Overlapping Marathons: What Happens to a Female Runner's Pace when the Men Catch Up? *

Erica G. Birk
Logan M. Lee
Glen R. Waddell

July 2018

Abstract

We exploit a highly competitive environment in which elite-female athletes are exposed to the presence of men, but without being in direct competition with them. Specifically, we use variation in the how fast the fastest man runs in the New York City Marathon to identify the potential influence of men on female performance while holding constant female-runners' marginal incentives to perform. Our results suggest that the presence of men negatively affects the performance of female runners differentially across ability, with performance declines concentrated among lower-ability runners.

Keywords: tournament, gender, competitiveness, sport

JEL Classification: J16, Z2

* Birk (erica.birk@analysisgroup.com) is an Associate at Analysis Group, Lee (leelogan@grinnell.edu) is an Assistant Professor at Grinnell College, and Waddell (waddell@uoregon.edu) is a Professor of Economics at the University of Oregon and a Research Fellow at IZA.

1 Introduction

A large literature suggests that men are more competitive than women, with men tending to enter into competition at higher rates (Niederle and Vesterlund, 2007; Azmat and Petrongolo, 2014) and perform better conditional on having competition imposed on them (Gneezy et al., 2003). Niederle and Vesterlund (2007) concludes that gender differences in competitiveness are driven by differences in confidence and underlying preferences for entering and performing in competitive environments. Ultimately, they call on economists to give more attention to what contributes to such gender differences, suggesting that "much may be gained if we can create environments in which high-ability women are willing to compete." To the extent economic gains are achieved through competition—from grades in the classroom to promotion and advancement on the job—the creation of such environments is crucial.

In an attempt to better understand competitive outcomes, we consider the elite runners of the New York City Marathon—a unique environment in which high-performing and competitive women are exogenously treated to the presence of men. Importantly, however, and, in a rare opportunity to identify part of the technology of competitiveness, this introduction of men occurs in a way that is without direct consequence to the competition itself. That is, while in competition against other females and rewarded only for performance relative to other female competitors, male runners just... show up.

They could just be ignored and the race would continue without effect. Yet, we will demonstrate that there are significant declines in the performance of females upon these interactions with male competitors. It is important, we believe, to have held constant the return to performance, even as the competitive environment changes with the introduction of men, and in this way elite marathoning presents a unique opportunity for us to potentially identify deep-seated aspects of competition and competitiveness. In the end, our results suggest that any tendency for women to underinvest in competition with men (of the sort one might anticipate given Niederle and Vesterlund (2007), for example) need not appeal to women anticipating different returns to their effort, or women merely updating anticipated probabilities of success when they are in competition with men. That is, even without a direct role in the competition, and among a selected group of elite competitors, we still find that the presence of men is costly to the performance of women.

Documenting performance declines in this setting has important implications for other settings in which men and women interact without entering into direct competition. For example, students often perceive themselves to be in competition with each other, even in the absence of strict curving of grades.[1] Similarly, many firms incentivize workers with bonuses or merit-based compensation—while bonuses can often be explicitly competitive, to the extent that the bonuses are achieved by reaching individual-performance thresholds, individuals are not actually in direct competition in terms of incentives. Yet, employees can remain very aware of others.[2] As the empirical regularity we find suggests that the mere presence of men may affect female performance, it would be too strong to conclude that environments without direct competition are also without systematic differentials in

[1] Buser et al. (2014) finds that differences in competitiveness account for 20 percent of the gender differences in choice of academic track.

[2] Grund (2015) finds that gender gaps in pay among more-educated professionals are driven by gaps in realized bonuses more than by salary differentials, consistent with realized gender gaps resulting from differentials in competitiveness or competitive payment schemes, even among the highly able.

performance.

As will become apparent in our discussion of identification, our environment affords no similar opportunity to cleanly identify the responsiveness of male competitors to the presence of females, nor the opportunity to retrieve the causal parameter on the responsiveness of females to other females. Instead, we contribute to the existing literature by identifying a particular sensitivity of female performance to the presence of men, and a systematic heterogeneity across ability, without having the potential confoundedness associated with men being in direct competition.

In Section 2 we describe other related literatures, developing context for the interpretation of our empirical results. We develop our empirical strategy in Section 3, spending time on the issue of identification as the associated challenges are real and it is important to articulate how we get around them. We then present our empirical results in Section 4—we also offer a descriptive account of what happens to the pace of male runners as they encroach on the female race. We offer our concluding remarks in Section 5.

2 Background

There are several examples in the existing literature that inform what we know of female competitiveness. For example, in an experiment in which fourth-graders are timed running short distances, Gneezy and Rustichini (2004) shows that competition enhances the performance of males, but not females. This "competitiveness gap" has since been found to vary across age (Larson, 2005; Samak, 2013), ethnic groups (Gong and Yang, 2012), urban and rural settings (Bjorvatn et al., 2016), geographic regions (Cárdenas et al., 2012; Khachatryan et al., 2015; De Paola et al., 2015), and across societies considered to be matriarchal or patriarchal (Andersen et al., 2013).[3]

Absent from the literature are empirical settings in which there are records of female performance with and without the presence of men while while holding constant the anticipated marginal return to female effort. In our setting, the competition in which women are engaged is held constant allowing for the identification of whether the mere *presence* of men is itself a determinant of performance. Of course, the workplace need not be explicitly separated by gender for there to be "presence effects"—our environment is unique in that it allows the identification of such effects—and we anticipate that any such effect may be contribute to general characterizations of the responsiveness of performance to competition.[4]

In this environment, it is reasonable to anticipate that there would be no measurable influence of males on female performance. For example, the presence of men has little or no informational value, as runners receive "split times" as they run, which inform them relative to their expectations. In the era of our sample, runners are also inclined to wear watches, giving them even-more-precise, real-time information on their performance and pace. The

[3] Gneezy et al. (2003) and Price (2008) also find that female performance suffers with competition with men. Other examples include Deci et al. (1981), Campbell (2002), Buser (2012), Wozniak et al. (2014), and Anbarci et al. (2014). Also relevant, both Booth and Nolen (2012) and Booth et al. (2014) find that girls in a single-sex classroom environment were less risk averse than were girls in mixed-gender classrooms, suggesting that repeated interactions with male competitors my increase competitiveness.

