They were robbed! Scoring by the middlemost to attenuate biased judging in boxing

Stuart Baumann Carl Singleton

February 9, 2024

Abstract

Boxing has a long-standing problem with biased judging, impacting both professional and Olympic bouts. "Robberies", where boxers are widely seen as being denied rightful victories, threaten to drive fans and athletes away from the sport. To tackle this problem, we propose a minimalist adjustment in how boxing is scored: the winner would be decided by the majority of round-by-round victories according to the judges, rather than relying on the judges' overall bout scores. This approach, rooted in social choice theory and utilising majority rule and middlemost aggregation functions, creates a coordination problem for partisan judges and attenuates their influence. Our model analysis and simulations demonstrate the potential to significantly decrease the likelihood of a partisan judge swaying the result of a bout.

JEL Codes: D91, L83, Z20, Z28
Keywords: Scoring Rules, Judgement Bias, Contests, Pugilism, Combat Sports

stuart@stuartbaumann.com; Corresponding author: carl.singleton@stir.ac.uk, Economics Division, Stirling Management School, University of Stirling, Stirling, FK9 4LA, Scotland, UK.
We are grateful for comments and advice from Anwesha Mukherjee.
Declarations of interest: none

1. Introduction

Boxing has a reputation for partisan and corrupt judging. At the amateur level, some decisions in Olympic gold medal bouts have attracted criticism and ridicule, becoming boxing folklore, such as Roy Jones Jr.'s defeat in the 1988 (Seoul) light heavyweight final to a South Korean fighter (Ashdown, 2012), and Joe Joyce's defeat in the 2016 (Rio de Janeiro) super heavyweight final (Ingle, 2021; Rumsby, 2021). In professional boxing, there is longstanding suspicion about the integrity of judges (e.g., US Senate, 2001). Recent perceived "robberies" include Haney Vs. Lomachenko (Wainwright, 2023) and the first two editions of Alvarez Vs. Golovkin (Reid, 2023).

This short paper models the decisions of boxing judges and proposes an alternative scoring method that has the potential to significantly attenuate judge bias. Currently, scoring at the elite level is on a per-judge basis, with three judges usually employed for elite professional bouts and five at the Olympic amateur level. Judges score each round individually and then award their entire "vote" to the boxer who wins a majority of rounds.[1] The bout is then awarded to the boxer receiving votes from a majority of judges. If neither boxer receives a majority, due to at least one tied scorecard among the judges, then the bout is a draw. In this system, "*aggregation over rounds and then judges*", or *"majority judges rule"*, it is relatively straightforward for a judge to ensure their vote goes to their favoured boxer. They just need to award them half the rounds (i.e., 7 of 12 for a world championship level men's professional bout). They can do this while minimising backlash, by choosing the best rounds for their favoured boxer.

The change we propose, "*aggregation over judges and then over rounds*", or "*majority rounds rule*", is for each round to be awarded based on the aggregate scores over all judges. Whoever wins the majority of rounds wins the bout, rather than whoever wins on a majority of the judges' scorecards. This represents a minimalist change to the scoring system in the sport, so that the aggregation of judges' scores is first between them within rounds, and then over rounds, rather than vice versa. The minor nature of this change is sufficient to introduce a significant coordination problem for a partisan judge, and may be acceptable among fans.[2]

We focus on modelling the simplest practical case, with three judges, one of whom is biased in favour of one boxer. Under majority judges rule, a partisan judge can substantially increase the probability of a boxer winning despite being outnumbered by unbiased judges. Under majority rounds rule, even if the partisan judge awards a majority of rounds to a favoured boxer, then this will have no impact on the final outcome unless those rounds align with the decisions of the other judges. This coordination problem implies that, to achieve a high probability of victory for their favoured boxer, the partisan judge would have to award more rounds to their favoured boxer than in the current system. This exposes them to scrutiny and potential backlash, as

[1] This is a slight simplification as judges can award additional points for a given round based on knockdowns, fouls, or particularly dominant performances by one fighter.

[2] Fans tend to scrutinise, oppose and criticise even quite small changes to the rules of their beloved sports. A notable example from cricket is the LBW rule, which has been continually 'improved' over the last century, often under opposition and criticism (Kumar, 2022).

boxing pundits and fans will often criticise poorly awarded rounds on judges' scorecards.[3] Our analysis and simulations of the model demonstrate that the scoring rule change could be highly effective in diminishing the incentives for biased judging in boxing and its influence.

Our proposed scoring rule is an application of majority rule and the middlemost aggregation function from social choice theory, which minimise the effective manipulability of outcomes by graders (e.g., Arrow, 1963; Balinski and Laraki, 2007; Young, 1974*b*,*a*). This principle is already applied somewhat to the scoring in boxing, since the decision of the middlemost judge now determines the bout result. We suggest awarding bouts based on the aggregated middlemost round-by-round votes instead.

The prevalence of judging bias in combat sports has been documented in a growing literature of empirical academic papers (e.g., Lee, Cork and Algranati, 2002; Holmes, McHale and Zychaluk, 2024). This behaviour was also described vividly in the recent judge-led independent investigation McLaren (2022) report, which examined unethical conduct in Olympic boxing after being commissioned by the Association Internationale de Boxe Amateur (AIBA). While the report did propose improved appointment processes and training of judges, it did not explore how to make the incentives inherent in the judging process more resilient to biases and corruption.

Our goal of improving the incentives of judges in boxing is closely related to the focus of Frederiksen and Machol (1988), who analysed the judging in sports like figure skating and dance, where judges need to decide between multiple competitors, a setting where Arrow's (1963) theorem implies that all possible ways to combine judge preferences have some undesirable characteristics. Frederiksen and Machol proposed a new method for aggregating judge scores for such situations that attenuates some of these issues. Their context though faced the problem of the Arrow Impossibility Theorem (social choice paradox), given there were more than two alternative outcomes in the contest. That theorem does not apply here for a boxing bout since it consists of just two competitors, only one winner, and potentially biased judges.

In general, this paper contributes to the vast operational research literature that either post hoc analyses changes to scoring rules and laws in sports or proposes new changes (for recent surveys see Wright, 2014; Kendall and Lenten, 2017). Our work falls into the latter type of study, particularly where minimalist changes have been proposed that could still in theory substantially improve the fairness of sports outcomes. For instance, in the world's most popular sport, association football, recent contributions have used simulations to explore whether incentives and outcomes could be altered significantly under different tie-breaking rules in round-robin tournaments (Csató, 2023; Csató, Molontay and Pintér, 2024), whether dynamic sequences in penalty shootouts could be fairer (Csató and Petróczy, 2022), and whether the allocation system for the additional slots of the expanded FIFA World Cup could be improved

[3]Criticism of judges in social and sports media is generally fiercest after bouts where a robbery is perceived to have happened. It often focuses on specific rounds where a judge's decision appears to be particularly poor. While authorities seldom intervene to order a rematch, judges may be stripped of their status and not employed again (e.g., Slavin, 2017).

according to the stated goals of the organisers (Krumer and Moreno-Ternero, 2023). Finally this paper builds on a growing sports economics and management literature studying various incentive issues in boxing and other combat sports (Akin, Issabayev and Rizvanoghlu, 2023; Amegashie and Kutsoati, 2005; Butler et al., 2023; Butler, 2023; Duggan and Levitt, 2002; Dietl, Lang and Werner, 2010; Tenorio, 2000). However, to the best of our knowledge, the incentives of boxing judges have not yet been studied, given the scoring rules they face, despite a well-developed literature on the influences and implications of biased decision making by the referees and judges in other sports (e.g., Dohmen and Sauermann, 2016; Bryson et al., 2021; Reade, Schreyer and Singleton, 2022), including other combat contests (Brunello and Yamamura, 2023)).

The remainder of our short paper proceeds as follows. In Section 2, we setup a styled model of potentially biased judging in a boxing contest. Section 3 describes our analysis and discussion of the model. The detailed proofs of the main propositions regarding the scoring rules are presented in the Online Appendix, as are variations on the main results from simulating the model.

2. The Model

Consider a contest between two boxers of equal ability, in the Blue and Red corners. We assume each sequential round $t \in \{1, 2, \ldots, N\}$ of the contest has a true result, $\tau_t \in \{B, R\}$, which is a binomial random variable with equal probability.

Each judge, $j \in \{1, 2, 3\}$, gets an i.i.d. signal, $x_{t,j} \in \{B, R\}$, about the result of a round. With probability $\alpha \in (0, \frac{1}{2})$ this signal is the incorrect result, $x_{t,j} \neq \tau_t$, while with probability $1 - \alpha$ it is correct, $x_{t,j} \equiv \tau_t$.

Judges have a utility of:

$$U = S\mathbf{1}_{\text{Blue wins}} + G\mathbf{1}_{\text{Red wins}} - L \ , \qquad\qquad L = \frac{\sum_{t=1}^{N} \mathbf{1}_{s_{t,j} \neq \tau_t}}{N} \ , \tag{1}$$

where $S \geq 0$ and $G \geq 0$ represents the a judge's value from Blue or Red winning respectively. L refers to a backlash cost.

