Can barely winning lead to losing? Evidence for a substantial
gender gap in psychological momentum

by

Mario LACKNER
Michael WEICHSELBAUMER

Can barely winning lead to losing? Evidence for a substantial gender gap in psychological momentum *

Mario Lackner[a] and Michael Weichselbaumer[b]

[a]*Department of Economics, Johannes Kepler University Linz (JKU)*
[b]*Vienna University of Economics and Business*

Revised version, May 9, 2022

Abstract

We use data from professional tennis to measure the causal effect of past on current performance for women and men. Identification relies on exogenous shocks to the probability of facing a contested game, which is a previous stage of competition with strong resistance. There are fundamental gender differences: whereas men's performance is unaffected by previously experiencing and winning a contested game, women experience a sizeable deterioration of performance after barely winning the previous stage. This result is linked to gender differences in negative psychological momentum. Detailed analysis reveals heterogeneous effects by experience, ability and contest progression.

JEL Classification: J16, M52

Keywords: Performance feedback, relative performance, process feedback, gender differences, psychological momentum.

I don't really see it as pressure. I mostly see it as a challenge for me. When he's up a break point, I see that as a challenge to overcome the difficulty.

Stefanos Tsitsipas, top-3 ranked ATP player, after beating Roger
Federer in the 2019 ATP World Tour Finals.[1]

1. Introduction

Intermediate performance feedback is a common feature of contest-like situations: in elections (Klumpp & Polborn 2006), patent races (Fudenberg et al. 1983), at the workplace (Casas-Arce & Martínez-Jerez 2009, Gill et al. 2019, Delfgaauw et al. 2014), in education (Azmat & Iriberri 2010, Tran & Zeckhauser 2012, Azmat et al. 2019, Hermes et al. 2021), and sports (Malueg & Yates 2010). How do contestants react to past performance? Do they raise or reduce effort?

One explanation for the causal effects of past on present performance is provided by psychological momentum (Adler 1981, Iso-Ahola & Dotson 2014), a concept describing how past performance can affect future performance of agents in contests. The empirical literature on psychological momentum relies on intermediate performance measures, which trigger either positive or negative momentum. Most estimates, however, will be distorted by strategic momentum (Klumpp & Polborn 2006, Malueg & Yates 2010, Dong & Huang 2019). When effort is costly, competitors will strategically adapt effort after intermediate outcomes changed the relative position, and the probability to win the contest.

Using data from top-level professional tennis, we investigate potential gender differences in the causal effect of past on future performance. We propose a novel empirical approach to measure negative psychological momentum resulting from near setbacks, ruling out any strategic performance effects. In tennis, only game wins matter, individual points are not counted towards winning a set or match. We estimate how winning a strongly contested game in $t-1$, compared to a decisive win, affects the probability to win the subsequent game t. Contested games are defined as games which need at least 8 points to be won by the player on serve, clearly exceeding the minimal game duration of 4 points. Winning a game either decisively, or experiencing a contested game, does not make a difference for the score going into game t, and, thus, no strategic effort adjustment of both players is plausible. This particular type of information on past performance is ambiguous in the sense that it can either cause positive or negative psychological momentum.

---

[1]Retrieved November 22, 2021. https://tennistonic.com/tennis-news/133295/
stefanos-tsitsipas-explains-how-he-saved-so-many-break-points-vs-federer-in-london/.

Our empirical approach relies on double faults as quasi-random exogenous shocks for causal identification of psychological momentum. To support this strategy, we extensively investigate and discuss the quasi-random nature of double faults. The benefit of our approach is that we measure a type of psychological momentum that is rooted in purely random changes of how the previous game was won, while keeping constant the outcome that is relevant for calculating win probabilities. This empirical approach rules out any other, potentially non-random changes in effort or temporary unobserved shifts in relative ability of both players.

Laboratory experiments found that the reaction to luck in contests is different for men and women (Mago et al. 2013, Gill & Prowse 2014). Our findings stress the gender gap in reaction to quasi-random feedback, in a real-world professional setting where men and women have strikingly similar career prospects. We estimate the negative psychological momentum effect to be particularly strong for young female contestants, as well as female contestants with low absolute and relative ability. The effect is, moreover, more pronounced when importance to perform is high.

Previous results on psychological momentum are presented by Berger & Pope (2011), who find that teams who are only slightly behind after the first stage are more likely to win the second stage and, thus, the overall contest, compared to other close scores. Morgulev et al. (2019) or Kniffin & Mihalek (2014), on the other hand, do not find such momentum effects for male basketball or ice hockey teams. Gauriot & Page (2019) study momentum effects in tennis and find that winning a point increases the probability to win the subsequent point. Other studies find empirical evidence for psychological momentum effects at set level in best-of-three (Depken et al. 2022) and best-of-five (Depken et al. 2021) matches in men's tennis. Meier et al. (2020) find a positive performance effect in men's tennis for games without a changeover interruption, compared to games with a changeover. They attribute the difference to psychological momentum, because the interruption would disrupt psychological momentum, but not strategic momentum. The fact that this momentum effect is not estimated for games after a regular interruption is interpreted as evidence of psychological (rather than strategic) momentum. Cohen-Zada et al. (2017) extend the empirical literature to potential gender differences. Positive momentum affects performance of male competitors in professional judo positively, but has no effect for women. They explain this result by changes in testosterone levels for men during contests. We propose an novel instrumental variable strategy to isolate the gender gap in (negative) psychological momentum, without potential distortions of strategic considerations.

Understanding gender differences in psychological momentum is relevant for a multitude of settings. Labor markets commonly feature competition, e.g. for job applications or promotions. Within organi-

zations, some form of competition is the rule rather than the exception. Empirical evidence suggests that random shocks during early careers can have long-lasting consequences on subsequent career developments (Oreopoulos et al. 2012, Schwandt & Von Wachter 2019, Von Wachter 2020). Similarly, performance feedback in educational contexts can have gender-specific effects (Jalava et al. 2015, Goulas & Megalokonomou 2021).

Our main results have implications for the understanding of the effects of feedback on performance. Previous work documents women's ability (Wozniak et al. 2014) or the duration of the effect (Wozniak 2012) define gender differences in the reaction to feedback on past performance. In an experimental setting, Jeworrek (2019) finds that the gender gap in tournament entry (i.e., competitiveness) is erased by relative performance feedback for women who are highly conscientious. Positive subjective feedback is found to have a positive impact on female participants of an online game, while affecting male players negatively (Lovász et al. 2022).

The type of information on intermediate relative performance we are interested in can also be interpreted as process-feedback. Process feedback is a form of feedback that focuses on how a performance is achieved rather than the outcome. Psychologists demonstrate that gender differences in the reaction to positive process-feedback are relevant among young children (Henderlong & Lepper 2002). Experimental results suggest that positive process-feedback can act as a motivational tool for work teams (Geister et al. 2006). The clear difference, however, is that our measure of past relative performance is a spurious type of feedback provided by the point-by-point record of the game, rather than external feedback.

Professional sports in general (Kahn 2000, Bar-Eli et al. 2020), and professional tennis in particular, provide a fruitful setting to study behavioral differences for men and women, and related research questions (Malueg & Yates 2010, Wozniak 2012, Banko et al. 2016). Top-level professional tennis provides high stakes (i.e. ranking points that affect the advertising value of a player, and prize money), detailed intermediate performance data, and rich heterogeneity in terms of relative abilities. Competition and profit opportunities are remarkably similar in men's and women's tennis, making professional tennis a profession with a high degree of gender equality. All major tournaments feature competitions for both genders, with identical facilities and, to a large degree, equal prize money. Equality in prize money is paralleled by off-court earnings, as illustrated by rankings of highest-paid tennis players, which count both on- and off-court earnings. For the 12 months until June 2019, 5 of the 10 highest paid tennis players were female (Forbes 2019b). Rankings of female athletes overall emphasize the outstanding role that tennis has for earning levels of women: 10 of the 10 highest paid female athletes are tennis players (12 of the top 15) (Forbes 2019a). Career opportunities for female and male players have become more

3

equal, including substantial opportunities to augment prize money with revenues from advertising and other activities.