[4] As there is evidence that the gender gap in competitiveness attenuates with repeated competition (Cotton et al., 2013), it is interesting to contemplate whether these presence effects (as a potential component partof competitiveness) change over time, though our priors are that the women in our sample, if anything, exhibit less responsiveness than we might expect in a more representative sampleof women. Evidence for learned behaviors tends also to be drawn from variation across age—10 and 11 year old girls perform worse in the presence of boys in *Jeopardy*, but the performance of women does not appear to be effected by men on the adult version of the show (Lindquist and Säve-Söderbergh, 2011; Jetter and Walker, 2017; Säve-Söderbergh and Sjögren Lindquist, 2017).

literature does suggest that performance-related information can impact performance in experimental subjects (Mitchell et al., 1985; Meyer and Gellatly, 1988). For example, negative feedback has been shown to reduce self-confidence, particularly among women (Roberts and Nolen-Hoeksema, 1989), which may induce performance declines (Woodman and Hardy, 2003). Consistent with this interpretation, De Paola et al. (2015) finds that women and men perform similarly in competitive settings where risk aversion, feedback, and self-confidence are not relevant.[5] Again, however, we do not anticipate appealing to information to explain out results, given the real-time information that is available to all competitors throughout the race.

Distraction itself may explain the empirical regularities we identify. There is some evidence, for example, that mobile phones lead to performance declines in drivers (Horberry et al., 2006) and in students (Ellis et al., 2010). However, we question whether the mid-race arrival of men is of the same sort of disruptive potential that these studies identify—that said, distraction would explain the empirical regularity we uncover.[6]

While somewhat tangential, there is a limited literature related to elite-female distance running itself that we would be remiss to neglect. Where relevant, the takeaway is that male races tend to be more competitive than female races.[7]

[5] Veldhuizen (2017) shows that 48 percent of the competition gap can be explained by differences in confidence, with an additional 37 percent of the gap explained by the interaction between confidence and risk preferences. Thus, concluding that the gender gap is driven by risk attitudes and overconfidence and not directly by a separate "competitiveness" trait.

[6] We do note, that there is evidence, also from phones and driving, that women suffer larger performance declines with distraction (Hancock et al., 2003; Lesch and Hancock, 2004).

[7] Frick (2011) used a cross section of ultra-marathons to argue that men's races are more competitive, while Deaner (2013) argues that the lower variance in times among the best male runners relative to the best female runners is evidence of a competitiveness gap.

3 Empirics

3.1 Data and identification

The New York City Marathon is an annual race with more than 50,000 participants in a typical year. Since 1976, the race has followed a 26.2-mile path that takes runners into all five of New York's boroughs. As one of six "World Marathon Majors," the New York City Marathon regularly draws in many of the most-elite marathon runners in the world. As with many marathons, prize pools are identical for both genders—in 2017, the fastest male and female runners each received $100,000 in prize money.[8]

For our analysis, we collect data from the 2007 through 2014 New York City (NYC) marathons, as the NYC Marathon facilitates the identification of an important parameter in the determination of competitiveness and the role of gender in competition.[9] As is typical in marathons, the rewards associated with performance are gender specific. Thus, even though both male and female runners can be on the course at the same time, female outcomes are not determined directly by male performance. This restricts the potential determinants of female performance in an important way, offering a rare opportunity to exploit variation in whether women experience the influence of men during their competition while holding constant the marginal return to female effort or performance. Thus, to the extent female runners systematically respond to male runners, there can be no appeal to within-tournament

[8] Prizes are also available to runners who complete the race within a pre-specified time. For example, in 2017, female (male) runners who finished in less than 2:27:00 (2:10:00) received an additional $10,000 bonus. U.S. citizens also compete for $58,000 in additional prize money, divided among the five fastest American runners in each race. Additional detail available from TCS New York City Marathon, at https://www.tcsnycmarathon.org/.

[9] Due to Hurricane Sandy, the 2012 marathon was not held.

mechanisms or incentives for explanation.

Also typical in elite marathons is the recording of each runner's "split-times"—cumulative running times at every five kilometers throughout the race. It is these split-times that allow for the inclusion of runner fixed effects, for example, and while our sample of runners is already elite, variation in runner ability may still be important to absorb, as might be other time-invariant runner attributes.

Before continuing, however, note a draw back to a "runner-fixed-effects" approach. Namely, runners often vary across ordinal ranking throughout the race, which then inadvertently exposes such an approach to identifying across observations that vary in the expected returns to performance. For example, if a runner responds to being passed by picking up pace, which triggers an increase in ordinal rank, the relationship between being passed and pace will be confounded with a coincident change in the returns to effort (as rewards increase in ordinal rank). This is threatening to the econometrician's ability to achieve clean identification through runner fixed effects, of course, and suggests that a "within-rank" approach to identification may well be informative. Thus, while we will report specifications inclusive of runner fixed effects, these challenges motivate an alternative source of identifying variation. (It is reassuring that our two approaches support the same interpretation of the data.)

Fortunately, there are two additional details of the NYC Marathon in particular that allow for an alternative identification of the responsiveness of female runners. First, we note that the New York City Marathon route is identical each year. While we still soak up any year-specific variation separately with year fixed effects (e.g., temperature, weather), route consistency will facilitate our identifying variation coming from comparisons of runners across years. In particular, we will compare runners who were at the same rank at the same distance in the race, across years. The second detail is our fundamental source of identifying information. In the NYC Marathon, "elite" runners are managed separately from the general population of entrants, and in a way that we exploit. Specifically, since 2002, the group of elite women—there are 32 in the typical NYC Marathon—begin the race approximately 30 minutes ahead of the next wave of runners, including a group of (typically 33) elite men. This sets up the prospect that the fastest male runners will begin passing female runners somewhere in the middle of the race. Unlike the elite runners, males and females in the general population start together. As such, female runners in the general population are not passed by male runners so much as they are continuously among them, endogenously separating and interacting with them throughout the race. "Treatment" is therefore much-less clear, and the conditionally exogenous variation required to identify the causal parameter of interest is not apparent. Thus, we restrict the sample to elite runners.[10]

The process for being selected as an elite runner by the NYC marathon is somewhat opaque, though the intent in inviting runners to run among the "elite" in the NYC Marathon is an attempt to include all runners with a realistic chance of winning the race (Lorge, 2010). The motivation for the 30-minute head start for these women is also related to this determination, as it is to prevent male pacesetters from aiding women during the race. Should a winner emerge from among those who did not come from this elite group, questions of

[10] We also exclude the 11 elite runners who did not finish, though the inclusion of these runners does not change our results in any meaningful way.

fairness would likely be raised (Lorge, 2010).

