]

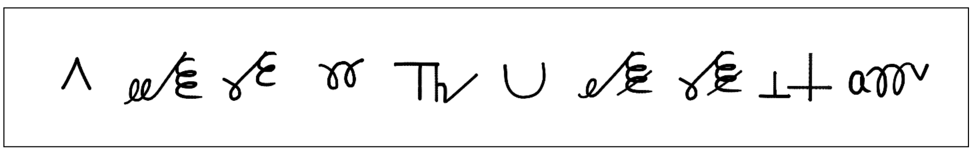


Figure 2 Example Men's Floor routine in gymnastics judging notation; the skills form a linear string (Source and explanation: [2]).

In this paper, we consider a range of routine-based judged Olympic sports and events. We find that the scoring functions are generally made up of four different types of rules, and that the NP-hardness of constructing the optimal routine within a scoring function depends on which types of rules are present.

2 Background

In a judged routine-based sport, the athlete chooses some sequence of skills, which taken together form a "routine". A scoring function then maps that sequence of skills to the "start value" of the routine, which is the maximum number of points achievable if the athlete performs the routine correctly with perfect form. Judges are employed to determine both (1) whether or not the athlete actually performed the skills of the routine and (2) what deductions to apply for imperfect form. The final score given is the start value of the routine that was performed, minus any deductions. At the end of the day, the athlete (or team of athletes) with the highest score wins.

In some sports such as figure skating and rhythmic gymnastics, routines are judged both on their technical difficulty and artistic dance merits. Typically, the final score will be a sum of a technical difficulty score and an "artistic presentation" score. As the latter is inherently subjective, we ignore all artistic presentation scores, and focus purely on technical difficulty and execution scores. However, do note that sports that place a higher emphasis on artistic presentation need not have as complicated technical scoring rules.

In the real world, there are two important optimization tasks in routine construction: (1) maximizing the expected score or (2) maximizing the probability of achieving some fixed score. The first task's relevance is obvious, as higher scores are more likely to win. There are also situations where achieving a fixed score is more useful, e.g.: in Olympic women's artistic gymnastics, an athlete who has accumulated a large lead on vault, beam, and uneven bars may only need a low score on floor to guarantee an overall win.

In this paper we consider a slightly simplified version of the routine construction where each athlete is able to reliably always perform any skill in their repertoire with a fixed deduction. More precisely, there is a Code of Points which provides the set of allowable skills and the scoring function, and the athlete knows the deductions they will incur for imperfect form every time they perform a skill. With this simplification, the complexity of routine construction depends on the scoring function.

Olympic sports have rules in the 100s of pages, and we shall see that we cannot capture every aspect of the sport in our mathematical models. Our work is limited to the objective technical difficulty scores, yet the choices of elements in a routine interact with the subjective artistic scores in sports that have them. Nonetheless, we do create a model that can express nearly all the intricacies of scoring, and most seeming-exceptions can still be encoded into the inputs using tricks similar to ones seen in some of the proofs.

2.1 Generalizing Sports

At the Olympics, each routine-based sport has a finite set of scorable moves and a fixed limit on routine length, so that optimizing a routine in a given year is trivially a finite problem. But the computational difficulty grows over time. In gymnastics, new skills can be invented at high-level competitions, which are then named after the first athlete to successfully perform that skill in competition – for example, on Balance Beam, the Biles is a double-twisting double-tucked salto backwards dismount, named after Simone Biles, who first performed the skill at the 2019 U.S. National Gymnastics Championships. It is theoretically possible to

#	Executed Elements	Info	Base Value	GOE
1	4Lz		11.50	4.76
2	4F		11.00	2.04
3	4T		9.50	3.39
4	3A		8.00	2.97
5	CCSp4		3.20	1.23
6	StSq4		3.90	1.67
7	4T+3T		15.07 X	3.39
8	3Lz+3T		11.11 X	2.11
9	3F+1Eu+3S		11.11 X	0.30
10	ChSq1		3.00	2.14
11	FCCoSp4		3.50	1.35
12	CCoSp4		3.50	1.50
			94.39	

Figure 3 Portion of scoring card from Nathan Chen's Free Skating routine at the 2019 World Championships. Explanation given at [1]

perform a gymnastics routine entirely consisting of skills not previously recognized. More broadly, for all sports under study, every few years their byzantine rulebooks are rewritten, and with each rewrite the roster of skills expands.

For each sport, it is thus natural to generalize to **arbitrary sets of skills** and with **new scoring rules similar to ones that already exist**. The hardness proofs in this paper will imply that e.g.: a SAT formula can be encoded into the scoring rules of a future version of the sport combined with the abilities of a particular athlete facing them. When the kinds of constraints found in scoring rules can encode NP-hard problems, it provides evidence of the practical difficulty faced by an athlete trying to optimize their routine in the finite problem posed by today's sports in a fashion analogous to how the gadgets found in hardness proofs of puzzle games give evidence of the cognitive challenge in solving levels that actually exist.

With this generalization in mind, we now investigate more formally the families of scoring functions found in various Olympic sports.

3 String Scoring Problems

In a routine, an athlete demonstrates a string of skills. These routines can be naturally modeled as a literal string, built from an "alphabet" Σ consisting of the skills they can physically perform that will be recognized by the judges. In real life, this set is a literal alphabet, as each sport defines a set of symbols notating each skill; Figs. 2 and 3 give example routines for gymnastics floor and figure skating in actual judging notation. The rules for scoring then induce a *scoring function* $f : \Sigma^+ \to \mathbb{R}$ mapping a string to a score. The goal is now to construct the highest scoring string that an athlete can perform.

In the remainder of this section, we present several families of scoring functions corresponding to the different types of scoring rules of different Olympic sports. We first start by introducing a simple family of scoring functions: assign a point value and deduction to each skill, and then sum the value of each skill in a routine, minus the deductions. We refer to this as the basic *compositional* scoring function, because the overall score is composed of the individual scores and deductions combined.

The scoring functions used in the Olympic sports we analyze in this paper all build on this foundation, but add complexities such as penalties for repeated moves or bonuses for difficult sequences. We also consider extra constraints on the set of allowed strings: some sequences of moves may be physically impossible (e.g.: doing a handstand on the high bar after already having dismounted it), or at least barred by the rules. We still call these modified scoring functions compositional, because although more complicated, they still depend on the point values and deductions of individual skills.

Table 1 Summary table of NP-hardness of different pairs of rule types, as proven in this manuscript. We consider four different types of rules: *Non-hierarchical* Anti-Repetition, *Hierarchical* Anti-repetition, Connection, and IncompleteGraph. Note that the ElementGroup penalty we describe is subsumed by the Anti-repetition rules. (Lemma 2) While some Olympic sports only have one type of rule, many include multiple types of rules. Note that we only show pairs, because all triples of rule types are NP-hard.

	Non-hierarchical Anti-Repetition	*Hierarchical* Anti-Repetition	Connection	Incomplete Graph
Non-hierarchical Anti-Repetition	NP-hard Thm 6	NP-hard Thm 6	NP-hard Thm 6	NP-hard Thm 6
Hierarchical Anti-Repetition		In P Thm 5	NP-hard Cor 10	NP-hard Thm 9
Connection			In P Thm 7	In P Cor 11
Incomplete Graph				In P Cor 8

Within variants of compositional approaches, the modifications different types of rules make to a simple summation greatly affect the computational complexity. In this paper, we consider four different types of rules commonly found in Olympic scoring codes of points, which will form part of the input to the string scoring problem.