We consider the case of two fair judges have $S = G = 0$. As these judge's utility does not depend on the bout's winner their optimal behaviour is to minimise backlash by awarding **fairly**, defined by choosing a round score of $s_{t,j} = B \iff x_{t,j} \equiv B$.[4] The third judge is **partisan** in favour of Blue and so has $S > 0$ and $G = 0$.

Under majority judges rule, the middlemost judge scorecard determines the bout. Under majority rounds rule, the middlemost judge determines each round's winner, and then the middlemost round determines the bout. Judges award rounds separately and simultaneously.

[4]We analyse the case where all three judges are fair in Online Appendix B.

3. Analysis, Results, and Discussion

The partisan judge ($j = 1$) can minimise backlash by awarding rounds fairly.[5] Under majority judges rule, they can maximally increase the chance of Blue winning the bout, while minimising backlash, by awarding $s_{t,1} = B$ in more than $\frac{N}{2}$ rounds. Under majority rounds rule, their problem is more complex; a judge could award more than $\frac{N}{2}$ rounds to a boxer who then does not win them because the other judges disagreed.

If S is low, however, then the expected backlash can be sufficient for the partisan judge to award rounds fairly. We can characterise the critical $\hat{S}$ where the partisan judge is indifferent between awarding fairly or gifting an additional round to Blue. We find that this critical value is higher under majority rounds than majority judges rule, indicating that the former is more resilient to judge bias.

Proposition 1. *For three-round bouts, in which Red won a majority of rounds according to the true realisations, $\boldsymbol{\tau} = [\tau_1, \tau_2, \tau_3]$, the critical $\hat{S}$ is higher under majority rounds than majority judges rule, $\forall \alpha \in (0, \frac{1}{2})$.*

Sketch of Proof: We can calculate the probability of each fair judge awarding a round for Blue, denoted by q, conditional on the signal seen by the partisan judge:

$$q|(x_{t,1} \equiv B) = (1-\alpha)^2 + \alpha^2 \ , \qquad\qquad q|(x_{t,1} \equiv R) = 2\alpha(1-\alpha) \ . \tag{2}$$

Under majority rounds rule, the number of fair judges awarding for Blue in a particular round can be represented as drawing from a binomial distribution with probabilities as in Equation 2. The survival function of this binomial, in conjunction with the decision of the partisan judge, is sufficient to infer the probability of Blue winning the round. From the probabilities for each round, we can derive the optimal number of rounds for the partisan judge to award for Blue. Under majority judges rule, we can infer the probability of another scorecard being in favour of Blue by combining the probabilities in Equation 2 across rounds. We can use these probabilities to evaluate whether the partisan judge should award additional rounds such that Blue wins on their card. Then we can derive, for each scoring rule, expressions for the critical $\hat{S}$ below which the partisan judge will award rounds fairly (see Online Appendix A). We find that there is a higher $\hat{S}$ under majority rounds rule, giving us the proposition. □

Proposition 1 establishes that the majority rounds rule is more robust to partisan judging than the majority judges rule for three-round bouts. We numerically solve the model to establish the robustness of this result in longer bouts.[6] We use a benchmark parametrisation of $\alpha = 0.1$, $S = 0.5$, three judges (one of whom is partisan), and $N = 12$ rounds.[7]

[5] If $S \equiv 0$, then this judge's actions will be congruent to the fair judges.

[6] This is done along the lines discussed in the sketch proof for Proposition 1. The code for calculating partisan judge best responses and bout simulations are included in the online supplementary material.

[7] Importantly, Online Appendix B shows that when all three judges are fair, the majority rounds rule is still more accurate than the majority judges rule in generating a deserving winner of the bout. Intuitively this occurs as

To demonstrate a partisan judge's decision making, Figure 1 shows the probability of Blue winning the bout, given they truly won 6 rounds, for each number of rounds the partisan judge awards them. Under majority judges rule, there is a sharp increase in the probability of Blue winning if the partisan judge awards them more than 6 rounds. If Blue truly deserved to win 4 or 5 rounds, then, to award Blue the win, the partisan judge only needs to risk the backlash associated with giving them 3 or 2 more rounds on their scorecard. In contrast, Figure 1 shows that under majority rounds rule, a judge cannot secure a sharp increase in the probability of Blue winning by giving them a small number of extra rounds; more rounds only gradually increase Blue's chances.

FIGURE (1) Simulated Probability of Blue winning, when both boxers truly won 6 of the 12 rounds, and 1 of the 3 judges favours Blue (in the absence of noise)

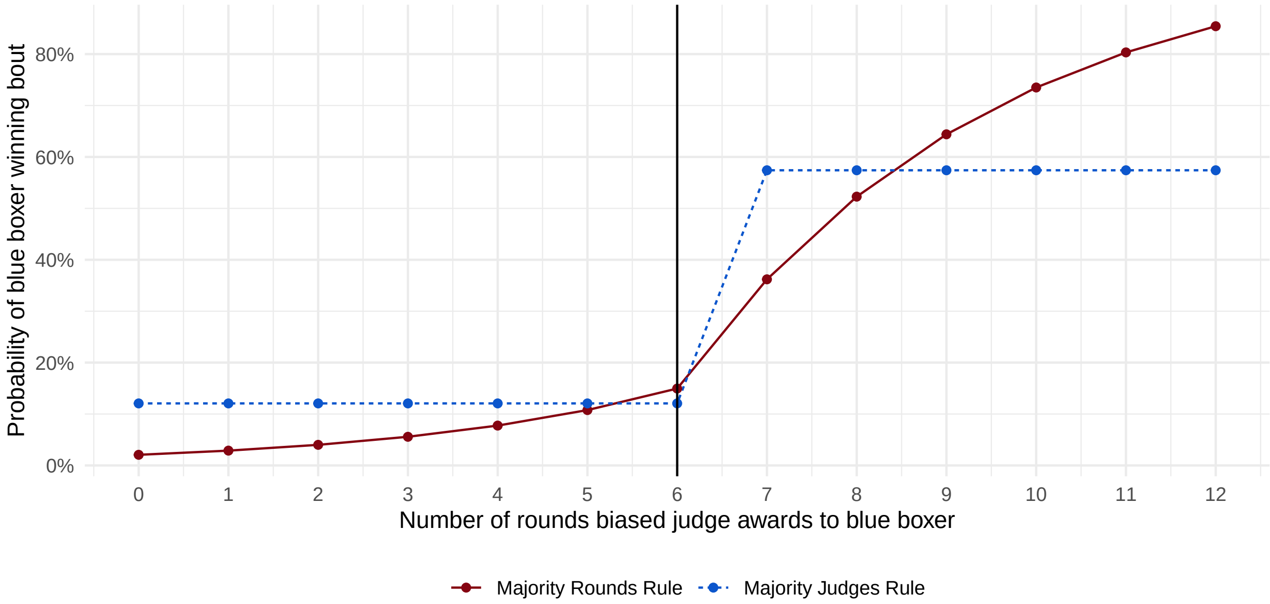


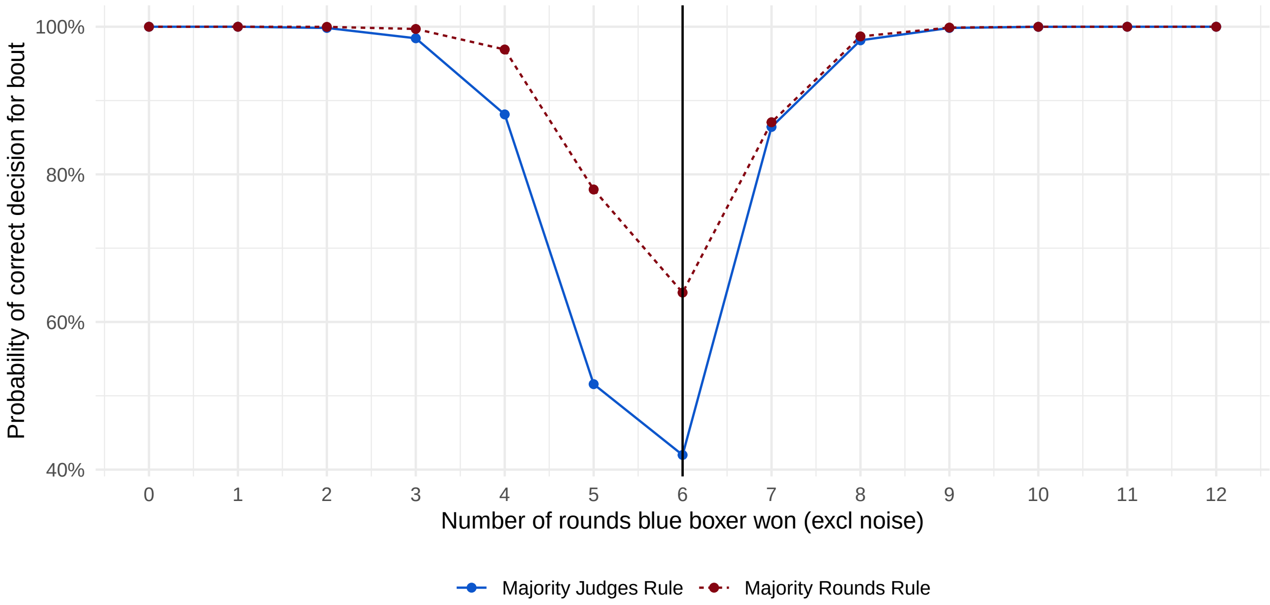
Figure 2 shows the impact of these differing incentives for the partisan judge, from running a series of simulations and counting the proportion of times each boxer wins under the two scoring systems, conditional on the true number of rounds won by Blue. When deciding the contest by majority judges, there is a high probability of erroneous results when Blue truly won only 4-6 rounds. When Blue truly wins the most rounds, the partisan judge unduly helps to lock in a deserved victory, so there is not a large difference in the number of incorrectly awarded bouts.