Specifically, then, we rank-order all elite-female runners at the end of each five-kilometer split and ask whether the pace of the r^{th}-ranked runners varies across years by whether they had just been passed by the fastest male runner or not. In short, as long as in some years runners of rank r had been passed by that split, while in other years runners of rank r had not been, there is identifying variation.[11] In the end, we are actually identifying off of variation in the performance of the fastest males across race-years, holding constant female ordinal rank and distance (and therefore the marginal returns to effort).[12] In lieu of controlling for unobserved runner heterogeneity, we include among our identifying assumptions that ordinal rank at given distances in the race is sufficient to absorb any level effects of ability on how runners respond to being passed by the fastest male. Importantly, however, by losing the runner fixed effects, our within-split-by-rank estimator does restrict identifying variation to runners who are projecting the same marginal returns to any change in ordinal rank in the tournament.

[11] It might seem evident that a kinked-regression-discontinuity design would be informative in identifying any change in pace as a runner is being passed. Unfortunately, because we observe only average pace over five-kilometer increments, and not a continuous measures of runner pace, we are not able to directly observe the change in pace that occurs at the precise moment a runner is passed. Due to this limitation, it is possible that runners begin to slow down before actually being passed when they first become aware of a male runner approaching. Alternatively, we cannot reject the hypothesis that women speed up to avoid being passed initially and then slow down more dramatically once they have been passed. Thus, we are estimating the net effect of any such within-split dynamics.

[12] For example, in 2008 and 2010, the 26th-fastest females in the 35-kilometer split had not been passed. However, in 2007, 2011, 2013, and 2014, the 26th-fastest female in 2007, 2011, 2013, and 2014 races was passed during the 35-kilometer split. Finally in 2009, the 26th fastest female runner had been passed in a previous split. In this setting, our rank-by-split fixed effect allows us to estimate whether the four runners passed in the 35-kilometer split changed paced differently then the two runners who had not yet been passed. The runner who had previously been passed does not contribute to our causal estimate because of the inclusion of a variable indicating whether runners had been passed in a previous split.

3.2 Model

Below, we formally consider whether deviations in runner i's pace systematically vary from the predicted pace in the interval in which she is first passed by male runners. Here we detail what is ultimately our preferred of two identified specifications, though we report variants when we present our empirical results. We define the unit of analysis as runner i in split s in year t, and—this is the important consideration—we identify off of variation within the same ordinal rank r and same split s, across years t. It is this within-rank estimator that restricts our identifying variation to that coming from within a group of competitors who are projecting the same return to performance at the same distance in the race when we observe them. In establishing a causal relationship between being passed and any resulting change in pace, our identifying assumption is therefore that after controlling for year and split-specific effects directly, the within-rank variation in pace across years is independent of average changes in pace. Specifically, we estimate

$$Pace_{isrt} = \beta_0 + \beta_1 \ \mathbb{1}(Passed_{srt}) + X_{sr}^{'}\Gamma + \gamma_{sr} + \delta_t + \epsilon_{isrt}, \tag{1}$$

where $Pace_{isrt}$ is the average pace at which runner i ran the "split" ending at $s \in [5, 10, ..., 40, 42]$. We allow $\hat{\beta}_1$ to absorb the average change in i's pace that is coincident with the first male runners passing i (in split s, while i is ranked r). That is, $Passed_{srt}$ is equal to one in the split s in which runner i is first passed, and equal to zero in all other splits.

We also control for split-by-rank effects, γ_{sr}, year-specific effects, δ_t, and, in $X_{sr}^{'}$, several runner attributes that may otherwise confound our estimate of $\hat{\beta}_1$.[13] In order to identify $\hat{\beta}_1$ cleanly (i.e., off of only the comparisons between runners caught in a given split and runners who had not yet been caught), among X'_{sr} we also absorb any difference in the performance of caught runners, subsequent to the split in which they were passed.[14] Unobservables are captured in ϵ_{isrt}, with all inference statements made with respect to the most-conservative standard-error estimates. Though statistical significance ends up not being sensitive to assumptions about the source of bias in standard-error estimates, in our preferred specifications we allow for clustering in split-by-rank cells. In the end, any systematic change in female-runner pace will thus be identified in $\hat{\beta}_1 \neq 0$.[15]

In identifying the causal parameter associated with being passed by the fastest male runners we assume that, conditional on observables, the five-kilometer split in which the female at rank r is passed by the fastest male is random with respect to the error in predicting her own split time. Important for our analysis is that the 30-minute head start afforded to elite female runners is sufficiently short that some female runners are passed by the fastest male runner. Elite-male runners are faster than their female counterparts across all distances. However, a large fraction of elite female runners—almost 40 percent in our sample—are not passed by even the fastest male runner. As these competitors are never "treated," they serve

[13] Runner attributes include quadratic age, whether the runner has been passed in an earlier split, is on team Nike, is an American citizen, and/or resides in the US.

[14] The pace of those passed continues to decline, though we hesitate to draw causal inference from the estimate as we lose identification in splits subsequent to the split in which the runner is first passed.

[15] As an example of the identifying variation, consider the seven runners (i.e., one from each year in our sample) who were in nineteenth place at the 35k splits. One of these seven had already been passed—we separately capture any difference in this runner's pace with a level shifter, as this runner is no longer able to contribute to cleanly identifying the causal parameter of interest. Of the other six, three were passed between 35k and 40k (i.e., $Passed_{srt} = 1$), while three were not (i.e., $Passed_{srt} = 0$). Any difference in the average pace of these two groups, then, identifies our variable of interest.

well to identify the effect of controls and to fit the average pace within each five-kilometer split.

