Working Paper in Economics No. 694

How long do you think it will take? Field Evidence on Gender Differences in Time Optimism

Mitesh Kataria

Department of Economics, February 2017

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

How long do you think it will take?
Field Evidence on Gender Differences in Time Optimism

Mitesh Kataria*

Department of Economics, University of Gothenburg, Box 640, SE-40530 Gothenburg, Sweden. mitesh.kataria@economics.gu.se

Abstract

Evidence from ten natural field studies comparing long-distance runners' incentivized predictions of race finishing time with their actual finishing time is reported. A modest but regular bias is found. Male runners are consistently found to be more time optimistic than female runners and finish slower than they predict to finish. Males are found to over-appreciate their physical fitness. To the extent this behaviour carries over to other contexts, such as the labor market, the tendency of men to overestimate their capacity could lead to distorted self-appraisals and give them advantages in terms of higher salaries and better positions.

JEL classification: C93, D01, D03

Keywords: Overconfidence, Time optimism, Gender differences

[0] I am most grateful for comments on early versions of this paper from Anna Dreber, Christina Gravert, Wolfgang Habla, Martin Kocher, Åsa Löfgren, Andreea Mitrut, Katarina Nordblom, Joseph Vecci and Conny Wollbrant. Any errors that remain are my sole responsibility.

1 Introduction

How long will it take? This is a very common question. It could be parents trying to schedule their children's weekly activities or a couple scheduling a lunch date. At work it could be a project manager trying to estimate the time it will take to complete a project. The project could be small-scale and private or large-scale and public. One example of the latter is the construction of Berlin Brandenburg Airport which began in 2006 after 15 years of planning. The airport was originally planned to open in 2010 but ran several delays and still has not been finished. Or suppose a manager deciding whom to recruit of the two candidates Anna and Bob. Bob is more confident about his productivity than Anna and claims to have a lower task completion time. Assuming that the candidates are honest, can we believe in their self-appraisal? More generally, can people be trusted to accurately predict their task completion times? Or are people likely to give optimistic predictions? And if so, do men provide more optimistic predictions than females? These are the questions that are addressed in this study.

The phenomenon we explore is known as overconfidence. The study of overconfidence originated in psychology and the trait expresses more strongly in relation to difficult tasks, tasks with low predictability, and for tasks that lack fast and clear feedback (Fischhoff et al., 1977; Griffin and Tversky, 1992). Many studies on the topic are based on survey data where overconfidence is measured by asking respondents to rank themselves compared with others. A common finding is that a majority of people overestimate their skills compared to the median individual for tasks such as driving a car which is known as overplacement and is an expression of overconfidence (e.g. Svenson, 1981). Another stream of studies uses objective measures to reach their conclusion. For example, Buehler et al. (1994) found that students expect to finish a task faster than their objectively measured completion time, which is known as overoptimism. Yet another stream of literature looks at whether people overestimate the precision of their knowledge (Fischhoff et al., 1977). Hence, the term overconfidence has a broad meaning referring to overplacement, overoptimism (e.g. time optimism) as well as overprecision. Overoptimism can either apply to own completion time, others' completion times or both. The focus in this study is on overoptimism regarding own completion time which we also refer to as time optimism, meaning that people overestimate their productivity. Time optimism is an issue that have not gained much attention by economists.[1] Biases in beliefs are, however, central to the understanding of behavior which in turn can have wider economic implications. For example, using a a mechanism design approach Palomino and Peyrache (2010) show that a gender gap in wages can emerge because of differences in self-confidence even if men and women are equally productive since overconfidence makes men to give more effort. Another reason for the gender gap in wages could be that women negotiate less frequently than men and when they do negotiate they ask for less money (Babcock and Laschever, 2009).

The interest of economist in understanding overconfidence was recently captured by a symposium in Journal of Economic Perspectives (Malmendier and Taylor, 2015).[2] An important issue in economics is whether biases that are found in individual decision-making are persistent or if they tend to vanish with increasing experience. The model of Gervais and Odean (2001) predicts that overconfidence decreases with financial-market experience and is corroborated by the experimental evidence of Menkhoff et al. (2013). While one could expect the overconfidence among rational Bayesian decision makers to disappear with sufficient experience (e.g. Van den Steen, 2004) there are also circumstances to the contrary where overconfidence in estimation precision increases with experience (Van den Steen, 2011). Buehler et al. (1994) found that decision-makers do not even consider past history when assessing their abilities. Hence, while overconfidence is claimed to be a robust finding the question of its existence among experienced people is not fully settled. As we will come back to, the data used in the present study provides a rare opportunity to learn about the behaviour of people with extraordinary access to immediate, precise, and objective feedback on their abilities. To find situations where people have access to better feedback in a natural context is, if not impossible, at least very hard.

The economic literature has shown that investors in general and males in particular tend to trade more than what is rationally granted, which in turn

[1] In psychology the phenomena have gained much more attention and relates to the literature about the planning fallacy.

[2] Grubb (2015) address how consumer overconfidence alters market outcomes. Daniel and Hirshleifer (2015) discuss the importance of understanding overconfidence in relation to financial decision-making and financial markets. Malmendier and Tate (2015) provides a framework to structure the literature on overconfidence of CEOs.

has been interpreted as a gender difference in overconfidence (e.g. Barber and Odean, 2001). Notably, however, the evidence found in the economic literature on overconfidence is not always based on economic decisions. For example, Bengtsson et al. (2005) found that males, especially young males, are more overconfident than females about their exam results, and Nekby et al. (2008) found evidence on overconfidence using data from a large 10 km running race in Sweden. Using a laboratory experiment where the participants solved math problems, Niederle and Vesterlund (2007) show that overconfidence is a key factor in explaining gender differences in willingness to compete. One reason that researchers take interest in evidence beyond economic markets is that overconfidence is considered a psychological trait that is relatively robust across situations (Borghans et al., 2008).[3] This does not mean that context is negligible in understanding overconfidence, e.g., Beyer and Bowden (1997) found significant gender differences in overconfidence in tasks that can traditionally be considered as male dominant such as knowledge in football, baseball, and basketball, but not in tasks considered to be traditionally female or gender-neutral.

The main aim of the present paper is to investigate time optimism and gender differences in time optimism. This is accomplished by comparing actual and predicted race times among Swedish long-distance runners. Sweden is known to do fairly well in gender equality compared with many other countries (Bekhouche et al., 2013), which makes it an interesting country for a study on gender differences. Participating in a long-distance race means months of extensive training.[4] Long-distance runners commonly measure the time, distance, and speed of their workouts to get feedback on their physical capacity which in turn helps them optimize their race pacing.[5] Weekly running distances for non-elite male marathon runners is estimated to 30-60 km while 25 percent run more than 60 km. Moreover, most non-elite runners who participate in a competition are not competing for the first time (Hölmich et al., 1989). Will the runners provide unbiased prediction

[3]Using a classic twin design Cesarini et al. (2009) estimate the genetic and environmental contributions to individual differences in overconfidence. The reported point estimates show that genes accounted for roughly three times the percentage of the variance in the overconfidence scores as compared to environment.

[4]While shorter races could attract a few participants without training experience, this would be very unlikely in long-distance (21.1, 30 and 42.2 km) races we use in this study.

[5]Optimal pacing means to run the race at a speed that minimizes the race finishing time.

or will they predict that they will run faster than they actually end up doing? Do males typically give more time optimistic predictions than females? These are the questions we wish to address. To this end a unique data set is used that is collected in the participants natural environment, consisting of experienced participants with extraordinary access to immediate, precise, and objective feedback on their abilities. Notably, the participants of this study are not claimed to be representative to the general population. To the contrary, they have access to extraordinary feedback and it is by observing behaviour in these extreme circumstances that offers an opportunity to learn whether the bias is sufficiently persistent. If runners are found overly optimistic about their abilities one can reasonable expect that people that do not have access to extraordinary feedback to be even more biased.