1. Anti-repetition: Are identical or similar skills allowed to be repeated for more points? If so, how many times? Note that deductions are always repeatable.
2. ElementGroup: Are skills partitioned into non-overlapping classes and routines penalized for not having a representative from each class?
3. Connections: Are contiguous pairs of particular skills worth more or fewer points than those skills separately?
4. IncompleteGraph: Are certain orderings of skills entirely disallowed (or physically impossible), disqualifying the entire routine?

Note that these are just a subset of the types of rules an actual Olympic sport has, but this taxonomy captures most of the technical non-subjective rules used.

How the rules are applied also makes a big difference to the hardness. We will find that the Anti-repetition and ElementGroup rules are almost equivalent, but the differentiating factor for NP-hardness is whether or not groups of similar skills are hierarchically arranged. We more formally define each rule type and give a few examples of Olympic sports with each of those classes of rules in the remainder of this section, but for the impatient, the full hardness results can be found in Table 1 and the full classification in Table 2.

Before going on though, let's formally define the problem:

- Let Σ be the set of possible skills, and $n = |\Sigma|$, the number of skills.
- Let $m = |S|$ be the maximum length of an allowed routine $S \in \Sigma^+$.
- Let q be the total number of scoring rules (of any type, defined in the next section).
- Let $z = n + m + q$, the size of the input.

When we speak of poly-time algorithms, we want it to be polynomial in the input size z. Note that in our definition of input size, we are effectively measuring the length of the routine m in unary. This is because we are considering the optimization variant of the scoring problem, where the goal is to determine the highest scoring routine subject to these constraints; without including the output size m, it would be easy to come up with a degenerate case where $n = 1$ and $q = 0$ but m is arbitrarily large.

Table 2 Olympic events categorized by rule types. Note that we consider *Non-hierarchical* to imply *Hierarchical*, because there is always a subset of the *Non-hierarchical* rules that is *Hierarchical*. See Sec. 6 for discussion of individual sports.

Sport/event	*Hierarchical* ANTI-REPETITION	*Non-hierarchical* ANTI-REPETITION	CONNECTION	INCOMPLETE GRAPH	Complexity
Skiing (4 events)	N	N	N	Y	P
Figure skating (Free Skate, Single and Pairs)	Y	N	N	N	P
Rhythmic gymnastics (Individual and Team)	Y	N	(?)	Y	NP-hard
Trampoline	Y	N	N	Y	NP-hard
MAG Floor	Y	N	Y	Y	NP-hard
Pommel Horse	Y	Y	N	Y	NP-hard
Rings	Y	N	N	Y	NP-hard
Parallel bars	Y	N	Y	Y	NP-hard
High Bar	Y	N	Y	Y	NP-hard
WAG Floor	Y	Y	Y	Y	NP-hard
Balance Beam	Y	N	Y	Y	NP-hard
Uneven bars	Y	Y	Y	Y	NP-hard

Moving on to scoring and rules,

- Let $M \in \mathbb{Z}^+$ be the maximum possible score for any individual skill, bonus, or penalty. For the purposes of our setup M should be polynomial in z. Note that in some of our reductions, we further increase M by a polynomial factor.
- Let $p : \Sigma \to [M]$ be the point value of a skill.
- Let $d : \Sigma \to [M]$ be the deduction the athlete will receive when performing a skill.

The other rules described blow will be encoded as modifications to a summed compositional score. Furthermore, note that in all of the codes of points, all of the non-integer scores are rational, so we simply scale them up to be integers.

In practice, although we include m, n, and q in the input size, the length of the routine m generally tends to be a small fixed constant. Instead, the natural generalization of Olympic scoring is in the number of skills n and the number of rules q because (1) each athlete adds more skills to their portfolio with training, and (2) new skills are invented and added to the code of points over time – indeed, many gymnastics skills are named after the first athlete to successfully perform that skill in a recognized high-level competition.

3.1 Compositional scoring with no additional rules

Without additional rules, the unmodified compositional score for a routine $S = s_1 \dots s_m$ is

$$f_{\text{BASIC}}(S) = \sum_{i=1}^{m}(p(s_i) - d(s_i)). \tag{1}$$

We were unable to find examples of basic compositional scoring with no additional rules in Olympic sports. This is entirely unsurprising, because as we will see later in Theorem 4, the optimal routine is not only trivially computable, but also entirely uninteresting – it is just the highest scoring move repeated over and over.. We present it here not because it is used, but simply because it is the base upon which all of the other scoring functions are built.

4. Special repetitions:
 a) Repeated elements (same Code Identification Number) cannot contribute to the "D" score. On Rings, this rule is extended so that a maximum of 1 final strength position in each EG may be recognized for difficulty. Thus, for example only two cross type elements (regular, L cross, or V cross) or support scale type elements (regular or straddled) are permitted in an exercise for difficulty value (one in Group II and one in Group III).
 b) A maximum of 2 Guczoghy type elements can be present in the exercise.

(a)

7. Special repetitions:
 a) A maximum of 2 strength elements (including strength handstands) may be performed in an exercise for content value. Elements I.1 to I.48 are considered to be strength elements, except:
 - Element I.19 - Handstand (2 sec).
 - Element I.31 - ½ or 1/1 turn in handstand or to handstand.
 b) A maximum of 2 circle, flair or Russian elements may be performed in an exercise for difficulty value.

(b)

Figure 4 Special repetition rules for rings (a) and men's floor (b) in artistic gymnastics [5].

3.2 Anti-Repetition rules

Many sports design their rules to force competitors to construct more diverse routines via Anti-repetition rules. A gymnast performing on the high bar may do a giant swing all around the bar many times to transition into various other skills. However, only one of these "giants" is itself scored as a skill (though the gymnast may still incur deductions for bad form). Sometimes these restrictions also apply to similar skills, such as only allowing two scored strength elements (such as handstands, v-sits, or support levers) in men's floor, as seen in Fig. 4.

Formally define an Anti-Repetition rule as a pair (ρ, k), where $k \in \mathbb{N}$ denotes the maximum number of skills in a subset $\rho \subseteq \Sigma$ that can be recognized for points. Given a routine $S = s_1 \ldots s_m$, we define a length-m bitstring $R = r_1 \ldots, r_m$ specifying whether or not each skill is recognized for points, where R has to satisfy all of the Anti-Repetition rules. Then the modified compositional score for a routine is

$$f_{\text{Anti-Repetition}}(S) = \max_{\text{valid } R} \sum_{i=1}^{m} \left(p(s_i) \cdot r_i - d(s_i)\right). \tag{2}$$

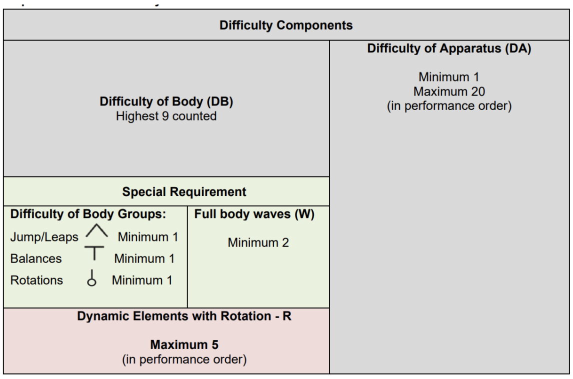
When all skills are recognized, such as when there are no Anti-Repetition rules, R is all 1's and this scoring function is equivalent to f_{Basic} from Equation 1.

