Economics Bulletin

Volume 30, Issue 4

The first shall be last: Serial position effects in the case contestants evaluate each other

Stefan D. Haigner
Gesellschaft für Angewandte Wirtschaftsforschung Innsbruck and University of Innsbruck

Stefan Jenewein
Gesellschaft für Angewandte Wirtschaftsforschung Innsbruck and University of Innsbruck

Hans-Christian Müller
Düsseldorf Institute of Competition Economics - University of Düsseldorf

Florian Wakolbinger
Gesellschaft für Angewandte Wirtschaftsforschung Innsbruck and University of Linz

Abstract

We analyze competitions where the contestants evaluate each other and find the first contestant to be disadvantaged. We suspect that this is due to information diffusion, Bayesian belief updating taking place in course of the contest and initial uncertainty about a contestant's relative quality.

We thank Matthias Sutter for helpful comments.
Citation: Stefan D. Haigner and Stefan Jenewein and Hans-Christian Müller and Florian Wakolbinger, (2010) "The first shall be last: Serial position effects in the case contestants evaluate each other", *Economics Bulletin*, Vol. 30 no.4 pp. 3170-3176.
Submitted: Oct 19 2010. **Published:** November 30, 2010.

1 Introduction

Many economic decisions involve the evaluation of alternatives which are presented in sequence. For example, job applicants are interviewed one after the other, contributions to architecture competitions are presented in a row, or decisions on the upcoming host of Olympic Games are made this way.

While from an efficiency point of view, best alternatives should be ranked first, the second-best ranked second, and so on, such evaluations have been found to be subject to a "serial position effect". For instance, data on synchronized swimming (Wilson, 1977), the Eurovision Song Contest (Haan et al., 2005, Bruine de Bruin, 2005), the Queen Elizabeth piano competition (Ginsburgh and van Ours, 2003, Glejser and Heyndels, 2001) or figure skating (Bruine de Bruin, 2006) have been employed to show that contestants who perform later are more likely to win, i.e. the last shall be first, although random assignment to the position in the series assured that the expected quality of all contestants is equal.

Beyond that, there is mixed evidence about the question whether this advantage rises linearly with the starting number (Bruine de Bruin, 2005) or whether there is a J-shaped relationship with the very first starter having a slight advantage compared to those with mid-level numbers, and greater advantages for later players (Page and Page, 2010, Haan et al., 2005). Contradicting both results, we find in analyzing the TV-cooking contest "The Perfect Dinner" that the first, and only the first, shall be last.

This cooking contest notably differs from other competitions in that competitors are both performers as well as jury members. Thus, each performance is judged by the rivals. Such scenarios might frequently be encountered if groups have to choose a leader from their midst, for example if party executives meet to find a new chairman after a sudden resignation of the predecessor.

We find the contestant to perform first to be disadvantaged. She has a lower probability of winning, although assignment to the position in the series is random. Thus, contestants should beware of performing first, while importantly, it makes no difference whether the contestant is second, third, fourth and so on in the row.

The paper is organized as follows. Section 2 introduces the contest and Section 3 describes the data and formulates some hypotheses. Section 4 presents the results and Section 5 summarizes.

2 The contest

The TV-cooking show "Das Perfekte Dinner" (The Perfect Dinner)[1] features five contestants who compete for the "best dinner of the week" award. From Monday to Friday, on each day one of the contestants cooks a three-course-meal which is evaluated by the four remaining contestants. It is possible to assign up to 10 points for a meal, such that the maximum total number of points to receive is 40[2].

Consumption and evaluation takes place immediately after cooking. However, the performing contestants are not informed about the number of points the others assigned. Importantly, the contestants are randomly assigned to the days on which they are asked to cook.

On Fridays, after each of the five meals has been evaluated, the winner of the week, who receives a price of € 1,500, is announced. In addition, everyone receives a lump-sum

[1] The show is broadcasted by VOX on evenings from Monday to Friday. Channel 4 (UK) previously broadcasted it under the name "Come Dine with Me", TLC (USA) called it "Dinner Takes All".

[2] For 22 weeks in 2009, the maximum total number of points is 80. We have normalized the results of these weeks to be consistent with the remaining data.

allowance of € 600 to cover the expenses. In case of a tie, the prize is shared equally among the winning contestants.

Within this setting, each contestant is both a performer (on one day) as well as an evaluator (on four days). From the point of view of contestant i, clearly the probability of winning decreases in the number of points she assigns to her competitors –i. For an egoistic and rational contestant it is thus a dominant strategy to assign zero points to each rival, and in the unique Nash-equilibrium each contestant assigns and receives zero points.