In Table 1 we offer summary statistics for both the pooled sample of runners and separately by whether runners were ever passed. In a typical year, the first of the elite female runners is passed by the fastest male between the 20km and 25km split. Being passed earlier is mechanistically related to pace, which is reflected in those ever passed having an average pace over the entire marathon that is 11-percent slower than those who are never passed. In Figure 1, we show that passed runners are slower even in the first-five kilometers of the race, suggestive of an ability to absorb part of this in runner fixed effects.[16] To illustrate some of what our identifying variation affords us, in columns (4) and (5) we report the predictive power in these observables in explaining $Passed_{srt}$, with and without the split-by-rank strategy we propose above. In a pooled sample (Column 4), passed runners tend to be older, less likely to appear in multiple years in our sample, less likely to be on team Nike, and more likely to have American citizenship or reside in the United States. As we will not be directly controlling for individual runner fixed effects in our preferred specification, any such imbalance in observables would be suggestive of a threat to identification. However, when limiting the variation to that which will identify our relationship between runner pace in response to being passed in our preferred specification (Column 5), these differences collapse in magnitude and significance, leaving no evidence of imbalance.[17]

[16] Differences in the rate at which passed and un-passed runners slow down are small but present, however, and something we will address in our preferred specifications.

[17] Such an exercise is not informative in our other well-identified model—that with runner fixed effects—as covariates are not time varying.

4 Results

4.1 Main results

In Table 2 we report estimates of the effect of being passed in a given split on a female athlete's average pace in that split. In Column (1) we report specifications with only year fixed effects and in Column (2) we include year and split fixed effects. In both models, point estimates are large and negative—our first indication that average female pace may fall when passed by males. In these models, average pace in the five-kilometer split in which females are first passed is roughly between 7- and 10-percent lower relative to the mean pace of 15.68 km/hr.

However, to the extent the fastest males passed females who are slower or are slowing down differentially more within splits, the point estimates in columns (1) and (2) will be biased down. Adding controls for runners' ordinal ranks in the race in Column (3) cuts the point estimate and effect size considerably, suggesting that with ordinal rank we way well be absorbing the confounding effects that we anticipated imparting negative bias to naive models. Of course, if slower females are the same as those who are passed, it is possible that bias remains in our Column (3) estimate of β_1. We therefore introduce individual fixed effects in Column (4), where being passed by the fastest men lowers female pace by just less than one percent (at the mean), explaining roughly 20 percent of a standard deviation in female pace.

Before continuing, note again that "runner fixed effects" identify $\hat{\beta}_1$ across runner-specific observations that will vary in the expected returns to performance, as pace determines rank which determines prize money. Thus, it is in Column (5) that we arrive at what is our most-preferred specification, though the two approaches yield qualitatively similar conclusions.

As we discussed earlier, the identifying variation in Column (5) comes from differences in pace across similarly ranked runners across race-years, some who had been passed between split $s-1$ and split s and some who had not. That is, we include rank-by-split fixed effects. Note, however, that year fixed effects take on a slightly elevated role, here, as part of the variation in whether female runners are passed or not (even within specific ranks at specific distances) could arise from whether the marathon was particularly slow or fast in a given year.[18] In the end, our preferred identification strategy is most conservative, but tells a similar story to that of the runner fixed effects model—being passed by the fasted men leads to a 1.7-percent decline in average pace.

While the point estimates reported in Table 2 are contributed to by within-rank-and-split estimates of treatment (and designed to absorb into the error structure any ability variation that is picked up by rank), to the extent those passed early in the race are lower ability, the average effect reported in Column (5) may be hiding systematic differences across splits. In Table 3, we therefore relax the constraint that the effect be constant across the entire race, adopting our preferred specification, but stratifying by split. This reveals important variation in the influence of men on female performance.[19] To the extent that runners who are passed later in the race tend to be of higher ability, heterogenous treatment effects across distance could be suggestive of differential responses to the presence of males by the ability of females.

[18] Of course, to the extent *individual* runners have a bad day, they fall in ordinal rank and we exploit only the within-rank variation for identification.

[19] Note that we are unable to identify heterogeneity across distance using the model of Column (4) in Table 2, as identifying variation is coming from variation across distance for given runners.

Using data from *Jeopardy*, Jetter and Walker (2017) suggest that high-ability females are more-positively responsive to the presence of male competitors, also suggesting the gender gap may be ability dependent.[20] As anticipated, we see larger falloffs in performance among those passed earlier in the race, with point estimates becoming less negative throughout the race—even becoming positive in the five-kilometer split before the final 2km of the race. The heterogeneity we find would also be consistent with the focus on finishing the NYC Marathon being increasingly salient toward the end of the race. Though distance fixed effects capture any such general tendency, female competitors are possibly able to realize increasing robustness to being passed as the race unfolds—ultimately, enough to yield what appears to be encouraging boosts in performance when passed late in the race.

In Figure 2, we plot the average number of men who pass females in the split in which the female was first passed. Male runners naturally tend to spread out throughout the race. As such, women who are passed in the 15km-20km split will tend to be passed by more men (roughly 24, on average) than those passed in the 35km-40km split (6, on average). In considering ability-related treatment-effect heterogeneity, then, we cannot rule out that the difference in point estimates across distance reflects the effect of treatment intensity. As point estimates in Table 3 do not separately identify whether it is that faster runners are less responsive or that slower runners are being passed more often, we offer a measure of "treatment per passes." Given the monotonicity presented in Figure 2, this reflects a similar

[20] Jetter and Walker (2017) finds that the gender gap in competitive behavior disappears when women are placed in competition with only men. To control for potential selection into competition with two males, they employ an individual fixed-effects specification using only women who competed against two men during one show and against at least one woman in another. As "only the player with the highest dollar score will receive that value in the form of cash and be eligible to compete in the subsequent episode," restricting the sample in this way may identify the effect of gender composition on only higher ability females.

pattern of decline—less-able runners are responding to being passed by male runners.[21]

One might imagine several mechanisms at play in the data-generating process we describe, relating to the psychology of competitiveness and the performance of elite-level female competitors. For example, one might assume roles for pride, shame, or feedback aversion (Weiner, 1985; Tiggemann and Boundy, 2008) in explaining the systematic declines in performance we observe.[22] To the extent any of these are operational, we note that their effect cannot be uniform across ability, as the performance of the very best among the elite competitors we describe is invariant to the influence of being passed.