2. Data and institutional setting

In order to win a set in tennis, a player must win a game when s/he is the returner (*"break serve"*).[2] Thus, a break difference (positive or negative) has a significant effect on the overall win probability. While a player with a break lead only has to hold serve —which is much easier than to break serve— and preserve this lead, a trailing player has to make up for the deficit in order to avoid losing the set (or match). The intermediate performance feedback that we focus on is when a player wins a contested game versus a "normal", non-contested game which is the more typical situation because the server usually enjoys an advantageous position. Both types of games have the same outcome, which is that the game is won. Experiencing a contested game, though, is a barely avoided major setback.

We use data from 12,839 top-level tennis matches to estimate the effect of experiencing a contested serve game on subsequent performance. In total, we observe 4,955 matches for male players competing in Association of Tennis Professionals (ATP) tournaments, and 7,884 matches for female players organized by the Women's Tennis Association (WTA).[3] The data provide detailed information about player characteristics, the WTA and ATP ranking and points of all players, and the exact information on how each point of the match was decided. We observe each game from the perspective of the return player. When a return player experienced a contested game —i.e., a game taking the serving player 8 or more points to hold serve in the previous serve game—, s/he has the feedback of past low performance, without losing the game. In Tennis, any game lasting 8 or more points will feature at least one situation where the server and returner are tied at 40 : 40 (deuce). In this situation, both players need to win 2 more points to decide the game. Figure A.1 illustrates the distribution of game durations for all return games. Very long games will have a strong effect on both players, potentially leading to biased estimates. Consequently, we drop outliers of games lasting longer than 16 points (less than 1% of all games).

Subsequent performance, which is the outcome of interest, is the ability to score a break in the game of the return player.. Singles matches in tennis are exclusively played against opponents of the same gender. This rules out potential distortions from variations in the gender compositions of contests (Booth & Yamamura 2018).

Figure 1 illustrates our main explanatory variable, which measures the probability of a contested

---

[2] In the case of a tie-break at score 6:6, it is necessary for one player to break serve of the opponent on at least one single point in the tie-break. This is the case because tie-breaks can only be won with a decisive point margin equal or greater than 2.

[3] All data are available at `github.com/JeffSackmann`.

game for all return games. The probability of a contested game for ATP matches is significantly lower at about 23% for return games 1 through 10. The corresponding probability for WTA matches is just below 36%. This difference is likely due to different physical prerequisites of female and male players who compete on a court with identical measurements. Our instrument circumvents the problems that arise from ex-ante differences in the probability of a contested game. In particular, Sections 3 and 4 will show that the instrument constitutes an almost identical exogenous shock for both genders.

FIGURE 1 — Probability of a contested game, by gender and game number
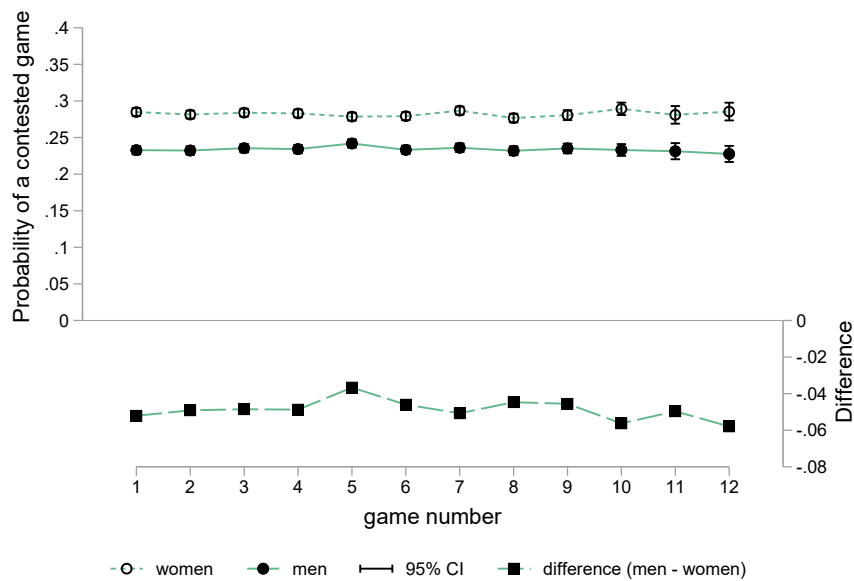
Notes: Probability (in %) that the player on return wins for all 12,839 games played by female and male players.

3. Empirical approach

Tennis, like almost all form of contests, features heterogeneous contestants. Heterogeneity in terms of ability will change over the course of a match. Relying on a simple OLS approach to estimate the effect of performance in game $t-1$ on subsequent performance in t will result in biased estimates from omitting ability (of both players) as a key variable. While WTA and ATP rankings provide a proxy for abilities, it is impossible to sufficiently control for all differences in ability.[4]

Winning a contested serve game —which may include successfully defending break points— does neither change the score that determines losing the set, nor for the overall match. In tennis, only game wins count, individual points are not counted towards winning a set or match. Consequently, it represents a form of spurious feedback which does not provide an incentive to strategically shift effort allocation

---

[4]Our sample includes matches where both contestants have a WTA or ATP ranking of 300 or better. This ensures a more homogeneous sample of players.

decisions of contestants. Cohen-Zada et al. (2017) identify an alternative channel through which a past event, which has no objective effect on the probability to win the contest, can affect future performance: psychological momentum (Adler 1981, Iso-Ahola & Dotson 2014). Put simply, they investigate how –all else equal– entering the next stage of a contest with a win compared to entering after just having lost. Their findings suggest that male contestants benefit from positive psychological momentum, while female athletes do not.

To obtain the causal effect of performance feedback across games, we estimate

$$break_{m,t} = \alpha \ + \ \gamma \, contested_{m,t-1} + \mathbf{X}'_{m,t}\beta + \delta_r + \xi_s + \theta_m + \nu_{m,t}, \tag{1}$$

where the dependent variable is equal to 1 if the game t results in a break, 0 otherwise. $\mathbf{X}_{m,t}$ is a vector including a set of continuous and binary control variables. We control for the average game duration before game $t-1$ to account for the intensity of the contest and potential tiring effects for players. We also control for the fraction of points scored by the server before game $t-1$ (of all points played) to control for the level of server dominance before game t. $\mathbf{X}_{m,t}$ includes binary variables indicating the score-difference in games for game t, dummies indicating the WTA/ATP ranking points bins and the ranking of the observed return player.

To control for unobserved heterogeneity, we include returner fixed-effects, δ_r, intermediate score when entering game t dummies ξ_s, and match-set fixed-effects θ_m. The variable $contested_{m,t-1}$ measures performance of game $t-1$: it is equal to 1 if game $t-1$ takes 8 or more points to find a winner, 0 otherwise. The coefficient γ measures the causal effect of experiencing (and winning) a contested serve game in $t-1$ on the performance in the observed return game t, compared to winning a "normal", uncontested game.

When investigating the causal effects of intermediate performance in contests, one faces the paramount endogeneity problem that falling back in a contest is related to the relative ability of the contestants. Like in other contests, relative performance is affected by short-lived factors influencing situational ability. In particular, we are not able to observe how competitive (or balanced in abilities) the match is throughout all observed games. The more balanced –i.e. the more equally able both contestants are– the more likely serve game $t-1$ will be contested, and the less likely a break in t will achieved. While we use match fixed-effects to hold match-time invariant factors constant, over the course of the match the momentum can shift multiple times, resulting in phases where players are relatively more or less competitive. Thus, we expect OLS estimates of Equation 1 to suffer from a positive omitted variable bias.
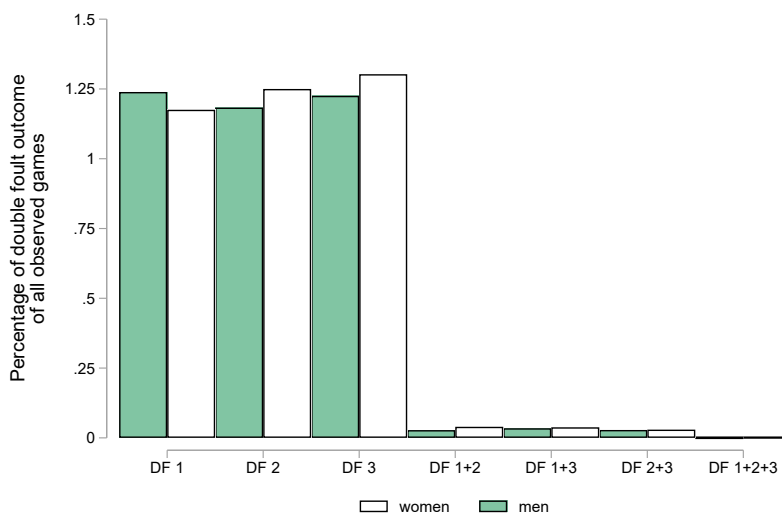
We use double faults during the first three points of game $t-1$ as an exogenous shock on the probability