3 Data and hypotheses

We use data from 186 competitions which took place between March 2006 and February 2010. For each contestant, we observe the day on which she has performed and how many points she assigned to, and received from, each of her competitors. This allows calculating her rank. Moreover, we observe the contestants' gender and age. The data are available online at the webpage of the producer.

Although it might be optimal for every contestant to assign zero points to each rival, there seem to be moral constraints or social norms which prevent them from doing so. Moreover, although those performing sooner and later do on average not differ in quality due to random assignment, it might be better not to perform first for three reasons.

First, later contestants have more information on their opponents since there is some informal chatting about the meal provided by the previous contestants. Chatting might transmit information on the opponents' preferences and thus make later contestants able to respond to those preferences, i.e. by providing a special vegetarian meal for vegetarian opponents.

Second, assigning at most a medium number of points to the very first meal might be a sign of uncertainty about its relative quality compared to the following (Bruine de Bruin, 2005). Such a rating might serve as a benchmark for the later starters which must leave space for possibly better cooks. Hence, the very first starters would face the disadvantage that the jurors could not estimate the overall quality at the beginning.

Third, some Bayesian belief updating might take place. While the first evaluation mirrors the consistency of the jury's prior beliefs with the performance, later evaluations might be based on updated beliefs. If, say, contestants first expect an extraordinarily delicious meal, they will rate a mediocre meal worse than if they expect a mediocre meal.

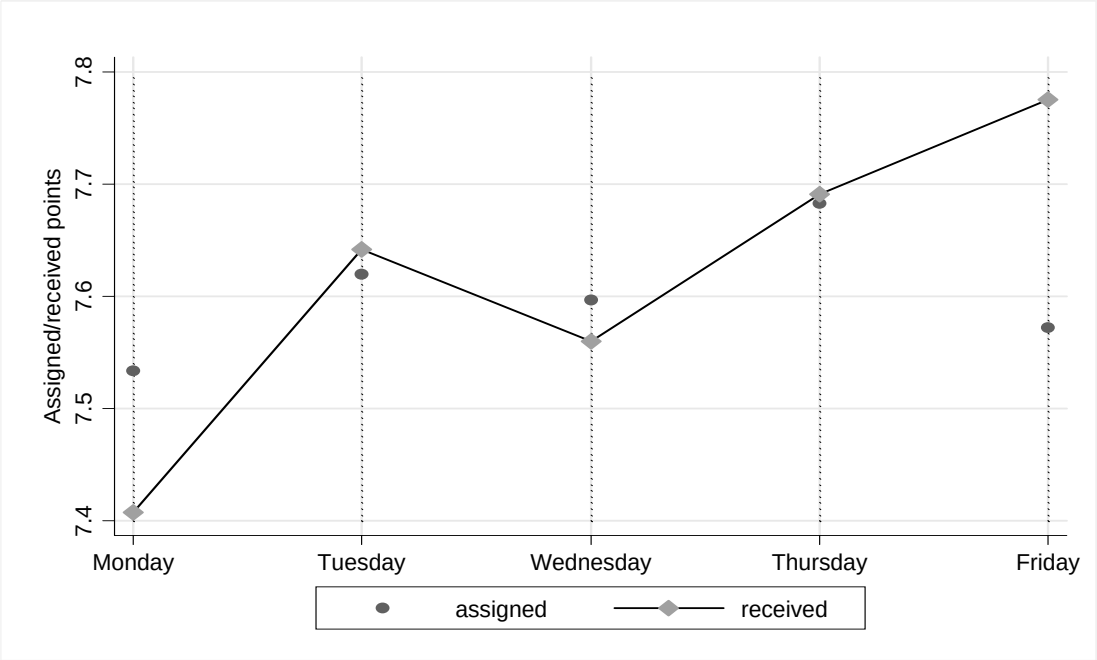
4 Results

4.1 Points

Figure 1 shows how many points (out of ten) contestants cooking on a specific day of the week on average received from and assigned to their competitors. Unlike standard economic theory would predict, average received and assigned points are far away from the lower bound of zero. Moreover, Monday-performances are on average rated significantly worse than performances on later days of the week[3].

[3] Mann-Whitney U-test p-values for a comparison of Monday's points with the points of other days of the week are <0.01 (Tuesday, Thursday and Friday) and 0.057 (Wednesday).