In plain language, all performed skills swept up in the Anti-Repetition rules will be worth 0 points, but still count for any deductions; however, if there are multiple satisfying bitstrings, the highest scoring one will be chosen. As an aside, how the ρ_j sets in the Anti-Repetition rules overlap will be a salient factor in our analysis later.

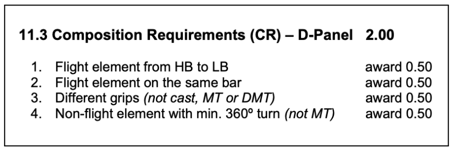
3.3 Element group penalty

Another way to encourage routine diversity is to partition skills into some number of non-overlapping *element groups*, and then penalize not having skills from every element group. For example, on men's floor, gymnasts will be penalized if they lack any of the three major disjoint element groups (non-acrobatic, acrobatic forwards, acrobatic backwards). Element group requirements are also present in other sports and can be more complicated; see Fig. 5.

More formally, an ElementGroup rule is a pair (ρ, p_ρ), where $\rho \subseteq \Sigma$ is the set of skills within an element group, and p_ρ is the point value associated with that element group. Given a set of q_{eg} non-overlapping ElementGroup penalties (ρ_j, p_{ρ_j}), and an indicator variable I_j which is 1 if $s_i \in \rho_j$ for some $i \in [m]$, we get a negative additive modification to the basic compositional score $-\sum_j (1 - I_j) p_{\rho_j}$. Of course, penalizing for lacking an element group is equivalent to giving a bonus for having an element group, so we can instead use a positive additive modification $\sum_k I_j p_{\rho_j}$, which is a bit easier to parse. This modification can of course



(a)

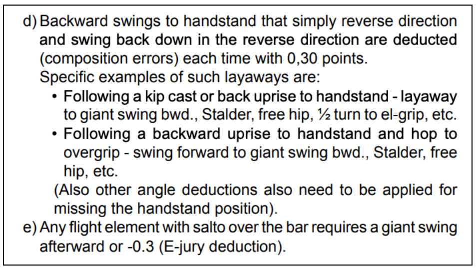


11.3 Composition Requirements (CR) – D-Panel 2.00

1. Flight element from HB to LB award 0.50
2. Flight element on the same bar award 0.50
3. Different grips *(not cast, MT or DMT)* award 0.50
4. Non-flight element with min. 360º turn *(not MT)* award 0.50

(b)

Figure 5 Element group rules for rhythmic gymnastics [7] and uneven bars [6].

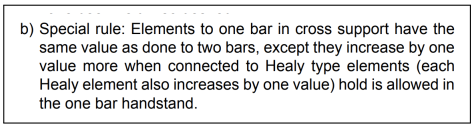


d) Backward swings to handstand that simply reverse direction and swing back down in the reverse direction are deducted (composition errors) each time with 0,30 points.
Specific examples of such layaways are:
- Following a kip cast or back uprise to handstand - layaway to giant swing bwd., Stalder, free hip, ½ turn to el-grip, etc.
- Following a backward uprise to handstand and hop to overgrip - swing forward to giant swing bwd., Stalder, free hip, etc.

(Also other angle deductions also need to be applied for missing the handstand position).

e) Any flight element with salto over the bar requires a giant swing afterward or -0.3 (E-jury deduction).

(a)



b) Special rule: Elements to one bar in cross support have the same value as done to two bars, except they increase by one value more when connected to Healy type elements (each Healy element also increases by one value) hold is allowed in the one bar handstand.

(b)

Figure 6 Skill-specific connection bonuses in high-bar and parallel-bars [5].

be simply added to either f_{BASIC} or $f_{\text{ANTI-REPETITION}}$ to get

$$f_{\text{ELEMENTGROUP}} = \sum_{i=1}^{m}(p(s_i) - d(s_i)) + \sum_{j=1}^{q_{eg}} I_j p_{\rho_j} \tag{3}$$

and

$$f_{\text{ANTI-REPETITION+ELEMENTGROUP}} = \max_{\text{valid } R} \sum_{i=1}^{m}(p(s_i)\cdot r_i - d(s_i)) + \sum_{j=1}^{q_{eg}} I_j p_{\rho_j} \tag{4}$$

3.4 Connections

The previous two classes of rules did not depend at all on the linear ordering of a string of skills in a routine. However, some events, like the balance beam, have "connection bonuses", whereby performing a specific set of skills in a row is worth more than each skill would individually. For example, a layout step-out smoothly transitioning into a back flip with piked legs is worth 0.1 more points than the sum of the two individual skills' values[1]. Because of these connection bonuses, the highest scoring routine may have repeated elements, even when these repeated elements themselves are not scored. There can also be negative connection

[1] Although fractional points are possible, all scores are rational in Olympic codes of points, so we simply rescale to integers in our model.

bonuses, such as the 0.1 point penalty for "illogical connections" in rhythmic gymnastics, or simply to model deductions for expected errors when an athlete attempts two difficult moves in sequence. While connection bonuses most commonly apply to all pairs in a broad class of skills, Fig. 6 gives examples of positive and negative connection bonuses being awarded for more specific combinations of skills.

More formally, a Connection rule can be modelled as a triple (s_1, s_2, c_{12}), where s_1 and s_2 are the two consecutive skills, and c_{12} is the amount of bonus points to be given. Alternately, it can also be thought of as a sparse asymmetric $n \times n$ matrix C, where c_{s_1,s_2} is the connection bonus between skills s_1 and s_2. Given such a matrix C, Connection bonuses amount to an additive modification to the compositional score $\sum_{i=1}^{m-1} c_{s_i,s_{i+1}}$. Thus,

$$f_{\text{Connection}}(S) = \sum_{i=1}^{m}(p(s_i) - d(s_i)) + \sum_{i=1}^{m-1} c_{s_i,s_{i+1}}. \tag{5}$$

Of course, it is straight-forward to construct the variant $f_{\text{Anti-Repetition+Connection}}$, or the variant $f_{\text{ElementGroup+Connection}}$, or $f_{\text{Anti-Repetition+ElementGroup+Connection}}$ by combining this with Equations 2, 3, and 4

3.5 Incomplete Graphs

An alternative view of connection scores is to cast the set of moves Σ as a complete directed graph with skills as nodes and connection bonuses as edge weights, where a routine is a path through the graph and its score is the sum of both node and edge weights. In this context, another type of constraint on compositional approaches can be making the graph graph incomplete, i.e.: forbidding transitions between some skills. This constraint can be imposed as a rule, but in practice is generally a physical constraint because some skills are impossible to follow by other skills. For example, the WAG Uneven Bars event apparatus consists of two different bars, a high bar and a low bar; some skills must begin on the high bar, which is physically impossible to start if an athlete ends the previous skill on the low bar. The graph of possible connections among 8 such uneven bars skills is shown in Fig. 1. More generally, the position, momentum, or idiosyncratic training of an athlete may disallow transitioning from one skill directly into another.