observing a contested game $t-1$, i.e. a game where the player on serve needs 8 or more points to hold serve. In professional tennis, a serving player has two attempts to place the serve inside the corresponding serve box on the opposite field. If both serves fail, the returner scores the point and a double fault is recorded. It is a plausible argument that very subtle mistakes will lead to double faults, supporting the argument of a rare and quasi-random event. We use only the first three points as any game is decided in at least a minimum of four points. Doing so, we can rule out using double faults that are a result of increased pressure at break points or excessive risk taking as a strategy to make up for large within-game deficits. Our binary instrumental variable (IV) is defined as

$$Z_{t-1} = \mathbb{1}\left(\sum_{p=1}^{3} double\,fault_{p,t-1} > 0\right) \tag{2}$$

where Z_{t-1} is equal to 1 if at least one of the first three points in game $t-1$ is decided by a double fault. Figure 2 illustrates the relative frequency of all cases with 1 or multiple double faults in our data. Double faults – while more frequent for female players[5] – are a rare event.

FIGURE 2 — Double faults during the first three points – all games

Notes: Relative frequency (in %) to observe double faults at the first three points (or combinations) by a serve game.

To solve the omitted variable problem arising from the unobserved component of situational ability, we propose a two-stage IV model to identify the causal effect of our specific past-performance measure on current performance. We instrument $contested_{m,t-1}$ by using information on double faults during

---

[5]Women make more double faults as a result of different physical endowments for female and male players who compete on a court with the same specifications. Krumer et al. (2016) show that the serve advantage is larger for male players, as long as the physical characteristics are not controlled for.

the first three points in game $t-1$. The first stage is specified as

$$contested_{m,t-1} = \pi_0 + \pi_1 Z_{m,t-1} + \mathbf{X}'_{m,t}\phi + \delta_r + \xi_s + \theta_m + \nu_{m,t}. \tag{3}$$

We focus on situations where the server is either one game behind or exactly even. These situations are defined by a set's initial serve order and no player has the advantage or disadvantage of a break. Including games with a larger intermediate score difference will threaten our identification strategy, as double faults may not be fully exogenous, but a result of large intermediate score differences. For example, a server may have a substantial lead and thus adapt risk taking by changing speed and placement of their serve. A serving player could also face a large deficit, reduce serve effort and concede the set. The double fault probability for female and male players could be affected by an increase of the intermediate score difference. Focusing on game difference of 0 or -1 avoids that the previous serve game being contested is the result of a situational decrease in the performance of any player. Figure 3 illustrates the mean value of our binary instrument: the probability to observe a double fault during the first three points is not statistically different at a game difference of 0 compared to -1. Consequently, we can ultimately use 166,328 (37.56%) of the total of 442,847 return games in the empirical analysis.

FIGURE 3 — Instrumental variable, by gender and intermediate game difference
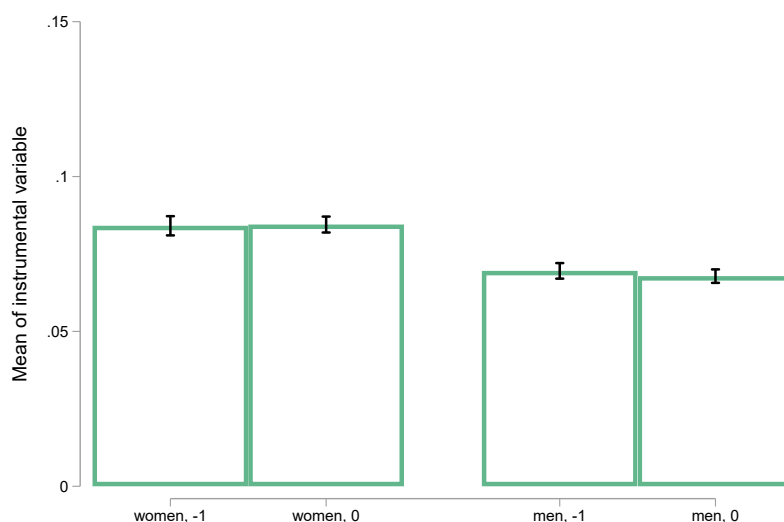
Notes: Probability (in %) that the return player wins for all 12,839 games played by female and male players.

Table 1 presents descriptive statistics for key variables of our analysis for the estimation sample. It is evident that games played by women feature more breaks (difference of 13 percentage points). However, the probability to observe a contested game or a defended break point in game $t-1$ for women is exceeding that for men by a smaller amount (5 and 3 percentage points). As a result of a more effective serve, men's games last slightly shorter (0.23 points) and female players on serve make 4 percentage points (of total

points) less than men. The gender differences in the mean rankings and ranking points are negligible and too small to be economically significant.

In addition to the intermediate standings or pre-match relative ability, other unobserved factors could pose a potential threat to our identification strategy. For example, a player could experience a period of physical discomfort or a temporary lack of concentration that will result in one or more double faults in her serve game while also affecting the performance in the subsequent return game as well as the following serve game. If this were the case, players would record multiple games with double faults during the first three points when on serve during such a period. In our data, however, more than 92% of all female and 88% of all male player-set combinations exhibit none or only one game where the instrumental variable is equal to 1. This rejects the hypothesis that double faults (and thus our IV) are results of extended periods of inferior performance which could induce double faults as well as poor play in subsequent return games.

TABLE 1 — Descriptive statistics for main variables, by player gender

	pooled		*women*		*men*		
	mean	(st.dev.)	mean	(st.dev.)	mean	(st.dev.)	*difference*
Break in game t ($yes = 1, no = 0$)	0.27	(0.45)	0.36	(0.48)	0.20	(0.40)	−0.16***
Contested game $t-1$	0.24	(0.42)	0.26	(0.44)	0.21	(0.41)	−0.05***
($\geq$ 8 points; $yes = 1, no = 0$)							
Break points defended $t-1$	0.15	(0.36)	0.17	(0.38)	0.14	(0.34)	−0.03***
Points played in game $t-1$	6.13	(2.24)	6.32	(2.32)	5.97	(2.15)	−0.35***
Fraction of point won by server	0.05	(0.11)	0.03	(0.08)	0.07	(0.13)	0.04***
before game $t-1$							
Avg. game duration in points	5.73	(2.07)	5.85	(2.23)	5.62	(1.92)	−0.23***
before game $t-1$							
Game number of game t	5.56	(3.10)	5.39	(3.09)	5.70	(3.10)	0.31***
WTA/ATP points (in 1,000)	1.39	(1.74)	1.48	(1.66)	1.33	(1.81)	−0.15***
WTA/ATP rank	79.02	(64.91)	80.04	(64.46)	78.17	(65.28)	−1.87***
Age	26.28	4.06	24.90	3.84	27.43	3.87	2.53***
Double faults in game t	0.21	(0.49)	0.26	(0.54)	0.18	(0.44)	−0.08***
Instrumental variable	0.08	(0.26)	0.08	(0.28)	0.07	(0.25)	−0.01***
N	166,328		75,670		90,658		

Notes: Reported difference measures the difference in means for the men minus the women sample; *, ** and *** indicate statistical significance at the 10%, 5%, and 1% level for a t-test.

Even within serve games, we do not see serial correlation of double faults during early points. We regress indicators of a double fault at points 2 and 3 of a particular serve game on a binary variables indicating whether a double fault was observed at point 1, 2, or both. The results are tabulated in panel *(a)* of Table A.1 in the Appendix. We estimate no statistically significant association between double faults during the first three points of a serve game for male players, irrespective of the particular timing. For female players we find a similar pattern, with the exception of the association between a double fault at the first and the second point. For this case, we estimate a negative association of −0.7 percentage points, significant at the 5-percent level. This estimate is likely economically insignificant and too small

to pose any threat to the identifying assumptions. The small and negative effect is also very short-lived, as there is no effect of a double fault at the first point beyond the second one. In order to avoid a second consecutive double fault, female players seem to increase focus for the second point when on serve and having recorded a double fault at the game's initial point.

