Strategic Thinking in Rupaul's Drag Race All Stars: a game theory analysis

By John Norell, 29/3-2019

Rupaul's Drag Race All Stars is an American reality competition show for drag performers. In simple terms, the format is this: a group of contestants each episode compete in different challenges related to drag, each week one gets voted of the show and when there's only one left, that contestant is crowned the winner. A feature of the All Stars game that is different from an ordinary season of Rupaul's Drag Race is that instead of the judge (Rupaul) deciding who's sent home, the winners of the episode challenges decide which of the two worst-performing contestants that gets eliminated. This introduces a strategic element into the competition, and this is what will be analysed in this text.

If you are the winner of the episode challenge, should you eliminate the strongest competitor or is there an advantage to gain from acting fair and eliminate the weakest one? Is it strategic to form alliances? Why is it, that in the show, the contestants tend to establish a norm of always sending home the weakest of the bottom contestants? These are some of the questions questions I will answer by using the tools of game theory.

In the next section, I outline the basic model and solve it to see what predictions it makes about how contestants will behave. I will then analyse the model and in particular different contestants probability of winning the whole game. Next, a number of modifications are proposed and analysed how they affect the predictions the model makes. These include asymmetric information, formation of alliances, a preference for being fair, and an agreement between all contestants to always eliminate weaker contestants. At the end of the paper is an appendix describing how to create a simulation model that can be used to calculate different contestants probability of winning and who benefits and loses from different types of collective agreements.

Constructing the All Stars-model

To start with, the game has players. In Rupaul's drag race, there's somewhere around 10 to 14 contestants at the start of the game/season. The number of contestants doesn't (for now at least) affect the model in any crucial way however.

A model is a tool to abstract from reality and simplify a strategic situation, so for now we will assume that the only thing the contestants care about is winning the game. We will loosen this assumption later, as we analyse the model more closely.

Next we assume that it is possible to rank all the contestant on their talent (or skill level), from most talented to least talented. Talent in this model is an underlying characteristic that is embedded within each player (a type of human capital) and will affect their chances of winning and losing the challenges each episode.

The game has different rounds, and in each round (episode), the players are given a task to perform. Their performance is evaluated and a winner is declared and two contestants end up in the bottom two, where they risk elimination from the game. In the model, the winner and losers are generated randomly, though the contestants don't have an equal chance of ending up in the top and bottom.

Talent level – their ranking in relation to the other contestants – translates into a higher probability of winning and a lower probability of losing in each round. In the context of Rupaul's Drag Race All Stars, talent for the craft of drag performance/artistry give contestants an advantage in the weekly challenges.

The ranking of talent is assumed to be static throughout the game. This is a simplification, in the real world different contestants have different abilities which gives them a comparative advantage depending on the nature of the challenge each episode (singing, acting, dancing, costume making etc.). But this is incorporated in the random element of the game. The talent ranking doesn’t perfectly translate into winning and losing, and the randomness concerning what the challenge will be each episode will be part of the random component of the model.

The winner then decides which of the bottom two contestants will be eliminated from the competition. The remaining contestants continue to the next round. The game continues until there is only one contestant left. This contestant the winner of the game.

In the real competition, things aren’t as simple as we’ve assumed here. The rules are more complex and some of the assumptions we’ve made may not seem true at all times. Have patience, we will add on more complexities further on when we analyse the model.

Solving the model

To solve the game, we start by looking at what will happen in the last round of the game. At that point, there are only two contestants left. The winners of the last round will send home the other contestant and become the winner of the whole game. There's not much of a strategic element here.

As long as there are more than two players from start, the game will have more than one round. What then happens in the second to last round?

There will then be three contestants. The winner of that round will get to decide which of the other two contestant will join him in the last round. Since everyone knows how the last round will play out, the only way the episode-challenge winner can affect their chance of winning the game is by eliminating the strongest competitor of the two other contestants so that they will have a greater chance of also being the episode-challenge winner in the last round.

What happens in the third to last round? Since it's already clear how the winner of the second to last round will act, the only way for the episode-challenge winner in the third to last round to improve their chance of winning the game is to eliminate the strongest competitor of the bottom contestants. The same will happen in the fourth to last round, the fifth to last round and all other rounds. The prediction the model makes is that in each round, whoever wins the episode-challenge will send home the strongest of the two bottom contestants. This is the only way they can affect the outcome of the game since the decisions by future episode-challenge winners will not be affected by what's happened in the past.

