The Score Reveal Problem: How do we Maximise Entertainment?

Aric Fowler and Richard Booth

School of Computer Science and Informatics, Cardiff University, UK

Abstract. In many elections or competitions, a set of voters assign points to the candidates in a way that indicates their preferences, with the winning candidate being the candidate with the highest total score. When it comes to revealing the result after all votes have been cast, some competitions proceed by having a roll call where each voter announces their vote in turn. This is often done for entertainment purposes, leading to the introduction of the *score reveal problem*: Which ordering of the voters should be chosen to maximise the entertainment value of the roll call? We define several *entertainment measures* and consider their properties, motivated by considerations such as avoiding early resolution of the outcome, focusing attention on the leading candidates, and catering towards preferences for surprise or suspense. We compare several approaches for finding optimal solutions, comparing the hardness of doing so with different entertainment measures and voting formats.

Keywords: computational social choice · entertainment · score reveal problem · voting · combinatorial optimisation

1 Introduction

Competitions with many candidates often rely on the assignment of points to decide a winner, for example the use of league tables in many team sports. In subjective contexts, judges may be used to award points to each candidate to decide a winner. After all votes have been cast the points are revealed in discrete stages such as rounds or votes and there exists a race between the candidates to earn the most points, yielding additional entertainment as candidates move up and down in the current rankings. This roll call format raises the question of whether entertainment can be increased by changing the order in which points are revealed.

A large real-world instance of the score reveal problem can be found in the Eurovision Song Contest [1]. During the competition, the winning candidate is decided by a combination of jury votes and public votes from each participating country. Since 2016, the jury and public votes have been revealed to the audience separately [2]. During the jury round, each country reveals the points awarded by their jury one by one, with candidate scores being updated after each vote. Due to the high number of participating countries (37, as of 2024) this requires a segment consisting of brief talks by relevant personalities from each country, wherein each presenter announces the receiver of that country's points. As points are incremented after each vote, we can graph the progress of the candidates over time to see how their scores interact during the jury voting phase. Figure 1 shows the candidate scores after each jury vote in the 2021 Eurovision final, with Switzerland's *Gjon's Tears* highlighted in blue and France's entry *Barbara Pravi* in red. We can see that for most of the reveal these are the only two countries battling for first place, with the ultimate winner becoming known after the 36th vote (with three votes yet to be revealed) causing the outcome to be known earlier than necessary. By changing the order that votes are announced it is possible to control the way candidates move during the reveal without changing the final outcome, leading us to our two main questions:

- How do we measure the entertainment of a reveal order?
- Given an entertainment measure, how do we find the most entertaining reveal order?

This paper serves as a starting point for research into this new problem, providing a formal definition in section 2 before establishing a common framework for discussing and solving the problem in later sections. We briefly explore existing models for entertainment from the fields of economics and consumer behaviour in section 3 and whether they can be applied directly to the problem, before making a case for the use of *entertainment measures* to measure the entertainment of a given roll call. We present some concrete entertainment measures in section 4, classifying them using using various *entertainment properties* and proving that one measure can be solved in polynomial time. In section 5 we explore the hardness of optimising the remaining measures, using an exact method and various optimisation approaches, before concluding in section 6.

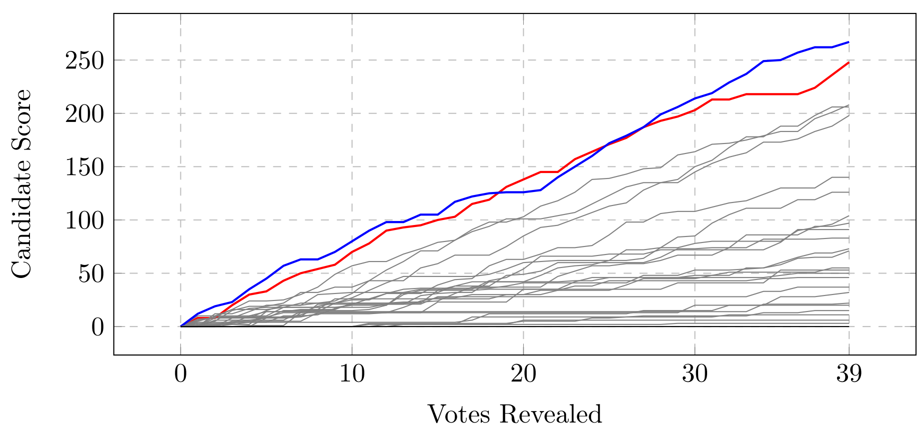


Fig. 1: Candidate scores as seen in the 2021 Eurovision final

2 The Score Reveal Problem

An instance of the score reveal problem consists of a score matrix S, representing the points assigned by a set of voters $V = \{v_1, v_2, \dots, v_N\}$ to a set of candidates $C = \{c_1, c_2, \dots, c_M\}$ such that each element $s_{ij} \in S$ represents the score given to c_i by v_j (we assume $V \cap C = \emptyset$). In most cases voters assign points according to some agreed-upon rule, assigning a decreasing series of values to candidates in order of their preference. In cases where each vote (the column $S_{*,j}$ for voter j) is a permutation of some shared non-increasing vote vector, we refer to this as *permutation voting* with the vector itself being the *voting format*. The voting format used in Eurovision is $(12, 10, 8, 7, \dots, 1, 0, \dots)$. We use permutation voting in our examples and experiments but proofs apply for any possible score matrix. A *reveal order* is some permutation ρ of the columns of S representing the order in which the votes are announced. We define ρS such that the j-th column of ρS is the $\rho(j)$-th column of S. Given the voters are numbered 1 to N, we assume them to be ordered and give ρ using one-line notation. Below is an example with 3 voters and 3 candidates using the voting format $(2, 1, 0)$, where voter 3 reveals first followed by voter 2 and then voter 1.