Finally, in the majority rounds case, it can be noted from Figure 2 that the probability of the

as there are more combinations of rounds that could be flipped to change the result. Consider a three-round bout with three fair judges and a τ realisation of $[\mathbf{B}, \mathbf{B}, \mathbf{R}]$. Consider that two mistakes happened in judging the bout (in that for two round-judges the $x_{t,j}$ realisation differs from τ). There are six possible pairs of $x_{t,j}$ values that can be flipped to change the result of the bout with the majority rounds rule. For one of the rounds where blue won, we need to flip two $x_{t,j}$ values and there are six combinations that achieve this. But there are twelve possible pairs of $x_{t,j}$ values that can be flipped to change the result of the bout with the majority judges rule. We need to flip two of the $x_{t,j}$ values awarded to blue on two different scorecards and there are twelve pairs of values that achieve this.

blue boxer winning always increases when the biased judge awards them more rounds. This is in contrast to the majority judges case, when a judge ceases to impact the result at the point at which they award a majority of their card to a boxer. For instance, consider a bout where the fair judge sees 10 rounds with $x_{t,1} \equiv B$ and only two rounds are seen to be won by red. In this case, the biased judge may award additional rounds to blue to lock in a blue victory, while they would have no such incentive under majority judges rule.

FIGURE (2) Probability of a "correct" result depending on the number of rounds truly won by Blue and how judges' scores are aggregated



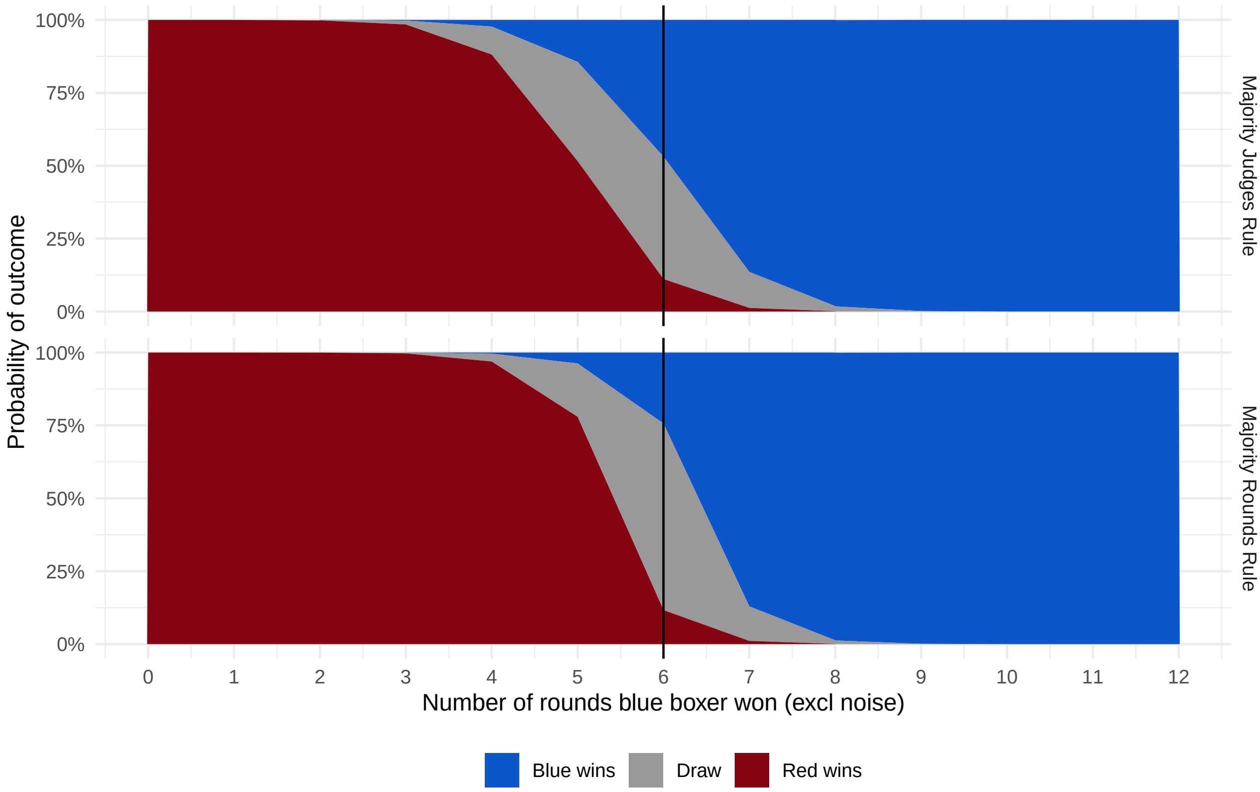
This effect, however, does not tend to lead to a greater probability of an erroneous result under the majority rounds rule. The main reason for this is that the effect occurs in a context where blue has likely won a large majority of rounds and is likely to win the bout. The more important case is when a bout is more even and there is a sharp increase, under the majority judges rule, in the winning probability at the 7 round level in Figure 2.

This point can be seen in Figure 3, which shows the probability of each possible outcome on the y-axis and the number of rounds Blue truly won (excluding noise) on the x-axis. Under majority judges rule (top panel), in evenly matched bouts, where the true result is a draw, Blue wins 46.9% and Red wins 11.1%. When evenly matched bouts are awarded under majority rounds rule (bottom panel), Blue wins 24.3% and Red wins 11.7%.

Figure 3 also shows the frequencies where one boxer wins despite the other deserving outright victory, e.g., the blue area to the left of the vertical black line. Under majority judges rule, it is more likely for an erroneous victory to be in favour of Blue than Red; in this parametrisation, a robbery in favour of Blue is 12.4 times more likely than a robbery in favour of Red. Under majority rounds rule, the likelihood of a robbery is still in Blue's favour, by 3.6 partisan, because there is still some incentive for the partisan judge to favour Blue. But this scoring system can substantially attenuate Blue's advantage from the presence of a partisan judge.

There are also fewer robberies in absolute terms.

FIGURE (3) Probability of each outcome depending on the number of rounds truly won by Blue and how judges' scores are aggregated



For robustness, the Online Appendices demonstrate extensions and checks on our analysis. Appendix C considers simulations with alternative parametrisations of the benchmark model, and, in Appendices D-F we repeat the analysis for setups consistent with women's professional, men's Olympic, and women's Olympic boxing, respectively (i.e., different numbers of rounds and judges). The results of all these extensions support our key findings: deciding bouts by majority rounds, compared with by majority judges, makes it less likely that a partisan judge sways the outcome of a bout.

References

Akin, Zafer, Murat Issabayev, and Islam Rizvanoghlu. 2023. "Incentives and Strategic Behavior of Professional Boxers." *Journal of Sports Economics*, 24(1): 28–49.

Amegashie, J. Atsu, and Edward Kutsoati. 2005. "Rematches in Boxing and Other Sporting Events." *Journal of Sports Economics*, 6(4): 401–411.

Arrow, Kenneth Joseph. 1963. *Social Choice and Individual Values.* Yale University Press.

Ashdown, John. 2012. "50 stunning Olympic moments No14: Roy Jones Jr cheated out of gold." *The Guardian*. https://bit.ly/3SQ5KhL.

Balinski, Michel, and Rida Laraki. 2007. "A theory of measuring, electing, and ranking." *Proceedings of the National Academy of Sciences*, 104(21): 8720–8725.

Brunello, Giorgio, and Eiji Yamamura. 2023. "Desperately Seeking a Japanese Yokozuna." Institute of Labor Economics (IZA) IZA Discussion Papers 16536.

Bryson, Alex, Peter Dolton, J. James Reade, Dominik Schreyer, and Carl Singleton. 2021. "Causal effects of an absent crowd on performances and refereeing decisions during Covid-19." *Economics Letters*, 198(C).