The most related papers to presented paper are Nekby et al. (2008) and Krawczyk and Wilamowski (2016). Nekby et al. (2008) use data from a large 10 km running race in Sweden to test for gender differences in self-selecting to start groups based predicted race times. Women were found more likely to self-select to faster starting groups than warranted by their performance compared to males. As discussed by the authors, whether this behavior can be explained by overconfidence or over-competitiveness cannot be concluded from their data. Strategic advantages of being in an early start group could potentially affects their results. Krawczyk and Wilamowski (2016) utilized data on 345 participants (20 percent women) of Warsaw Marathon to study overconfidence and found men to be more overconfident. In an extension they use mid-point split-time as an indirect proxy measure for overconfidence to test for determinants of overconfidence.

There are several advantages of the research presented in this paper compared with other studies. First of all, we make use of a unique data set. The data is collected in the participants natural environment and the participants have extraordinary access to feedback on their abilities. This allows us to assess test the persistence of overconfidence. Moreover, the data is collected as part of a marketing campaign without any ambitions to use the data for scientific purposes. This could potentially help to reduce the problem of demand artifacts (Zizzo, 2010). Second, instead of relying on only one running race we use data from ten samples consisting of near 5000 observations. The samples consist of data from races arranged in different years and with different subject pools. This allows us to assess the robustness of our results.

Third, a direct measure of overconfidence is used to capture gender differences and hence we do not confound our analysis with proxy variables. Using proxy variables would make it potentially harder to distinguish whether any gender differences are caused by physiological or behavioral differences.[6] Finally, the the runners' predictions were incentivized.[7] While most of the experimental studies of overconfidence in psychology do not provide monetary incentives for truthful responses, Cesarini et al. (2006) found that the level of measured overconfidence was largely reduced when using monetary incentives in several confidence interval tasks.

The main result of this study is that men are found more time optimistic compared to women. A positive gender difference is found in all samples and is significant in most of the samples. Moreover, a pooled analysis that includes the influence of weather confirms the existence of time optimism as a general phenomena among runners.

The remainder of the paper is organized as follows. In the next section the conceptual framework is introduced. After detailing the field studies (Section 3), we present the empirical findings (Section 4). We summarize the main points of our study and offer concluding remarks in Section 5.

2 Conceptual Framework and Identification

Two types of information are important for successful predictions, singular and distributional information (Kahneman and Tversky, 1977). Singular information consists of distinguishing knowledge about the particular case under consideration. For runners, singular information may include an as-

[6]Deaner et al. (2015) study gender differences in pacing using data from 14 marathons in the U.S. While male runners are found to be more likely to slow down in pace compared to female runners, several reasons are mentioned for this behaviour. It could reflect physiological differences, such as men's greater susceptibility to muscle glycogen depletion, or women's skeletal muscle which is shown to suffer from lesser fatigability. The gender difference in pacing may also reflect differences in decision making. As discussed by Hayner and Mercer (2011) many marathon runners simply want to complete the event and for them an increased variability of pace would be optimal. Others might be willing to take a risk of running too fast in the beginning in an attempt to break some personal record. For these reasons using split time as a proxy for overconfidence could be problematic.

[7]Incentivizing could also make things worse as runners could slow down to finish on the predicted time. We do not believe this to be a problem in practice. As suggested by Allen et al. (2016) "for the overwhelming majority of runners, race times are a source of internal pride and fulfillment and not an extrinsic reward."

sessment the particular race compares with other races.[8] Distributional information may include statistics of the distribution of outcomes in similar situations. It could be based on previous experience of similar tasks, e.g. times in comparable races and training sessions. Runners collect distributional information primarily during their training using either a simple stopwatch that only measures time or using more advanced gadget that measures both time and distance.

What could lead to unsuccessful predictions and time optimism? Suppose a marathon-runner who completed 18 training sessions in the month leading up to the race. Fourteen of these were ran under ideal conditions, i.e., with perfect weather and after eating the right food and getting good sleep. The remaining four were run under non-ideal conditions e.g., bad weather or an aching knee, yielding slower running times. A runner could ignore the undesirable outcomes and only consider the faster running times as representative of his or her capacity. Hence, runners may truncate the right tail of the distribution of running time and hence reduce the expected race time. According to Taylor and Brown (1988) people exaggerate their own abilities because it improves their mental health. This is of course only one possible reason why some runners are overoptimistic about their race times. Using a think-aloud protocol Buehler et al. (1994) found that decision-makers focused primarily on future scenarios when predicting their completion times.

There is also increasing evidence of gender differences in competitiveness in distance running (for a review see Deaner et al., 2015). Although most non-elite runners run for affiliation, health and life-meaning, important gender differences have been noted. While male runners are more goal-oriented and motivated by competition, female runners are motivated more by the affiliation running offers them, and to use running as strategy to cope with stress and control their body weight (Deaner et al., 2015). This could make males more likely to suppress distributional data that conflicts with their desired race time, which in turn can potentially result in a gender difference in time optimism. Moreover, males could have some biological advantages in running (Cheuvront et al., 2005). Therefore, to be as good as male runners, female runners could need to compensate with increased training. More

[8]Details of the race course and elevation is usually available on the web-page of the race.

training means more feedback on performance capacity which in turn can make female runners less time optimistic. We will now turn to the question of how time optimism is identified in our study.

A systematic deviation between the predicted and actual running time is identify as bias and distinguished from a random error. Hence, to sometimes be better and sometimes worse than predicted is not to be biased. To systematically perform worse than predicted is, however, to be biased.[9] To give some examples, suppose that there is a group of runners where each runner either finish in 240 or 250 minutes with equal probability and that they have knowledge about this particular distribution. If half of the runners would predict 240 and 250 minutes respectively, their predictions would be unbiased as a group. On the contrary, if more than half of the runners would predict a race time of 240 or 250 minutes they would be biased. Also note that if all of the runners believe that they will finish in 245 minutes their beliefs would be distorted but their predictions unbiased. Hence, if predictions are biased, beliefs are distorted but distorted beliefs does not necessary imply that predictions are biased. Alternatively, suppose that 95 percent of the runner finish in 240 minutes and 5 percent finish in 300 minutes. If everyone believes that they will finish in 240 minutes they clearly have disordered beliefs. It is even possible that each runner knows the probabilities but prefers to be exactly correct 95 percent of the time compared to being correct on average, and therefore predict a race time of 240 minutes. Notably the average race time is 243 minutes. Beliefs are distorted, since despite knowing the true distribution none of the runners believe that they will belong to the 5 percent with a slower race time. This form of optimistic bias concerning personal risk is well-known (Weinstein, 1989).

The following two research hypotheses will be tested: 1)Do people typically predict that they will run faster than they actually end up doing. 2) Do males typically make more time optimistic predictions than females?

While our sample consists of experienced participants, we do expect within-sample heterogeneity in experience. More specifically, more experienced people are expected to be less biased. Training increases experience, which in turn improves both performance and knowledge about physical lim-

[9]More specifically, suppose the time bias is symmetrically distributed with a mean of zero. Given this outcome we will not conclude half of the participants to be overconfident(and the other half to be under-confident). Instead we simply conclude that the null hypothesis of no bias cannot be rejected.

its. Therefore people that are faster are expected to be less overconfident and a positive relationship between bias and race time is expected.[10] [11]

3 The Field Studies

The data examing runners' predictions of their race time was collected by SJ, a government-owned passenger train operator in Sweden, as part of a marketing campaign, and JOGG [12], a Swedish online community for runners. Participants could take part in the prediction challenge either on the webpage of SJ or JOGG, or at runners' expeditions, starting two days before the respective races. Notably, at the time the campaigns were on there was no ambition to make use of the data for scientific purposes. The runners' predictions were subsequently matched with actual race time data that is available on the web page of the races.