More formally, an IncompleteGraph rule can be modelled as an edge (s_i, s_j). Alternately, all of the IncompleteGraph rules can naturally be encoded into an $n \times n$ graph adjacency matrix C, where $c_{s_i,s_j} = 0$ if (s_i, s_j) in an IncompleteGraph rule and $c_{s_i,s_j} = 1$ otherwise. Unlike the previous rules, which amounted to modifications to the score, this is a hard constraint on whether a routine is allowed at all. Equivalently, we deem that any routine that breaks this rule is scored 0, which can be written as a multiplicative modification of the existing rules

$$f_{\text{IncompleteGraph}}(S) = \left(\sum_{i=1}^{m}(p(s_i) - d(s_i))\right) \prod_{i=1}^{m-1} c_{s_i,s_{i+1}}. \tag{6}$$

Naturally, this also can be combined with any of the other rules-sets.

4 Relationship between rule classes and polynomial reductions

You may notice that although the rule classes are written differently in the Olympic codes of points, the first two and the latter two seem to cover different kinds of conditions. Anti-Repetition and ElementGroup rules do not care about the ordering of skills, but just how many times each skill is performed. On the other hand, Connection and IncompleteGraph are both solely concerned with consecutive orderings of skills. In this section, we show that when considering the NP-hardness of rule sets, those rule sets are very closely related. Indeed, we'll find that for the ordering-free rules, whether the similarity classes overlap is substantially more important than the difference between Anti-Repetiton and ElementGroup.

▶ **Lemma 1.** *IncompleteGraph can be reduced to Connection.*

Proof. In Connection, the scoring function can be modelled as a fully connected weighted directed graph $G = (V, E)$ with skills for vertices and connections for edges. Let $w(s)$ and $w(s_1, s_2)$ be the respective weights on a vertex $s \in V$ or directed edge $(s_1, s_2) \in E$, corresponding to the effective scores after deductions. A routine is precisely a path in G, and the score of that routine is precisely the sum of the vertex and edge weights along that path, including the start and end nodes.

In IncompleteGraph, because all scores are bounded by M, IncompleteGraph can be reduced to Connection by setting the weights for existing edges to 0, and then adding edges with weight $-10Mm$ for each missing edge to make a complete directed graph. Clearly, these edges cannot be traversed in any optimal path, so solving Connection would also solve IncompleteGraph. This means that IncompleteGraph is no harder than Connection. Furthermore, IncompleteGraph+Connection can still be reduced to just Connection by keeping existing edge weights and adding $-10Mm$ weight edges for all the missing edges. ◀

▶ **Lemma 2.** *ElementGroup can be reduced to Anti-Repetition.*

Proof. First, notice that we have defined the ElementGroup rules to be non-overlapping; this is crucial for this transformation. As before, a penalty for missing element groups is equivalent to a bonus for the very first time a skill within a group is performed. Let's duplicate the set of skills $s_1, \ldots, s_n$ to include $s'_1, \ldots, s'_n$, where the s'_i skills are used to denote the first time a skill is done in routine. Thus, given that element groups are disjoint, for a skill $s \in \rho$ we can simply add the element group penalty p_ρ as a bonus to the point value $p(s') = p(s) + p_\rho$ of a skill s, while keeping the deduction $d(s') = d(s)$. Then, we add an Anti-Repetition rule $(\rho', 1)$, where ρ' is the copy of ρ within the duplicated skills. As a result, it never makes sense to use a particular duplicated skill more than once in a routine, as replacing it with the original skills is always better (or at least as good). We have thus encoded the ElementGroup penalty as an Anti-Repetition rule.

In fact, we further claim that ElementGroup+Anti-Repetition can be reduced to just Anti-Repetition. Importantly, unlike ElementGroup rules, which are disjoint, Anti-Repetition rules are allowed to be overlapping, which is crucial for the reduction. We start by duplicating the skills, as above, and encoding all of the ElementGroup rules as new Anti-Repetition rules. However, for each existing Anti-Repetition rule in the original version of the problem, (ρ, k), we replace it with a new Anti-Repetition rule in the duplicated skills world of $(\rho \cup \rho', k)$. That completes the encoding. ◀

Furthermore, although Anti-Repetition rules are allowed to overlap, most of the time, similarity classes are either subsets of one another, or entirely disjoint. For example, there are often Anti-Repetition rules for each individual skill, as well as ElementGroup penalties. Even when there are other similarity classes, very often, they will be subsets of ElementGroups. It turns out that this hierarchical case is much easier than when arbitrary overlaps are allowed in the Anti-Repetition rules, so we will consider the *Hierarchical* and *Non-hierarchical* cases separately.

▶ **Definition 3** (*Hierarchical* Anti-Repetition structure)**.** *Consider a set of Anti-Repetition rules* $\{(\rho_1, k_1), \ldots, (\rho_q, k_q)\}$*. If there exists a pair* (ρ_i, ρ_j) *where* $\rho_i \cap \rho_j \neq \varnothing$ *and* $\rho_i \cap \rho_j \neq \rho_i$ *and* $\rho_i \cap \rho_j \neq \rho_j$*, then the Anti-Repetition rules are Non-hierarchical.*

Note that we consider *Non-hierarchical* to imply *Hierarchical*, because there is always a subset of the *Non-hierarchical* rules that is *Hierarchical*.

5 Hardness proofs

5.1 Pure rule sets

We organize our proofs in increasing order of hardness, in terms of what types of rules are used. In this subsection, we discuss pure rule sets, where only a single type of rules is used. As an aside, instead of separating ElementGroup and Anti-Repetition rule sets, which are largely mathematically equivalent, we instead prove hardness results for *Hierarchical* Anti-Repetition (which includes ElementGroup) and *Non-hierarchical* Anti-Repetition, as that's where the complexity changes.

▶ **Theorem 4.** *Basic compositional string scoring is in P.*

Proof. There is a trivial poly-time algorithm to construct the highest scoring routine of length m for f_{Basic}. Choose the skill s that maximizes $p(s) - d(s)$. If $p(s) - d(s) \geq 0$, then the optimal routine is to simply repeat s a total of m times. Otherwise, $p(s) - d(s) < 0$, so the optimal routine is the empty string. ◀

▶ **Theorem 5.** *Hierarchical Anti-Repetition is in P.*

Proof. First, notice that our limit m on the total length of the routine can be encoded as an Anti-Repetition rule (Σ, m), and that this is compatible with *Hierarchical* Anti-Repetition. In the absence of specific Anti-Repetition rules on individual skills, we can just add Anti-Repetition rules for each skill s of the form $(\{s\}, m)$.

We now transform the problem into a minimum-cost maximum flow problem, which can be solved in polynomial time using e.g. linear programming [24]. We can write down all of the Anti-Repetition similarity classes ρ_i's into a tree T encoding their hierarchical structure, with the top-level root being Σ, and the leaves being the individual skills $\{s\}$ for all $s \in \Sigma$. On each of the nodes of the tree T, we have a capacity k corresponding to the number of times that similarity class can be repeated due to the corresponding Anti-Repetition rule. Furthermore, on the leaves of the tree, we also have a point value $p(s)$.