Butler, David, Robert Butler, Joel Maxcy, and Simon Woodworth. 2023. "Outcome Uncertainty and Viewer Demand for Basic Cable Boxing." *Journal of Sports Economics.*

Butler, Robert. 2023. "An Introduction to the James Quirk Special Issue and the Economics of Combat Sport." *Journal of Sports Economics.*

Csató, László. 2023. "How to avoid uncompetitive games? The importance of tie-breaking rules." *European Journal of Operational Research*, 307(3): 1260–1269.

Csató, László, and Dóra Gréta Petróczy. 2022. "Fairness in penalty shootouts: Is it worth using dynamic sequences?" *Journal of Sports Sciences*, 40(12): 1392–1398. PMID: 35675384.

Csató, László, Roland Molontay, and József Pintér. 2024. "Tournament schedules and incentives in a double round-robin tournament with four teams." *International Transactions in Operational Research*, 31(3): 1486–1514.

Dietl, Helmut M., Markus Lang, and Stephan Werner. 2010. "Corruption in Professional Sumo: An Update on the Study of Duggan and Levitt." *Journal of Sports Economics*, 11(4): 383–396.

Dohmen, Thomas, and Jan Sauermann. 2016. "Referee Bias." *Journal of Economic Surveys*, 30(4): 679–695.

Duggan, Mark, and Steven D. Levitt. 2002. "Winning Isn't Everything: Corruption in Sumo Wrestling." *American Economic Review*, 92(5): 1594–1605.

Frederiksen, Jesper, and Robert Machol. 1988. "Reduction of paradoxes in subjectively judged competitions." *European Journal of Operational Research*. https://www.sciencedirect.com/science/article/abs/pii/037722178890375X.

Holmes, Benjamin, Ian McHale, and Kamila Zychaluk. 2024. "Detecting individual preferences and erroneous verdicts in mixed martial arts judging using Bayesian hierarchical models." *European Journal of Operational Research)*.

Ingle, Sean. 2021. "Judges 'used signals' to fix Olympic boxing bouts, McLaren report finds." *The Guardian*. https://bit.ly/3sGXk1M.

Kendall, Graham, and Liam J.A. Lenten. 2017. "When sports rules go awry." *European Journal of Operational Research*, 257(2): 377–394.

Krumer, Alex, and Juan D. Moreno-Ternero. 2023. "The Allocation of Additional Slots for the FIFA World Cup." *Journal of Sports Economics*, 24(7): 831–850.

Kumar, Manish. 2022. "Explained: LBW rules and the controversial umpire's call in DRS." *The Times of India*. https://bit.ly/47F6ef4.

Lee, Herbert, Daniel Cork, and David Algranati. 2002. "Did Lennox Lewis beat Evander Holyfield?: methods for analysing small sample interrater agreement problems." *Journal of the Royal Statistical Society. Series D (The Statistician)*.

McLaren, Richard. 2022. "Independent investigation of the AIBA." https://bit.ly/3sHR5e1.

Reade, J. James, Dominik Schreyer, and Carl Singleton. 2022. "Eliminating supportive crowds reduces referee bias." *Economic Inquiry*, 60(3): 1416–1436.

Reid, Alex. 2023. "Boxing's biggest robberies." *talkSPORT*. https://bit.ly/47BO6T8.

Rumsby, Ben. 2021. "Joe Joyce demands Rio 2016 gold medal from IOC after boxing corruption report." *The Telegraph*. https://bit.ly/46ndxqw.

Slavin, Harry. 2017. "Gennady Golovkin and Canelo Alvarez judge Adalaide Byrd disciplined for lopsided scorecard as she is stood down from major title fights." *Mail Online*. https://bit.ly/46nC8vh.

Tenorio, Rafael. 2000. "The Economics of Professional Boxing Contracts." *Journal of Sports Economics*, 1(4): 363–384.

US Senate. 2001. "Committee on commerce, science and transportation - A review of the professional boxing industry - is further reform needed? Senate Hearing 107-1090."

Wainwright, Anson. 2023. "Fair or foul? Experts weigh in on Devin Haney-Vasiliy Lomachenko result." https://bit.ly/49IQUzO.

Wright, Mike. 2014. "OR analysis of sporting rules – A survey." *European Journal of Operational Research*, 232(1): 1–8.

Young, H. 1974*a*. "A Note on Preference Aggregation." *Econometrica*, 42(6): 1129–1131.

Young, H. 1974*b*. "An axiomatization of Borda's rule." *Journal of Economic Theory*, 9(1): 43–52.

They were robbed! Scoring by the middlemost to attenuate biased judging in boxing

Online Appendix

Appendix A. Full proof of Proposition 1

To simplify, we assume there are three rounds. Consequently, there are four information sets that the partisan judge can receive before they choose how many rounds to award to Blue. These are { BBB, BBR, BRR, RRR }, which are the signals of each round of the bout after sorting round results.[2]

We start by calculating the probability of a fair judge awarding a round for the Blue conditional on the partisan judge observing a result of B (we call this conditional probability $p|B$) and in the complementary case ($p|R$).

$$p|B = 1 - (2\alpha - 2\alpha^2)\ , \qquad\qquad p|R = 2\alpha - 2\alpha^2 \tag{A.1}$$

To simplify, we use the notation $\nabla = 2\alpha - 2\alpha^2$. We use this to calculate the probability of another judge's scorecard being in favour of Blue for each information set that the partisan judge observes. We use C to denote this scorecard probability :

$$C|BBB = (1-\nabla)^2(2\nabla + 1) \qquad\qquad C|RRR = \nabla^2 + 2(1-\nabla)\nabla^2 \tag{A.2}$$

$$C|BBR = (1-\nabla)^3 + (1-\nabla)\nabla^2 + (1-\nabla)\nabla \qquad C|BRR = \nabla^3 + (1-\nabla)^2\nabla + (1-\nabla)\nabla \tag{A.3}$$

A.1 Majority Judges Rule

The utilities available at each information set (where the first column shows the information set for each row) and action (the second to fifth columns) are shown in Table A1.

	Award BBB	Award BBR	Award BRR	Award RRR
BBB	$S(1-(1-C\|BBB)^2) - \alpha$	$S(1-(1-C\|BBB)^2) - \frac{2\alpha+(1-\alpha)}{3}$	$S(C\|BBB)^2 - \frac{\alpha+2(1-\alpha)}{3}$	$S(C\|BBB)^2 - (1-\alpha)$
BBR	$S(1-(1-C\|BBR)^2) - \frac{2\alpha+(1-\alpha)}{3}$	$S(1-(1-C\|BBR)^2) - \alpha$	$S(C\|BBR)^2 - \frac{2\alpha+(1-\alpha)}{3}$	$S(C\|BBR)^2 - \frac{\alpha+2(1-\alpha)}{3}$
BRR	$S(1-(1-C\|BRR)^2) - \frac{\alpha+2(1-\alpha)}{3}$	$S(1-(1-C\|BRR)^2) - \frac{2\alpha+(1-\alpha)}{3}$	$S(C\|BRR)^2 - \alpha$	$S(C\|BRR)^2 - \frac{2\alpha+(1-\alpha)}{3}$
RRR	$S(1-(1-C\|RRR)^2) - (1-\alpha)$	$S(1-(1-C\|RRR)^2) - \frac{\alpha+2(1-\alpha)}{3}$	$S(C\|RRR)^2 - \frac{2\alpha+(1-\alpha)}{3}$	$S(C\|RRR)^2 - \alpha$

Table (A1) Utilities for each action and information set in the basic model

The best responses to seeing BBB and BBR are to award BBB and BBR, respectively. This minimises backlash while still awarding the card of the partisan judge to the favoured Blue. If the partisan judge sees BRR, however, then they need to choose between BBR (which awards the card to Blue) and BRR (which minimises expected backlash). The condition for awarding

[2] Note, in both scoring systems, no distinctions are made as to when in the bout a particular round result occurred. Therefore, a bout with true result BRB is the same as a bout with true result BBR, and we reorder round results to simplify the analysis.

rounds BRR as preferred to awarding rounds BBR is:

$$\begin{aligned} \widehat{S}_{\text{Maj. Jgs, BRR}} &\leq \frac{1-2\alpha}{6(C|BRR)(1-(C|BRR))} \\ &\leq \frac{1-2\alpha}{24\alpha\left(1-\alpha\right)\left(2\alpha^2-2\alpha+1\right)\left(4\alpha^4-8\alpha^3+7\alpha^2-3\alpha+1\right)\left(8\alpha^4-16\alpha^3+10\alpha^2-2\alpha+1\right)} \end{aligned} \tag{A.4}$$

For a simple example, when $\alpha = 0.1$, this expression becomes approximately 1.09.