Analysing the prediction of the model: "The least talented contestant is the most likely winner"

One implication of this is that the worst (least talented) contestant in the group actually have a decent chance of winning the whole game. The reason is that every time he is in the bottom two, and at the risk of being eliminated, he will be saved since he is never the strongest competitor. (Think Chi Chi DeVayne in All Stars 3).

In fact, if everyone acts according to the model, we can be certain that the least talented contestant will stay in the game until the last round. The winner of that round will then eliminate the other one and win the whole game. Depending on who that other person is, and how much talent affects your chance of winning the last episode-challenge, the probability of the least talented to win the whole game can be determined.

The most talented on the other hand will not have as great odds of winning as one might think. This is because if they ever happen to be in the bottom two, they will surely be eliminated as they're always the strongest and most dangerous competitor to the other contestants.

For every contestant, the more talent they have, the less likely they are to end up in the bottom, though if they happen to do so, they are more likely to be the stronger of the bottom two and hence be eliminated. To know the exact probabilities of winning for each contestant, we would need to know how talent translates into probabilities of winning and losing in the challenges. In the appendix, I describe how to make a simulation model to calculate this[1].

Out of all possible pairs in the last round of the game (with 14 contestants there are 91 possible ending pairs), I believe the most likely one is that the most talented and least talented players are the ones to go head to head. Though this might depends on how much of an advantage talent is so I'm not really sure. The least talented will always be around until the end, though it's not clear who will join him.

If talent gives only a slight advantage, then the least talented person is both more likely to have a chance at winning that last round and ending up as the winner of the whole game. But he will also be more likely to stand head to head in the last round with a competitor that's also not very talented. This is because when talent only gives a small advantage, very talented players will still very often end up in the bottom, and as soon as they do, they are booted off the competition.

Anything that increases the role of chance in this game (opposed to skill) will benefit the least talented players in the group. If talent isn't strongly tied to success in the challenges, then the most likely winner of the whole game is the least talented contestant. This may seem like an odd and somewhat counterintuitive outcome.

[1] It's not very difficult, though I don't think it can be done in excel, and I haven't had time to set up the simulation in python. If anyone would do it, I'd love to see the results!

Talent vs. Luck: “Without the element of randomness, there would be no need for strategizing”

What is key is how talent translates into probabilities of winning and losing each round. If talent completely determined winning and losing (i.e. the most talented contestant in the group always win and the least talented are always be in the bottom), then there would be no point in strategizing. The most talented player would win every single challenge and the whole game. There is no point for the winner to have a strategy for who to eliminate in each round, because there's nothing to gain (or lose) from it.

The other extreme: if talent has nothing to do with who wins challenges each round (essentially saying that everyone is equally talented), we'll then there's no need for strategizing either since there is no way to affect your chances of winning future challenges or the game anyway. If the outcomes of each round is completely random, then we would have no idea of who would win in the end since its then all a game of chance. They might as well play dice about who should win.

To conclude, without the element of randomness, there would be no need for strategizing. Remember though that we were able to determine the optimal strategy in the previous section without taking into account exactly how talent translates into higher/lower probability of winning and losing challenges. If talent only gives a slight advantage, it's better for the winner of each round to always send home the strongest competitor. If talent gives a large advantage, so that the most talented contestant almost always wins in the challenges, it's still the best strategy to send the strongest competitor packing.

Can players gain by pretending to be bad?

Now that we’ve analysed the predictions the standard model makes, let’s look at what happens when we make some modifications. One assumption we’ve made is that the episode-challenge winner always knows who's the strongest competitor of the bottom contestants. If that's not the case, the least talented might be eliminated by chance. It's also possible that some players will fake being bad (not doing their very best early on) so that the winner will perceive them as less of a threat and eliminate the wrong person.

There is a trade-off when it comes to deciding whether or not to try and fool the other players however. Though it decreases the risk of being eliminated when being one of the bottom two, not doing your best will increase the risk of being in the bottom. For this reason, it's not clear if this would actually be a smart thing to do.

It's worth noting that the contestants that compete in the All Stars competition are not unknown to the public. They have all been on the show before and competed in one of the regular seasons, so their talents are common knowledge, both to the tv-audience and to the other contestants.