$$S = \begin{pmatrix} 2 & 0 & 1 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix} \qquad \rho = \begin{pmatrix} 3 & 2 & 1 \end{pmatrix} \qquad \rho S = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 2 & 1 \\ 2 & 1 & 0 \end{pmatrix} \tag{1}$$

The votes are revealed one by one to a spectator, whose entertainment we wish to maximise. Let $\mathbb{E}_S(\rho)$ denote some measure of the spectator's entertainment given a reveal order ρ on S. Throughout this paper we will assume that the spectator is neutral over the sets of candidates and voters, meaning they have no preferences or beliefs regarding the identities of either group. In particular, they are not a member of either group. A suitable measure of $\mathbb{E}$ must be defined that adheres to certain properties of the spectator (addressed in section 4) after which we can define the score reveal problem as follows:

SCORE REVEAL
INSTANCE: A score matrix S and spectator with entertainment measure $\mathbb{E}$
QUESTION: What is the reveal order ρ such that $\mathbb{E}_S(\rho)$ is maximal?

After each vote is announced, the scores from that voter are added to the current candidate scores. Thus, in many cases we will need to refer to the cumulative scores of the candidates at each moment in time. Let $\overline{\Sigma}$ denote the *cumulation* of a matrix, where the j-th column of $\overline{\Sigma}(S)$ is the sum of the first j columns in S. The cumulative scores of the candidates given some reveal order ρ are given by $\overline{\Sigma}(\rho S)$, where each element $\overline{\Sigma}(\rho S)_{i,t}$ is the score of candidate i at time t.

$$\overline{\Sigma}(\rho S)_{i,t} = \sum_{j=1}^{t} (\rho S)_{i,j} \qquad \overline{\Sigma}\begin{pmatrix} 1 & 0 & 2 \\ 0 & 2 & 1 \\ 2 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 3 \\ 0 & 2 & 3 \\ 2 & 3 & 3 \end{pmatrix} \tag{2}$$

The reader may note that in many applications the score matrix is not known in advance, particularly the scheduling of sports leagues. Given some knowledge base of score predictions (e.g., probability of a win, draw, or loss for each pairwise comparison of teams) we can estimate the entertainment utility of a reveal order by repeatedly sampling the score matrix to obtain a distribution of entertainment utilities for that reveal order. One can then approach the problem by changing the question to "What is the reveal order ρ such that $E(\mathbb{E}_S(\rho))$ is maximal?", or using a threshold value K to ask "What is the reveal order ρ such that $\mathbb{E}_S(\rho) > K$ with a 95% confidence interval?". Such variations of the problem allow us to tackle cases with unknown score matrices using the same techniques we would use for a determined score matrix, due to the central utility function $\mathbb{E}_S(\rho)$ being unchanged.

3 Background: Entertainment Utility

Before we can maximise entertainment we need to determine how to measure it, and we can refer to existing work in the fields of economics and consumer behaviour. Dobni's conceptual model of *entertainment value* [6] considers the spectator's characteristics, the medium of the entertainment, and the *benefit* and *sacrifice* components prior, during, and after the consumption of entertainment. We consider this model too broad for the score reveal problem, since we only consider entertainment during the revealing of the votes and do not consider any sacrifice components like monetary cost. Ely et al.'s model of *suspense and surprise* [7] considers entertainment of a spectator derived from non-instrumental information (information that they have no control over), making it more appropriate for use here. Suspense and surprise are defined in terms of a set of *outcomes* and a discrete number of *world states* over time. A world state is the probability of each outcome $\omega \in \Omega$ at a moment in time. *Surprise* is defined as a moment where the world state has changed greatly, meaning the difference between the current outcome probabilities and the previous outcome probabilities is large, for example, a team has just scored in a football match. Its utility function is the sum of squared differences between consecutive world states over time. *Suspense* is defined as a moment of high uncertainty where there is high variance in the possible next world states, for example, a team is close to scoring and there is a high difference between the world state given they score and the world state given they miss. The suspense utility function is the sum of differences between each world state and the believed next world state given some prior belief. Suspense is shown to improve entertainment value [5] and audience retention [10], and the suspense and surprise model is supported by data from the Wimbledon tennis championships [4] and live e-sports competitions [11].

Applying suspense and surprise to the score reveal problem, we must define a set of outcomes, Ω. We consider the outcome of a score reveal instance to be the final ranking of the candidates. However, there are $M!$ unique rankings of M candidates not including non-strict rankings where ties are present, making evaluating the entertainment of even a single reveal order intractable. Smaller sets of outcomes such as $\{\omega_{i,p} \mid \text{candidate } i \text{ finishes } p\text{-th}\}$, $\{\omega_{i,j} \mid \text{candidate } i$ beats candidate $j\}$, or even $\{\omega_i \mid \text{candidate } i \text{ wins}\}$ each have polynomial size with respect to M, however, accurately estimating outcome probabilities for each world state and set of believed world states remains difficult, with approaches such as a Monte Carlo method requiring very high sample sizes to obtain stable estimates while still being subject to random noise affecting the scores of solutions, making optimisation inconsistent. This is our motivation for the use of *entertainment measures* to estimate suspense and surprise.

4 Entertainment Measures and Properties

The purpose of an entertainment measure is to measure entertaining features of reveal orders with minimal computational cost, allowing optimisation algorithms to run very quickly. The entertainer can then define exactly what features they want in a reveal order and quickly find permutations that maximise those features. In this section we define properties including Late Resolution, Anonymity, and k-Tail Independence and use them to design entertainment measures.