Let's construct an auxiliary graph G based off of T, but with the following transformations:

1. We start by making G a directed copy of T, where all the edges point towards the root, and have capacity m.
2. We use a standard trick to convert node capacities into edge capacities. For each vertex $v =$ with capacity k, split it into two vertices v^{in} and v^{out}, where v^{in} has all of the incoming edges and v^{out} has all the outgoing edges.
3. Furthermore, there is a directed edge from v_i^{in} to v_i^{out} with capacity k
4. For each new edge $(\{s\}_{in}, \{s\}_{out})$ corresponding to a leaf $\{s\}$ of T, we assign a cost $d(s) - p(s)$, and we assign a cost of 0 for every other edge.
5. Now augment the graph with a source node with directed edges pointing at each of the leaves, all with capacity m and cost 0.
6. Also augment the graph with a target node with a single edge coming from the root of the tree, with capacity m and cost 0.

7. Lastly, we add another n edges directly from the source to the target corresponding to performing an unrecognized skill; each edge corresponds to a single skill, and has capacity m and cost $d(s)$.

Importantly, this new graph G is still a directed acyclic graph, so there are no negative cycles. After constructing this graph G, integer solutions to the minimum-cost maximum flow problem from the source to the target will correspond to an optimal scoring routine. One way to ensure that we get integer solutions to put an epsilon perturbation on all of the costs, so that no two paths have the exact same cost, preventing flow from being split in non-integer ways. ◀

▶ **Theorem 6.** *Non-hierarchical* Anti-Repetition *is NP-hard.*

Proof. We prove this via reduction from the positive one-in-three 3-SAT problem (1-in-3-SAT+), which is NP-hard [18, 26]. In this problem, we are given a family of Boolean variables $\Sigma = \{s_1, \ldots, s_n\}$ and a collection of triples $\tau_i \subseteq \Sigma$, where $|\tau_i| = 3$. The goal is to determine whether or not there is an assignment of all the variables such that every triple has exactly one true value.

We now show that we can encode this into a string scoring problem with *Non-hierarchical* Anti-Repetition rules. Specifically, we first encode the problem of finding a solution with exactly m variables set to true. Repeating the construction for all m completes the proof.

1. Let Σ be the family of skills, so we have one skill for each Boolean variable.
2. For every skill s, assign a point value $p(s) = 1$ and deduction $d(s) = 0$.
3. For every skill s, add an Anti-Repetition rule $(\{s\}, 1)$, so that each skill can only be performed once for credit.
4. Add another Anti-Repetition rule corresponding to each triple in the 1-in-3-SAT+ problem, $(\tau_i, 1)$. Thus, at most one skill in each triple can be performed for credit.
5. We also need to ensure that at least one skill in each triple is performed, so add another Anti-Repetition rule with the complement of a triple, $(\Sigma \setminus \tau_i, m - 1)$.

Notice that successfully performing a skill for credit gives only 1 point, so the maximum possible score is m points for m skills. However, if even one of those skills is not recognized, or for any routine that is length $< m$, then the maximum possible score would be $< m$. Thus, if an optimal routine could be found that scores m points, then that implies that there is a 1-in-3-SAT+ solution with m variables set to true.

Since there are n variables, the number of true variables in a solution must be between 1 and n. We thus run n instances of the reduction, and if 1-in-3-SAT+ has any satisfying variable assignment with $0 < m' \leq n$ Trues, that will be found in the reduction where we let $m = m'$. Thus, we have a polynomial-time reduction from 1-in-3-SAT+ to *Non-hierarchical* Anti-Repetition routine scoring. ◀

▶ **Theorem 7.** Connection *is in P.*

Proof. We exhibit a dynamic programming algorithm below to find an optimal routine of fixed length m. We model the scoring function as a fully connected directed graph with vertices Σ and edge weights given by the matrix C, where c_{s_i,s_j} is the connection value. Assign weights on the vertices by $w(s) = p(s) - d(s)$. A routine is precisely a path in G, and the score of that routine is precisely the sum of the vertex and edge weights along that path, including the start and end nodes.

For any given starting node, we can use dynamic programming to find the maximum scoring routine of length m in polynomial $O(n^3m)$ time:

1. Compute an $n \times n \times m$ tensor D, where $D(x, y, z)$ is the weight of highest scoring routine of length z starting at skill x and ending at skill y:
   a. Initialize D with $D(x, x, 0) = w(x)$ for all nodes x, and $D(x, y, 0) = 0$ for any $x \neq y$.
   b. Recursively compute D by $D(x, y, z) = \max_{y' \in \Sigma} D(x, y', z-1) + c_{y',y} + w(y)$
2. The actual optimal routines themselves can be stored in an auxiliary data structure.

The reason we were able to do this is because revisiting nodes is allowed without penalty. ◀

▶ **Corollary 8.** *IncompleteGraph is in P.*

Proof. Note that IncompleteGraph reduces to Connection in Lemma 1. ◀

5.2 Mixing rule classes

Non-hierarchical Anti-Repetition is already NP-hard, so clearly combining it with anything else would still be NP-hard. Additionally though, we show below that combining together *Hierarchical* Anti-Repetition with either Connections or Incomplete graph makes things NP-hard as well.

▶ **Theorem 9.** *Hierarchical Anti-Repetition+IncompleteGraph is NP-hard.*

Proof. We proceed by reduction from the Hamiltonian path problem, which is NP-complete [18]. In the Hamiltonian path problem, the goal is to find a simple path through a directed graph $G = (V, E)$ such that every node is visited exactly once. Encode each node as a skill s. For each skill s, set its point value $p(s) = 1$ and its deduction $d(s) = 0$. Add an Anti-Repetition rule for each skill. Then, the first time a skill is performed, it will be worth 1 point, but every subsequent time, it will be worthless. Set $m = n$, so we are therefore looking for a routine that is as long as there are skills. Clearly, if there is a Hamiltonian path, that corresponds to precisely a routine that does every single skill exactly once, which would be worth m points. Thus, if there is an optimal solution in *Hierarchical* Anti-Repetition+IncompleteGraph, then there is also a Hamiltonian path in the underlying graph. ◀

▶ **Corollary 10.** *Hierarchical Anti-Repetition+Connection is NP-hard.*

Proof. Recall Lemma 1, which showed that we can encode IncompleteGraph as Connection. Then apply Theorem 9 above. ◀

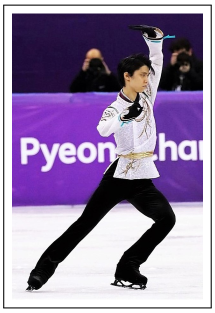
▶ **Corollary 11.** *Connection+IncompleteGraph is in P.*

Proof. This follows from Lemma 1 and Theorem 7, because without an Anti-Repetition rule, Connection+IncompleteGraph is equivalent to just Connection, which we proved to be in P. ◀

▶ **Corollary 12.** *Any scoring function combining three of Non-hierarchical Anti-Repetition, Hierarchical Anti-Repetition, Connection, and IncompleteGraph is NP-hard.*

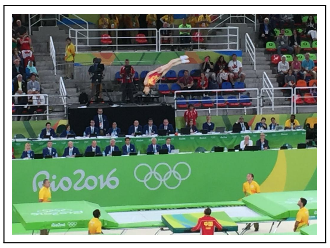
Proof. This is clear from the fact that every combination of three rule types includes at least one of the NP-hard combinations above. ◀

(a) Freestyle Skiing Men's Aerials Final (2010 Winter Olympics) [25].



(b) Free skating (2018 Winter Olympics) [15].

(c) Rhythmic gymnastics (2016 Olympics) [20].