Loosening this assumption doesn't change the prediction about optimal strategy (to always try and eliminate the strongest of the bottom two) so I will not go deeper into how to incorporate asymmetry in information in the model. In the real game, it doesn't seem to be the case that people would fake being bad either. In fact, in the real show, winners tend not to eliminate the strongest competition, they do the exact opposite! So next, let's look at how this could be explained.

Why the model fails in its predictions

Episode-challenge winners tend to send home the weakest competitor, not the strongest one, the opposite of what the constructed model predicts. The reasoning they usually give for this is that they want to play fair. Is this the truth or might there be less nobel reasons for acting this way? Remember that there's a lot on the line here for contestants. Winning the show doesn't just get you a crown, there's $100 000 in prize money, and a substantial career boost which comes with the title of winner of Rupaul's Drag Race. Do contestants really care *that* much about fairness?

I can imagine that a lot of people would argue that the assumption we made in the beginning about players only caring about winning is wrong. And that this is the reason the model makes predictions that don't match with reality. This might be correct, but I'm not convinced this tells the whole story. What we observe in the TV-show is the emergence of a pact. In the beginning of the game, there's a discussion about how they should all play the game, and the outcome of these discussions is that each contestant more or less explicitly declares that if they win a challenge, they will "play fair" and eliminate the weakest of the bottom two.

Why the pact emerges

To understand why such an agreement emerges, let's look at who benefits and who loses from it. To start with, it's worth noting that for anyone to gain from the agreement (in terms of higher chance of being the winner of the whole game in the end), someone else must lose from it. The game is what's called a zero-sum game, the gains and losses will balance out completely.

It's clear to see that the least talented will be the biggest loser from such a pact. Recall that this person was sure to make it to the last round if winner's always eliminated the strongest of the bottom contestants. With this new elimination rule, suddenly, the least talented will be the least likely to win. In fact, for them to be the winner in the end, they'd have to avoid getting in the bottom for every single round and then win that last round. They certainly have the odds against them.

At the opposite side of the talent spectrum, we have the most talented contestant. Under this new elimination rule, this person will always stay until the last round. Though it's not certain that they will win in that last round, because of the element of randomness discussed previously.

Contestants with relatively high talent will gain from the pact and contestants of relatively low talent will lose by it. It's obvious why the contestants with relatively high talent will propose and advocate this pact, it's less clear why the rest of the group agrees to go along with it instead of proposing that they play according to the rule where the strongest of the bottom two is sent home.

The difference between these two possible agreements is that while there are contestants that have an incentive to advocate both pacts, only the relatively more talented have a way to enforce their pact. The highly talented are more likely to win the episode-challenges and hence get to decide about who to eliminate. Don't want to play by our rules? Guess what, you'll be the first one out. (Remember Morgan McMichaels being the first one sent home in season 3 because of this).

Any contestant planning to not uphold the agreement better not speak aloud of it, since this puts a target on their back. This explains why this subject is discussed so extensively throughout the show, to find out if anyone can be suspected of wanting to defect from the pact. If it was just about fairness, then it wouldn't matter much whether everyone else planned on also playing fair.

And if fairness was the reason the challenge winner didn't send home the strongest competitor when they got the chance, then why would they deviate from that promise (by eliminating defecters) just because someone say they don't plan on upholding the agreement. (When BenDeLaCreme eliminated Morgan in season 2, it was even explicitly said that it because Morgan wasn't trusted to keep the agreement).

How stable is the pact?

The reason that everyone agrees on sending home the weakest competition when they're the challenge winner is because there's a credible threat that if they don't, they will be the next one to go. But is this threat always credible?

As the game progress, the contestants are eliminated one by one. The relatively more talented will tend to be the ones to eliminate, and the relatively less talented will tend to be the ones to be sent home. This will continue for as long as the threat is credible. The closer the end we get however, the more unstable the pact gets. Recall that we already know what's going to happen in the last round of the game. This means that in the second to last round, the threat to the challenge winner of being eliminated in the last round (would they choose to defect) will therefore not be credible. Knowing this, is the elimination threat credible in the third to last round?