4.1 Resolution

A general principle is that the outcome should be in doubt to the spectator for as long as possible. Our first measure considers that the eventual winner should not be known until the last possible moment, which was not satisfied by the reveal order shown in figure 1. We consider a reveal order ρ to be *resolved with respect to* S if at some time t less than the final time N it is impossible for the current leader to be overtaken by another candidate. In instances using permutation voting the format is known by the audience, so we will consider h to be the highest amount of points a candidate can receive from a single voter, and l to be the lowest[1]. At any moment in time, if any candidate is close enough to overtake the current leader it follows that the second-placed candidate can overtake the current leader. If the current second-placed candidate cannot overtake the leader, it follows that no candidate can overtake the current leader. Therefore, to measure resolution we need only consider the candidate in second place at that moment in time. Let $\overline{\Sigma}^1_{*,t}, \overline{\Sigma}^2_{*,t}$ denote the greatest and second-greatest candidate scores at time t (in the case of a tie they are equal). Assuming the leader receives l points from all remaining voters and the runner-up receives h points from all remaining voters, we can state that a reveal order ρ is resolved on S at time t if $\overline{\Sigma}(\rho S)^1_{*,t} - \overline{\Sigma}(\rho S)^2_{*,t} > (N-t)(h-l)$. We define the level of resolution in terms of the number of unrevealed votes, so a reveal order is k-resolved if it is resolved at time $N-k$. Intuitively, if a solution is resolved on S with k votes remaining then it is resolved with $k-1$ votes remaining. A solution is "indecisive" if it is not resolved for any value of k. This occurs if and

[1] In other voting cases, h and l may not necessarily be known, so a reveal order cannot be deemed resolved.

only if the final outcome is a tie as $\overline{\Sigma}^1_{*,N} - \overline{\Sigma}^2_{*,N} \leq 0$ can only be satisfied when $\overline{\Sigma}^1_{*,N} = \overline{\Sigma}^2_{*,N}$. All permutations on a tied instance will yield indecisive solutions since the sum of all scores is independent of their order.

We call any reveal order that is k-resolved for some $k > 0$ *resolved* since the winner is known with some unrevealed votes remaining, while we refer to solutions that are 0-resolved only as *unresolved* since the winner is not guaranteed until after the final vote. The reveal order in the 2021 Eurovision final (shown previously in figure 1) is 3-Resolved, however, the difference in final points between the winner (Switzerland) and runner-up (France) is only 19. Both Belgium and Denmark gave 12 points to Switzerland and zero points to France, so if either of those countries had announced their vote last the difference in points prior to the final vote would have been a mere 7 points, making the reveal order unresolved. Traditionally in Eurovision the host country announces their jury vote last. In this case it was the Netherlands who gave 5 points to Switzerland and 12 points to France, meaning the gap prior to their vote was 26 points. If we choose to keep their vote as the final vote, we can put either Belgium or Denmark immediately before the Netherlands, bringing the difference down to 14 points with 2 votes remaining and hence achieving a solution that is only 1-resolved. This demonstrates that, while it is very easy to find a reveal order that is as unresolved as possible, we often only need to fix one or two votes to minimise resolution and the remaining votes can be freely placed in any order allowing plenty of room for further optimisation. Let $\mathrm{RES}_S(\rho)$ denote the greatest value of k for which ρS is k-resolved, giving us the Late Resolution property.

Property 1 (Late Resolution (LATE)). *We say $\mathbb{E}$ satisfies LATE if, for all matrices S and reveal orders ρ_1, ρ_2, $RES_S(\rho_1) < RES_S(\rho_2)$ implies $\mathbb{E}_S(\rho_1) > \mathbb{E}_S(\rho_2)$.*

One may consider Late Resolution to be a minimum requirement for choosing an entertaining reveal order, but once a partition of the votes has been found such that Late Resolution is satisfied we may choose any permutation of each subset while maintaining the optimality of the solution (demonstrated in Theorem 1). If we want to guarantee Late Resolution we can use $-\mathrm{RES}_S(\rho)$ as an entertainment measure of its own (made negative so that maximising the value results in minimal resolution), which we call the Resolution measure.

Theorem 1. *If $\mathbb{E} = RES$, **SCORE REVEAL** can be solved in polynomial time.*

Proof. A polynomial-time algorithm to find an optimal reveal order is given as follows:

procedure LATE RESOLUTION(S)
  $V_u \leftarrow \{1, 2, \ldots, N\}$ ▷ Unfixed votes
  $V_f \leftarrow \{\}$ ▷ Fixed votes
  Let i, j be the indices of the final winner and runner-up, respectively.
  $\Delta \leftarrow \overline{\Sigma}(S)_{i,N} - \overline{\Sigma}(S)_{j,N}$

while $\Delta > (N - |V_f|)(h - l)$ **do**
  $k \leftarrow \arg\max_{g \in V}(S_{i,g} - S_{j,g})$
  $V_u \leftarrow V_u \setminus \{k\}$ ▷ Move k from V_u to V_f
  $V_f \leftarrow V_f \cup \{k\}$
  $\Delta \leftarrow \Delta - S_{i,k} + S_{j,k}$ ▷ New difference at time $N - |V_f|$
Let π_1, π_2 be any permutations on V_u, V_f
return $\rho \leftarrow \pi_1 \circ \pi_2$ ▷ Concatenate π_1 and π_2

This algorithm calculates Δ, the score difference between the overall winner and runner-up, and checks if the solution is resolved at time N. If not, it selects the vote that maximises the points given to the winner and minimises the points given to the runner-up and fixes that vote so that it will be revealed last, therefore closing the gap by the greatest amount. If the new value for Δ at time $N-1$ is still too large, then no reveal order exists on S that is unresolved at time $N-1$ as we have fixed the optimal vote to reveal last. The algorithm continues to fix votes until the size of the gap at time $N - |V_f|$ is small enough to be unresolved and then exits the *while* loop. Upon exiting the loop, we will have a set V_u of unfixed votes and a set V_f of fixed votes. As long as all unfixed voters announce their scores before the fixed voters then the reveal order will be optimal, as no solution exists that is non-$|V_f| - 1$-resolved and all permutations of the votes in V_u will yield a non-$|V_f|$-resolved solution. Hence, the algorithm returns any permutation ρ given by the concatenation of $\pi_1(V_u)$ and $\pi_2(V_f)$. This algorithm provides us with a reveal order that has optimal Resolution score, and has worst-case time complexity $O(MN + N^2)$. □