Here is an example of a slogan used by Swedish national train company SJ to attract runners to predict their race time:

Predict your race time: At SJ we know how important it is to run on time. Now we challenge you to predict your race time. The runners with the best predictions will have a chance to win a free trip with us.

JOGG, the online community for runners, used a text like the following:

Dear Runners, go to the race page and guess your time. The winners are those that guess closest to the actual race time. At stake are several very good travelers' checks from SJ and Apollo Travel. Here we go![13]

JOGG's campaign could be one way to strengthen members ties and keep members active and engaged in the online community.

The runners' predictions were encouraged with material incentives such as chance to win train tickets. As already mentioned, Cesarini et al. (2006)

[10] Tanda (2011) found that mean distance per week and the mean training pace were strongly correlated with marathon race time.

[11] Unexpected shocks like an injury could result in a slower race time and a positive bias in predictions. However, as long as the unexpected shocks are randomly distributed among all runners, prediction bias and race time will be uncorrelated.

[12] JOGG collected data in collaboration with SJ.

[13] Apollo is one of Scandinavia's largest travel organisation companies.

found that the level of measured overconfidence was largely reduced when using monetary incentives in several confidence interval tasks. Material incentives give people incentives to stick to the truth and can also make them exert more cognitive effort to make more accurate predictions. This is of course important as it directly relates to the hypothesis of the present study. As already mentioned, too high incentives could potentially make things even worse as runners could slow down to finish on the predicted time which implies that we if anything would underestimate time optimism. This is, however, not deemed to be a problem in practice as race times are a source of internal pride and fulfillment and not an extrinsic reward (Allen et al., 2016).

All in all we have data from ten field studies. These are gathered from four races (Göteborgsvarvet Half Marathon, Lidingöloppet, Stockholm Half Marathon and Stockholm Marathon) using two different sampling frames (SJ and JOGG) and two different years (2015 and 2016). All of these races are held annually. The field studies are summarized in Table 1.

Gothenburg Half Marathon, i.e., Göteborgsvarvet has been recognized as the world's largest half marathon. The first race was held in June 7, 1980, and it usually takes place in early May every year. Stockholm Half Marathon is in early September and has been arranged since 2001. Lidingöloppet is a 30 km cross-country race held in Lidingö, Sweden in late September. The race attracts 15,000 participants per year and is one of the largest cross-country events in the world. It is held the last weekend of September each year. Stockholm Marathon has been arranged since 1979 and serves as the Swedish marathon championship race. It is usually held in late May or early June.

More male than female runners participate in the races. For the half-marathons the share of males is typically 60 - 65 percent, while the share of males in the longer races is about 70 - 75 percent. The mean age of the participants is about 40 years. As we will come back to when discussing the results, only a small share of all runners predicted their race time. Those that participated were in general faster than those that did not participate and therefore potential selection effects will be handled in the regression analysis.

Table 1: Summary of the field studies

Field Study	Race	Sampling Frame	Year	Distance (km)	Population size	Sample size
1	Gothenburg Half Marathon	JOGG	2015	21.1	46287	1463
2	Gothenburg Half Marathon	SJ	2015	21.1	46287	330
3	Gothenburg Half Marathon	SJ	2016	21.1	44995	531
4	Stockholm Half Marathon	SJ	2015	21.1	12232	71
5	Stockholm Half Marathon	SJ	2016	21.1	10622	171
6	Lidingöloppet	JOGG	2015	30	17482	180
7	Lidingöloppet	SJ	2015	30	17482	1036
8	Lidingöloppet	SJ	2016	30	15432	308
9	Stockholm Marathon	SJ	2015	42.2	17482	474
10	Stockholm Marathon	SJ	2016	42.2	12854	304

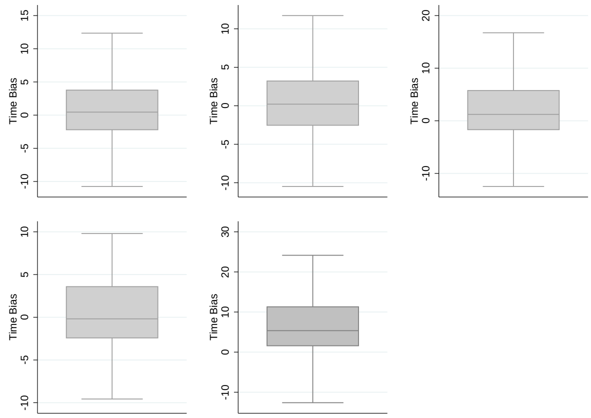
4 Results

We will discuss the results first for the Gothenburg and Stockholm Half Marathon (21.1 km race), then for Lidingöloppet (30 km race) and finally for the Stockholm Marathon (42.2 km race). Since the subjects were asked to enter their predictions into a computer without technical assistance the presence of incorrectly entered data cannot be excluded. To reduce the influence of incorrectly entered data the following measures were taken. For the Half-Marathon (21 km) we do not include subjects who expected a race time in excess of 300 minutes. Note that it takes less than 300 minutes to walk this distance at a comfortable speed (5 km/hour) hence these participants do not in any case belong to the targeted population of runners. Similarly, for the 30 km race and the full marathon, the thresholds were set to 400 minutes and 480 minutes, respectively. In total 26 observations were excluded of 4873 (0.5 percent).[14]

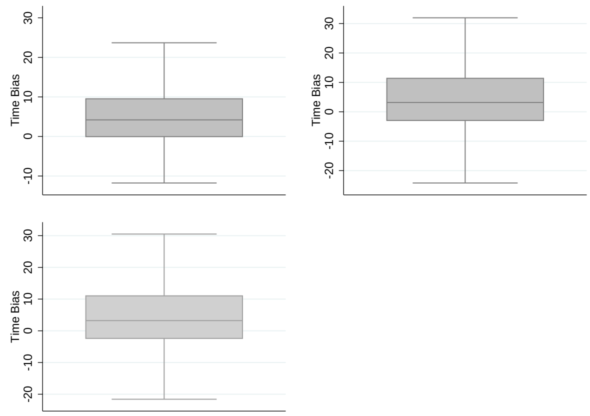
In figure 1 we graphically summarize the numerical data through their quartiles using box-plots of the different samples. Initially we want to get a general impression of the accuracy of the predictions and therefore we do not divide the sample between male and females. Bias is measured as the difference between actual and predicted race time. Hence, a positive value indicates time optimism and a negative value the opposite. Starting with the half marathons (Panel A) we see that the runners predictions are relatively well-calibrated and that the median bias indicated by the horizontal line inside the boxes do not exceed 5 minutes for any of the samples. For most of the samples the bias is even smaller. The box (which contains 50 % of the observations) is often located relatively more on the positive part of the Y-axis which reveals a tendency of the subjects to be time optimistic. Nevertheless, a non-negligible fraction of the runners underestimate their race time. If we look at the 30 km races (Panel B) we observe that the boxes are now more on the positive axis. Finally, in Panel C we see that the typical runners of the full marathon are also relatively well-calibrated, however, the quality of the predictions is more scattered than for the shorter

[14]Three of 2566 observations were excluded in the half marathon samples, one observation each for three of the samples, while all observations were kept in the remaining 2 samples. For three 30 km samples, 12 of 1524 observations were excluded. Turning to the full marathon race, 11 observations were deleted from the year 2015 sample, while no observations were excluded from the 2016 sample.