In every round, the challenge-winner have to decide whether they should uphold the pact and send home the weakest contestant or if they should better their own odds by draining the talent pool and risking retribution. Ideally they would like to send home the strongest competitor but get credit for doing the fair thing. This introduces another type of fraudulent behaviour, this time by the challenge winners. Though there's a risk of pushback if they try to do this (i.e. eliminate the strongest competitor but claiming they thought it was the weaker one).

The pact described is likely to arise initially but becomes less stable as the game draws near the end. If contestants care about how the tv-audience perceive them, this is likely to strengthen the pact additionally. Even if players don't care at all about fairness, they might still want to be perceived as fair if the tv-audience punish bad sportsmanship. It will also make it less likely that contestants will be open about what the agreement really is about. If you can claim to have noble motives (being fair) while acting strategically (improving your own odds of winning), then there's no reason to reveal your true motives.

A second feature of the real game that increases the strength of the agreement is the fact that in the real competition, instead of there being one episode-challenge winner that gets to decide on who is eliminated each round, there are two winners, of which only one (chosen randomly) decides about who's eliminated. Though both winners have to reveal who they'd pick to send home. This makes it even more risky to not uphold the pact, since it's possible that you will not be the one to send home your pick, but be revealed to be a traitor.

Unstable alliance formations

The pact described in the previous section is an agreement that dictate a shared strategy for all contestants. A similar kind of agreement is an alliance, where a subgroup of contestants get together to promise not to eliminate each other. It might be two people making this promise to each other, or it might be a larger subgroup. Such agreements are not common, and using the same reasoning as above, we can explain why that is.

Again, the all-stars game is a zero-sum game. Any agreement that benefits one contestant or a subgroup of contestants will hurt someone else. It might sound like a good idea to form alliances to try and improve your odds. The problem is that everyone who isn't part of the alliance will see it as a priority to break it up if they get a chance as it hurts their chances in the game. For that reason, alliances will not be likely to survive for long.

If the members of the alliance are among the most talented contestants however, it's possible that it can survive for longer. It's possible to imagine a divide among the contestants and we get two large alliances forming. It would then be a question of which group will succeed in eliminating all members of the other group first.

Conclusion

In this text, I've constructed a sequential game theory model to describe the strategic interaction in the reality competition show Rupaul's Drag Race All Stars. The model predicts that contestants, when being episode-challenge winners, will seek to drain the talent from the competition in order to improve their own odds of winning. This doesn't fit which what's observed in the real competition, where people do the complete opposite arguing it's because they wish to play fair.

It turns out however that honorable motives isn't needed to explain why an agreement to "play fair" might arise. Such a pact to eliminate the worst contestant each week is a way for the relatively talented to improve their odds at the expense of the least talented. The pact is likely to arise since there are contestants with both the incentive to advocate it and the tools to enforce it.

As the end of the game draws near, the pact is likely to break apart since the enforcement mechanism no longer is credible. This was seen in the fourth season of the show when Naomi Smalls broke the untold agreement in episode 8 (out of 10) and eliminated the frontrunner Manilla Luzon. It was after watching that episode that I started thinking about this model to try and explain why it happened.

The conventional view is that the optimal decision (strategy wise) is for the episode-challenge winner to eliminate the strongest competitor when given the opportunity, but that the wish to be fair leads people to refrain from the strategically best decision. What I've argued in this text however is that even though the contestants might claim high motives, the formation, and actions to play in accordance with the more or less untold agreement of "being fair" is just as strategic as any other.

Game theory is a subfield of economics, it's the study of strategic behaviour. Often it's applied to understand how individuals, firms or political parties compete for resources, votes and customers. The models and tools used in game theory however are very generalizable and can be applied to understand any strategic interaction. Game shows such as Rupaul's Drag Race All Stars is a perfect example of where game theory can be applied to understand human behaviour. The reason I wrote this text was mostly as a way to practice writing and economic modelling. The text doesn't contain any mathematical expressions, graphs or tables. I've written it with the intention that anyone (non-economists) should be able to follow along.

In the appendix below, I describe how one could make a simulation model to describe how the degree of talent vs randomness in determining probabilities of winning and losing challenges influence the game, and in particular each contestants probability of winning. It's also possible to use the simulation to exactly calculate how much each contestant benefits or loses by the pact (of eliminating the weakest contestants) described above compared to the default where it's the strongest contestant that is eliminated.