If the partisan judge sees RRR, then they need to choose between BBR (which awards the card to Blue) and RRR (which minimises backlash). The condition for RRR to be perferred to BBR is:

$$\begin{aligned} \widehat{S}_{\text{Maj. Jgs, RRR}} &\leq \frac{1-2\alpha}{3(C|RRR)(1-(C|RRR))} \\ &\leq \frac{1-2\alpha}{12\alpha^2\left(\alpha-1\right)^2\left(2\alpha^2-2\alpha+1\right)^2\left(1+4\alpha-4\alpha^2\right)\left(4\alpha^2-4\alpha+3\right)} \end{aligned} \tag{A.5}$$

For a simple example, when $\alpha = 0.1$, the right-hand-side here becomes approximately 12.14.

It makes intuitive sense that the critical level of blue winning utility will be higher here than in Equation A.4, as they need to award two more rounds than they believe Blue won (so higher backlash) and there is lower odds of at least one other judge awarding in favour of Blue (so less chance that partisan judging will deliver a victory).

A.2 Majority Rounds Rule

We can derive the following probabilities of Blue winning a round conditional on what the partisan judge observes and does:[3]

Partisan judge observes B and does B:	$q_1 = 1 - \nabla^2$	(A.6)
Partisan judge observes R and does B:	$q_2 = \nabla(2 - \nabla)$	(A.7)
Partisan judge observes B and does R:	$q_3 = (1 - \nabla)^2$	(A.8)
Partisan judge observes R and does R:	$q_4 = \nabla^2$	(A.9)

At this point, we define the function f to denote the probability of getting at least two realisations from 3 binomial distribution trials with probabilities p_1, p_2, p_3:

$$f(p_1, p_2, p_3) = p_1p_2(1-p_3) + p_1(1-p_2)p_3 + (1-p_1)p_2p_3 + p_1p_2p_3 \tag{A.10}$$

Using this function, we can write the utilities available at each information set (where the first column shows the information set for each row) and action (the second to fifth columns), shown in Table A2.

Similar to the case in Table A1, the utility on the diagonal is better than the utility from

[3] For instance, we can work out q_1 and q_3 as follows. If we see B, then there is $(1-\alpha)$ chance that B is the true state and α chance that R is the true state. If B is the true state, then there is a chance α that a fair judge sees R and a $1-\alpha$ chance they see B. If R is the true state, then there is a chance α that a fair judge sees B and a $1-\alpha$ chance they see R. Adding up the probabilities that B wins on the fair judge's card we come to $1-\nabla$. For q_3, we need both fair judges to score it this way, the odds of which are $(1-\nabla)^2$. For q_1 we need one (or both) of the judges to score it this way, the odds of which are $1-\nabla^2$. We can work out q_2 and q_4 analogously.

	Award BBB	Award BBR	Award BRR	Award RRR
BBB	$Sf(q_1,q_1,q_1)-\frac{3\alpha+0(1-\alpha)}{3}$	$Sf(q_1,q_1,q_3)-\frac{2\alpha+1(1-\alpha)}{3}$	$Sf(q_1,q_3,q_3)-\frac{1\alpha+2(1-\alpha)}{3}$	$Sf(q_3,q_3,q_3)-\frac{0\alpha+3(1-\alpha)}{3}$
BBR	$Sf(q_1,q_1,q_2)-\frac{2\alpha+1(1-\alpha)}{3}$	$Sf(q_1,q_1,q_4)-\frac{3\alpha+0(1-\alpha)}{3}$	$Sf(q_1,q_3,q_4)-\frac{2\alpha+1(1-\alpha)}{3}$	$Sf(q_3,q_3,q_4)-\frac{1\alpha+2(1-\alpha)}{3}$
BRR	$Sf(q_1,q_2,q_2)-\frac{1\alpha+2(1-\alpha)}{3}$	$Sf(q_1,q_2,q_4)-\frac{2\alpha+1(1-\alpha)}{3}$	$Sf(q_1,q_4,q_4)-\frac{3\alpha+0(1-\alpha)}{3}$	$Sf(q_3,q_4,q_4)-\frac{2\alpha+1(1-\alpha)}{3}$
RRR	$Sf(q_2,q_2,q_2)-\frac{0\alpha+3(1-\alpha)}{3}$	$Sf(q_2,q_2,q_4)-\frac{1\alpha+2(1-\alpha)}{3}$	$Sf(q_2,q_4,q_4)-\frac{2\alpha+1(1-\alpha)}{3}$	$Sf(q_4,q_4,q_4)-\frac{3\alpha+0(1-\alpha)}{3}$

Table (A2) Utilities for each action and information set in the basic model under majority rounds rule

awarding more rounds than this to Red. This implies that if the partisan judge sees BBB, then they should award rounds as BBB. We can solve for the levels of S at which the partisan judge prefers to award rounds fairly rather than giving additional rounds to Blue. Starting at the BBR information set:

$$\widehat{S}_{\text{Maj. Rds, BBR}} \leq \frac{1-2\alpha}{96\alpha^3\left(8\alpha^9-48\alpha^8+124\alpha^7-180\alpha^6+158\alpha^5-80\alpha^4+15\alpha^3+7\alpha^2-5\alpha+1\right)} \tag{A.11}$$

For the BRR information set, the partisan judge could do BBR or BBB rather than the fair result of BRR. We derive the critical S for both and can determine that a partisan judge will deviate to BBR at a higher S value than they would deviate to BBB. Hence, we below report the threshold above which the partisan judge will not deviate to BBR:

$$\widehat{S}_{\text{Maj. Rds, BRR}} \leq \frac{2\alpha-1}{12\alpha\left(64\alpha^{11}-384\alpha^{10}+992\alpha^9-1440\alpha^8+1264\alpha^7-640\alpha^6+120\alpha^5+56\alpha^4-38\alpha^3+4\alpha^2+3\alpha-1\right)} \tag{A.12}$$

Finally, for the RRR information set, the partisan judge could do BRR, BBR or BBB, rather than the fair result of RRR. We can establish that if α is near zero, then the partisan judge will deviate to BBB at a higher S than they would deviate to the other options. When α is near (but below) 0.5, then they will deviate to BBR at a higher S than they would deviate to the other options. As a result, we have the critical S value:

$$\widehat{S}_{\text{Maj. Rds, RRR}} \leq \min\Big[\frac{1-2\alpha}{16\alpha^2\left(16\alpha^{10}-96\alpha^9+264\alpha^8-440\alpha^7+504\alpha^6-432\alpha^5+294\alpha^4-162\alpha^3+69\alpha^2-20\alpha+3\right)}, \frac{2\alpha-1}{24\alpha^2\left(16\alpha^8-80\alpha^7+172\alpha^6-208\alpha^5+152\alpha^4-64\alpha^3+11\alpha^2+2\alpha-1\right)}\Big] \tag{A.13}$$

Now we can summarise Equations A.4, A.5, A.11, A.12 and A.13 with a chart of α against the critical S ratio below which the partisan judge does not mis-award rounds. This is shown in Figure A1.

In all cases where Red wins more rounds, the majority rounds rule has a higher S value at which the partisan judge is indifferent to awarding fairly and giving more rounds to Blue. This indicates that the majority rounds rule is more robust to partisan judging.