(d) Trampoline (2010 Olympics) [29].

(e) Pommel Horse (2016 Olympics) [23].

Figure 7 Images of each of the Olympic sports (though not all events) we examine.

6 Classification of Olympic Sports

6.1 A Note on Counting "Events"

At the Olympics, an "Event" refers to a contest in which individuals or teams accumulate points over a series of games or performances, and for which a single set of medals is awarded. Under this definition, Free Skate in figure skating (the main component of figure skating with a significant routine construction component) is not an event, but is instead a component of four other events (Men's and Women's Single Skating, Pair Skating, and Team Event), all of which contain other performances as well. This is confusing for the purposes of analyzing events with routine construction, since Free Skate is the only component of figure skating featuring difficult routine construction with objective scoring.

For lack of a better term, we use "event" to mean a program during the Olympics in which an athlete or team is asked to perform athletic feats and receives a distinct score, where two events are the same if the activities performed and scoring rules are both substantially the same. Under this definition, Single and Pair Free Skate are both their own event, but Men's and Women's Free Skate are the same event.

6.2 Figure Skating

There are seven main events in figure skating: (Men's and Women's) Short Program, (Men's and Women's) Free Skate, Pair Short Program, Pair Free Skate, Rhythm Dance, Pattern Dance, and Free Dance. Of these, the two Free Skate events are most relevant to our analysis. The dance events are scored primarily for artistry and forbid many of the difficult tricks that characterize the rest of figure skating. In Short Program, athletes perform a sequence of 7 required skills in any order. While this is technically a Hamiltonian Path problem, it is trivially attainable, as skaters reset their position between skills and thus can complete any skill in any order. We focus our analysis on Single and Pair Free Skate, in which skaters perform a routine of up to 12 technical skills, separated by artful skating around the ice. In the technical component of a routine's score, each recognized skill is scored according to a Level of Difficulty for the skill, plus a Grade of Execution (GoE) between -5 and $+5$ determined by the judges. While the rules do not bar individual skills from being repeated, routines in the Singles program are limited to a maximum number of skills in each of four categories. [9] The technical Elements Score is joined by a more subjective Program Components score, consisting of both technical (e.g.: Skating) and artistic (e.g.: Composition) components. The relative weights of these scores differes by event.

Focusing only on the more objective Elements score, Free Skate is most naturally modeled as basic compositional scoring with *Hierarchical* ANTI-REPETITION. But there are two complexities in modeling figure skating scoring into our framework. First, some skills actually can be connected to each other, in either a "jump combination" or a "jump sequence." Fig. 3 contains several examples of these, such as the Triple Flip-euler-triple toe loop combination (3F+1Eu+3S). However, in many ways, such combined elements are treated as a single element: they are given only a single GoE score, and only count as one element towards the maximums. Even if an athlete can combine any of their skills into a jump sequence or combination, so long as their maximum length is bounded, encoding these possibilities into an expanded output is at most a polynomial blowup in the input size.

Second, there is a 10% "fatigue bonus" for up to 3 jumps performed in the second half of a routine in Single Skating, such as the elements marked with an X in Fig. 3. However, since all other routine construction rules are indifferent to element order, this can be optimized by using the algorithm of Thm. 5 to choose the set of elements in the highest scoring routine, and then placing the hardest jumps in the second half, as it appears the athlete has actually done in Fig. 3. An alternative modeling choice is to create a directed graph encoding both the position of skills in a routine and how competitors may not be able to perform their most difficult skills at the end, though it is unclear whether this is actually realistic, and in any case we are not modeling fatigue in any of the other sports.

6.3 Skiing

While most events in Olympic skiing are races scored purely on time to complete a course, there are five "freestyle" events which are judged based on aerial tricks performed: Aerials, Big Air, Slopestyle, Halfpipe, and Mogul. In Aerials and Big Air, skiers complete a series of twists and turns during a single jump off a large ramp; in Slopestyle and Halfpipe, skiers execute a series of tricks as they progress down a course. Mogul skiing consists primarily of athletes constantly skiing around bumps in the snow ("moguls") while occasionally jumping off a ramp and performing a trick; we exclude it from our analysis as this this "Air" component is only 20% of their score, although the Air component is identical to the other events for complexity modeling purposes.

In each of the included freestyle events, competitors perform a sequence of tricks across one or multiple jumps or on other obstacles. Each trick is scored by its Degree of Difficulty along with deductions. This score is then multiplied or augmented by other judged factors. For example, in Aerials, the Degree of Difficulty is multiple by a number between 0 and 10 built out of components for the take-off, height-and-distance, form, and landing for the jump[3]. Although the tricks performed may impact this multiplier, which then applies to all tricks in the jump, we shamelessly ignore this complexity.

The possible sequences of skills in all skiing events are an Incomplete Graph: it is not possible to flip both forwards and backwards in a single jump in aerials, and an athlete may not be able to chain together arbitrary tricks across consecutive jumps in Halfpipe.

While competitors may perform the same trick multiple times in a single jump, and receive separate Degree of Difficulty scores for them, the Anti-Repetition rules become muddied for aerials and big air: they are actually given two jumps across two separate runs, and are forbidden from performing the exact same "routine" in both jumps. However, this is most easily modeled by saying that the goal of a skier is actually to construct the two highest scoring routines, justifying the decision to model skiing events as not containing an Anti-Repetition rule. Although commentators discuss that competitors are encouraged to perform different tricks in Slopestyle and Halfpipe, we have been unable to find anything in the rules to that effect. We therefore model all skiing events as basic Compositional Scoring with IncompleteGraph.

6.4 Rhythmic Gymnastics

In rhythmic gymnastics, competitors manipulate and juggle a small "apparatus" object such as a ball or hoop while performing a series of dance and contortionist moves, all in time to music. The major emphasis is on simultaneously performing a "Difficulty of Body" (DB) with a "Difficulty of Apparatus" (DA). While DAs recognized do differ with the apparatus, and athletes are required to give multiple performances with different apparatus, the routine construction and scoring rules are the same. However, Team differs significantly from Individual, as e.g.: team members are required to hand or throw their apparatus to another (a "Difficulty of Exchange" [DE]). We thus follow the Olympic classification and deem there to be only two events in rhythmic gymnastics: Individual and Team.

As they are performed simultaneously, we have a choice of modeling DAs and DBs either as a single element (on the grounds they must be practiced together) or as separate elements. However, the latter choice is a more natural fit for the scoring rules: both the DB and DA components have ElementGroup constraints with minimum and maximum requirements, and both have their own Anti-Repetition rules for individual elements. Even without this choice, skills and positions flow smoothly into each other, forming an IncompleteGraph. The scores of each element are summed into the Difficulty Score, while their deductions are used to compute the Execution Score. These are summed with an Artistry score to obtain the final score. Of note is that the Artistry Score contains a 0.1 point deduction for "illogical connections," which can presumably be transformed into a known penalty per pair of elements. Even without including this Connections rule, rhythmic gymnastics scoring contains *Hierarchical* Anti-Repetition and IncompleteGraph, and is thus NP-HARD.