Appendix: Simulating Randomness vs. Talent

In the following section I will illustrate different ways talent can be tied to winning and losing each round. I have previously noted that in the two extreme cases (when talent completely determine outcomes and when chance completely determine outcomes) the game seize to be a strategic game, and hence there is no need for strategizing. As soon as there is influence of both, the optimal strategy is for the winner to always eliminate their fiercest competition. I will here show how the probability of winning the whole game can be calculated using a simulation and how the ratio of talent/randomness affect these outcomes.

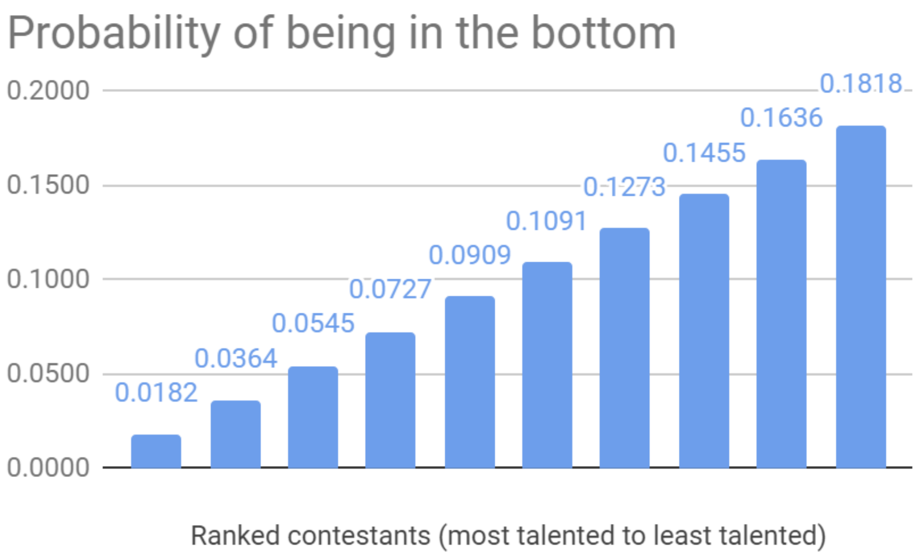
Let’s start by assuming the game has 10 players. These can all be ranked by talent, from the most talented to the least talented. Let’s have a look at the first round. Since we are interested in how the winning strategy affect the game, we only need to model how the bottom contestants are generated in each round.

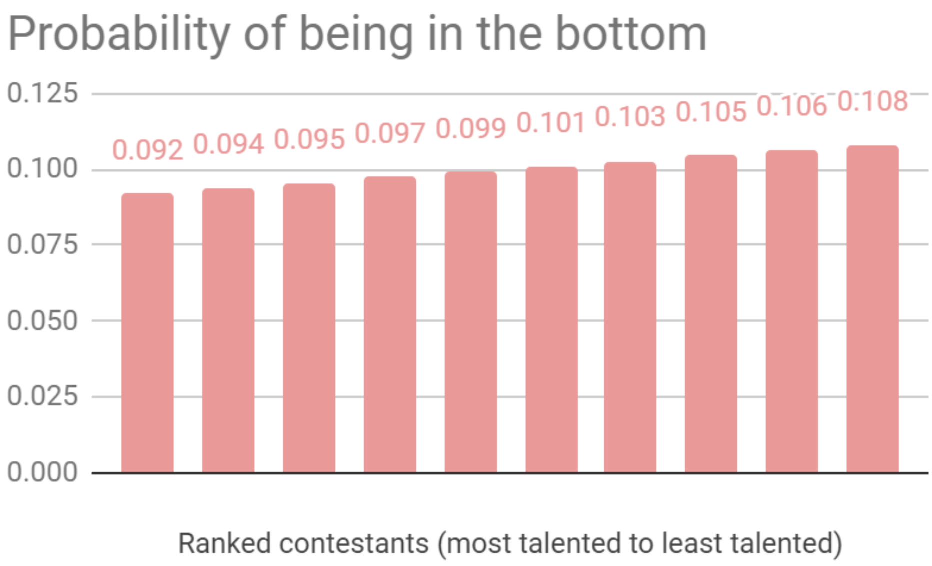
Imagine that we have a large basket in which we put one ball for each contestant. We then draw two balls at random to decide which contestants are in the bottom. If every contestant have the same number of balls in the basket, they will att have the same probability of ending up in the bottom.