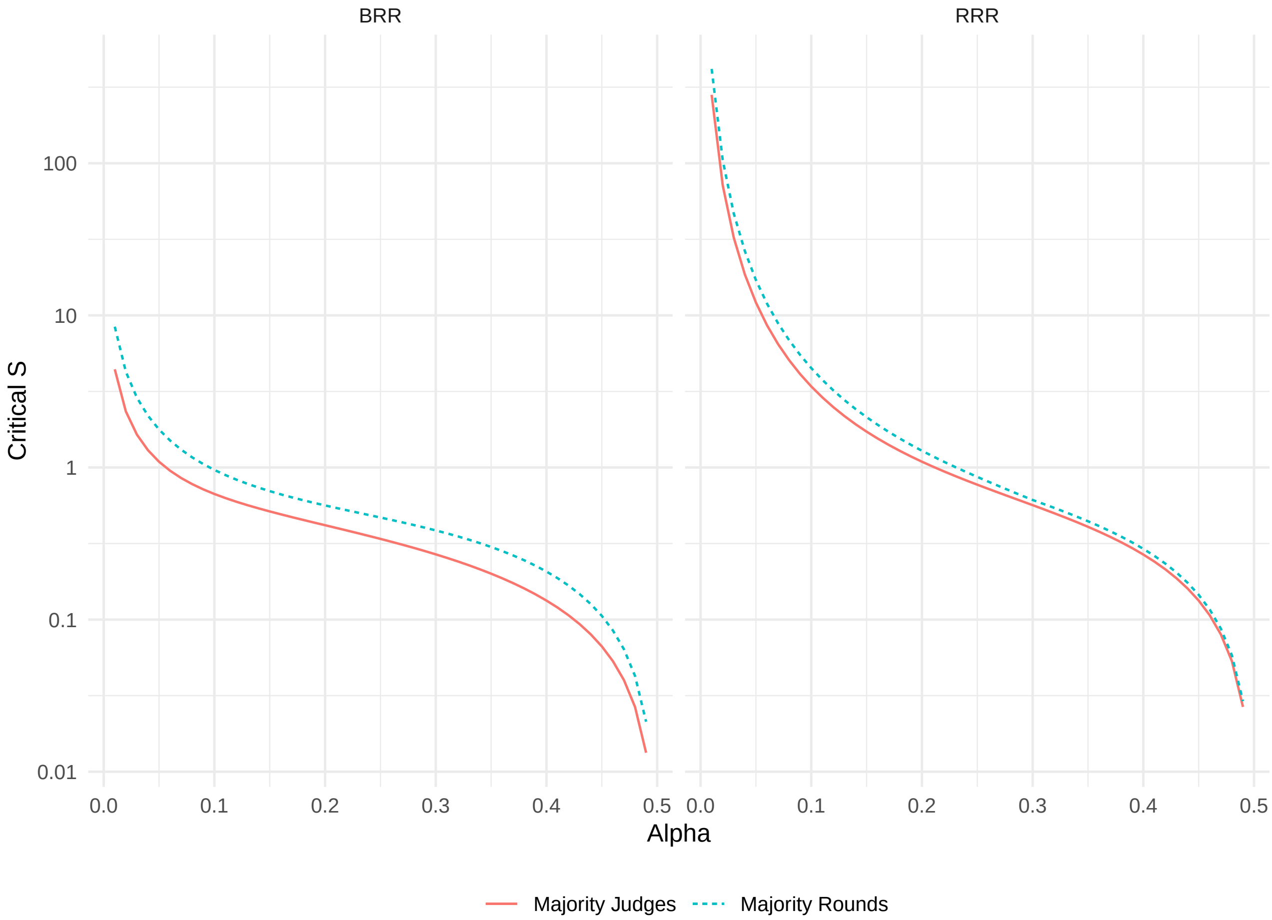
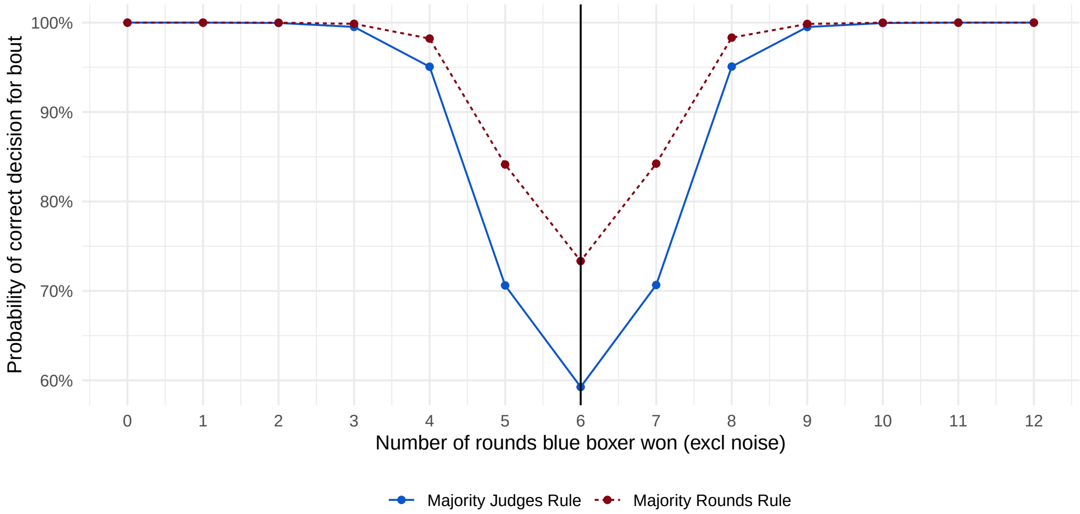


FIGURE (A1) Critical S below which partisan judges do award fairly for three-round bouts with three judges in the BRR and RRR information sets

Appendix B. No judges are partisan

When no judges are partisan, the proposed change, from awarding bouts by majority judges to majority rounds, reduces the probability of bouts being wrongly decided. This can be seen in Figure B1, which is comparable to Figure 2, with the same parametrisation, but reflects the case where all three judges are fair.

FIGURE (B1) Probability of a "correct" result depending on the number of rounds truly won by blue and how judges' scores are aggregated: the case of no partisan judges



Appendix C. Other parametrisations

We consider different model parametrisations, to demonstrate the extent to which the qualitative results of this paper may vary.

High disagreement between different judges

We increase the noise variance, such that the different judges disagree more often about the outcome of a round. Specifically, we increase $\alpha = 0.2$ and leave the other parameters as they were in the main body of the paper. Figures 1-3 are reproduced below for this new parametrisation as Figures C1- C3.

FIGURE (C1) Probability of favoured boxer winning in one bout where both boxers won 6 rounds (in the absence of noise)

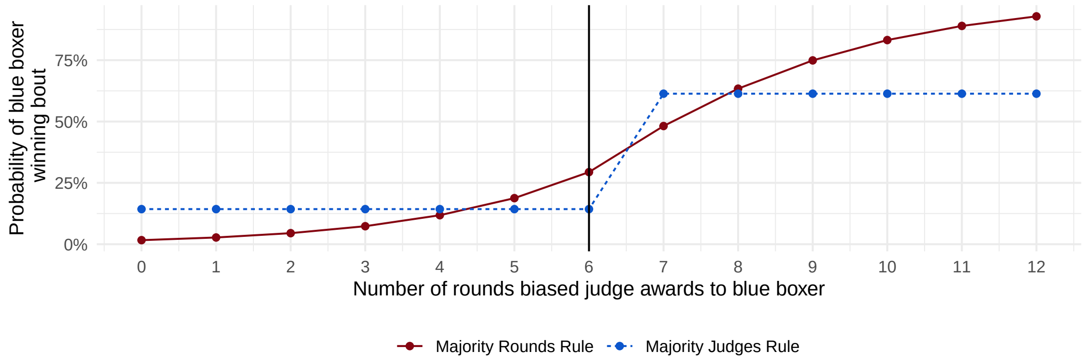


FIGURE (C2) Probability of correct result for bout

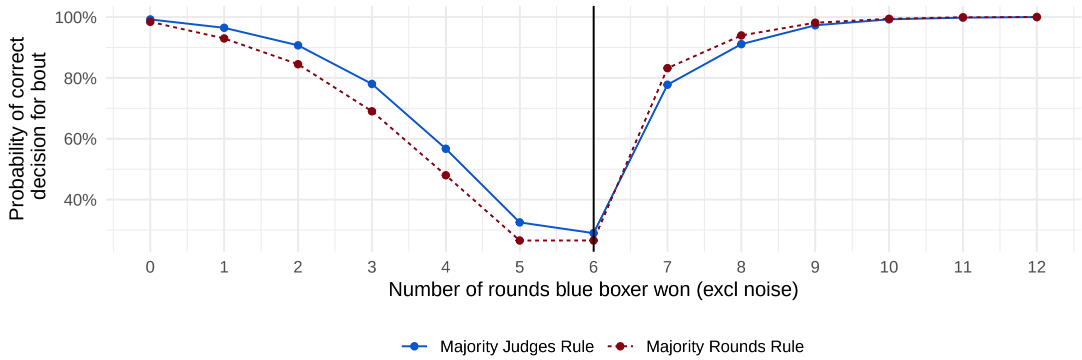
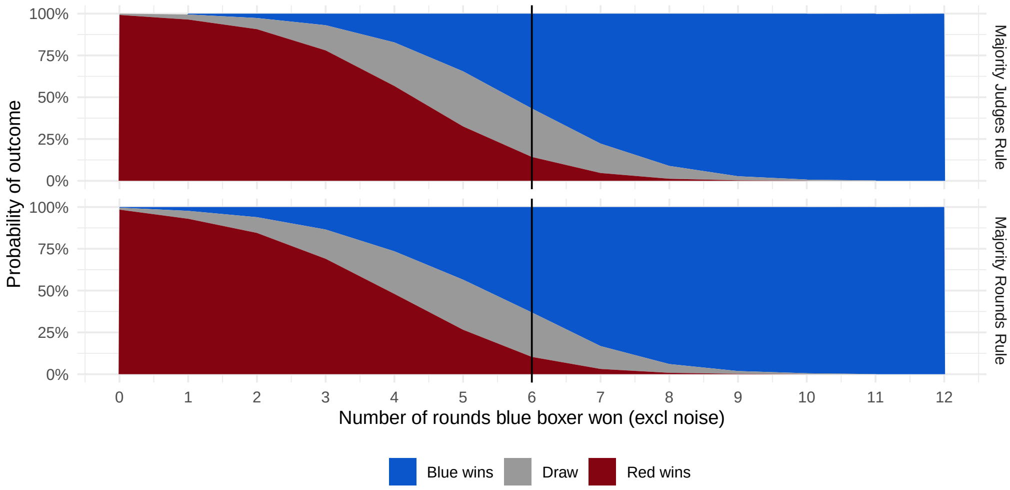