The exclusion restriction is violated if double faults in game $t-1$ directly affect performance in game t, beyond acting as an exogenous shock to increase the probability to observe contested games $t-1$. A player on serve in $t-1$, for example, could feel lingering mental effects by being upset over having committed a double fault. Although game t will feature the observed player in a completely different task as the returner, the general level of play could be affected. We test this by regressing the probability to win points 4 through 8 on our instrument, i.e. a variable indicating if at least one of the 3 points before were decided by a double fault. The results are presented in panel *(b)* of Table A.1 in the Appendix. We estimate no association of the instrumental variable and the probability that the server wins any of the subsequent points for male players. For female players, there is a short-lived relationship, as the probability to win points 4 and 5 decrease by 1.6 and 1.0 percentage points. No significant association is found beyond point 5 of the same game. While double faults are associated with a short-lived decrease in performance, this association is not reaching beyond serve game $t-1$.

Figure 1 demonstrates that contested games are more likely for female players during all 12 games of a set. This corresponds to findings by Krumer et al. (2016) who investigate gender differences in the overall break probability, relating them to differences in physical endowments. This should not pose a threat to our identification, as we compare contestants of equal gender.

Increased importance to perform and hold serve could lead to strategic changes in risk-taking. Change in risk during serve (e.g. by adapting speed or placement) might affect our IV and outcome variable simultaneously. Figure A.2 rejects this potential threat to our identification strategy: double faults during the first three points of any serve game relevant in our sample (2-11) are not more likely in later stages of a set compared to early games. Risk-taking might vary as sets progress, but our IV is unaffected.

4. Main results

We measure the causal effect of experiencing a contested game on relative performance on the next game. The results are tabulated in Table 2, with the main IV-results in Column (3) for women and Column (8) for men. For female players we estimate a significant and sizeable effect of -0.150: Female players suffer from a reduced probability of breaking the opponent after a contested game by -15 percentage points, which is significant and a sizeable effect. However, we estimate no significant effect of winning a

contested game $t-1$, compared to a decisively won game, on the probability to observe a break in game t. The point estimate of γ is small at 0.016 and not significant.

The first-stage results for both genders confirm that the probability to observe a contested game of 8 or more points in game t increases if one or more of the first three points in the game are decided by a double fault: 14.6 for female and 14.2 percentage points for male players. Again, our instrument is sufficiently strong at first-stage F-statistics of 365 for the women's sample, and 395 for the men's sample. By constituting an almost quantitatively identical exogenous shock for both genders, our instrument solves the problems that arise from gender differences in the probability of a contested game.

In a second step, we change the type of performance feedback slightly and focus on successfully defending one or multiple break points. Defending break points is an alternative way to win a contested game in which the opposing player had at least one chance to win the game and break serve. This measure carries even more ambiguity, as defending break points can either be interpreted negatively (as a break almost being conceded) or as positive feedback, as a break by the opposing player has been successfully prevented by winning a crucial point.

Columns (4) and (9) of Table 2 present the estimation results. The point estimate of the break-point effect is quantitatively and qualitatively similar to the main results presented above. The likelihood of a break for women after a game with defended break points decreases by 18.3 percentage points. Male players are unaffected, as we measure an insignificant coefficient of 0.016.[6] We also use a continuous measure of the number of points observed in game $t-1$. We estimate that women's probability to break serve in t decreases by 2.3 percentage points with each point played in $t-1$. A quantitatively small and not significant effect for male players is estimated at 0.2 percentage points (columns 5 and 10). The first stages for both alternative feedback measures are again sufficiently strong.

The significant estimates for women cannot be attributed to gender-specific shifts in server dominance due to previous long and severely contested games in $t-1$. Players on serve, according to this explanation, could suffer more from physical or mental exhaustion if the game consisted of a relatively high number of points. Figure A.3 rejects this alternative explanation, as we do not see an increase in the probability to observe a contested game (i.e. a decrease in server dominance) as the match progresses. This suggests that the advantage a player on serve has, for female and male players, is not affected by the physical tiring of players. The confrontational structure of tennis supports this, as both players should tire comparably as they both experience contested games in a similar way as direct opponents.

---

[6]As an alternative, we combine our main measure of contestedness of game $t-1$ and the break point variable. Both variables a correlated at 0.77. The estimates confirm our main findings.

TABLE 2 — Main results: effect of having successfully defended at least one break point in $t-1$ on performance in t

	women					men				
	(1) OLS[a]	(2) RF[b]	(3) IV-main[c]	(4) IV-alt I[d]	(5) IV-alt II[e]	(6) OLS[a]	(7) RF[b]	(8) IV-main[c]	(9) IV-alt I[d]	(10) IV-alt II[e]
Contested game in $t-1$	−0.042***		−0.150***	−0.183***	−0.023***	−0.022***		0.016	0.020	0.002
(1 = yes, 0 = no)	(0.005)		(0.048)	(0.059)	(0.007)	(0.003)		(0.038)	(0.048)	(0.005)
Instrumental variable		−0.022***					0.002			
		(0.007)					(0.005)			
Fraction of point won by server	0.546***	0.517***	0.625***	0.613***	0.620***	0.231***	0.215***	0.204***	0.206***	0.205***
before game $t-1$	(0.019)	(0.018)	(0.040)	(0.037)	(0.039)	(0.013)	(0.013)	(0.029)	(0.027)	(0.027)
Avg. game duration before $t-1$	−0.042***	−0.038***	−0.051***	−0.049***	−0.050***	−0.021***	−0.019***	−0.017***	−0.018***	−0.018***
	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.004)	(0.003)	(0.003)
Add. binary controls[f]	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Returner FEs	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Match×set FEs	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
First stage coef.[g]			0.146***	0.120***	0.960***			0.142***	0.114***	1.001***
			(0.007)	(0.007)	(0.037)			(0.007)	(0.006)	(0.033)
First stage F. stat.[h]			392.581	323.039	687.071			440.004	348.647	926.247
N	75,670					90,658				

Notes: Robust standard errors clustered on the match-set level in parentheses. *, ** and *** indicate statistical significance at the 10-percent level, 5-percent level, and 1-percent level. Singleton variables are not used to calculate standard errors. The dependent variable (mean 0.359, std. dev. 0.480 for women, mean 0.205, std. dev. 0.404 for men) is 1 if the observed return game t ends in a break, 0 if it is won by the player on serve. Only observations are used where the score differences in terms of breaks is equal to 0 for t and $t-1$. Consequently, the player on return in t has won the previous game in all cases. Singleton observations are dropped and not reported in the number of observations. [a] Results obtained by estimating Model 1 with simple OLS. [b] Results for estimating the reduced-form specification of Model 1. [c] IV estimates of Model 1. The dependent main explanatory variable is equal to 1 if the serve game $t-1$ (from the perspective of the returner in t) is categorized as contested, 0 otherwise. A contested serve game is defined by the number of points needed to find a winner: contested means ≥ 8 points. [d] IV estimates of Model 1. The main explanatory variable (mean 0.171, std. dev. 0.376 for women, mean 0.137, std. dev. 0.344 for men) is equal to 1 if the serve game $t-1$ (from the perspective of the returner in t) is won after defending at least one break point. [e] IV estimates of Model 1. The main explanatory variable is equal to the number of points played during serve game $t-1$ (from the perspective of the returner in t) needed to find a winner. [f] Additional binary controls include dummy variables indicating the score difference before game t in terms of games from the returning player's perspective. In addition, we include binary variables indicating bins of absolute WTA or ATP ranking points, as well as bins for the WTA/ATP ranking. [g] Coefficient of the instrumental variable in the first stage (Model 1). [h] Cragg-Donald Wald F statistic test for weak identification.

5. Potential channels

The channels, in terms of particular situations or individual characteristics, affecting how agents react to past performance concerning future performance or effort provision are mostly unknown (Meier et al. 2020). As introduced in Sections 1 and 3, the treatment we investigate consists of ambiguous feedback in the sense that it can either cause positive or negative psychological momentum. In this section, we dig deeper to find evidence for either a positive or negative effect. Our setting also provides an ideal environment to study heterogeneous effects and shed some light on the driving forces behind the causal effect of past on future performance.

Relative ability   So far, we can't unambiguously link our main result to a negative momentum effect for the returner or a positive effect for the returner. Relative ability, however, can provide some insights. Our data captures relative ability in a match by the ranking and the associated ranking points of both contestants. When distinguishing between clear underdogs and favorites in WTA matches, we estimate a negative (albeit only significant at the 10% level) momentum effect only for underdog players who are expected to have a lower probability to win the match than their opponents (Figure 4, panel *a*).