4.2 Suspense and Surprise Approximations

Our next entertainment measure is Position Shifting (POS), inspired by Ely et al.'s surprise function. If large changes in the world state result in high surprise, this can be approximated by considering changes in the ranking of the candidates. POS compares each pair of successive candidate rankings in $\overline{\Sigma}(\rho S)$ and counts the total difference in rank between each candidate's current and previous position. The rank of a given candidate c_i at time t can be found by counting the number of candidates with greater score at that time. This is done by the function $\text{rank}_{\overline{\Sigma}(\rho S)}(i,t) = |\{g \mid \overline{\Sigma}(\rho S)_{g,t} > \overline{\Sigma}(\rho S)_{i,t}\}| + 1$, such that the candidate with the greatest score is given a rank of 1 representing first place.

$$\text{POS}_S(\rho) = \sum_{i;t>1} \left|(\text{rank}_{\overline{\Sigma}(\rho S)}(i,t) - \text{rank}_{\overline{\Sigma}(\rho S)}(i,t-1)\right| \tag{3}$$

Proximity (PROX) takes the spirit of the suspense function. Suspense is a moment where a large change in world state is possible, which is most likely to occur when candidates are close in points (the likelihood of a candidate overtaking another, or pulling away to solidify their place, is higher). The Proximity measure approximates suspense by considering each candidate at each point in time, and measuring the distance between that candidate and their closest rival. Small distances to rivals indicate higher suspense, as there is a higher chance of the world state changing, while greater distances indicate lower suspense. The measure is calculated as

$$\mathrm{PROX}_S(\rho) = -\sum_{i,t} \min_{k \neq i} \left\{ \left| \overline{\Sigma}(\rho S)_{i,t} - \overline{\Sigma}(\rho S)_{k,t} \right| \right\} \tag{4}$$

where the sum is negative, so smaller distances result in a greater output value.

4.3 Neutrality and Anonymity

As mentioned in section 2 we assume that the spectator has neutral preference over the set of candidates, and no prior knowledge about the voters. In this case, any measure of entertainment value should consider the points revealed only, with no reference to the identities of the candidates or voters. In this section we define several properties that entertainment measures should satisfy for a neutral spectator and show that POS and PROX satisfy them. The first of these properties is Point Equivalence (PE), which requires that if the same points are read out in the same order, regardless of instance, the entertainment should remain the same. Point Equivalence is given as

Property 2 (Point Equivalence (PE)). *We say $\mathbb{E}$ satisfies PE if, for all score matrices S_1, S_2 and reveal orders ρ_1, ρ_2 on them, $\rho_1 S_1 = \rho_2 S_2$ implies $\mathbb{E}_{S_1}(\rho_1) = \mathbb{E}_{S_2}(\rho_2)$.*

We can also model the intuition that permuting the score matrix and reveal order such that the scores are given in the same order but by different voters has no effect on the entertainment:

Property 3 (Voter Neutrality (VN)). *We say $\mathbb{E}$ satisfies VN if, for all score matrices S and reveal orders ρ, π on V, $\mathbb{E}_S(\rho) = \mathbb{E}_{\pi S}(\pi^{-1}\rho)$.*

Voter Neutrality explicitly assumes that the second score matrix is a permutation of the first, however, some readers may note that Point Equivalence implicitly assumes the same, as it only applies when $S_2 = [\rho_1 \rho_2^{-1}] S_1$.

Proposition 1. *Let $\mathbb{E}$ be an entertainment measure. Then $\mathbb{E}$ satisfies Point Equivalence iff it satisfies Voter Neutrality*

Proof. First, by letting $\pi = \rho$ it follows that any $\mathbb{E}$ satisfying VN also satisfies $\mathbb{E}_S(\rho) = \mathbb{E}_{\rho S}(\rho^{-1}\rho) = \mathbb{E}_{\rho S}(\iota)$. We can use this to prove VN implies PE as, if $\rho_1 S_1 = \rho_2 S_2$ then by VN it follows that $\mathbb{E}_{S_1}(\rho_1) = \mathbb{E}_{\rho_1 S_1}(\iota) = \mathbb{E}_{\rho_2 S_2}(\iota) = \mathbb{E}_{S_2}(\rho_2)$ satisfying PE. We can verify that PE implies VN by assigning $\rho_1 = \rho$, $\rho_2 = [\pi^{-1}\rho]$, $S_1 = S$, and $S_2 = \pi S$. $\rho S = [\pi\pi^{-1}\rho]S$ so PE applies, and by PE $\mathbb{E}_S(\rho) = \mathbb{E}_{\pi S}(\pi^{-1}\rho)$ as required. $\square$

Both Point Equivalence and Voter Neutrality describe an entertainment measure's relationship with the voters, however, neither of them describes the anonymity of the candidates. We will use σ to represent any row-wise permutation of the score matrix, to distinguish candidate-wise permutations from the voter-wise permutations ρ and π. Candidate Neutrality requires that permuting the rows of the score matrix has no effect on the entertainment value.

Property 4 (Candidate Neutrality (CN)). *We say $\mathbb{E}$ satisfies CN if for all score matrices S, permutations σ on C, and reveal orders ρ on V, $\mathbb{E}_S(\rho) = \mathbb{E}_{\sigma S}(\rho)$.*

Now that we have definitions for Voter Neutrality and Candidate Neutrality, a combined definition that requires both the candidates and voters to be anonymous can be given.