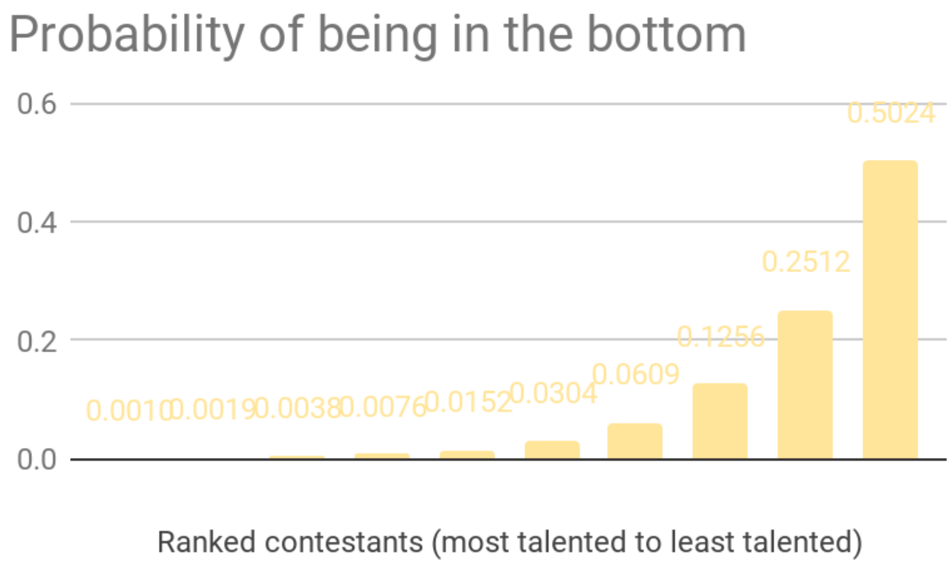
If on the other hand, each contestant has the same number of balls as their rank, then the most talented one will have one ball in the basket while the least talented will have 10 balls in the basket. The basket will contain 55 balls, which means that the probability of the most talented to end up in the bottom is 1.8% while it 18% for the least talented.

We can adjust the number of balls each contestant have in the basket to change their relative probability of ending up in the bottom. Below, I’ve shown a table with three ways talent may affect the probability of ending up in the bottom two. I have constructed three corresponding graphs showing each contestants probability of ending up in the bottom. In the blue version, the probability of ending up in the bottom is a linear function of your ranking. In the red version, each person have gotten an additional 50 balls in the basket. Now, talent have very little effect on the outcome. It’s almost only chance that decides outcome. In the yellow table, talent and loosing-probability is related exponentially. In this case, talent have a lot of influence on the outcome.

Talent ranking	nr of balls	P(being in the bottom)	nr of balls	P(being in the bottom)	nr of balls	P(being in the bottom)
1	1	0.0182	51	0.092	1	0.0010
2	2	0.0364	52	0.094	2	0.0019
3	3	0.0545	53	0.095	4	0.0038
4	4	0.0727	54	0.097	8	0.0076
5	5	0.0909	55	0.099	16	0.0152
6	6	0.1091	56	0.101	32	0.0304
7	7	0.1273	57	0.103	64	0.0609
8	8	0.1455	58	0.105	132	0.1256
9	9	0.1636	59	0.106	264	0.2512
10	10	0.1818	60	0.108	528	0.5024
Total	55		555		1051	







The next step is to determine the bottom two. One of these will be eliminated depending on what elimination rule we apply. After that, we continue to the next round. When it's only two individuals left, we have to change the approach however since this is when the winner of the game is declared. Determining the winner of the game can however be done in a similar way as we determined the bottoms.

After running the simulation of the game a couple of thousand times, we can get numbers on how likely each contestant is to win, depending on how much of an advantage talent is, as well for different elimination rules. We can find out exactly how each contestants chance of winning is affected by the agreement to always send home the weakest of the two compared to the default situation when they act in accordance with the models prediction.

Since there’s only one outcome observed in the real game show, we cannot find out how much of an advantage talent is. What we can do however is look at the odds for each contestant produced by betting markets before the game starts (these can easily be translated into probability statistics), find out how the audience believe talent relative to randomness influence the competition.