We interpret this as negative momentum for the observed underdog player on return. It is counterintuitive to expect favorites to benefit from positive momentum after almost breaking serve in $t-1$. Favorites should definitely not view such a form of feedback as positive, as they should expect to contest the serve of the opposing underdog. Underdogs, however, should feel their initial disadvantage in ability even reinforced and suffer from negative psychological momentum. We estimate no effect for male favorites and underdogs.

Experience   The experience a competitor has in terms of interpreting intermediate performance feedback can play a role when reacting to pseudo setbacks. Our data provide an ideal setting to investigate the role of experience, as the career paths of female and male professional tennis players are highly similar. All players enter professional tournaments around an age of 18, typically without practicing any other profession beforehand. They will compete as long as they are able to rank below a certain threshold of the WTA or ATP rankings long enough, in order to have a chance to finance their playing activities and stay eligible for large tournaments.
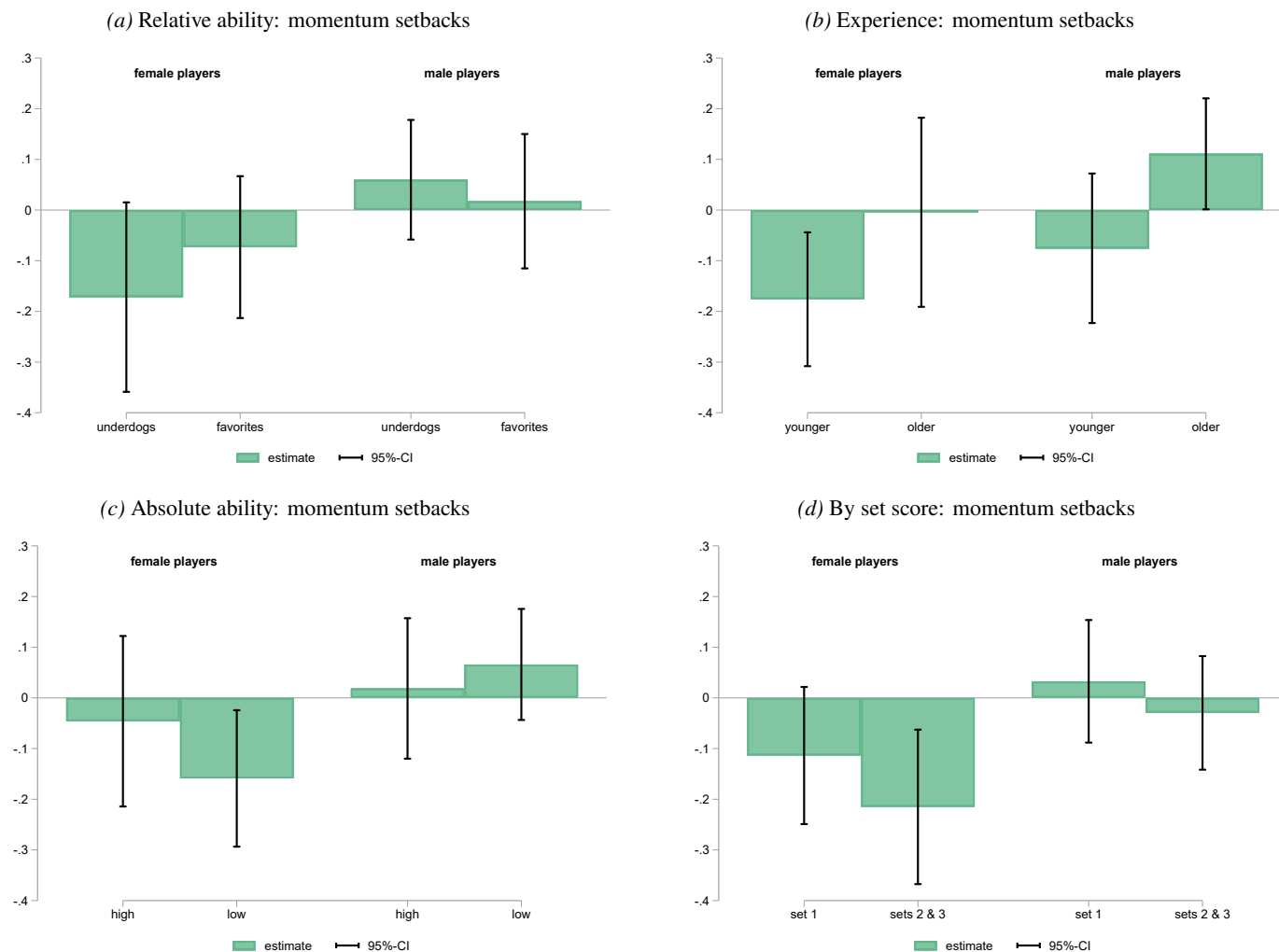
To account for heterogeneity in experience, we split the overall female and male sample along the overall median age of 26. The results are illustrated in panel *(b)* of Figure 4. For female players, we estimate a negative performance effect only for the younger players. For male players, we measure no

significant effect for younger players. For older male competitors, however, we estimate a positive and significant effect of winning a contested game on subsequent chances to break serve. Whereas experience for female players seems to mitigate negative psychological momentum, older male players benefit from "surviving" a contested serve game.

Absolute ability   A competitor's ability is a crucial mediating factor for how performance feedback is interpreted. But it is ambiguous what the interaction is between ability and experiencing a spurious type of performance feedback. For example, high ability contestants could be unaffected by psychological momentum, as their expectations to perform are always high. On the other hand, players with high ability could make the strongest reduction in effort, as they should not expect to experience substantial competition and react stronger to psychological momentum. Indeed, we estimate that low-ability female players show a significant momentum effect; high ability players do not. Again, we find no effect for male players (Figure 4, panel *c*). This contrasts the findings of Shastry et al. (2020): their findings suggest that low-ability men do react to negative feedback, while high- and low-performing women attribute negative feedback —albeit a result of pure luck— to lack of ability.

Intermediate score   An important form of heterogeneity is the intermediate score or the progression of the contest in general. In all first sets, a negative performance (e.g. conceding a break) is least consequential: while a player leading by 1:0 has an incentive to avoid conceding a break because s/he has a chance to win the match in the current set, a player down 0:1 or tied at one set each also finds herself in a situation of increased importance to perform (note that all matches in our sample are decided by winning two sets, often referred to "best-of-three"). Any form of intertemporal effects of past on future performance should therefore be larger when the scare of almost going down a break is more pronounced. Players get an ambiguous type of feedback that the previous game was very close and contested, feedback that is potentially stronger the higher the importance to perform. Estimates illustrated in panel *(d)* of Figure 4 confirm this hypothesis: female players competing in sets 2 and 3 experience a significant momentum that is even more pronounced.

FIGURE 4 — Identification of potential channels

(a) Relative ability: momentum setbacks

(b) Experience: momentum setbacks

(c) Absolute ability: momentum setbacks

(d) By set score: momentum setbacks

Notes: Panel *(a)* presents the result from estimating Model 1 with split samples. We split the sample into young (≤ 26) and older players (> 26), which is roughly the sample median age for all observed female and male players. Panel *(b)* plots the estimation results for splits along relative ability. We define favorites as players who have more ranking points than their opponents. Underdogs have fewer points. Panel *(c)* illustrates estimation results for samples of high and low absolute ability. Low ability is defined as having less than 1,000 WTA or ATP ranking points before the observed match, high ability as having more than 1,000 points. Finally, panel *(d)* plots the different estimates along the course of the game for all first sets, and later sets.

6. Robustness

Table A.2 provides a series of robustness checks to support our main results. First, we estimate our Model excluding game number 12 from the sample. Game 12 is special, as it is the only game that can end the set despite a break difference of 0 in t. The pressure to perform for the player on serve is higher than in all other games, as losing serve will directly imply losing the set. A player under such competitive pressure could perform worse (Apesteguia & Palacios-Huerta 2010, Böheim et al. 2019). The estimates presented in columns (1) of panels A and B demonstrate that the omission does not change our main results.

Second, we test whether the inclusion of overly long games in $t-1$ change the psychological momentum effects. Estimation results from this extended sample do confirm our main results (see columns 2 of panels A and B). Third, we omit observations from Grand Slam tournaments from the women's sample.[7] Again, our main results are confirmed by the results presented in columns 3 of panels A and B.