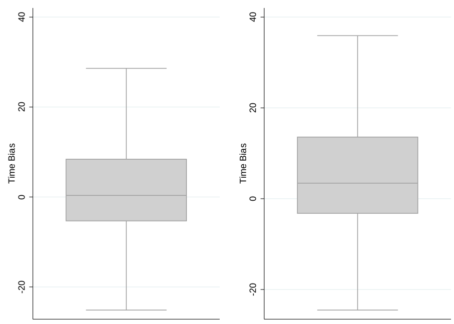
races. Throughout most of the samples we observe that the larger part of the box is above the median, meaning that the data is skewed right, and this is particularly true for the full marathon race. Across all the samples we find the runners' beliefs and predictions to be fairly accurate. Extreme values located beyond the whiskers of the box-plots are not shown yet are present in all of the samples. For this reason we will focus on the median instead of the mean when assessing judgment errors. The median is more robust to extreme values and is the preferred measure or central tendency for skewed distributions. Also note that we do not distinguish between males and females when constructing the box-plots. To gain a first impression of gender differences we turn to the descriptive statistics.



(a) Half-marathon



(b) Lidingöloppet



(c) Stockholm marathon

Figure 1: Field-study 1 (top left panel) to 10 (bottom right panel)

Descriptive statistics for the samples are presented in Appendix A, B and C.[15] For males the median bias is positive for all ten samples. For female runners the median bias is negative for two of the samples (field-studies 4 and 9) and positive for the remaining samples. The general impression based on these samples is that runners typically underestimate their race time and that they suffer from time optimism.[16] The point-estimate for the bias is greater for the 30 km and 42.2 km races, than for the 21.1 km races.

Turning to gender differences, the median bias is greater for males than for females in all but one of the samples (field-study 2).[17] We will not, however, conclude based on tests that compares the medians of two populations but instead turn to regression analysis. Using regression analysis allows to compare gender differences for equally fast runners and to control for observable selection effects. Notably, although we have a considerable number of observations in our samples only a small share of all runners made a prediction about their race time. Moreover, the average race times are in generally lower for runners who participate in the prediction contest compared to the population of runners (see Appendix A, B and C) which makes it important to control for these selection effects.

Another interesting result is that comparing the shares of males and females who ran slower than predicted, the share of males who are time optimistic ranges from 53 to 82 percent with a average of 66 percent. For women, the share ranges from 38 to 78 percent with a average of 55 percent. Hence, the tendency to be time optimistic is again found to be stronger among males.[18]

---

[15] Appendix A present descriptive statistics for the half-marathon races, Appendix B for the 30 km races, and Appendix C for the full marathon races.

[16] The median bias is statistically significant for seven of the samples for male runners (the samples with non-significant median values are field-studies 2, 4 and 9) and two of the samples for female runners (field-studies 3 and 5) at a 5 percent significance level.

[17] The difference is statistically significant for three of the populations using the Wilcoxon two-sided rank-sum test at the 5 percent significance level and additional three at the more generous 10 percent significance level. The field-studies with the significant differences are field-studies 1, 7, 8 at 5 percent significance level and 4, 6, and 10 at a 10 percent significance level.

[18] The share of males and females that ran slower than predicted were: Field-study 1: 58 % males, 51% females. Field-study 2: 56 % males, 52% females. Field-study 3: 63 % males, 58% females. Field-study 4: 65 % males, 38% females. Field-study 5: 82 % males, 78% females. Field-study 6: 77 % males, 58% females. Field-study 7: 68 % males, 51% females. Field-study 8: 67 % males, 57% females. Field-study 9: 53 % males, 48% females. Field-study 10: 68 % males, 58% females

4.1 Regression Results

For simplicity the same model specification is used for all samples (see equation 1).[19] OLS and Quantile regression models were estimated for each of the samples separately. Since the observations between the samples cannot be assumed to be independent and individuals that are included in several samples cannot be identified the samples are initially not pooled. Once the main results are established a pooled analysis is presented to explore how time optimism is effected by weather together with some additional control variables. One needs to be cautious, however, when interpreting the pooled analysis since the observations are not necessarily independent which in turn can effect the standard errors.

$$Bias = \alpha + \beta_1 \cdot Male + \beta_2 \cdot Race\,time + \epsilon \tag{1}$$

The dependent variable is the difference between actual and expected race time. The main parameter of interest is β_1 which captures gender differences. The second variable in equation (1) is a control variable that facilitates a comparison of time optimism bias between male and females with the same race time. As already mentioned, it also captures a within-sample heterogeneity in experience and a positive β_2 means that faster runners are less time optimistic.

While both ordinary least square (OLS) and quantile regressions are presented the focus is on the quantile regressions. The quantile regressions are more robust to outliers and therefore better reflects typical bias in the population. OLS regressions are still included for comparison. While OLS estimates the mean of the dependent variable conditional on the covariates, the quantile regression estimates the effect of the covariates on different quantiles of the distribution of the dependent variable that captures the

[19] An interaction term between the gender dummy and race time was included in addition to the main effects in an initial analysis. Because of biological differences one could expect that females have to train more to run as fast as males. More training in turn makes them more experienced. Therefore, an improved finishing time could be expected to effect time optimism for females more than males which in turn justifies inclusions of interaction terms. None of the samples did support this conjecture, however, and on the contrary, we find that that a change in finishing time effects males more than females in two of the samples (field studies 6 and 7). Therefore we proceed without including the interaction terms that was insignificant in most of the samples.

bias. This gives us a more detailed understanding of gender differences of time optimism. The 0.25, 0.50 and 0.75 quantile regressions are presented but the main focus is on the 0.50 quantile which is also known as the median regression function. The location of these quantiles are shown in the box plots where we see that for most of the samples the 0.25 quantile is negative while the 0.50 and 0.75 quantile is positive. Robust standard errors that are asymptotically valid under heteroskedasticity are used for the OLS (White, 1980) and for the quantile regressions (Machado et al., 2015).

4.1.1 Half-marathon

The results of the median regressions for the 21.1 km races are presented in Table 2 for the first three field-studies and in Table 3 for the remaining two field-studies. Gender differences are found in all samples at a 5 percent significance level except in field study 2. The positive values for β_1 across all samples mean that males are more confident than females with the same race time. The average of β_1 across the five samples is about 2.4 minutes meaning that male runners prediction error is 2.4 minutes larger than that of female runners. Also note that the time optimism significantly increases with race time ($\beta_2 > 0$) and the point estimates range from 0.08 to 0.21. Finally, if we focus on people with stronger bias (i.e. on the 0.75 quantile) we see statistically significant gender differences for all of the samples at the 5 percent significance level. The magnitude of the coefficient is greater at the 0.75 quantile for all samples except field study 4. For the 0.25 quantile the opposite pattern with lower β_1 coefficient is observed and statistically significance is only found in two of the samples at a 1 percent significance level. Hence, we find stronger gender differences if we compare people who are more time optimistic instead of people who are less time optimistic.

In a next step we use equation (1) to calculate the median time bias adjusted for selection effects for males. Table 4 summarizes the results, time biases for both male and female runners are evaluated at the population median finishing time for male runners. First of all we note that the bias is positive for all of the samples for males which is indicative that they in particular are time optimistic while the results for female runners are more mixed. Taking an examples, for *Field study 1* the bias for male runners is estimated to be 1.65 minutes. The bias for females runners with the same finishing time is -0.25 minutes (-15 seconds). The gender difference

Table 2: Regression results for the Gothenburg half-marathons (21.1 km)

	Median	Quantile_0.25	Quantile_0.75	OLS
		Field Study 1		
Intercept	-10.8680***	-5.9397***	-20.3252***	-22.1292***
	(1.3506)	(1.4356)	(1.8464)	(3.3695)
Male	1.9***	1.3586***	3.5658***	3.7865***
	(0.3337)	(0.3834)	(0.4641)	(0.5679)
Race time	0.0923***	0.0264*	0.1960***	0.1781***
	(0.0114)	(0.0120)	(0.0156)	(0.0280)
R^2	0.04	0.01	0.14	0.13
Number of obs.	1463	1463	1463	1463
		Field Study 2		
Intercept	-9.3140***	-6.6661*	-15.09691***	-7.7823
	(3.1760)	(3.0058)	(2.6964)	(4.7252)
Male	0.6913	1.3459	1.9506**	0.1854
	(0.8301)	(1.2001)	(0.7731)	(1.1156)
Race time	0.082***	0.0279	0.1481***	0.0655*
	(0.0252)	(0.0255)	(0.0231)	(0.0370)
R^2	0.04	0.01	0.13	0.03
Number of obs.	330	330	330	330
		Field Study 3		
Intercept	-14.0496***	-7.2308***	-23.19871***	-20.5841***
	(3.1595)	(2.4036)	(2.3559)	(3.3151)
Male	2.3675***	1.9862***	3.2973***	2.8241***
	(0.7978)	(0.6798)	(0.8508)	(0.8259)
Race time	0.1221***	0.0389**	0.2305***	0.1770***
	(0.0261)	(0.0206)	(0.0183)	(0.0267)
R^2	0.07	0.01	0.21	0.17
Number of obs.	531	531	531	531

[1] *, **, and *** denote significance at 10%, 5%, and 1%, respectively for a two-sided test.