[21] We explored one additional source of potential heterogeneity, inquiring into whether there is evidence that runners respond differently when passed by a male runner who may be in her peer group. The phycology literature suggests there is significant potential for a differential response. For example, Christenfeld et al. (1997) documents differential physiological responses to evaluation by friends or strangers and (Ravaja et al., 2006) demonstrates that people playing video games with friends are significantly more engaged and had higher levels of anticipated threat than people who played with strangers. In Appendix Table A1, we allow the effect of being passed by a male runner to vary by three such measures: if the passing runner is a teammate, a citizen of the same country, or lives in the same country. If we have controlled well for runner ability, team affiliation should not matter, as marathon "teams" are more an indication of common sponsorship than anything more substantive, and no expectation that teammates would know each other or train together. For example, the 32 female Team Nike runners in our data reside and train in nine different countries. Similarly, the second most common team affiliation, Adidas, fields 14 female runners who reside in eight different countries. Allowing different parameter estimates for those who share the same citizenship or same country of residence as the fastest male, we have limited variation to exploit. Only 40 runners (over all years) are passed by a male with whom they share a nationality and only 44 runners are passed by a male who lives in the same country. (39 runners of these are the same runners.) Moreover, a strong majority of the same-nationality (78 percent) and same-residence (86 percent) cases come from runners in the United States. While we find suggestive evidence that familiarity with the passing runner causes a larger decrease in performance, the limited variation available prevents us from attributing a causal interpretation to this result.

[22] In this context, feedback aversion implies female runners may change their pace to avoid the additional negative feedback of trying and failing to keep up with a male runner who has recently passed them. For example, (Zeelenberg, 1999) find that decision making responds to whether or not individuals expect feedback, and more if the feedback is expected to be negative. Fellner and Sutter (2009) shows that less frequent feedback can influence investment decisions through myopic loss aversion. Ellingsen and Johannesson (2008) shows that the potential for feedback that elicits pride or shame can motivate changes in behavior.

4.2 What we identify, and what we do not identify

In order to identify an effect of males on female performance, the ideal experiment randomizes the arrival of males into the competitive environment in which females are competing. In such an experiment, the performance of those who experience the arrival of men would then be compared to the performance of those who don't. We have argued that we have two approximations of that ideal experiment, in both cases demonstrating systematic variation in female performance associated with the quasi-random arrival of men: 1) comparing a runner's pace in the split in which she is passed to her pace in other splits, and, 2) comparing the pace of just-passed runners' within rank-by-split cells, to runners who were in the same rank-by-split cells (i.e., in the same rank at the same distance in their race, but in a different year) and had not just been passed.

There are two additional parameters that would be of interest, neither of which are identified. First, we are unable to identify the effect of these elite competitors being passed by other female runners. As a matter of course, these athletes are best responding to each other, and we have no exogenous variation off of which to identify any individual runner's responsiveness to another. Thus, our inference statements should not be extended to us arguing that males influence females *differentially* more or less than females might influence other female runners. That said, we will suggest that any systematic differences in female performance attributable to the exogenous arrival of men should be thought of as over and above those mutually determined peer effects being determined within the normal running of the NYC Marathon.

Second, as there is no available variation allowing one to consider the effect of females on male performance, we are not able to speak to the general equilibrium effects of mixed-gender competition. Thus, our inference statements should not be extended to us arguing that males influence females more or less than females would influence male runners. In supplemental analysis, we find that the average pace of male runners increases in the 5km split prior to first passing a female runner, and decreases in the 5km split that follows their first passing of a female runner—the *ex post* decrease is evidently three times the magnitude of the *ex ante* increase, suggestive of some effort being expensed in that passing. We maintain that this is descriptive, however, and without clean variation affording the identification of a causal parameter we choose not to produce these results.

Third, we are unable to speak to the role of elite status directly, though we imagine that we have likely identified a lower-bound on the effect one might identify in a sample of less-competitive women.

5 Conclusion

We exploit an opportunity presented in the New York City Marathon to consider the potential influence of men on an elite group of female competitors. For identification, we exploit the initial passing of these female runners by the fastest male runners, with such passings creating the rare opportunity to change nothing about the anticipated rewards to female performance or the returns to effort in the competitive environment, which removes within-tournament mechanisms or incentives as potential confounders in explaining how female runners systematically respond to being passed by men.

We offer two alternative specifications to identify the parameter of interest, each supporting similar inference. In a runner fixed-effect specification, we find that the pace of the average elite female falls a modest 1.9 percent relative to the mean when overtaken by the fastest man. Our preferred specification identifies off of (across year and distance in the race) variation in whether equally ranked females had experienced being passed by the fastest men—in some years a runner of a given rank had been overtaken by men by that distance and in other years she had not yet been. This suggests an average decline of 1.7 percent. However, these average declines in performance hide significant heterogeneity revealed by considering runner ability. Specifically, declines in performance are concentrated in lower-ability runners. This is suggestive that gains associated with creating environments in which the highest-ability women choose to compete are less likely to be dissipated directly through the influence of men. However, we also learn that any elimination of direct, competitive pressure from men would be an incomplete solution to narrowing gender differentials in competitive outcomes, as the indirect effect itself is significant in its influence on female performance, outside of those at the very highest levels of ability. With performance reductions evident even within a highly competitive environment, one might worry all the more among more-broadly representative competitive environments.

References

Anbarci, Nejat, Jungmin Lee, and Aydogan Ulker, "Win at All Costs or Lose Gracefully in High-Stakes Competition? Gender Differences in Professional Tennis," *Journal of Sports Economics*, 2014, *17* (4), 323–353.

Andersen, Steffen, Seda Ertac, Uri Gneezy, John A List, and Sandra Maximiano, "Gender, Competitiveness, and Socialization at a Young Age: Evidence From a Matrilineal and a Patriarchal Society," *Review of Economics and Statistics*, 2013, *95* (4), 1438–1443.

Azmat, Ghazala and Barbara Petrongolo, "Gender and the Labor Market: What Have we Learned from Field and Lab Experiments?," *Labour Economics*, 2014, *30*, 32–40.

Bjorvatn, Kjetil, Ranveig Falch, and Ulrikke Hernæs, "Gender, Context and Competition: Experimental Evidence from Rural and Urban Uganda," *Journal of Behavioral & Experimental Economics*, 2016, *61*, 31–37.

Booth, Alison L and Patrick Nolen, "Gender Differences in Risk Behaviour: Does Nurture Matter?," *The Economic Journal*, 2012, *122* (558).

Booth, Alison, Lina Cardona-Sosa, and Patrick Nolen, "Gender Differences in Risk Aversion: Do Single-Sex Environments Affect Their Development?," *Journal of Economic Behavior & Organization*, 2014, *99*, 126–154.