Property 5 (Anonymity (ANON)). *We say $\mathbb{E}$ satisfies ANON if for all score matrices S, permutations σ on C, and reveal orders ρ, π on V, $\mathbb{E}_S(\rho) = \mathbb{E}_{[\sigma\pi]S}(\pi^{-1}\rho)$.*

We can show that VN and CN are jointly equivalent to ANON, verifiable by substituting identity permutations into the definition of Anonymity. We can also show that Position Shifting and Proximity satisfy Anonymity by inserting the permutations into their definitions and showing that this has no effect on the value.

4.4 Tail Independence

So far we have defined entertainment measures that consider every candidate equally, however, we propose that even a neutral observer with no preference over the set of candidates will still have some bias towards candidates that are winning, meaning the candidates that are closer to first place contribute more to suspense and surprise. Intuitively, these candidates have a greater effect on the most important outcome: "which candidate will win?".

For a set of candidate scores at time t, given by $\overline{\Sigma}(\rho S)_{*,t}$, let the *k-head* of ρS at time t be the set of candidates where each $\text{rank}_{\overline{\Sigma}(\rho S)}(i,t) \leq k$. We can determine if a candidate is in this set by counting the number of candidates with greater score, and checking if this number is less than k. The candidates outside the k-head will be referred to as the k-tail. Given a matrix $\overline{\Sigma}$, the k-head mask $\triangle_k \overline{\Sigma}$ is given as

$$(\triangle_k \overline{\Sigma})_{ij} = \begin{cases} \overline{\Sigma}_{ij} & \text{if } c_i \in k\text{-head} \\ * & \text{otherwise} \end{cases} \tag{5}$$

where any candidate scores that are part of the k-tail are replaced with the non-numerical symbol $*$, ($* < x$ for all $x \in \mathbb{R}$). k-Tail Independence requires that for any two cumulative matrices $\overline{\Sigma}_1, \overline{\Sigma}_2$, if their k-heads are equal the entertainment must be equal. This has the result that the entertainment measure is independent of low-ranking candidates.

Property 6 (k-Tail Independence (KTI)). *We say $\mathbb{E}$ satisfies KTI if, for all score matrices S and reveal orders ρ_1, ρ_2 on V, $\triangle_k \overline{\Sigma}(\rho_1 S) = \triangle_k \overline{\Sigma}(\rho_2 S)$ implies $\mathbb{E}_S(\rho_1) = \mathbb{E}_S(\rho_2)$.*

If a measure is independent of the k-tail, it follows that it must also be independent of the $k+1$-tail. For an instance with M candidates, the M-tail is nonexistant, so while we could state that POS and PROX satisfy k-Tail Independence for $k = M$, this is the only case they satisfy. We will use the term Tail Independence to refer to measures that are independent of some non-zero tail.

Property 7 (Tail Independence (TI)). *We say $\mathbb{E}$ satisfies TI if and only if, given a score matrix S with M candidates, it satisfies k-Tail Independence for some $k < M$.*

We can generalise our previous measures to make them tail independent by introducing a rank bias. Let $w(p)$ denote a weight in the range $[0, 1]$ of a position p in the rankings, representing the level of interest or importance given to candidates in that position. For Rank-Biased Proximity (R-PROX) we can introduce this weight as shown in equation 6 where each difference is multiplied by the weight of that candidate's position. This has the effect that any candidates in a position p where $w(p) = 0$ will be ignored. For Rank-Biased Position Shifting (R-POS) each position change is weighted using the candidate's current position, however, to maintain k-Tail Independence we must also consider if the candidate has moved up into the k-head from the k-tail. In these cases the value must be equal regardless of the candidate's previous position. We enforce this by passing $\triangle_{\overline{p}}\overline{\Sigma}(\rho S)$ into the rank function, where $\overline{p}$ is the greatest value of p for which $w(p)$ is non-zero. This has the effect that candidates moving up into the k-head are treated as if they moved up from position $k+1$, regardless of their true previous position.

$$\text{R-PROX}_S(\rho) = -\sum_{i,t}\left(\min_{k\neq i}\{\left|\overline{\Sigma}(\rho S)_{i,t} - \overline{\Sigma}(\rho S)_{k,t}\right|\} \times w(\text{rank}_{\overline{\Sigma}(\rho S)}(i,t))\right) \quad (6)$$

$$\text{R-POS}_S(\rho) = \sum_{i;t>1}\left(\left|\text{rank}_{\overline{\Sigma}(\rho S)}(i,t) - \text{rank}_{\triangle_{\overline{p}}\overline{\Sigma}(\rho S)}(i,t-1)\right| \times w(\text{rank}_{\overline{\Sigma}(\rho S)}(i,t)\right)$$
$$\text{where } \overline{p} = \arg\max_p(w(p) \neq 0) \quad (7)$$

Given there are many possible weight functions that could be used, we consider each instance of a rank-biased measure to be a different *configuration* of the measure. If a measure satisfies a property in all configurations, we say that it necessarily satisfies the property; if a measure does not satisfy a property in any configurations, we say that it never satisfies the property; and if a measure satisfies a property in some configurations, but does not in others, we say that it possibly (but not necessarily) satisfies the property. Both the rank-biased Position Shifting and Proximity functions possibly satisfy k-Tail Independence and necessarily satisfy Anonymity.