Table 3: Regression results for the Stockholm half-marathons (21.1 km)

	Median	Quantile_0.25	Quantile_0.75	OLS
		Field Study 4		
Intercept	-17.1075***	-14.0688	-16.5107***	-38.4238*
	(5.4153)	(8.8099)	(4.4250)	(20.4163)
Male	3.1376**	3.5114	4.7679**	0.0352
	(1.5097)	(1.1512)	(2.0518)	(3.2902)
Race time	0.1338***	0.0928	0.1560***	0.3165*
	(0.0457)	(0.0775)	(0.0349)	(0.1657)
R^2	0.07	0.05	0.11	0.07
Number of obs.	71	71	71	71
		Field Study 5		
Intercept	-21.1230***	-19.8553***	-23.3965***	-21.26471***
	(3.7408)	(5.6648)	(4.7639)	(5.1138)
Male	4.1493***	2.3369	3.6130***	2.3427*
	(1.1567)	(1.6335)	(1.2361)	(1.3691)
Race time	0.2100***	0.1710***	0.2604***	0.2139***
	(0.0312)	(0.0428)	(0.0410)	(0.0391)
R^2	0.17	0.08	0.23	0.24
Number of obs.	171	171	171	171

[1] *, **, and *** denote significance at 10%, 5%, and 1%, respectively for a two-sided test.

[2] The β_1 coefficient for OLS in Field study 4 raises to 2.91 and is significant if one extremely under-confident respondent is removed that ran about 170 minutes faster than predicted.

Table 4: Time Bias [95% confidence intervals]

Field Study	Male	Female
1	1.65 [1.21; 2.08]	-0.25 [-0.70; 0.20]
2	0.64 [-0.20; 1.48]	-0.05 [-1.24; 1.13]
3	2.40 [1.38; 3.41]	0.03 [-0.97; 1.02]
4	0.68 [-1.51; 2.88]	-2.45 [-4.59; -0.31]
5	7.56[6.10; 9.02]	3.41 [1.68; 5.14]

corresponds to 1.9 minutes. The 95 percent confidence intervals in Table 4 show that males are found overconfident in three out of five samples, and females are found under-confident in one and overconfident in one of the five samples. Using equation (1) we can also calculate the median time bias adjusted for selection effects for females. These results are given in Table D1 in Appendix. Notably, we see that if both male and female runners are evaluated at the population median finishing time for female runners the point estimate of time bias is higher. This is a reoccurring pattern for all of the races presented in this study. In summary, we observe a systematic difference between male and female runners in their predictions of their own race times, where males tend to be more time optimistic.

4.1.2 Lidingöloppet

Table 5 summarizes the regression results for the 30 km races. The β_1 coefficient for the median regressions that capture gender differences in time optimism range from 4.2 to 7.5 minutes with an average of 5.35 minutes. Hence, compared with the half-marathons the β_1 coefficients are generally larger. The coefficient is statistically significant for all populations at the 1 percent significance level except field study 6 which contains substantially fewer observations. Focusing on the 0.75 quantile we observe same pattern as observed earlier with higher point estimates of the β_1 coefficients. The coefficient is statistically significant at the 5 percent significance level for all populations. For the 0.25 quantile the opposite pattern, i.e., lower β_1 coefficient is observed. The effect is statistically significant in one of the populations at the 1 percent significance level and in one population at the 10 percent significance level.

Equation (1) is used to calculate the median bias for the 30 km races for male runners adjusted for selection effects. Time biases for both male and female runners are evaluated at the population median finishing time for male runners. Table 6 shows that male runners are typically overconfident while female runners are less biased. The median time biases adjusted for selection effects for females are given in Table D2 in the Appendix.

Table 5: Regression results for the Lidingöloppet (30 km) races

	Median	Quantile_0.25	Quantile_0.75	OLS
		Field Study 6		
Intercept	-15.2802	-11.0831*	-34.3646***	-38.6812***
	(9.77258)	(5.6466)	(10.6437)	(13.3571)
Male	4.3796	5.7354	6.4853**	11.0444***
	(2.895959)	(3.5563)	(2.8896)	(3.4136)
Race time	0.0976*	0.0398*	0.2436***	0.2129***
	(0.0553)	(0.0239)	(0.0590)	(0.0701)
R^2	0.06	0.02	0.19	0.26
Number of obs.	180	180	180	180
		Field Study 7		
Intercept	-24.4552***	-9.3432***	-33.5416***	-28.7306***
	(2.9052)	(3.4946)	(3.2594)	(3.4384)
Male	7.4943***	6.2033***	9.1821***	9.4892***
	(0.9565)	(1.1876)	(1.0559)	(1.1553)
Race time	0.1312***	0.0099	0.2115***	0.1432***
	(0.0150)	(0.0180)	(0.0175)	(0.0175)
R^2	0.06	0.03	0.15	0.13
Number of obs.	1036	1036	1036	1036
		Field Study 8		
Intercept	-28.7506***	-7.6299	-37.2152***	-32.16133***
	(6.7294)	(6.3888)	(3.5709)	(5.1955)
Male	4.1905***	3.5421*	7.4929***	7.037***
	(1.6184)	(1.8818)	(1.3616)	(1.7847)
Race time	0.1752***	0.0172	0.2450***	0.1804***
	(0.0376)	(0.0358)	(0.0192)	(0.0285)
R^2	0.11	0.02	0.22	0.17
Number of obs.	308	308	308	308

[1] *, **, and *** denote significance at 10%, 5%, and 1%, respectively for a two-sided test.

Table 6: Time bias [95% confidence intervals]

Field Study	Male	Female
6	6.39 [3.03; 9.74]	2.00 [-2.98; 6.98]
7	6.28 [5.27; 7.29]	-1.21 [-2.64; 0.21]
8	6.81 [4.56; 9.07]	2.62 [0.29; 4.95]

4.1.3 Marathon

Table 7 summarizes the results from the full marathons (42.2 kilometers). The β_1 coefficients range from 2.7 to 5.0 minutes for the median regressions with an average of 3.85 minutes and are statistically significant at the 5 percent level for both of the samples. In line with previous results, the β_1 coefficient at the 0.75 quantile regressions are consistently higher. In line with previous field studies, lower β_1 coefficients are observed for the 0.25 quantile. The coefficient is only statistically significant in one of the samples (field study 10) at the 10 percent significance level.

Equation 1 is used to estimate the bias for males evaluated at population median race time for males. Table 8 summarizes the confidence intervals. While males are again shown to suffer from time optimism the bias in female runners' predictions are more in line with their actual race time. Women's median time biases adjusted for selection effects are given in Table D3 in Appendix.