6.5 Trampoline Gymnastics

Trampoline Gymnastics, sometimes just "Trampoline," is almost self-explanatory: athletes jump on a trampoline and perform a series of twists and flips. The only Olympic events are Men's and Women's Individual Trampolining, consisting of both compulsory and voluntary (free) routines (called "exercises"). In Voluntary Exercise, competitors design a routine of 10 elements. Each element consists of a jump with a series of flips, twists, or other acrobatic body manipulations. Repetitions are disallowed. The Difficulty score is then the sum of the scores for each element, while the Execution score is 20 minus the deductions for the elements. These are then added to electronically-determined scores for horizontal displacement and time of flight, which are effectively another form of per-element execution deduction. [8]

It is difficult to connect arbitrary skills in trampolining. A trampoline athlete who can do both front and back flips will need additional training to perform them in sequence. Thus, we model trampoline scoring as *Hierarchical* Anti-Repetition with IncompleteGraph. However, it may effectively form a complete graph to sufficiently advanced athletes, as all moves are supposed to start and end in identical straight jump positions.

Outside of Individual Trampolining at the Olympics, Trampoline Gymnastics features three other non-Olympic events: Synchronized Trampoline, where two athletes perform the same routine on separate trampolines; Power Tumbling (aka "Tumbling"), where athletes execute a series of handsprings and flips down a 25-meter sprung track; and Double Mini-Trampoline, where competitors run at a small trampoline, and perform a skill off it onto a second trampoline, whence they perform a second skill onto a mat. Synchronized Trampoline and Power Tumbling have the same structure as Individual Trampoline with regards to our analysis. Double Mini-Trampoline contains exactly two skills, and would be excluded from our analysis. Also of note is that the requirements for junior age groups pose a rare example of *Non-Hierarchical* ElementGroup requirements, requiring e.g.: one element *to* front or back and one element *from* front or back, although the same element may not count for both.

6.6 Artistic Gymnastics

Artistic gymnastics, usually just referred to as "gymnastics," is among the most recognizable Olympic sports, being one of the original 9 sports featured in the 1896 Summer Olympics. Artistic Gymnastics is split into Men's Artistic Gymnastics (MAG) and Women's Artistic Gymnastics (WAG). The MAG events are Floor, Pommel Horse, Rings, Vault, Parallel Bars, and Horizontal Bar (a.k.a.: High Bar), while the WAG are Vault, Uneven Bars, Balance Beam, and Floor. Unlike other sports such as figure skating with separate men's/women's divisions, the shared apparatus between MAG and WAG (Floor and Vault) have very different scoring rules and sets of recognized moves; most saliently, WAG floor routines are set to music and have dance requirements, while MAG floor routines are silent and have none.

In both MAG and WAG Vault, the athlete performs only one skill, and hence we exclude both from our analysis. The other 8 events all demand lengthy routines subject to complicated scoring rules. We summarize below the official Code of Points for both disciplines. [5, 6]

All MAG events share some common rules of scoring and routine construction, as do all WAG events. We describe the common principles behind both here. A routine consists of a sequence of elements. The Code of Points provides a long list of recognized elements per event, split into a handful of element groups. All MAG events have 3 or 4 element groups and require one skill from each; Repeated elements are never counted. All events except for Rings and Pommel Horses feature Connections, with a small Connection Bonus awarded for performing two difficult skills in succession. Repeated skills are never counted. Routines are capped at 10 scored skills,

The total score of a routine is the sum of a Difficulty ("D-jury") and Execution ("E-jury") score. The Difficulty score consists primarily of the sum of the difficulty value of each element, along with any Connection Bonuses, and, in WAG, awards for Composition Requirements as described below. Each element is given a difficulty between A and G, with A skills being worth 0.1 points, B skills being worth 0.2, etc. The Execution score is 10 minus the deductions for all the skills performed, where each skill has an enumerated set of possible flaws and the deduction associated with each.

Position and momentum are very important in initiating skills, and so all events form an IncompleteGraph. For example, in both Parallel Bars and Rings, most moves can only be initiated from either the position of handstand, of torso above the apparatus (support), or of torso below the apparatus (hang).

The WAG events feature cross-cutting Composition Requirements; see, e.g.: Fig. 5b. Because of these, both Uneven Bars and WAG Floor have *Non-hierarchical* Anti-Repetition rules. Pommel horse is the only MAG event to have *Non-hierarchical* Anti-Repetition rules, thanks in large part to Article 11.2.2 paragraph 3b, limiting the competitor to two "Russian

Wendeswing" moves, a set that crosses multiple element groups. The requirement to use all parts of the pommel horse can also be modeled as an extra element group which cuts across the other element groups, as certain moves are only done off the center or sides of the horse. In classifying these events but not the others as *Non-hierarchical* Anti-Repetition, we make the somewhat arbitrary extrapolation that additional Anti-Repetition rules added to these events but not others may intersect non-hierarchically with existing ElementGroup and Anti-Repetition rules. However, all of these events would be NP-hard even without any Anti-Repetition rules.

Outside of these, gymnastics events feature a staggering number of special scoring and composition rules (see e.g.: Figs 4, 5, 6), and our modeling captures most but not all of them. Two notable rules not captured are the Short Routine penalty, a deduction for any routine under 6 moves, and the requirement that any MAG floor routine pass through all four corners of the floor. But any high-scoring routine will already satisfy these, and we have shown the problem is NP-hard even without considering these constraints.

7 Discussion

7.1 Other Sports

We decided to restrict the scope of this paper to Olympic sports, lest we be compelled to investigate competition rules for every form of dance. Several other Olympic sports feature judged routines, but were excluded for various reasons. Equestrian (Dressage) and snowboarding (Big Air, Halfpipe, and Slopestyle) lack an objective Code of Points. [4, 10] Diving, Vault in artistic gymnastics, and Double Mini-Trampoline in trampoline gymnastics all feature "routines" of at most one or two skills, and are trivially computationally tractable.

Artistic Swimming was a close call for inclusion. Its Technical Routine event features objective scoring assigning a "Degree of Difficulty" and deductions to a series of acrobatic elements, but the set and sequence of skills is fixed. Its "Free Routine" and "Highlight Routine" events allow competitors to construct their own routines (with ElementGroup constraints for Highlight Routine). However, these events lack an objective scoring component. Indeed, the authors confused these events and originally wrote this paper with Artistic Swimming included as a 6th sport. However, there is a new ruleset under proposal that will add element-based objective scoring to all Artistic Swimming events. [22] This new ruleset bases the difficulty score on the sum of per-element scores for both acrobatic and "hybrid" elements (long sequences of underwater leg movements), where each hybrid is composed of a large number of individually-scored leg movements. Whether the new rules make routine construction NP-hard depends on the tractability of hybrid scoring, and whether Free Routine and Highlight Routine keep their current lack of AntiRepetition rules.

We should emphasize that our results here depend on a generalization where we allow the routine length, number of potential skills, and number of rules to grow. Thus, we are not measuring directly the hardness of routine construction in a specific sport, but rather instead the hardness of routine construction in a hypothetical sport that is allowed to use the rule types we consider.

Furthermore, although we attempted to choose our rule types to cover as much of the technical scoring functions given in the codes of points, many sports have one-off rules that do not exactly fit into any of our classes directly. In the previous section, we discuss some of these idiosyncrasies. In the interest of full disclosure, we should note that this article is written by two club gymnasts, and there is possibly a gymnastics-centric bias in which aspects of the sports we chose to model and how we chose to generalize them, in particular in our choice to ignore "subjective" artistic scores, which also often encourage a diversity of skills and interesting connections.