FIGURE (C3) Probability of each outcome for bout



High degree of favouritism

We increase S to 0.8 and leave other parameter values as they are in the main body of the paper. Figures 1-3 are reproduced below for this new parametrisation as Figures C4-C6

FIGURE (C4) Probability of favoured boxer winning in one bout where both boxers won 6 rounds (in the absence of noise)

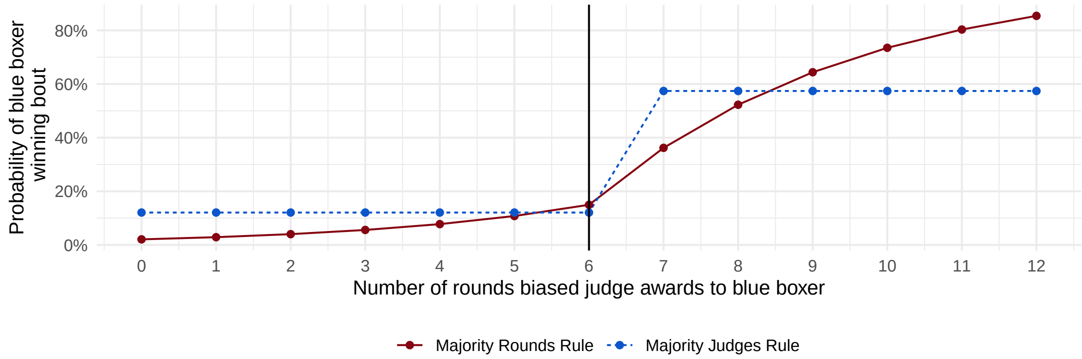


FIGURE (C5) Probability of correct result for bout

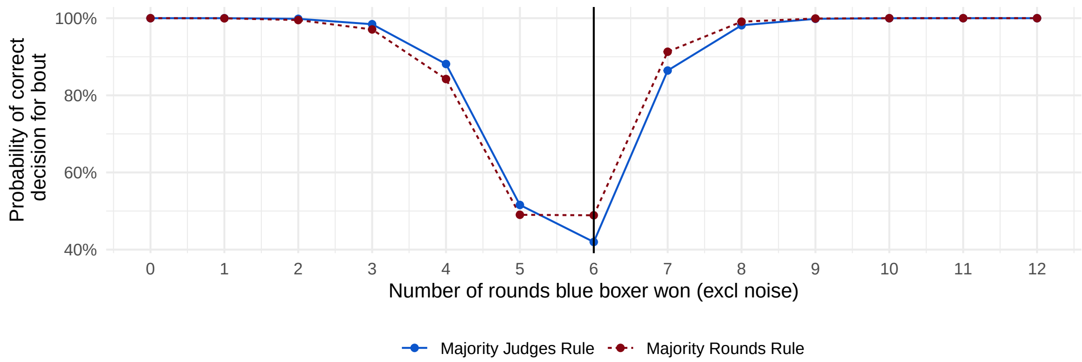
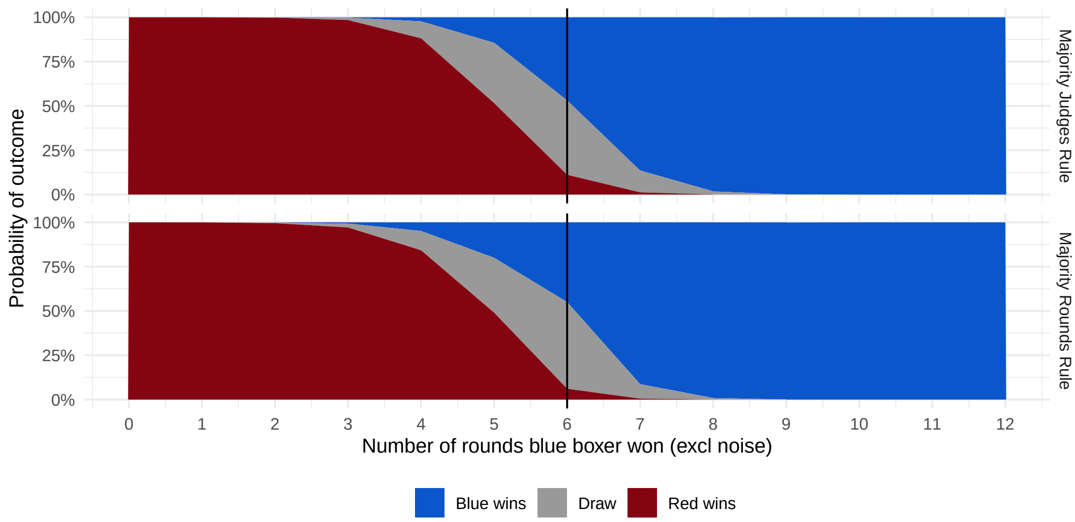


FIGURE (C6) Probability of each outcome for bout



Appendix D. Women's professional boxing

In women's professional boxing, there are generally 10 rounds and 3 judges. Figures 1- 3 are reproduced below for women's professional boxing as Figures D1-D3, with otherwise identical parametrisations.

FIGURE (D1) Probability of favoured boxer winning in one bout where both boxers won 6 rounds (in the absence of noise) - Women's professional boxing

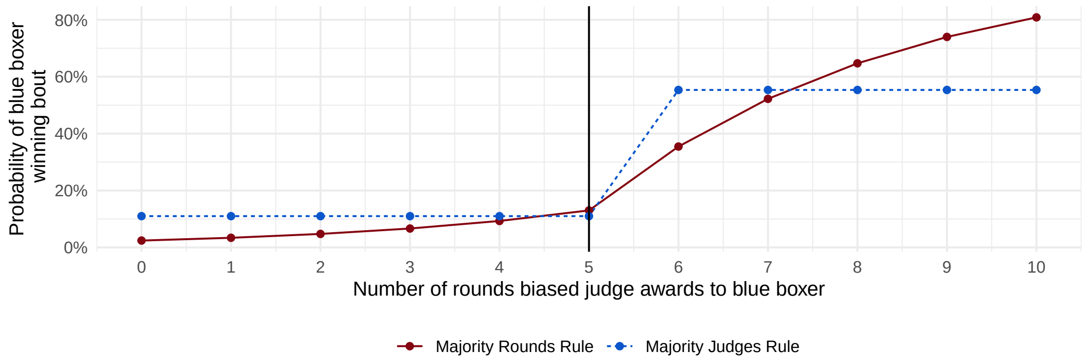


FIGURE (D2) Probability of correct result for bout - Women's professional boxing

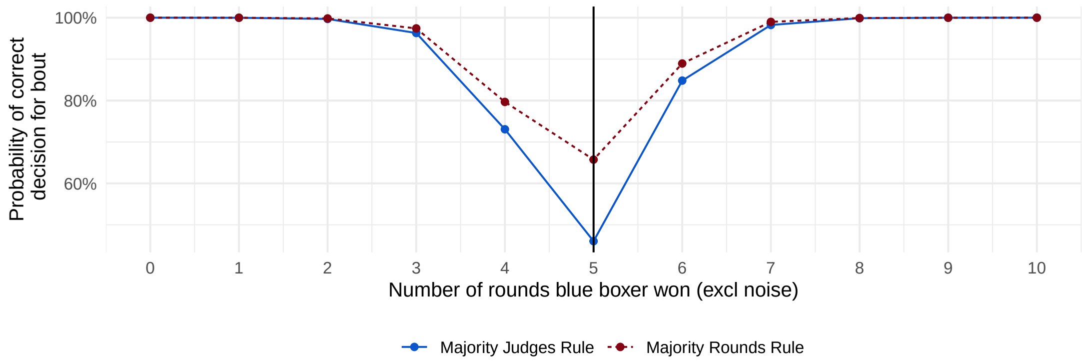
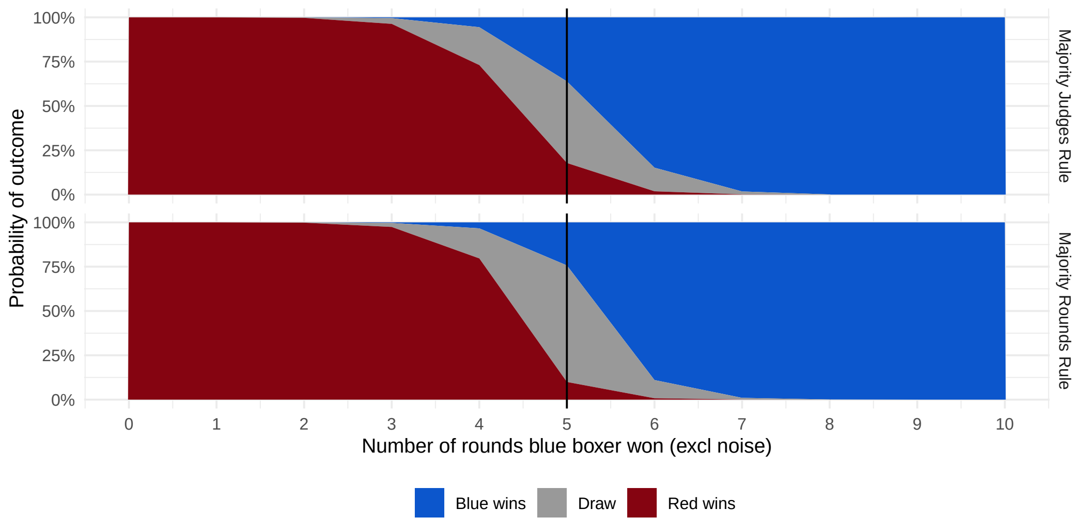


FIGURE (D3) Probability of each outcome for bout - Women's professional boxing



Appendix E. Men's Olympic amateur boxing

In men's Olympic amateur boxing, there are generally 3 rounds and 5 judges. Figures 1-3 are reproduced below for men's Olympic amateur boxing as Figures E1-E3.