Buser, Thomas, "The Impact of the Menstrual Cycle and Hormonal Contraceptives on Competitiveness," *Journal of Economic Behavior & Organization*, 2012, *83* (1), 1–10.

_ , Muriel Niederle, and Hessel Oosterbeek, "Gender, Competitiveness, and Career Choices," *The Quarterly Journal of Economics*, 2014, *1409*, 1447.

Campbell, Anne, *A Mind of Her Own: The Evolutionary Psychology of Women*, Oxford University Press, 2002.

Cárdenas, Juan-Camilo, Anna Dreber, Emma Von Essen, and Eva Ranehill, "Gender Differences in Competitiveness and Risk Taking: Comparing Children in Colombia and Sweden," *Journal of Economic Behavior & Organization*, 2012, *83* (1), 11–23.

Christenfeld, Nicholas, William Gerin, Wolfgang Linden, Mara Sanders, Jennifer Mathur, James D Deich, and Thomas G Pickering, "Social Support Effects on Cardiovascular Reactivity: Is a Stranger as Effective as a Friend?," *Psychosomatic medicine*, 1997, *59* (4), 388–398.

Cotton, Christopher, Frank McIntyre, and Joseph Price, "Gender Differences in Repeated Competition: Evidence from School Math Contests," *Journal of Economic Behavior & Organization*, 2013, *86*, 52–66.

Deaner, Robert O, "Distance Running as an Ideal Domain for Showing a Sex Difference in Competitiveness," *Archives of Sexual Behavior*, 2013, *42* (3), 413–428.

Deci, Edward L, Gregory Betley, James Kahle, Linda Abrams, and Joseph Porac, "When Trying to Win: Competition and Intrinsic Motivation," *Personality and Social Psychology Bulletin*, 1981, *7* (1), 79–83.

Ellingsen, Tore and Magnus Johannesson, "Anticipated Verbal Feedback Induces Altruistic Behavior," *Evolution and Human Behavior*, 2008, *29* (2), 100–105.

Ellis, Yvonne, Bobbie Daniels, and Andres Jauregui, "The Effect of Multitasking on the Grade Performance of Business Students," *Research in Higher Education Journal*, 2010, *8*, 1.

Fellner, Gerlinde and Matthias Sutter, "Causes, Consequences, and Cures of Myopic Loss Aversion—An Experimental Investigation," *The Economic Journal*, 2009, *119* (537), 900–916.

Frick, Bernd, "Gender Differences in Competitive Orientations: Empirical Evidence from Ultramarathon Running," *Journal of Sports Economics*, 2011, *12* (3), 317–340.

Gneezy, Uri and Aldo Rustichini, "Gender and Competition at a Young Age," *The American Economic Review Papers and Proceedings*, 2004, *94* (2), 377–381.

_ , Muriel Niederle, and Aldo Rustichini, "Performance in Competitive Environments: Gender Differences," *Quarterly Journal of Economics*, 2003, *118* (3), 1049–1074.

Gong, Binglin and Chun-Lei Yang, "Gender Differences in Risk Attitudes: Field Experiments on the Matrilineal Mosuo and the Patriarchal Yi," *Journal of Economic Behavior & Organization*, 2012, *83* (1), 59–65.

Grund, Christian, "Gender Pay Gaps Among Highly Educated Professionals—Compensation Components do Matter," *Labour Economics*, 2015, *34*, 118–126.

Hancock, Peter A, Mary Lesch, and Lisa Simmons, "The Distraction Effects of Phone Use During a Crucial Driving Maneuver," *Accident Analysis & Prevention*, 2003, *35* (4), 501–514.

Horberry, Tim, Janet Anderson, Michael A Regan, Thomas J Triggs, and John Brown, "Driver Distraction: The Effects of Concurrent In-Vehicle Tasks, Road Environment Complexity and Age on Driving Performance," *Accident Analysis & Prevention*, 2006, *38* (1), 185–191.

Jetter, Michael and Jay K Walker, "The Gender of Opponents: Explaining Gender Differences in Performance and Risk-Taking?," *European Economic Review*, 2017.

Khachatryan, Karen, Anna Dreber, Emma Von Essen, and Eva Ranehill, "Gender and Preferences at a Young Age: Evidence from Armenia," *Journal of Economic Behavior & Organization*, 2015, *118*, 318–332.

Larson, Lia Christine, *When Girls Stop Competing Against Boys: An Experimental Analysis of the Competitive Behavior of Young Children*, Harvard University, 2005.

Lesch, Mary F and Peter A Hancock, "Driving Performance During Concurrent Cell-Phone Use: Are Drivers Aware of Their Performance Decrements?," *Accident Analysis & Prevention*, 2004, *36* (3), 471–480.

Lindquist, Gabriella Sjögren and Jenny Säve-Söderbergh, ""Girls will be Girls", especially among Boys: Risk-Taking in the "Daily Double" on Jeopardy," *Economics Letters*, 2011, *112* (2), 158–160.

Lorge, Abigail, "For Some Women, Elite Start Could Mean a Lonely Race," *The New York Times*, November 2010.

Meyer, John P and Ian R Gellatly, "Perceived Performance Norm as a Mediator in the Effect of Assigned Goal on Personal Goal and Task Performance," *Journal of Applied Psychology*, 1988, *73* (3), 410.

Mitchell, Terence R, Miriam Rothman, and Robert C Liden, "Effects of Normative Information on Task Performance.," *Journal of Applied Psychology*, 1985, *70* (1), 48.

Niederle, Muriel and Lise Vesterlund, "Do Women Shy Away from Competition? Do Men Compete too Much?," *The Quarterly Journal of Economics*, 2007, *122* (3), 1067–1101.

Paola, Maria De, Francesca Gioia, and Vincenzo Scoppa, "Are Females Scared of Competing With Males? Results From a Field Experiment," *Economics of Education Review*, 2015, *48*, 117–128.

Price, Joseph, "Gender Differences in the Response to Competition," *Industrial & Labor Relations Review*, 2008, *61* (3), 320–333.

Ravaja, Niklas, Timo Saari, Marko Turpeinen, Jari Laarni, Mikko Salminen, and Matias Kivikangas, "Spatial Presence and Emotions During Video Game Playing: Does it Matter With Whom You Play?," *Presence: Teleoperators and Virtual Environments*, 2006, *15* (4), 381–392.