4.1.4 Summary of the results

In summary, gender differences in time optimism are found in all except two samples for the median regressions. Males are more time optimistic and more likely to overestimate their capacity. Comparing gender differences between runners with the same race time, the average optimism bias across the five samples of half-marathon runners is about 2.4 minutes greater for males than for females. For the 30 km races, it is about 5.35 minutes and for the full marathons 3.86 minutes. We also find a within-sample heterogeneity in all of our samples, faster runners being less time optimistic. Using the regression models it possible to estimate how fast a runner has to be for him or her not to be overconfident. Based on these calculations we find

Table 7: Regression results for the Stockholm Marathons (42.2 km)

Regressions	Median	Quantile_0.25	Quantile_0.75	OLS
		Field Study 9		
Intercept	-13.9132***	-1.6296	-26.1250***	-26.3259***
	(4.1270)	(7.1635)	(5.4554)	(5.9721)
Male	2.7272**	1.2971	4.3630**	3.6680**
	(1.2117)	(1.4971)	(1.7330)	(1.5698)
Race time	0.05625***	-0.0203	0.1331***	0.1015***
	(0.0175)	(0.0314)	(0.0224)	(0.0238)
R^2	0.02	0.002	0.08	0.04
Number of obs.	474	474	474	474
		Field Study 10		
Intercept	-19.5027***	-5.6132	-39.67823***	-35.52923***
	(7.4728)	(5.6662)	(7.3692)	(7.7731)
Male	4.9810**	3.9168*	5.7447***	5.826**
	(2.1393)	(2.0447)	(2.2198)	(2.4813)
Race time	0.0864***	-0.0033	0.2057***	0.152***
	(0.0313)	(0.0227)	(0.0316)	(0.0313)
R^2	0.04	0.01	0.14	0.10
Number of obs.	304	304	304	304

[1] *, **, and *** denote significance at 10%, 5%, and 1%, respectively for a two-sided test.

Table 8: Time bias [95% confidence intervals]

Field Study	Male	Female
9	2.12 [0.73; 3.51]	-0.61 [-2.58; 1.36]
10	6.18 [3.67; 8.69]	1.20 [-2.26; 4.66]

that the median runner is overconfident in all of the samples. For example, for field study 10, males that run slower than 168 minutes are found to be overconfident, while the median finishing time for males is about 240 minutes.[20]

The average time optimism bias for males is 2.59, 6.49 and 4.15 minutes for the half marathons, 30 km races and full marathons respectively. For the females it is 1.81, 3.78 and 2.01 minutes.[21]

4.2 Pooled Regression Results

In a final analysis we pool the ten samples and estimate regression models for the pooled data. In addition to the two main variables, gender and race time, temperature and wind is also included to control for the influence of weather on time optimism. Furthermore, a dummy variable is included to control for time specific effects, distance is controlled using a dummy for the 30 km ($Distance30$) and full-marathon race ($Distance42.2$) respectively. A dummy variable SJ is included to distinguish whether data was collected by SJ, government-owned passenger train operator in Sweden, or JOGG, a Swedish online community for runners. Table 7 summarizes the results. Most of the coefficients are significant at the 1 percent level. Gender differences in time optimism is confirmed, males are more time optimistic. In line with previous results, faster runners are less time optimistic. Considering the effect of the weather, higher temperatures and wind speed is positively correlated with time optimism. Moreover, members of the online running community JOGG are more time optimistic. It could be that members of JOGG are more competitive which in turn makes them more biased. These results are all in line with what one could expect.[22] In general the pooled model is consistent with models presented in the previous sections but due to the increased sample size the precision of the estimates have substantially increased.

[20](-19.5027+4.9810/-0.0864)=168

[21]These estimates are based on predictions from the median regression model using the median race time for the male and female population respectively.

[22]Distance is important to control, however, the interpretation is less important since one cannot reasonably expect the same finishing time for the half marathons, 30 km races and full marathons respectively.

Table 9: Pooled Regression Analysis

	Median	Quantile_0.25	Quantile_0.75	OLS
Intercept	-17.8290***	-9.8561***	-25.2792***	-23.6058***
	(1.2511)	(1.3966)	(1.1693)	(1.9000)
Male	2.9632***	2.5075***	4.2061***	4.7843***
	(0.2632)	(0.3135)	(0.2992)	(0.4127)
Race time	0.1079***	0.0195**	0.1972***	0.1482***
	(0.0071)	(0.0076)	(0.0067)	(0.0100)
Temperature	0.1977***	0.1526***	0.1594***	0.1537***
	(0.0375)	(0.04651)	(0.0397)	(0.0561)
Wind	0.3353***	0.3306***	0.4617***	0.4599***
	(0.1188)	(0.1222)	(0.1208)	(0.1767)
Year2016	2.4936***	2.0612***	2.6006***	3.4993***
	(0.3411)	(0.3946)	(0.3591)	(0.5330)
Distance30	-2.9344***	-1.0140*	-5.0043***	-4.9719***
	(0.4774)	(0.5276)	(0.5046)	(0.6520)
Distance42.2	-10.8245***	-4.3382***	-16.9279***	-16.1563***
	(0.8656)	(1.0657)	(0.8699)	(1.3428)
SJ	-0.9099***	-0.7976**	-1.1865***	-1.9103***
	(0.2975)	(0.3404)	(0.3181)	(0.5089)
R^2	0.12	0.03	0.11	0.12
Number of obs.	4868	4868	4868	4868

[1] *, **, and *** denote significance at 10%, 5%, and 1%, respectively for a two-sided test.

5 Discussion

How long will it take? This is a question familiar to most people from daily life and the workplace. In this paper we explore whether long-distance runners give an unbiased answers to a question about how they will perform in an upcoming race. Participating in a long-distance race usually means months of extensive training and therefore also of important experience and performance feedback. Hence, we have a unique set of data on experienced participants collected in the participants natural occurring environment. The data is used to address the important question of whether overconfidence disappear with sufficient experience. The strength of our data is that we observe people with extraordinary access to immediate, precise, and objective feedback. Furthermore, instead of relying on only one race we use data from ten different samples. Making use of several studies instead of one allows us to test the robustness of our findings.

Based on evidence from ten field studies we find that long-distance runners often give biased predictions. That is, although their predictions are relatively accurate they tend to be overly optimistic about their race time. Moreover, the bias is found to be greater for males than for females. Hence, the results show that if two individuals with different genders but equal performance capacity are asked how long it will take them to finish a task males will typically report a faster finishing time. This is not an expression of their productivity but rather of their bias. Based on earlier findings in the literature this could be a result of males being more goal-oriented and motivated by competition (Deaner et al., 2015). These traits in turn make them suppress past experience of how fast they actually are if it conflicts with their desired race time. Moreover, these characteristics to overestimate their capacity could be typical for not only male runners but for males in general.

To be aware of the limitations of one's capacity could be beneficial in many situations, like in the prediction task in this study. In other situations, e.g., the labour market, overconfidence could even offer some benefits (Palomino and Peyrache, 2010). Overconfidence is an personality trait, and traits are known to be sufficiently stable across situations (Borghans et al., 2008). Hence, it is tempting to generalize the behaviour found in this study that males overestimate their capacity to carry over to other situations as well e.g. the labour market. One could actually even expect overconfidence to be stronger in many other contexts since the immediate, precise, and objective feedback that is available to long-distance running is rare in other situations. But there are reasons to be cautious when generalizing the findings of one study to other contexts and populations. Ideally, it would be fruitful to test for time optimism in other contexts such as the labour market before taking any major conclusions. This is, however, a non-trivial task, and it requires access to a data base that could match both psychological parameters of individuals and their salary (Palomino and Peyrache, 2010).

An interesting feature of the labor market is that an employer can only gain a true understanding of the employees' output and productivity of employees with a time delay, if at all. This can make the employer to turn to self-appraisals. Since self-appraisals could be biased and favourable for males that overestimate their capacity compared to females it would be better for females, however, if less attention were be paid to self-appraisals in the recruitment of labor and setting wages.[23]

The most important lesson learned from this study is that even in situations where people are experienced gender differences in time optimism can be expected. If we want to judge how productive people are, it is better to let them "walk the walk" than to "talk the talk". People do not act in agreement with what they say and therefore we should judge them by their behavior. Judging them by their talk will favour males over females.