Proposition 2. *Rank-Biased Proximity satisfies k-Tail Independence if $w(p) = 0$ for all $p \geq k$.*

Proof. Let R-PROX$_S(\rho)$ be such that $w(p) = 0$ for all $p \geq k$ and let S, ρ_1, ρ_2 be such that $\triangle_k \overline{\Sigma}(\rho_1 S) = \triangle_k \overline{\Sigma}(\rho_2 S)$. We must show that R-PROX$_S(\rho_1)$ = R-PROX$_S(\rho_2)$. Since $w(p) = 0$ for all $p \geq k$, the bottom $m - k$ candidates are ignored. Since the k-heads are equal, their utility according to the rank bias will be equal. It follows that R-PROX$_S(\rho_1)$ = R-PROX$_S(\rho_2)$ as required. □

Proposition 3. *If $M \geq 2$ and $w(M) \neq 0$, Rank-Biased Proximity does not satisfy Tail Independence.*

Proof. Given any $M \geq 2$, let S be such that the first $M-1$ rows are $(2, 0, 0)$ and the final row is $(0, 0, 1)$. Let $\rho_1 = \begin{pmatrix}1 & 2 & 3\end{pmatrix}$ and $\rho_2 = \begin{pmatrix}1 & 3 & 2\end{pmatrix}$. Let $w(M) = \alpha \neq 0$ and assume that R-PROX satisfies Tail Independence. Since $\triangle_{M-1} \overline{\Sigma}(\rho_1 S) = \triangle_{M-1} \overline{\Sigma}(\rho_2 S)$ it follows that R-PROX$_S(\rho_1)$ = R-PROX$_S(\rho_2)$. Since the top $M - 1$ candidates are all tied at 2 points throughout, the proximity measure will calculate a sum of zero for those candidates. The bottom candidate has a proximity of 5 on ρ_1 and 4 on ρ_2. The R-PROX scores of the two solutions are therefore R-PROX$_S(\rho_1) = 4\alpha$ and R-PROX$_S(\rho_1) = 5\alpha$. By Tail Independence $4\alpha = 5\alpha$, however, we have defined $\alpha \neq 0$. Hence our assumption that R-PROX satisfies Tail Independence reaches a contradiction, and therefore R-PROX cannot satisfy TI when $w(M) \neq 0$. □

Since we have shown cases where R-PROX satisfies and does not satisfy TI, R-PROX possibly satisfies Tail Independence. The same result for R-POS can be proven using similar methods. Both measures remain anonymous, verifiable by inserting permutations into their definitions and simplifying.

Lead and Hotseat. At the extreme, each measure has a rank-biased variant that values only the winning candidate, given by the rank bias $w(p) = 1$ if $p = 1$, 0 otherwise. For Proximity this is the Lead (LEAD) measure, which satisfies 2-Tail Independence as it measures only the difference between 1st and 2nd place. For Position Shifting this is the Hotseat (HOT) measure which satisfies 1-Tail Independence as, even with a 1-high mask, we can differentiate which candidate(s) are winning from candidates that are not since $*$ is defined as being less than any other real-valued score. Our rank-biased variants of POS and PROX combined with RES give us the seven entertainment measures presented in this paper, all of which satisfy ANON as shown in table 1. RES satisfies 2-Tail Independence as it considers only the scores of the top two candidates in order to check if a reveal order is resolved, and never satisfies 1-Tail Independence. Its anonymity can be shown by the scores of the winner and runner up being independent of permutations of the voters or candidates.

5 Comparison of Solving Approaches

In this section we compare several methods for finding optimal and near-optimal solutions for each entertainment measure. We initially used the Gurobi solver,

	ANON	1TI	2TI	TI	LATE
PROX	□	×	×	×	×
R-PROX	□	×	⋄	⋄	×
LEAD	□	×	□	□	×
POS	□	×	×	×	×
R-POS	□	⋄	⋄	⋄	×
HOT	□	□	□	□	×
RES	□	×	□	□	□

Table 1: Cross-table of entertainment measures and their properties. □ means that the measure satisfies a property in all configurations, ⋄ means that the measure possibly (but not necessarily) satisfies a property, and × means that no configuration satisfies the property.

however, we found that the number of variables and constraints scales poorly with the size of the score matrix resulting in excessive execution times. An alternative tree-based approach was implemented that can quickly find exact solutions within a second on score matrices up to 8x8, however, it remains too slow for instances of a realistic size. We then compared several optimisation approaches on randomly generated 8x8 instances including tabu search, simulated annealing, and greedy construction heuristics. All experiments were run using Java version 21 on an Intel Core i7-1165G7 with 16GB of memory. We have excluded Resolution from this experiment as it was proven to be solvable in polynomial time in section 4.1. For R-POS and R-PROX we chose a rank bias of $w(p) = \max(\frac{5-p}{4}, 0)$, giving the top 4 candidates a decreasing weight from 1 to 0.25 and the remaining candidates a weight of zero. This provides a middle ground between the measures that consider all candidates equally and the measures that consider only the winner.

5.1 Experiment Design

Our comparison of solving approaches was run on randomly generated score matrices. To achieve realistic instances, voting is correlated to some "true" ranking of the candidates, randomly generated for each instance. Each voter's order of preference over the set of candidates is decided by sampling a set of normal distributions representing the performance of each candidate. Analysis of the points awarded in the 2019 and 2021 Eurovision finals showed that low-scoring candidates tend to have lower variance in the points received, while high-scoring candidates tend to have higher variance, as shown in figure 2. This method allows the generation of instances where points are realistically distributed among the candidates, however, it does not model the correlation between points received by candidates (e.g., in the case of Eurovision it has been shown that countries tend to vote for each other in blocs [12], so a candidate's expected score would be higher given a candidate in the same bloc receives points from the current voter).

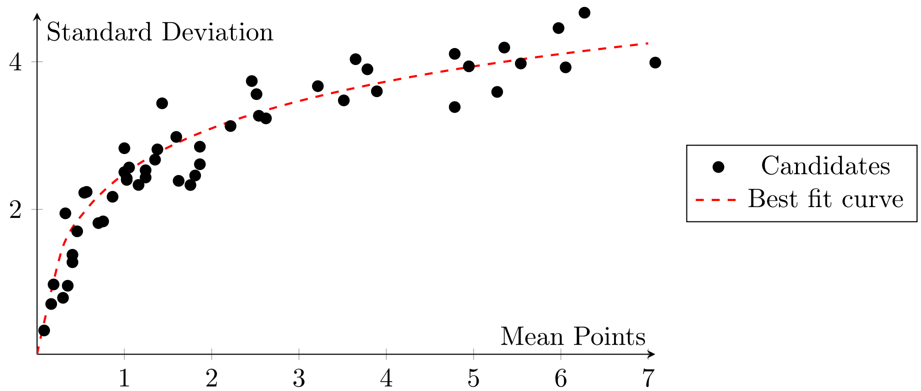


Fig. 2: Mean and standard deviation of points received by candidates in the 2019 and 2021 Eurovision finals.