Roberts, Tomi-Ann and Susan Nolen-Hoeksema, "Sex Differences in Reactions to Evaluative Feedback," *Sex Roles*, 1989, *21* (11-12), 725–747.

Samak, Anya C, "Is There a Gender Gap in Preschoolers' Competitiveness? An Experiment in the US," *Journal of Economic Behavior & Organization*, 2013, *92*, 22–31.

Säve-Söderbergh, Jenny and Gabriella Sjögren Lindquist, "Children Do Not Behave Like Adults: Gender Gaps In Performance And Risk Taking In A Random Social Context In The High-Stakes Game Shows Jeopardy and Junior Jeopardy," *The Economic Journal*, 2017, *127* (603), 1665–1692.

Tiggemann, Marika and Michelle Boundy, "Effect of Environment and Appearance Compliment on College Women's Self-Objectification, Mood, Body Shame, and Cognitive Performance," *Psychology of Women Quarterly*, 2008, *32* (4), 399–405.

van Veldhuizen Roel, "Gender Differences In Tournament Choices: Risk Preferences, Overconfidence or Competitiveness?," 2017. Rationality and Competition Discussion Paper Series 14.

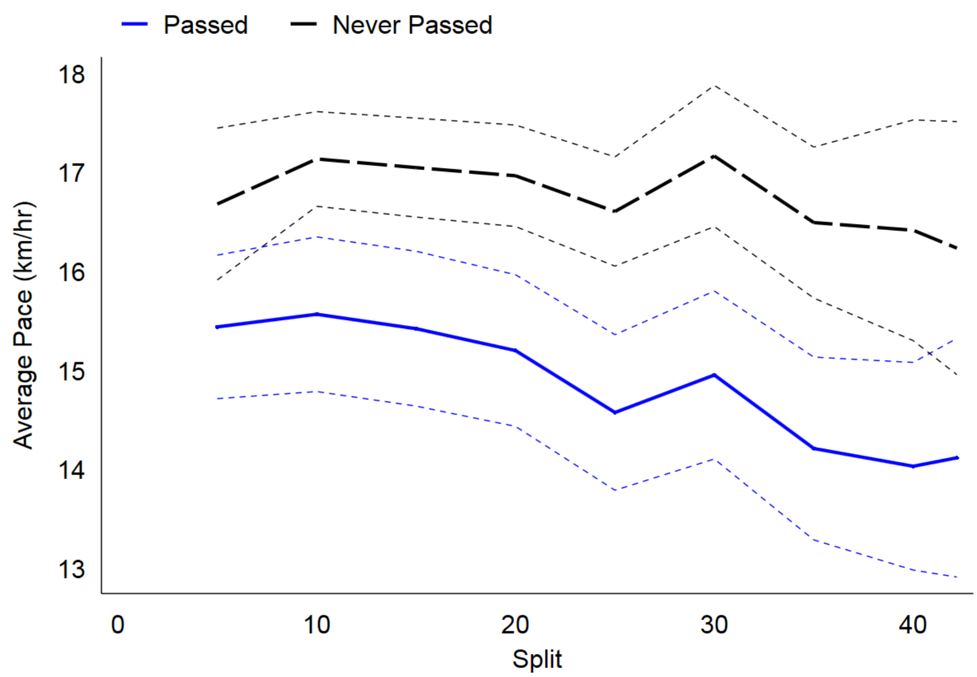
Weiner, Bernard, "An Attributional Theory of Achievement Motivation and Emotion," *Psychological Review*, 1985, *92* (4), 548.

Woodman, Tim and Lew Hardy, "The Relative Impact of Cognitive Anxiety and Self-Confidence Upon Sport Performance: A Meta-Analysis," *Journal of Sports Sciences*, 2003, *21* (6), 443–457.

Wozniak, David, William T Harbaugh, and Ulrich Mayr, "The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices," *Journal of Labor Economics*, 2014, *32* (1), 161–198.

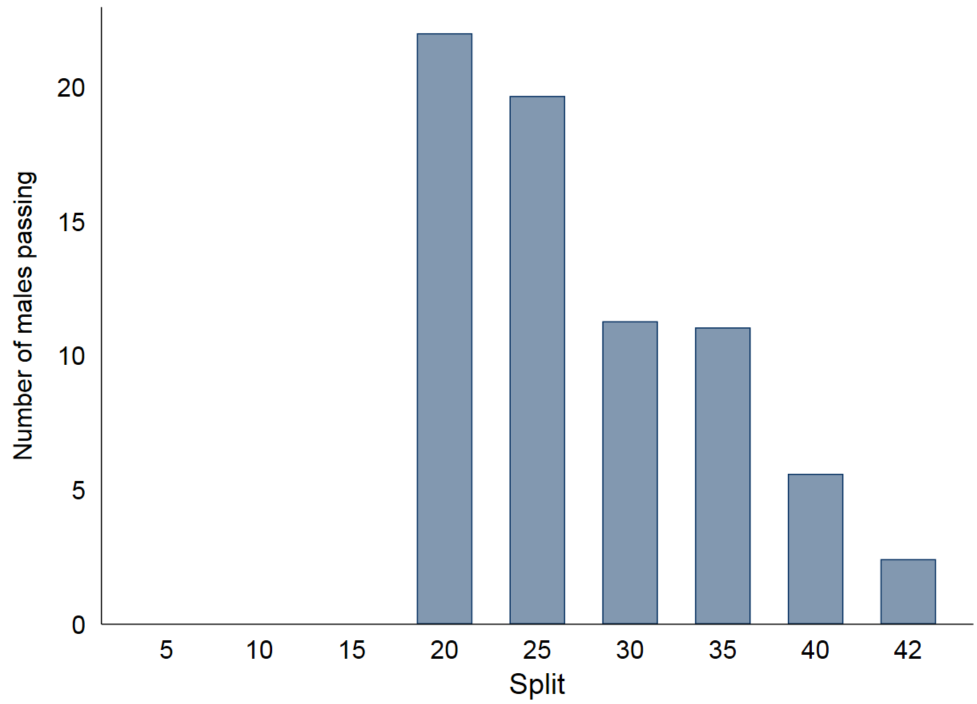
Zeelenberg, Marcel, "Anticipated Regret, Expected Feedback and Behavioral Decision Making," *Journal of Behavioral Decision Making*, 1999, *12* (2), 93.

Figure 1
Average pace, by split (km)



Notes: Data represent the average pace of runners in each 5km split. Dashed lines indicate one standard deviation above and below the mean values.