In a next step, we use observations from the main draws only. These matches are all within the tournament tree, without initial qualifying matches. All main results are confirmed (see columns (4) of panels A and B).

The threshold of 8 points to define a contested game—despite its special position within the rules of tennis by indicating deuce— may seem arbitrary. Alternative definitions of >6 or >10 points played in game $t-1$ for contested games confirm our main results. These estimates are presented in columns 5 (>6) and columns 6 (>10) of panels A and B in Table A.2. We estimate that the causal effect in game t is stronger, the more contested game $t-1$ (i.e., longer) was.

Table A.3 in the Appendix presents alternative specifications of Model 1, step-by-step including additional control variables. The estimates indicate that no specific inclusion of control variable or fixed-effect is driving our main results.

7. Discussion and Conclusion

Our estimates indicate a causal effect of past on future performance for female professional athletes, and no feedback effect for men. The causal effect emerges from performance feedback that a player had difficulties in winning a stage of the contest, without actually losing the stage. We call such a stage a "contested game". Losing the stage would directly and substantially impact winning probabilities, as losing a serve game is the fundamental ingredient to lose a tennis match.

---

[7]We do not use observations from men's Grand Slam matches, as these matches are played in best-of-5 format. Best-of-5 matches will likely have too many underlying structural differences from best-of-3 matches as used in our analysis.

Women experience a negative causal effect of 12 percentage points to score a break in the game following a contested game. Male players do not experience any effect. Identifying potential channels, mostly inexperienced and low-ability female competitors are affected negatively. Women who are underdogs in a match experience a decrease in performance after contested games, which demonstrates that relative ability of the contestants influences the effect of intermediate performance. The effect is also more pronounced in stages of the contest where importance to perform is high.

We interpret our results as negative psychological momentum. Women might interpret a contested game in the past as a specific form of negative feedback, subsequently triggering negative momentum and decreasing performance. Consequently, our findings complement those of Cohen-Zada et al. (2017), who find that only men are affected by positive momentum, while women remain unaffected. Men, in contrast, do not react with performance decrements. For both genders, experience leads to a more positive effect of a contested game. Experienced women have no feedback effect, and experienced men have a positive feedback effect. Our findings could also be explained by gender differences in the reaction to certain types of performance feedback. Female and male children are found to react differently to positive procedure feedback (Henderlong & Lepper 2002). Experiencing a contested game when on serve is related to procedure feedback: competitors receive feedback on how they won the game. Our results suggest a gender gap in the reaction to negative procedure feedback for adult competitors who are experts in their domain.

The results of our analysis have large implications for a multitude of different settings. For example, we shed new light on the optimal design of contests and, in particular, potential means of providing information on intermediate relative performance for both genders. Gender specific long-lasting consequences of feedback on past performance for career development are also of concern (Oreopoulos et al. 2012, Schwandt & Von Wachter 2019, Von Wachter 2020). Contrasting previous results, our estimates suggest different reactions of women to negative shocks on past performance.

Applied to an educational setting, our results mean that male students should remain unaffected by psychological momentum triggered by performance feedback. However, we would expect undesired detrimental effects on effort of female students. Revealing to them that they were close to failing a midterm exam due to some purely random exogenous shock (e.g. illness) could have significant negative effects for subsequent exams. This undesirable effect would even be stronger for low-ability individuals, both for people having low absolute ability and for people that just are underdogs in the respective competition.

References

Adler, P. (1981), *Momentum, a Theory of Social Action: A Theory of Social Action*, Sage Pubns.

Apesteguia, J. & Palacios-Huerta, I. (2010), 'Psychological pressure in competitive environments: Evidence from a randomized natural experiment', *American Economic Review* **100**(5), 2548–64.

Azmat, G., Bagues, M., Cabrales, A. & Iriberri, N. (2019), 'What you don't know... can't hurt you? a natural field experiment on relative performance feedback in higher education', *Management Science* **65**(8), 3714–3736.

Azmat, G. & Iriberri, N. (2010), 'The importance of relative performance feedback information: Evidence from a natural experiment using high school students', *Journal of Public Economics* **94**(7-8), 435–452.

Banko, L., Leeds, E. M. & Leeds, M. A. (2016), 'Gender differences in response to setbacks: Evidence from professional tennis', *Social Science Quarterly* **97**(2), 161–176.

Bar-Eli, M., Krumer, A. & Morgulev, E. (2020), 'Ask not what economics can do for sports-ask what sports can do for economics', *Journal of Behavioral and Experimental Economics* **89**(C).

Berger, J. & Pope, D. (2011), 'Can losing lead to winning?', *Management Science* **57**(5), 817–827.

Böheim, R., Grübl, D. & Lackner, M. (2019), 'Choking under pressure–evidence of the causal effect of audience size on performance', *Journal of Economic Behavior & Organization* **168**, 76–93.

Booth, A. & Yamamura, E. (2018), 'Performance in mixed-sex and single-sex competitions: What we can learn from speedboat races in japan', *Review of Economics and Statistics* **100**(4), 581–593.

Casas-Arce, P. & Martínez-Jerez, F. A. (2009), 'Relative performance compensation, contests, and dynamic incentives', *Management Science* **55**(8), 1306–1320.

Cohen-Zada, D., Krumer, A. & Shtudiner, Z. (2017), 'Psychological momentum and gender', *Journal of Economic Behavior & Organization* **135**, 66–81.

Delfgaauw, J., Dur, R., Non, A. & Verbeke, W. (2014), 'Dynamic incentive effects of relative performance pay: A field experiment', *Labour Economics* **28**, 1–13.

Depken, C. A., Gandar, J. M. & Shapiro, D. A. (2021), 'Set-level strategic and psychological momentum in best-of-five matches in professional tennis', *Applied Economics Letters* pp. 1–5.

Depken, C. A., Gandar, J. M. & Shapiro, D. A. (2022), 'Set-level strategic and psychological momentum in best-of-three-set professional tennis matches', *Journal of Sports Economics* p. 15270025221085715.

Dong, L. & Huang, L. (2019), 'Is there no 'i'in team? strategic effects in multi-battle team competition', *Journal of Economic Psychology* **75**, 102070.

Forbes (2019*a*), 'The highest-paid female athletes 2019: Serena and Osaka dominate'. Ret. November 1 2021, https://www.forbes.com/sites/kurtbadenhausen/2019/08/06/the-highest-paid-female-athletes-2019-serena-and-osaka-dominate/?sh=45bafa0f2fcc.

Forbes (2019*b*), 'The highest-paid tennis players 2019'. Ret. July 1 2021, https://www.forbes.com/sites/kurtbadenhausen/2019/08/26/highest-paid-tennis-players-roger-federer/?sh=1750b1c36cc6.

Fudenberg, D., Gilbert, R., Stiglitz, J. & Tirole, J. (1983), 'Preemption, leapfrogging and competition in patent races', *European Economic Review* **22**(1), 3–31.

Gauriot, R. & Page, L. (2019), 'Does success breed success? A quasi-experiment on strategic momentum in dynamic contests', *The Economic Journal* **129**(624), 3107–3136.

Geister, S., Konradt, U. & Hertel, G. (2006), 'Effects of process feedback on motivation, satisfaction, and performance in virtual teams', *Small Group Research* **37**(5), 459–489.

Gill, D., Kissová, Z., Lee, J. & Prowse, V. (2019), 'First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision', *Management Science* **65**(2), 494–507.

Gill, D. & Prowse, V. (2014), 'Gender differences and dynamics in competition: The role of luck', *Quantitative Economics* **5**(2), 351–376.

Goulas, S. & Megalokonomou, R. (2021), 'Knowing who you actually are: The effect of feedback on short-and longer-term outcomes', *Journal of Economic Behavior & Organization* **183**, 589–615.

Henderlong, J. & Lepper, M. R. (2002), 'The effects of praise on children's intrinsic motivation: A review and synthesis.', *Psychological bulletin* **128**(5), 774.

Hermes, H., Huschens, M., Rothlauf, F. & Schunk, D. (2021), 'Motivating low-achievers—relative performance feedback in primary schools', *Journal of Economic Behavior & Organization* **187**, 45–59.

Iso-Ahola, S. E. & Dotson, C. O. (2014), 'Psychological momentum: Why success breeds success', *Review of General Psychology* **18**(1), 19–33.