We found that the distribution of mean points received has a mean of 2.27, standard deviation 3.70, and skewness of 0.90. The distribution is curved, so we calculated a curve of best fit $f(x) = 2.41 + 0.94 \ln x$ using exponential regression. The standard deviation of the residuals is 0.31. Therefore, random candidate performance distributions can be generated by first getting the mean score using a skewed normal distribution $X \sim N(2.27, 3.70, 0.90)$ and the standard deviation is given by $Y \sim N(2.41 + 0.94 \ln X, 0.31)$.

Several voting formats were chosen to represent varying levels of preference information given, listed in table 2. The strongest format is Borda count, where candidates are ranked in complete order of preference and assigned a decreasing number of points from M to 1; the weakest format is single point voting, where the assignment of a single point by each voter represents the least information that could be given; and the two "half" formats were chosen to represent intermediate levels of strictness. The half Borda format was chosen to mimic Eurovision's format, where some candidates are strictly ranked with decreasing amounts of points and the remaining candidates receive zero points. The half approval format was chosen to mimic sports tournaments, where half of teams will win a match on any given round and the other half will lose. For instances where M is odd, $\frac{M}{2}$ is rounded up meaning the middle candidate receives 1 point in the half approval and half Borda formats.

Name	Example	Rule for position p
Borda	$(6, 5, 4, 3, 2, 1)$	$M + 1 - p$
Half Borda	$(3, 2, 1, 0, 0, 0)$	$\max(\frac{M}{2} + 1 - p, 0)$
Half Approval	$(1, 1, 1, 0, 0, 0)$	if $p \leq \frac{M}{2}$ then 1 else 0
Single Point	$(1, 0, 0, 0, 0, 0)$	if $p = 1$ then 1 else 0

Table 2: The voting formats used.

A search algorithm was implemented to find exact solutions, allowing us to find the optimal and least-optimal scores on sufficiently small instances. Let

$\mathbb{E}_S(\max)$ and $\mathbb{E}_S(\min)$ denote these, and let the *optimality* of a reveal order ρ on S be given as $\frac{2(\mathbb{E}_S(\rho)-\mathbb{E}_S(\min))}{\mathbb{E}_S(\max)-\mathbb{E}_S(\min)} - 1$ such that optimal reveal orders have a score of 1, least-optimal reveal orders have a score of -1, and other values are linearly interpolated in that range. Four optimisation algorithms were considered. Tabu search [8] and simulated annealing [3] were implemented using the set of permutations ρ on S and a neighbourhood function that swaps any two elements in ρ. A basic greedy heuristic was implemented using the same tree structure as the exact algorithm, choosing an optimal voter to fix at each time t without exploring the subtree. GC L begins by fixing the first voter and working from left to right, while GC R begins with the final voter and works from right to left. We generated 1000 random 8x8 score matrices (250 of each voting format) and calculated the mean optimality of each algorithm on each entertainment measure, shown in figure 3.