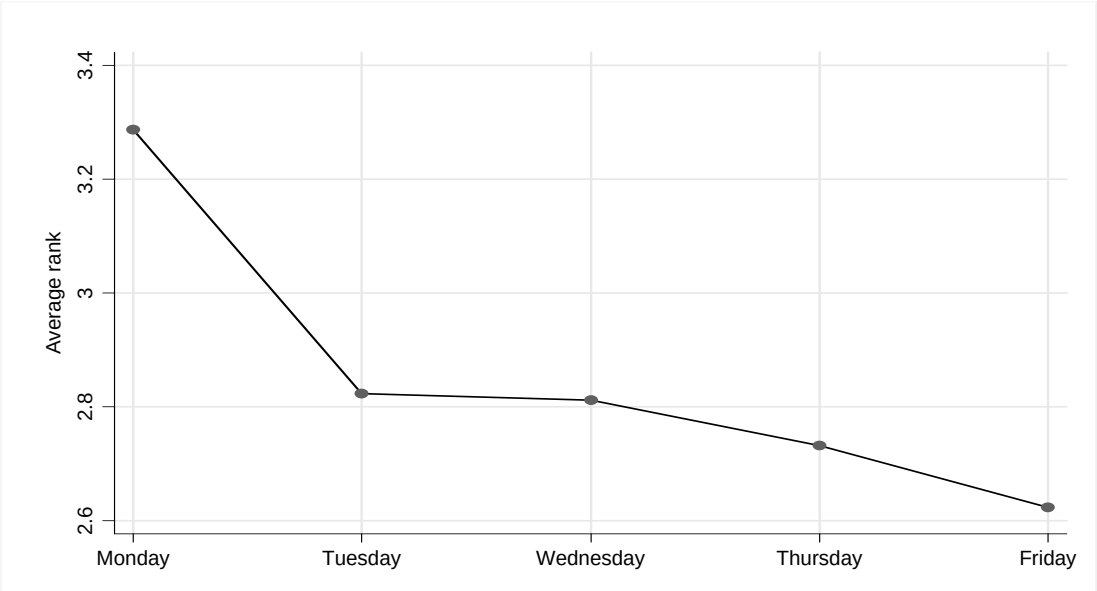
Figure 1: Average number of assigned and received points on each day of the week



4.2 Ranks

Figure 2 shows the average rank which the contestants of each of the five days of the week achieved. The average rank of Monday-contestants is higher by more than 0.4 than that of Tuesday to Friday-contestants, which differ in average ranks by less than 0.2 at most. Thus, we find a serial position effect where only the first performer seems to be disadvantaged.

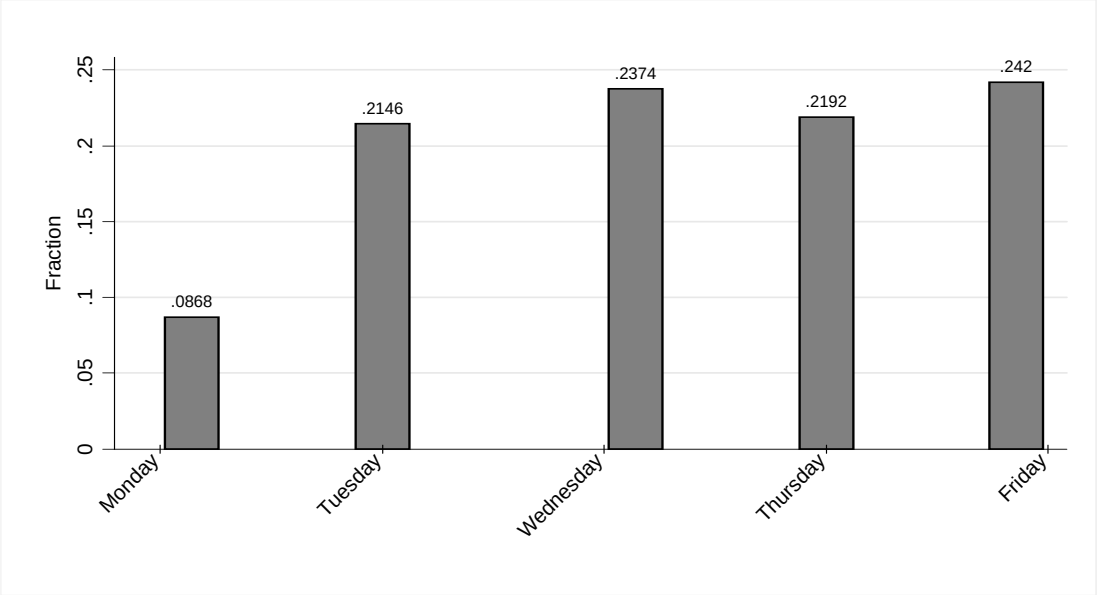
Figure 2: Average rank per day



4.3 Winning

Figure 3 shows that contestants cooking on Mondays have won in only 8.6%, while contestants performing on Tuesdays to Fridays have won in more than 20% of competitions. Moreover, there is surprisingly little variation in the probability of winning among Tuesdays' to Fridays' contestants. Figure 3 presents further evidence for the serial position effect with only the first contestant to be disadvantaged.