Jalava, N., Joensen, J. S. & Pellas, E. (2015), 'Grades and rank: Impacts of non-financial incentives on test performance', *Journal of Economic Behavior & Organization* **115**, 161–196.

Jeworrek, S. (2019), 'Gender stereotypes still in mind: Information on relative performance and competition entry', *Journal of Behavioral and Experimental Economics* **82**, 101448.

Kahn, L. M. (2000), 'The sports business as a labor market laboratory', *Journal of Economic Perspectives* **14**(3), 75–94.

Klumpp, T. & Polborn, M. K. (2006), 'Primaries and the New Hampshire effect', *Journal of Public Economics* **90**(6-7), 1073–1114.

Kniffin, K. M. & Mihalek, V. (2014), 'Within-series momentum in hockey: No returns for running up the score', *Economics Letters* **122**(3), 400–402.

Krumer, A., Rosenboim, M. & Shapir, O. M. (2016), 'Gender, competitiveness, and physical characteristics: Evidence from professional tennis', *Journal of Sports Economics* **17**(3), 234–259.

Lovász, A., Cukrowska-Torzewska, E., Rigó, M. & Szabó-Morvai, Á. (2022), 'Gender differences in the effect of subjective feedback in an online game', *Journal of Behavioral and Experimental Economics* **98**, 101854.

Mago, S. D., Sheremeta, R. M. & Yates, A. (2013), 'Best-of-three contest experiments: Strategic versus psychological momentum', *International Journal of Industrial Organization* **31**(3), 287–296.

Malueg, D. A. & Yates, A. J. (2010), 'Testing contest theory: evidence from best-of-three tennis matches', *The Review of Economics and Statistics* **92**(3), 689–692.

Meier, P., Flepp, R., Ruedisser, M. & Franck, E. (2020), 'Separating psychological momentum from strategic momentum: Evidence from men's professional tennis', *Journal of Economic Psychology* **78**, 102269.

Morgulev, E., Azar, O. H. & Bar-Eli, M. (2019), 'Does a "comeback" create momentum in overtime? analysis of nba tied games', *Journal of Economic Psychology* **75**, 102126.

Oreopoulos, P., Von Wachter, T. & Heisz, A. (2012), 'The short-and long-term career effects of graduating in a recession', *American Economic Journal: Applied Economics* **4**(1), 1–29.

Schwandt, H. & Von Wachter, T. (2019), 'Unlucky cohorts: Estimating the long-term effects of entering the labor market in a recession in large cross-sectional data sets', *Journal of Labor Economics* **37**(S1), S161–S198.

Shastry, G. K., Shurchkov, O. & Xia, L. L. (2020), 'Luck or skill: How women and men react to noisy feedback', *Journal of Behavioral and Experimental Economics* **88**, 101592.

Tran, A. & Zeckhauser, R. (2012), 'Rank as an inherent incentive: Evidence from a field experiment', *Journal of Public Economics* **96**(9-10), 645–650.

Von Wachter, T. (2020), 'The persistent effects of initial labor market conditions for young adults and their sources', *Journal of Economic Perspectives* **34**(4), 168–94.

Wozniak, D. (2012), 'Gender differences in a market with relative performance feedback: Professional tennis players', *Journal of Economic Behavior & Organization* **83**(1), 158–171.

Wozniak, D., Harbaugh, W. T. & Mayr, U. (2014), 'The menstrual cycle and performance feedback alter gender differences in competitive choices', *Journal of Labor Economics* **32**(1), 161–198.

8. TABLE AND GRAPHICAL APPENDIX

FIGURE A.1 — Relative frequency of game durations, by gender
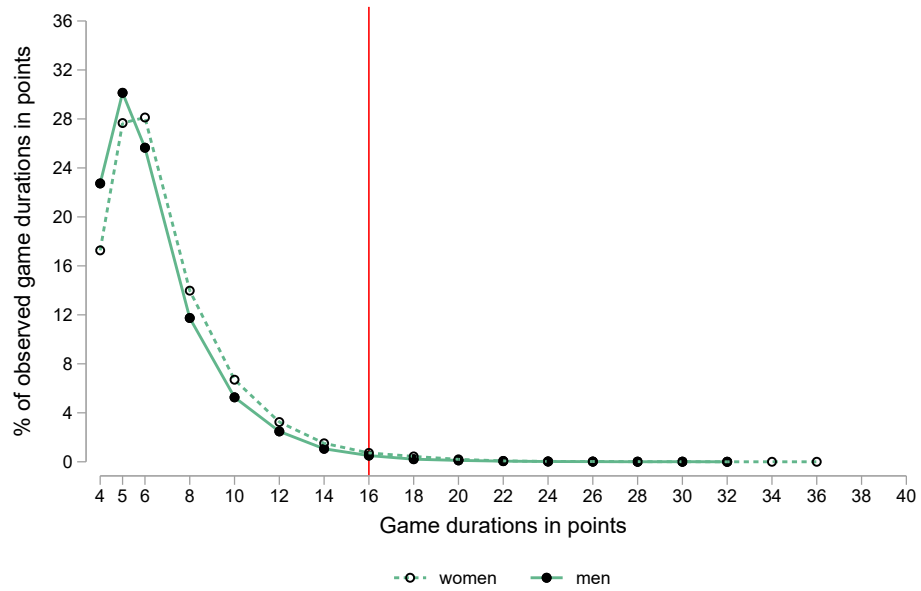
Notes: Relative frequency (in %) of observed game length for game *t* for female and male players. Only games shorter than 18 points are used for the empirical analysis (as indicated by the vertical red line).

TABLE A.1 — Testing the validity of the instrumental variable: association of past and future double faults

(a) Association of past and future double faults

|	women				men			
	(1) DF-P2	(2) DF-P3	(3) DF-P3	(4) DF-P3	(5) DF-P2	(6) DF-P3	(7) DF-P3	(8) DF-P3
Double fault P1[a]	−0.004 (0.004)		−0.004 (0.004)	−0.004 (0.004)	−0.007** (0.004)		−0.002 (0.004)	−0.002 (0.004)
Double fault P2[b]		−0.006 (0.004)		−0.006 (0.004)		−0.001 (0.004)		−0.001 (0.004)
Serving player FEs	yes	yes	yes	yes	yes	yes	yes	yes
Game score FEs	yes	yes	yes	yes	yes	yes	yes	yes
Match×set FEs	yes	yes	yes	yes	yes	yes	yes	yes
N	118,655				117,135			
R^2	0.222	0.217	0.217	0.217	0.182	0.183	0.183	0.183

Notes: Robust standard errors clustered on the match-set level in parentheses. *, ** and *** indicate statistical significance at the 10-percent level, 5-percent level, and 1-percent level. The dependent variable for columns (1) and (5) is a binary variable equal to 1 if the second point of the observed serve game t is decided by a double fault, 0 otherwise. The dependent variable for columns (2) through (5), and (6) through (8) is a binary variable equal to 1 if the third point of the observed serve game t is decided by a double fault, 0 otherwise. [a] Binary variable equal to 1 if the first point in the observed serve game is decided by a double fault, 0 otherwise. [b] Binary variable equal to 1 if the second point in the observed serve game is decided by a double fault, 0 otherwise.

(b) Association of instrumental variable on probability to win subsequent points

|	women's matches					men's matches				
	point 4	point 5	point 6	point 7	point 8	point 4	point 5	point 6	point 7	point 8
IV	−0.016*** (0.005)	−0.010* (0.006)	−0.008 (0.007)	−0.007 (0.012)	−0.019 (0.012)	−0.006 (0.006)	−0.003 (0.006)	0.001 (0.008)	0.013 (0.014)	0.009 (0.013)
Serving player FEs	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Game score FEs	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Match×set FEs	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.205	0.226	0.287	0.398	0.390	0.176	0.206	0.285	0.394	0.402
N	118,655	98,995	64,841	25,855	25,855	117,135	92,368	56,434	20,982	20,982

Notes: Robust standard errors clustered on the match-set level in parentheses. *, ** and *** indicate statistical significance at the 10-percent level, 5-percent level, and 1-percent level. The dependent variable is equal to 1 if one of points number 4 through 8 is won by the player on serve, 0 otherwise.