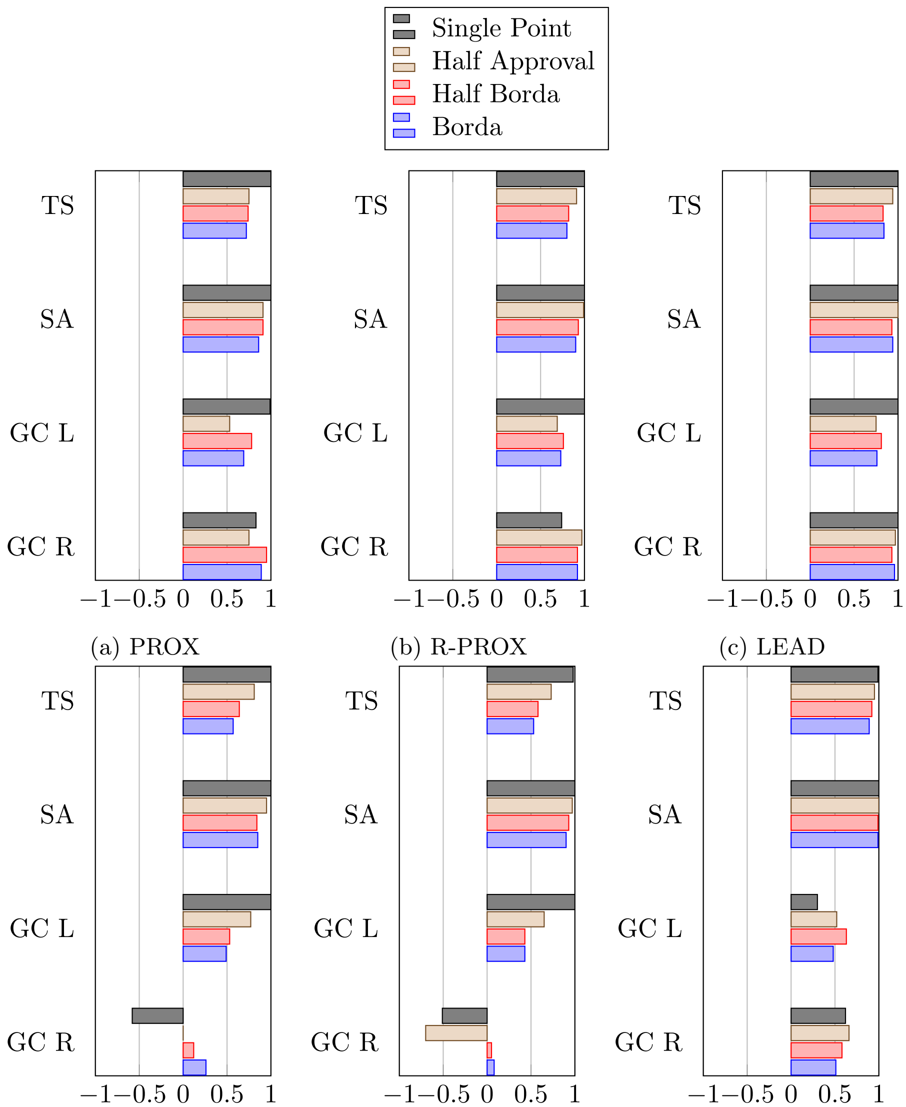
5.2 Results

Across the board we see that single point voting is easy to solve, while stricter voting formats become increasingly more difficult. A notable exception to this is the Lead measure (figure 3c), where both optimisation algorithms achieved slightly better results with the borda format compared with half borda. This may be due to less consistency in point incomes on half borda: it is more likely for a leading candidate to receive a zero and be overtaken, creating a large gap which is less optimal.

We also see that restricting entertainment measures with a rank bias makes them easier to solve. The consistently good performance of greedy construction on the Lead measure indicates that this may even be polynomially solvable, while the unrestricted Proximity measure is best tackled with simulated annealing. The Position Shifting and Rank-Biased Position Shifting measures seem harder to solve, with the runtimes of SA and TS averaging 4.56ms per instance, more than double the 1-2ms per instance required for any other measure. GC R performs exceptionally poorly here, which we believe is due to the greedy algorithm being unable to move the candidates together while working backwards from the end of the reveal leading to poor-quality solutions. The greedy algorithms struggled with the Hotseat measure, most likely due to the arbitrary breaking of ties when multiple voters could result in a change in leader. A more sophisticated heuristic for breaking ties may result in much better performance here, so we have not ruled out Hotseat being a polynomial-time problem.

6 Conclusion

The score reveal problem is a new problem with the unique challenge of optimising a subjective function. We have formally defined a framework for tackling the problem and describing the properties of a given spectator's entertainment utility, including presenting the properties of Late Resolution, Anonymity, and k-Tail Independence. The Position Shifting and Proximity functions provide baseline entertainment measures that can be modified with a rank bias to focus on



(a) PROX (b) R-PROX (c) LEAD

(d) POS (e) R-POS (f) HOT

Fig. 3: Average optimality achieved by each alorithm on each entertainment measure, split by voting format. Algorithms: tabu search, simulated annealing, greedy construction (left to right), and greedy construction (right to left).

higher-ranked candidates. Our experiment showed that the hardness of finding optimal reveal orders depends on the strictness of the voting format used and the level of information required by the entertainment measure. Simulated annealing was the most effective algorithm by average solution optimality for all measures, with Position Shifting seeming to be the hardest to optimise. While small instances were used within the experiment for comparison to known optima, high-quality solutions on large matrices ($>$100 voters) are typically found within minutes, demonstrating the ease of application to real contexts.

A clear area for future work is the consideration of non-neutral spectators; this opens opportunities for modelling candidate and voter bias, which we believe could improve entertainment value by playing into the spectator's expectations. There is also the natural extension to an audience of spectators, where a combination of preferences can be considered using multi-objective optimisation or compound entertainment measures. The computational complexity of solving the Position Shifting, Proximity, Hotseat, and Lead entertainment measures remain open questions and there remains a need to test their effectiveness in real contests or through research studies. Another consideration is the use of alternate scoring systems such as *rank dependent* rules [9].

Acknowledgements. The authors would like to thank Joe Singleton for his precise proofreading and the reviewers for their helpful suggestions.

References

1. How the Eurovision Song Contest works | Eurovision Song Contest (May 2019), https://eurovision.tv/about/how-it-works
2. Voting systems in Eurovision history. EUROVISIONWORLD (Nov 2022), https://eurovisionworld.com/esc/voting-systems-in-eurovision-history
3. Bertsimas, D., Tsitsiklis, J.: Simulated annealing. Statistical science **8**(1), 10–15 (1993)
4. Bizzozero, P., Flepp, R., Franck, E.: The importance of suspense and surprise in entertainment demand: Evidence from Wimbledon. Journal of Economic Behavior & Organization **130**, 47–63 (2016)
5. Bryant, J., Rockwell, S.C., Owens, J.W.: "Buzzer beaters" and "barn burners": The effects on enjoyment of watching the game go "down to the wire". Journal of Sport and Social Issues **18**(4), 326–339 (1994)
6. Dobni, D.: Entertainment value: The concept and its dimensions. Journal of Hospitality & Leisure Marketing **15**(4), 5–23 (2007)
7. Ely, J., Frankel, A., Kamenica, E.: Suspense and surprise. Journal of Political Economy **123**(1), 215–260 (2015)
8. Glover, F.: Tabu search—part I. ORSA Journal on computing **1**(3), 190–206 (1989)
9. Goldsmith, J., Lang, J., Mattei, N., Perny, P.: Voting with rank dependent scoring rules. In: Proceedings of AAAI. vol. 28 (2014)
10. Kaplan, S.: Entertainment utility from skill and thrill. SSRN 3888785 (2021)
11. Simonov, A., Ursu, R.M., Zheng, C.: Suspense and surprise in media product design: Evidence from Twitch. Journal of Marketing Research (2023)
12. Yair, G., Maman, D.: The persistent structure of hegemony in the eurovision song contest. Acta sociologica **39**(3), 309–325 (1996)