Figure 3: Fraction of victories on each day of the week



4.4 Determination of ranks and probabilities of winning

Table I below presents the results from ordered probit as well as probit regressions explaining the rank and probability of winning[4]. For the regressions given in columns 2 and 3, we use dummies for the day of the week (Monday is the basis) as well as indicators for gender and age as explanatory variables. The coefficient estimates show that the probability of winning when performing on Tuesdays to Fridays is by 21 to 26 percentage points and significantly higher than when performing on Mondays. However, the coefficients for the day-dummies do not significantly differ from one another.

Moreover, we test for the influence of a linear time trend and the first-performer disadvantage in an additional regression given in column 4. The latter disadvantage persists, while the time trend is insignificant, which indicates that apparently all relevant information diffuses at discussion of the first performance and that the competitors update their beliefs after the first performance while later on there is no essential updating anymore.

[4] We allow for correlation among those performers who compete against each other by clustering the observations by weeks.

Table I: Regression results explaining rank and probability of winning

	Ordered Probit Rank	Probit Winner	Probit Winner
Monday			-0.16
			[0.039]***
Tuesday	-0.327	0.217	
	[0.115]***	[0.062]***	
Wednesday	-0.347	0.248	
	[0.118]***	[0.063]***	
Thursday	-0.398	0.226	
	[0.111]***	[0.061]***	
Friday	-0.495	0.261	
	[0.118]***	[0.064]***	
Time			0.009
			[0.015]
Female	0.144	-0.049	-0.049
	[0.077]*	[0.030]	[0.030]
Age	0	0	0
	[0.004]	[0.001]	[0.001]
Observations	902	902	902
Clusters	181	181	181

Robust standard errors in brackets

Clusters by week

* significant at 10%; ** significant at 5%; *** significant at 1%

5 Conclusion

Previous research has found that in contests with independent juries, competitors performing later have advantages over those performing sooner, i.e. the last shall be first. This holds for both end-of-sequence (all contestants are evaluated at the end of the competition) as well as step-by-step (each contestant is evaluated right after her performance) judgements (Bruine de Bruin, 2006).

For the TV-cooking contest “The perfect dinner” in which the competitors evaluate each other we find, however, that only the very first shall be last. The first contestant has a significantly higher rank and a significantly lower probability of winning than her competitors performing afterwards. However, the following four cooks do not have any disadvantages relative to each other.

Why is the first performer, and only the first, last if competitors evaluate each other? We suspect that this is due to the following three reasons. First, the first performer might set a reference point for further performers, and beliefs on the performances to be expected might be updated after the first performance. Consistency (leading to higher evaluations) of updated beliefs and performance might be easier to achieve than consistency of prior beliefs and performance. Second, jurors might assign only a medium number of points to the first performer out of uncertainty about the relative quality of her meal compared to the upcoming ones. As such, the first rating might be a benchmark which must leave space for possibly

better meals. Third, in course of the contest, information on the contestant's preferences could be revealed, which disadvantages earlier performers, in particular the first contestant.

6 Literature

Bruine de Bruin, W. (2005) "Save the last dance for me: unwanted serial position effects in jury evaluations" *Acta Psychologica* **118**, 245-260.

Bruine de Bruin, W. (2006) "Save the last dance II: Unwanted serial position effects in figure skating judgments" *Acta Pychologica* **123**, 299 – 311.

Ginsburgh, V., van Ours, J. C. (2003) "Expert opinion and compensation: Evidence from a musical competition" *American Economic Review* **93**, 289 - 296.

Glejser, H., Heyndels, B. (2001) "Efficiency and Inefficiency in the ranking in competitions: The case of the Queen Elisabeth music contest" *Journal of Cultural Economics* **25**, 109 – 129.

Haan, M., Dijkstra, S.G., Dijkstra, P.T. (2005) "Expert Judgment Versus Public Opinion – Evidence from the Eurovision Song Contest" *Journal of Cultural Economics*, **29**, 59 – 78.

Page, L., Page, K. (2010) "Last shall be first: A field study of biases in sequential performance evaluation on the idol series" *Journal of Economic Behaviour and Organisation* **73**, 186 – 198.

Wilson, V. (1977) "Objectivity and effect of order of appearance in judging of synchronized swimming meets" *Perceptual and Motor Skills* **44**, 295 – 298.