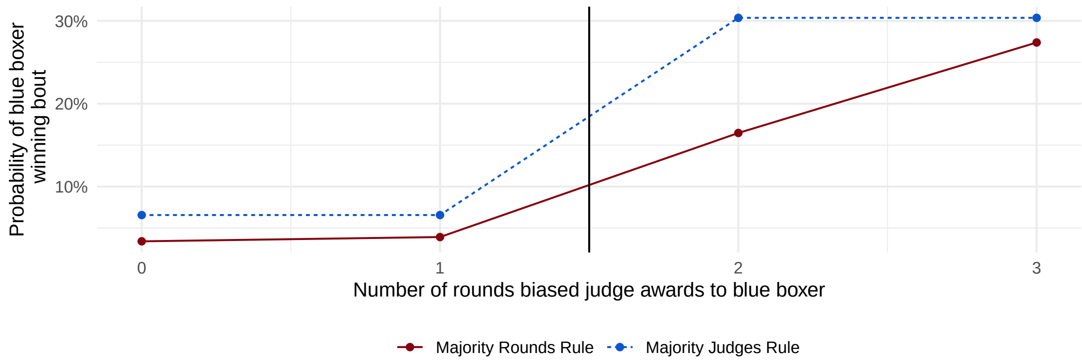


FIGURE (E1) Probability of favoured boxer winning in one bout where the blue boxer won 1 round and the red boxer won 2 (in the absence of noise) - Men's Olympic amateur boxing

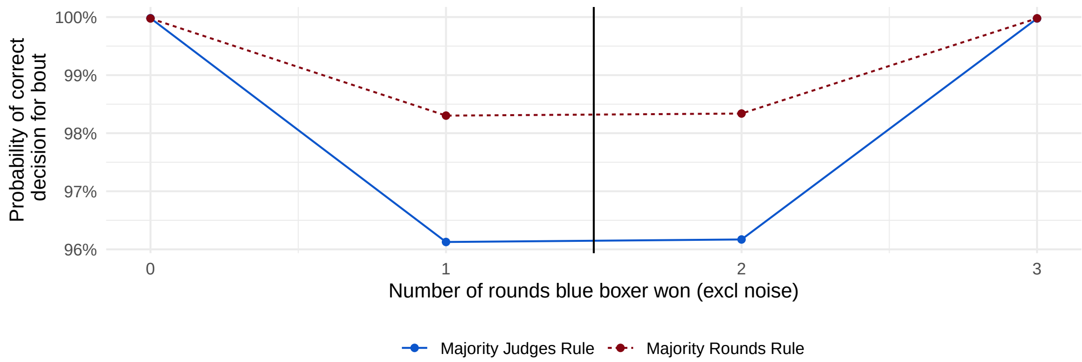


FIGURE (E2) Probability of correct result for bout - Men's Olympic amateur boxing

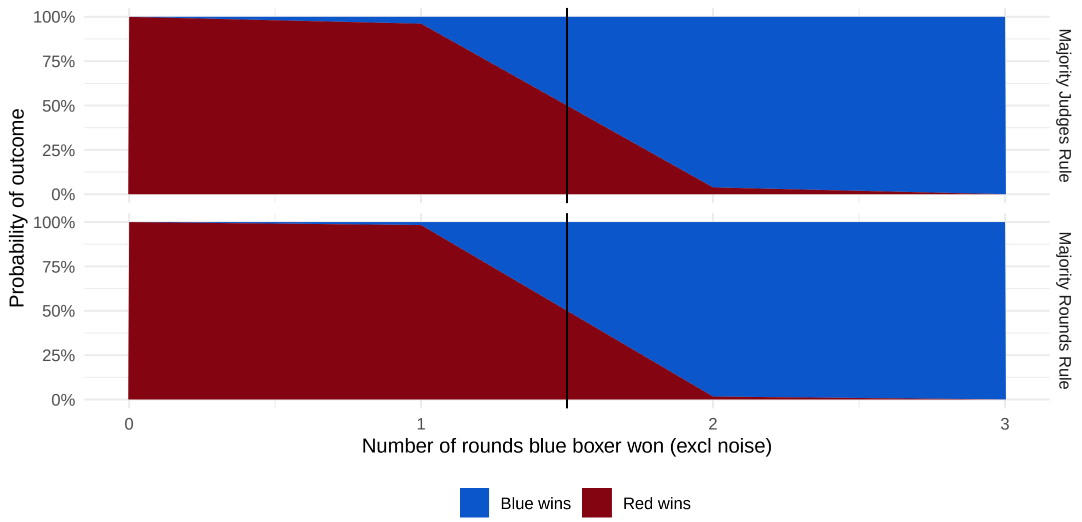


FIGURE (E3) Probability of each outcome for bout - Men's Olympic amateur boxing

Appendix F. Women's Olympic amateur

In women's Olympic amateur boxing, there are generally 4 rounds and 5 judges. Figures 1-3 are reproduced below for women's Olympic amateur boxing as Figures F1-F3.

FIGURE (F1) Probability of favoured boxer winning in one bout where both boxers won 6 rounds (in the absence of noise) - Women's Olympic amateur boxing

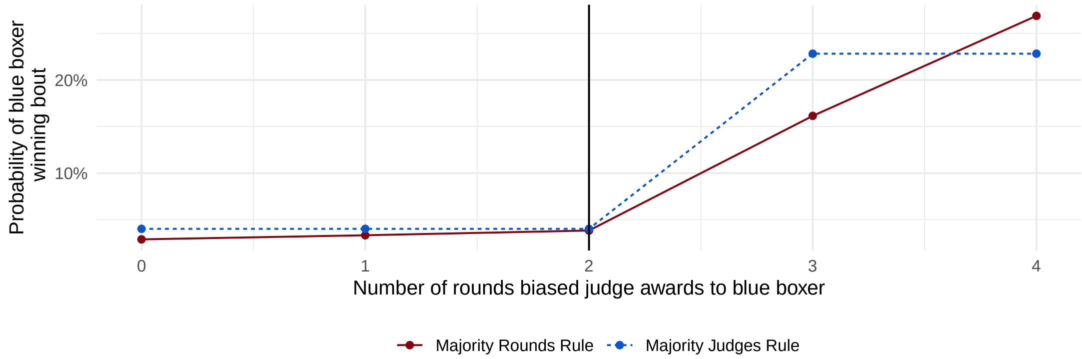


FIGURE (F2) Probability of correct result for bout - Women's Olympic amateur boxing

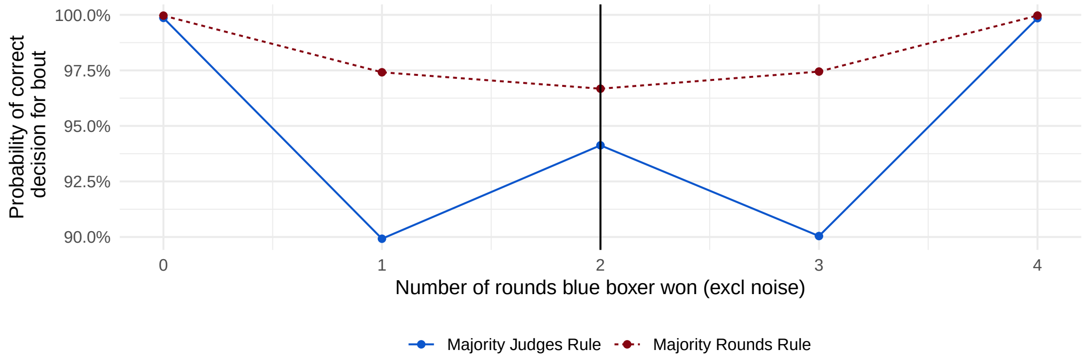


FIGURE (F3) Probability of each outcome for bout - Women's Olympic amateur boxing