[23] Interestingly, Leibbrandt and List (2014) show that simple statements of whether the wage is negotiable or not in a job announcement can make a huge difference. They find that when there is no explicit statement that wages are negotiable, men are more likely to negotiate for a higher wage, whereas women are more likely to signal their willingness to work for a lower wage. Explicitly mentioning the possibility of negotiating wages, however, eliminates the gender differences in the share of candidates that negotiate for higher wage and share of candidates that are willing to work for lower wage.

References

Allen, E. J., P. M. Dechow, D. G. Pope, and G. Wu (2016). Reference-dependent preferences: Evidence from marathon runners. *Management Science*.

Babcock, L. and S. Laschever (2009). *Women don't ask: Negotiation and the gender divide*. Princeton University Press.

Barber, B. M. and T. Odean (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, 261–292.

Bekhouche, Y., R. Hausmann, L. Tyson, and S. Zahidi (2013). The global gender gap report 2013. Geneva Switzerland World Economic Forum 2013.

Bengtsson, C., M. Persson, and P. Willenhag (2005). Gender and overconfidence. *Economics Letters 86*(2), 199–203.

Beyer, S. and E. M. Bowden (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin 23*(2), 157–172.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Ter Weel (2008). The economics and psychology of personality traits. *Journal of Human Resources 43*(4), 972–1059.

Buehler, R., D. Griffin, and M. Ross (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology 67*(3), 366.

Cesarini, D., P. Lichtenstein, M. Johannesson, and B. Wallace (2009). Heritability of overconfidence. *Journal of the European Economic Association 7*(2-3), 617–627.

Cesarini, D., Ö. Sandewall, and M. Johannesson (2006). Confidence interval estimation tasks and the economics of overconfidence. *Journal of Economic Behavior & Organization 61*(3), 453–470.

Cheuvront, S. N., R. Carter, K. C. DeRuisseau, and R. J. Moffatt (2005). Running performance differences between men and women. *Sports Medicine 35*(12), 1017–1024.

Daniel, K. and D. Hirshleifer (2015). Overconfident investors, predictable returns, and excessive trading. *The Journal of Economic Perspectives 29*(4), 61–87.

Deaner, R. O., R. E. Carter, M. J. Joyner, and S. K. Hunter (2015). Men are more likely than women to slow in the marathon. *Medicine and Science in Sports and Exercise 47*(3), 607.

Deaner, R. O., A. Lowen, W. Rogers, and E. Saksa (2015). Does the sex difference in competitiveness decrease in selective sub-populations? a test with intercollegiate distance runners. *PeerJ 3*, e884.

Fischhoff, B., P. Slovic, and S. Lichtenstein (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance 3*(4), 552.

Gervais, S. and T. Odean (2001). Learning to be overconfident. *Review of Financial Studies 14*(1), 1–27.

Griffin, D. and A. Tversky (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology 24*(3), 411–435.

Grubb, M. D. (2015). Overconfident consumers in the marketplace. *The Journal of Economic Perspectives 29*(4), 9–35.

Hölmich, P., S. Christensen, E. Darre, F. Jahnsen, and T. Hartvig (1989). Non-elite marathon runners: health, training and injuries. *British Journal of Sports Medicine 23*(3), 177–178.

Kahneman, D. and A. Tversky (1977). Intuitive prediction: Biases and corrective procedures. Technical report, DTIC Document.

Krawczyk, M. and M. Wilamowski (2016). Are we all overconfident in the long run? evidence from one million marathon participants. *Journal of Behavioral Decision Making*.

Leibbrandt, A. and J. A. List (2014). Do women avoid salary negotiations? evidence from a large-scale natural field experiment. *Management Science 61*(9), 2016–2024.

Machado, J. A. F., P. M. Parente, J. S. Silva, et al. (2015). qreg2: Stata module to perform quantile regression with robust and clustered standard errors. *Statistical Software Components*.

Malmendier, U. and G. Tate (2015). Behavioral CEOs: The role of managerial overconfidence. *The Journal of Economic Perspectives 29*(4), 37–60.

Malmendier, U. and T. Taylor (2015). On the verges of overconfidence. *Journal of Economic Perspectives 29*(4), 3–8.

Menkhoff, L., M. Schmeling, and U. Schmidt (2013). Overconfidence, experience, and professionalism: An experimental study. *Journal of Economic Behavior & Organization 86*, 92–101.

Nekby, L., P. S. Thoursie, and L. Vahtrik (2008). Gender and self-selection into a competitive environment: Are women more overconfident than men? *Economics Letters 100*(3), 405–407.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Palomino, F. and E.-A. Peyrache (2010). Psychological bias and gender wage gap. *Journal of Economic Behavior & Organization 76*(3), 563–573.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica 47*(2), 143–148.

Tanda, G. (2011). Prediction of marathon performance time on the basis of training indices. *Journal of Human Sport and Exercise 6*(3), 511–520.

Taylor, S. E. and J. D. Brown (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin 103*(2), 193.

Van den Steen, E. (2004). Rational overoptimism (and other biases). *The American Economic Review 94*(4), 1141–1151.

Van den Steen, E. (2011). Overconfidence by bayesian-rational agents. *Management Science 57*(5), 884–896.

Weinstein, N. D. (1989). Optimistic biases about personal risks. *Science 246*(4935), 1232–1234.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817–838.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics 13*(1), 75–98.

6 Appendix 1

6.1 Descriptive Statistics

Table A1: Field Study 1

Population	Mean	Median	Min	Max	St.Dev	#obs
Race time	122.25	119.38	60.73	298.83	20.85	46287
Race time:Male	117.72	114.98	60.73	271.07	20.00	30197
Race time:Female	130.76	128.15	68.22	298.83	19.75	16090
Sample						
Race time	115.79	113.60	60.73	235.27	20.72	1463
Race time:Male	110.03	107.75	60.73	207.03	18.82	963
Race time:Female	126.89	124.46	84.53	235.27	19.68	500
Guessed time	114.81	112.8	69.92	250.00	19.89	1463
Guessed time:Male	108.78	107.6	74.17	230.00	17.28	963
Guessed time:Female	126.43	122.54	69.92	250.00	19.43	500
Bias:Male	0.55	1.25	-76.50	43.35	8.41	963
Bias:Female	0.13	0.46	-110.47	122.17	11.33	500

Table A2: Field Study 2

Population	Mean	Median	Min	Max	St.Dev	#obs
Race time	122.25	119.38	60.73	298.83	20.85	46287
Race time:Male	117.72	114.98	60.73	271.07	20.00	30197
Race time:Female	130.76	128.15	68.22	298.83	19.75	16090
Sample						
Race time	118.04	115.52	71.77	206.72	20.04	330
Race time:Male	114.17	111.83	71.77	206.72	20.04	244
Race time:Female	129.01	127.74	87.28	191.57	19.99	86
Guessed time	119.95	115.77	72.00	232	21.28	330
Guessed time:Male	114.28	110.17	72.00	232	20.89	244
Guessed time:Female	128.34	127.47	88.98	185.45	18.91	86
Bias:Male	-0.12	0.11	-66.60	26.53	8.69	244
Bias:Female	0.67	0.42	-17.47	22.44	6.62	86

Table A3: Field Study 3

Population	Mean	Median	Min	Max	St.Dev	#obs
Race time	122.41	119.67	60.73	309.20	20.99	44955
Race time:Male	117.67	115.27	60.73	309.20	20.05	29310
Race time:Female	131.28	128.70	68.22	254.20	19.78	15645
Sample						
Race time	120.39	117.60	73.22	233.85	24.44	531
Race time:Male	111.48	113.66	73.22	218.25	21.58	334
Race time:Female	131.80	127.02	76.45	233.85	24.81	197
Guessed time	117.89	116.20	68.15	236.20	22.52	531
Guessed time:Male	111.30	109.78	69.98	184.25	19.43	334
Guessed time:Female	129.05	125.25	68.15	236.20	23.04	197
Bias:Male	2.36	1.29	-87.42	43.25	9.58	334
Bias:Female	2.74	0.68	-27.42	67.28	10.39	197