FIGURE A.2 — Instrumental variable by game number
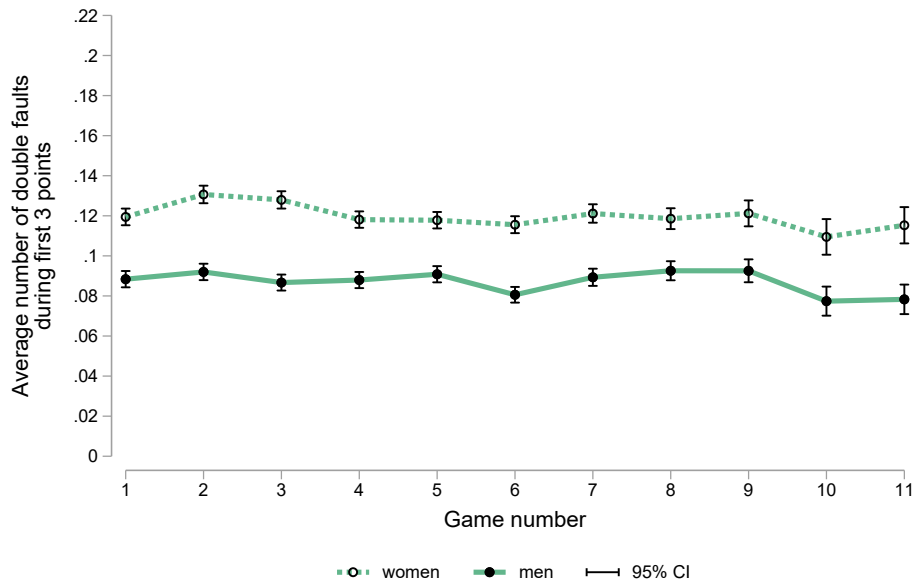
Notes: Probability to observe at least one double fault during the first three points for all games that were used to construct our instrumental variable (games 1 through 11 of a set).

FIGURE A.3 — Relative frequency of contested game along the progression of a match
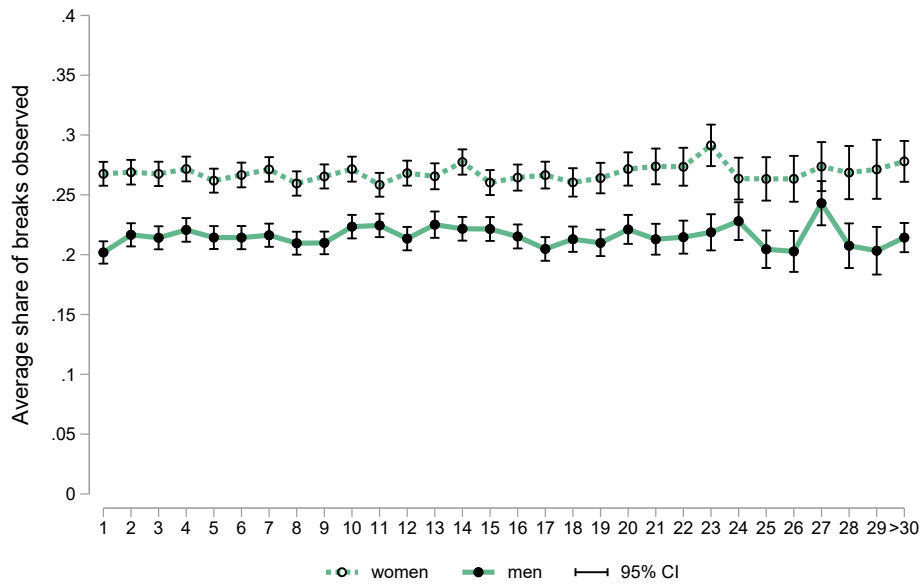
Notes: Relative frequency of a contested game over the course of a match. The maximum possible match duration is 36 games, games 30 through 36 are pooled together due to low number of observations.

PANEL A. WOMEN

	(1)	(2)	(3)	(4)	(5)	(6)
Contested game in $t-1$	−0.157***	−0.137***	−0.129**	−0.134**	−0.094***	−0.299***
	(0.050)	(0.047)	(0.051)	(0.054)	(0.030)	(0.098)
Frac. of points won by	0.624***	0.593***	0.634***	0.608***	0.588***	0.662***
server up to game $t-1$	(0.042)	(0.040)	(0.043)	(0.044)	(0.030)	(0.052)
Avg. game duration	−0.051***	−0.046***	−0.050***	−0.049***	−0.046***	−0.055***
before game $t-1$	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)	(0.006)
Add. binary controls	yes	yes	yes	yes	yes	yes
Returner FEs	yes	yes	yes	yes	yes	yes
Match×set FEs	yes	yes	yes	yes	yes	yes
N	72,252	76,274	63,432	61,158	75,670	75,670
first stage F. stat.	366.329	408.048	356.587	298.692	1,088.536	154.375

PANEL B. MEN

	(1)	(2)	(3)	(4)	(5)	(6)
Contested game in $t-1$	0.026	0.015	0.016	0.027	0.009	0.039
	(0.039)	(0.038)	(0.038)	(0.041)	(0.021)	(0.091)
Frac. of points won by	0.190***	0.200***	0.204***	0.200***	0.208***	0.199***
server up to game $t-1$	(0.031)	(0.029)	(0.029)	(0.031)	(0.022)	(0.041)
Avg. game duration	−0.016***	−0.017***	−0.017***	−0.018***	−0.018***	−0.017***
before game $t-1$	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.005)
Add. binary controls	yes	yes	yes	yes	yes	yes
Returner FEs	yes	yes	yes	yes	yes	yes
Match×set FEs	yes	yes	yes	yes	yes	yes
N	86,247	91,048	90,658	75,267	90,658	90,658
first stage F. stat.	416.306	452.293	440.004	376.604	1,409.843	142.004

Notes: Robust standard errors clustered on the match-set level in parentheses. *, ** and *** indicate statistical significance at the 10-percent level, 5-percent level, and 1-percent level. Singleton variables are not used to calculate standard errors. The dependent variable is equal to 1 if the observed game t results in a break, 0 otherwise. For columns (1) only games 2–11 for each set were used, game 12 was excluded. Columns (2) includes also overly long games exceeding a total of 16 points played. Columns (3) uses a sample excluding Grand Slam, while columns (4) only use matches from tournaments' main draws (excluding qualifying matches). Columns (5) present an alternative specification defining contested games in $t-1$ as ≥ 6 points played, while contested games in (6) are defined as ≥ 10.

TABLE A.3 — Aditional robustness checks

PANEL A. WOMEN

	(1)	(2)	(3)	(4)	(5)
Contested game in $t-1$	−0.103***	−0.102***	−0.091**	−0.086**	−0.150***
	(0.039)	(0.039)	(0.039)	(0.039)	(0.048)
Frac. of points won by		−0.147***	−0.151***	−0.118***	0.625***
server up to game $t-1$		(0.012)	(0.013)	(0.013)	(0.040)
Avg. game duration		0.011***	0.010***	0.008***	−0.051***
before game $t-11$		(0.001)	(0.001)	(0.001)	(0.004)
Add. binary controls	*no*	*no*	*yes*	*yes*	*yes*
Returner FEs	*no*	*no*	*no*	*yes*	*yes*
Match×set FEs	*no*	*no*	*no*	*no*	*yes*
N	75,670	75,670	75,670	75,670	75,670
first stage F. stat.	635.240	635.637	632.088	639.317	392.581

PANEL B. MEN

	(1)	(2)	(3)	(4)	(5)
Contested game in $t-1$	−0.006	0.000	0.008	0.025	0.016
	(0.036)	(0.036)	(0.036)	(0.035)	(0.038)
Frac. of points won by		−0.148***	−0.154***	−0.112***	0.204***
server up to game $t-1$		(0.009)	(0.010)	(0.010)	(0.029)
Avg. game duration		0.014***	0.014***	0.010***	−0.017***
before game $t-11$		(0.001)	(0.001)	(0.001)	(0.004)
Add. binary controls	*no*	*no*	*yes*	*yes*	*yes*
Returner FEs	*no*	*no*	*no*	*yes*	*yes*
Match×set FEs	*no*	*no*	*no*	*no*	*yes*
N	90,658	90,658	90,658	90,658	90,658
first stage F. stat.	570.580	574.542	568.591	586.167	440.004

Notes: Robust standard errors clustered on the match-set level in parentheses. *, ** and *** indicate statistical significance at the 10-percent level, 5-percent level, and 1-percent level. Singleton variables are not used to calculate standard errors. The dependent variable is equal to 1 if the observed game t results in a break, 0 otherwise.