Figure 2
How many men passed the average female runner, by the split (km) in which she was passed



Notes: Data represent the average number of male runners who passed an elite female runner in all of the 5km splits of the New York City (NYC) marathons from 2007 through 2014.

Table 1
Summary statistics

		Stratified samples		$Passed_{srt} = 1$	
	Pooled sample	Ever passed	Never passed	OLS with covariates	Identifying variation only
	(1)	(2)	(3)	(4)	(5)
Pace	15.68	14.84	16.75	-1.914***	-0.122
	(1.36)	(1.05)	(0.84)	(0.043)	(0.097)
Age	32.53	33.69	31.05	2.638***	-0.072
	(6.56)	(7.52)	(4.67)	(0.286)	(0.728)
Appearances	1.70	1.52	1.93	-0.407***	0.064
	(1.00)	(0.90)	(1.07)	(0.044)	(0.114)
Team Nike	0.14	0.02	0.30	-0.248***	-0.026
	(0.35)	(0.12)	(0.46)	(0.014)	(0.022)
American	0.43	0.57	0.25	0.320***	0.031
	(0.50)	(0.50)	(0.43)	(0.021)	(0.062)
Lives in USA	0.56	0.68	0.40	0.280***	0.032
	(0.50)	(0.47)	(0.49)	(0.021)	(0.054)
Year FE				No	Yes
Split×Rank FE				No	Yes
Observations	2,052	1,152	900	2,052	2,052
Unique runners	228	128	100	228	228

Notes: Observations are at the level of split-runner. In (1)–(3) we report standard deviations in parentheses, while in (4) and (5) we report standard errors. In (4), we report the results of univariate regressions on the complete sample, each with a variable indicating whether the runner was passed for the first time by a male runner in the given split as the only explanatory variable. In (5), we replicate our preferred specification by also including year and rank-by-split fixed effects, suggestive of our strategy eliminating significant differences that may otherwise confound.

Table 2
Average responsiveness of female pace (km/h) when passed by the fastest males

	(1)	(2)	(3)	(4)	(5)
				Preferred specifications	
Passed	-1.438***	-1.317***	-0.463***	-0.301***	-0.266**
	(0.120)	(0.116)	(0.105)	(0.066)	(0.105)
Year FE	Yes	Yes	Yes	Yes	Yes
Split FE		Yes	Yes	Yes	Yes
Rank FE			Yes	Yes	Yes
Runner FE				Yes	
Split×Rank FE					Yes
Observations	2,052	2,052	2,052	2,052	2,052
Unique runners	228	228	228	228	228
Mean	15.68	15.68	15.68	15.68	15.68
% change	-9.17	-8.40	-2.96	-1.92	-1.70
Effect size	1.06	0.97	0.34	0.22	0.20

Notes: Observations ($n = 2,052$) are in units of years-by-runners-by-splits. Reported coefficients are estimates of the effect of being passed on the average pace between split $s-1$ and split s, where $s \in \{5, 10, .., 40, 42\}$. In all specifications, controls for age, age squared, having been passed in an earlier split, and whether the runner is on team Nike, an American citizen, and/or living in the United States are included but not reported. Robust standard errors are reported in parentheses, allowing for clustering at the level of identifying variation. Percent impacts are relative to the unconditional mean pace. *** significant at 1%; ** significant at 5%; * significant at 10%.

Table 3
The differential responsiveness of female pace (km/h), by distance

	⇐ "ability" decreasing			"ability" increasing⇒		
	(1)	(2)	(3)	(4)	(5)	(6)
	15k-20k	20k-25k	25k-30k	30k-35k	35k-40k	40k-42k
Passed	-2.567***	-0.982***	-0.507*	-0.493***	0.114	0.074
	(0.082)	(0.304)	(0.282)	(0.169)	(0.138)	(0.271)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Split×Rank FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	228	228	228	228	228	228
Mean	15.98	15.47	15.92	15.21	15.08	15.05
% change	-16.08	-6.35	-3.19	-3.24	0.75	0.49
Effect size	2.33	0.80	0.37	0.35	0.07	0.05
Males passing	22.00	19.67	11.28	11.05	5.60	2.42
Effect / pass	-0.12	-0.05	-0.04	-0.04	0.02	0.03

Notes: Observations ($n = 228$) are units of years by runners. Reported coefficients are estimates of the effect of being passed on the average pace between split $s-1$ and split s, where $s \in \{5, 10, .., 40, 42\}$. Controls for age, age squared, having been passed in an earlier split, and whether the runner is on team Nike, an American citizen, and/or living in the United States are included but not reported. Robust standard errors are reported in parentheses, allowing for clustering in split-by-rank bins. Percent impacts are relative to the unconditional mean pace. *** significant at 1%; ** significant at 5%; * significant at 10%.

6 Appendix

Table A1
The differential responsiveness of female pace (km/h), by "closeness"

	(0)	(1)	(2)	(3)
Passed	-0.266***	-0.247**	-0.116	-0.045
	(0.105)	(0.115)	(0.130)	(0.126)
× Same team		-0.215		
		(0.198)		
× Same citizenship			-0.430***	
			(0.157)	
× Same residence				-0.571***
				(0.141)
Year FE	Yes	Yes	Yes	Yes
Split×Rank FE	Yes	Yes	Yes	Yes
Observations	2,052	2,052	2,052	2,052
Unique runners	228	228	228	228
Mean	15.68	15.68	15.68	15.68
% change (Same=0)	-1.70	-1.58	-0.74	-0.29
Effect size (Same=0)	0.20	0.18	0.09	0.03
% change (Same=1)		-2.95	-3.48	-3.93
Effect size (Same=1)		0.34	0.41	0.45

Notes: Observations are units of years by runners by splits. Reported coefficients are estimates of the effect of being passed on the average pace between split $s-1$ and split s, where $s \in \{5, 10, .., 40, 42\}$. Controls age, age squared, having been passed in an earlier split, and whether the runner is on team Nike, an American citizen, and/or living in the United States are included but not reported except team Nike is excluded as a control in (1), American citizen is excluded in (2), and living in the United Stated is excluded in (3). Robust standard errors are reported in parentheses, allowing for clustering in split-by-rank bins. Percent impacts are relative to the unconditional mean pace. *** significant at 1%; ** significant at 5%; * significant at 10%.