Table A4: Field Study 4

Population	Mean	Median	Min	Max	St.Dev	#obs
Race time	115.05	116.28	70.68	189.12	17.23	12232
Race time:Male	111.39	109.52	70.68	189.12	16.35	7641
Race time:Female	124.41	123.18	77.33	181.02	15.51	4591
Sample						
Race time	117.19	116.07	74.1	174.4	18.80	71
Race time:Male	108.03	107.17	74.1	157.35	16.99	31
Race time:Female	124.30	122.74	99.48	174.4	17.15	40
Guessed time	118.51	116.45	74.28	260.48	18.21	71
Guessed time:Male	112.23	106.73	74.28	260.48	31.61	31
Guessed time:Female	123.39	122.11	76.88	164.6	16.93	40
Bias:Male	-4.21	0.38	-171.38	9.47	31.30	31
Bias:Female	0.91	-1.18	-14.28	56.83	10.58	40

Table A5: Field Study 5

Population	Mean	Median	Min	Max	St.Dev	#obs
Race time	122.82	121.91	64.85	194.05	18.69	171
Race time:Male	118.11	116.83	64.85	191.93	18.13	93
Race time:Female	130.30	129.33	76.05	194.05	17.04	78
Sample						
Race time	125.22	123.25	74.27	184.02	22.09	10622
Race time:Male	118.38	116.62	74.27	171.62	19.57	6512
Race time:Female	133.37	130.56	84.15	184.02	22.27	4110
Guessed time	118.43	117.08	60.00	171.85	19.52	171
Guessed time:Male	111.99	110.00	76.00	169.35	17.43	93
Guessed time:Female	126.11	124.00	60.00	171.85	19.18	78
Bias:Male	6.40	5.38	-24.10	26.98	8.53	93
Bias:Female	7.26	5.26	-10.80	42.85	9.93	78

Table B1: Field Study 6

Population	Mean	Median	Min	Max	St.Dev	#obs
Race time	187.04	182.12	97.60	396.42	33.90	17482
Race time:Male	181.97	177.10	97.60	368.52	33.04	12768
Race time:Female	200.75	196.63	118.23	396.42	32.38	4714
Sample						
Race time	169.58	163.30	109.62	290.60	32.61	180
Race time:Male	162.78	158.12	109.62	290.60	28.16	144
Race time:Female	196.04	186.24	141.63	277.93	36.07	36
Guessed time	163.30	157.49	109.82	296.30	29.84	180
Guessed time:Male	155.73	152.50	109.82	257.12	22.09	144
Guessed time:Female	192.95	179.31	135.87	296.30	37.81	36
Bias:Male	7.09	4.48	-8.77	104.55	12.73	144
Bias:Female	3.08	1.93	-19.95	49.92	12.45	36

Table B2: Field Study 7

Population	Mean	Median	Min	Max	St.Dev	#obs
Race time	187.04	182.12	97.60	396.42	33.90	17482
Race time:Male	181.97	177.10	97.60	368.52	33.04	12768
Race time:Female	200.75	196.63	118.23	396.42	32.38	4714
Sample						
Race time	183.54	179.62	104.88	336.47	33.38	1036
Race time:Male	177.34	174.95	104.88	336.47	31.11	756
Race time:Female	200.28	198.35	118.23	318.58	33.61	280
Guessed time	179.06	175.73	91.78	315.00	33.27	1036
Guessed time:Male	171.18	169.98	91.78	315.00	28.77	756
Guessed time:Female	200.32	195.16	102.22	305.38	35.28	280
Bias:Male	6.15	4.21	-66.70	66.33	14.22	756
Bias:Female	-0.05	0.45	-71.35	56.72	16.22	280

Table B3: Field Study 8

Population	Mean	Median	Min	Max	St.Dev	#obs
Race time	189.73	184.98	97.85	396.42	35.73	15432
Race time:Male	184.50	179.04	97.85	368.52	34.89	11174
Race time:Female	203.46	198.78	116.88	396.42	34.24	4257
Sample						
Race time	182.25	177.22	105.05	362.58	36.47	308
Race time:Male	176.99	169.35	105.05	362.58	36.83	209
Race time:Female	193.37	189.13	133.08	351.52	33.24	99
Guessed time	176.77	172.48	81.58	315.00	34.07	308
Guessed time:Male	170.18	165.02	81.58	315.00	33.58	209
Guessed time:Female	190.65	189.13	133.47	308.13	30.92	99
Bias:Male	6.80	3.57	-64.77	82.93	16.43	209
Bias:Female	2.71	2.73	-41.57	56.95	15.15	99

Table C1: Field Study 9

Population	Mean	Median	Min	Max	St.Dev	#obs
Race time	246.25	241.5	138.37	376.53	38.77	17482
Race time:Male	240.47	236.49	138.37	376.53	37.99	12768
Race time:Female	262.29	260.48	154.23	368.76	36.31	4714
Sample						
Race time	338.88	236.28	141.15	352.95	38.15	474
Race time:Male	232.43	227.93	141.15	352.95	36.64	342
Race time:Female	255.58	251.70	176.63	352.37	37.02	132
Guessed time	238.31	233.94	137.68	414.98	39.60	474
Guessed time:Male	231.49	225.53	137.68	414.98	38.21	342
Guessed time:Female	255.96	252.20	177.33	350.00	37.76	132
Bias:Male	0.94	0.53	-204.02	61.58	21.20	342
Bias:Female	-0.37	-0.67	-35.63	51.00	13.37	132

Table C2: Field Study 10

Population	Mean	Median	Min	Max	St.Dev	#obs
Race time	250.72	247.24	130.97	377.15	39.91	12854
Race time:Male	245.28	239.65	130.97	377.15	39.61	9460
Race time:Female	265.87	263.83	151.77	375.45	36.74	3394
Sample						
Race time	245.61	239.11	165.97	351.42	38.60	304
Race time:Male	238.02	233.38	165.97	332.50	34.68	210
Race time:Female	262.55	265.96	172.07	351.42	41.61	94
Guessed time	239.74	236.03	159.98	358.00	37.93	304
Guessed time:Male	231.51	228.00	159.98	358.00	33.82	210
Guessed time:Female	258.13	261.33	173.17	347.53	40.27	94
Bias:Male	6.52	4.98	79.58	66.02	16.82	210
Bias:Female	4.42	1.52	-61.97	72.60	21.72	94

6.2 Time bias

Table D1: Time Bias [95% confidence intervals]

Field Study	Male	Female
1	2.86 [2.18; 3.55]	0.96 [0.44; 1.48]
2	1.70 [0.34; 3.06]	1.01 [-0.11; 2.12]
3	4.04 [2.43; 5.64]	1.67 [0.60; 2.74]
4	2.51 [0.07; 4.96]	-0.63 [-2.77; 1.51]
5	10.19 [8.47; 11.90]	6.04 [4.14; 7.93]

Table D2: Time bias [95% confidence intervals]

Field Study	Male	Female
6	8.29 [2.92; 13.67]	3.91 [-1.77; 9.60]
7	8.85 [7.41; 10.28]	1.35 [-0.06; 2.76]
8	10.27 [6.78; 13.76]	6.08 [3.21; 8.95]

Table D3: Time bias [95% confidence intervals]

Field Study	Male	Female
9	3.47 [1.54; 5.40]	0.74 [-1.49; 2.97]
10	8.27 [4.60; 11.94]	3.29 [-0.66; 7.23]