7.2 Related Work

Although we are to our knowledge the first to examine the complexity of routine construction, there does exist a body of literature on the strategy surrounding multiteam tournament [21, 16]. Many sports that are structured as 2-player competitive games; in soccer or hockey, the objective is to score more points than the opposing team, but the opposing team can (and should!) interfere with your attempts to score. However, often, there are more than 2 countries competing in a tournament, so there is some complexity in the means through which pair-wise games are used to determine a single gold medalist; often, there may not be a well-defined ordering such that the winner would necessarily beat every other player in a direct competition. E.g., some variants of FIFA soccer rules make it NP-hard to ask if a specific team still has a chance of winning or coming in a particular place [21].

Outside of sports, string scoring functions have also played a role in bioinformatics, such as in predicting the melting temperature of nucleic acids [28] or the prediction of cleavage sites in protein sequences [14]. However, most of these scores are based on k-mer composition, rather than explicitly considering ordering in the way we do here. The more complex bioinformatics string applications tend to be in directly comparing two similar strings, such as in sequence alignment [27], rather than in scoring individual strings.

References

1 U.S. Figure Skating Scoring Guide. URL: `https://usfigureskating.org/sites/default/files/media-files/Scoring%20Cheat%20Sheet.pdf`.

2 Element Symbols for Men's Artistic Gymnastics, May 2015. URL: `https://www.gymnastics.sport/publicdir/rules/files/en_MAG%20Element%20Symbols%20Booklet.pdf`.

3 FIS Freestyle Skiing Judging Handbook, October 2018. URL: `https://assets.fis-ski.com/image/upload/v1540187845/fis-prod/Freestyle_Skiing_Judging_Handbook.pdf`.

4 Judges Handbook, Snowboard & Freeski, June 2019. URL: `https://assets.fis-ski.com/image/upload/v1610953451/fis-prod/assets/FIS_SB_FK_JudgesHandbook_20_21.pdf`.

5 2022 – 2024 Code of Points: Men's Artistic Gymnastics, February 2020. URL: `https://www.gymnastics.sport/publicdir/rules/files/en_MAG%20CoP%202022-2024.pdf`.

6 2022 – 2024 Code of Points: Women's Artistic Gymnastics, February 2020. URL: `https://www.gymnastics.sport/publicdir/rules/files/en_WAG%20CoP%202022-2024.pdf`.

7 2022 – 2024 Code of Points: Rhythmic Gymnastics, December 2021. URL: `https://www.gymnastics.sport/publicdir/rules/files/en_2022-2024%20RG%20Code%20of%20Points.pdf`.

8 2022 – 2024 Code of Points: Trampoline Gymnastics, May 2021. URL: `https://www.gymnastics.sport/publicdir/rules/files/en_TRA%20CoP%202022-2024.pdf`.

9 Special Regulations & Technical Rules: Single & Pair Skating and Ice Dance, 2021, June 2021. URL: `https://www.isu.org/figure-skating/rules/fsk-regulations-rules/file`.

10 Dressage Rules; 25th edition, effective 1st January 2014, January 2022. URL: `https://inside.fei.org/sites/default/files/FEI_Dressage_Rules_2022_Clean_Version_V2.pdf`.

11 Aviv Adler, Jeffrey Bosboom, Erik D Demaine, Martin L Demaine, Quanquan C Liu, and Jayson Lynch. Tatamibari is NP-complete. *arXiv preprint arXiv:2003.08331*, 2020.

12 Aviv Adler, Erik D Demaine, Adam Hesterberg, Quanquan Liu, and Mikhail Rudoy. Clickomania is Hard, Even With Two Colors and Columns. *The Mathematics of Various Entertaining Subjects: Research in Games, Graphs, Counting, and Complexity*, 2:325, 2017.

13 Greg Aloupis, Erik D Demaine, Alan Guo, and Giovanni Viglietta. Classic Nintendo games are (Computationally) Hard. *Theoretical Computer Science*, 586:135–160, 2015.

14 Christina Backes, Jan Kuentzer, Hans-Peter Lenhof, Nicole Comtesse, and Eckart Meese. GraBCas: a Bioinformatics Tool for Score-Based Prediction of Caspase-and Granzyme B-Cleavage Sites in Protein Sequences. *Nucleic Acids Research*, 33(suppl_2):W208–W213, 2005.

15 David W. Carmichael. File:2018 Winter Olympics - Yuzuru Hanyu FS (1).jpg. `https://commons.wikimedia.org/wiki/File:2018_Winter_Olympics_-_Yuzuru_Hanyu_FS_(1).jpg`, February 2018.

16 Jan Christensen, Anders Nicolai Knudsen, and Kim S Larsen. Soccer is Harder than Football. *International Journal of Foundations of Computer Science*, 26(4):477–486, 2015.

17 Erik D Demaine, Martin L Demaine, Sarah Eisenstat, Adam Hesterberg, Andrea Lincoln, Jayson Lynch, and Y William Yu. Total Tetris: Tetris with Monominoes, Dominoes, Trominoes, Pentominoes,... *Journal of Information Processing*, 25:515–527, 2017.

18 Michael R Garey and David S Johnson. *Computers and Intractability*, volume 174. Freeman San Francisco, 1979.

19 Olympics Gymnastics. Women's Uneven Bars Final | Tokyo Replays. `https://www.youtube.com/watch?v=I84it6P5_i8`, October 2021.

20 Ilgar Jafarov. File:Rhythmic gymnastics at the 2016 Summer Olympics, Marina Durunda 30.jpg. `https://commons.wikimedia.org/wiki/File:Rhythmic_gymnastics_at_the_2016_Summer_Olympics,_Marina_Durunda_30.jpg`, November 2016.

21 Walter Kern and Daniël Paulusma. The New FIFA Rules are Hard: Complexity Aspects of Sports Competitions. *Discrete Applied Mathematics*, 108(3):317–323, 2001.

22 Christina Marmet. What to Expect From the New Scoring System? URL: `https://insidesynchro.org/2021/04/08/what-to-expect-from-the-new-scoring-system/`.

23 Javid Nikpour/Tasnimnews. Gymnastics at the 2016 Summer Olympics. `https://www.tasnimnews.com/en/news/2016/08/12/1155951/gymnastics-at-the-2016-summer-olympics/photo/3`, August 2016.

24 James B Orlin. A Polynomial Time Primal Network Simplex Algorithm for Minimum Cost Flows. *Mathematical Programming*, 78(2):109–129, 1997.

25 Duncan Rawlinson. Freestyle Skiing Men's Aerials Final 19 . `https://www.flickr.com/photos/44124400268@N01/4392687057`, February 2010.

26 Thomas J Schaefer. The Complexity of Satisfiability Problems. In *Proceedings of the Tenth Annual ACM Symposium on Theory of Computing*, pages 216–226, 1978.

27 Temple F Smith, Michael S Waterman, et al. Identification of Common Molecular Subsequences. *Journal of Molecular Biology*, 147(1):195–197, 1981.

28 Masamichi Tsuboi. On the Melting Temperature of Nucleic Acid in Solution. *Bulletin of the Chemical Society of Japan*, 37(10):1514–1522, 1964.

29 Sander van Ginkel. Rio 2016 Summers Olympics. `https://flickr.com/photos/139991533@N04/28558139233`, August 2016.