University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 1/2015 (149)

Michał Krawczyk
Maciej Wilamowski

Are we all overconfident in the long run? Evidence from one million marathon participants.

Warsaw 2015

University of Warsaw
Faculty of Economic Sciences

Are we all overconfident in the long run? Evidence from one million marathon participants.

Michał Krawczyk
Faculty of Economic Sciences
University of Warsaw
e-mail: mkrawczyk@wne.uw.edu.pl

Maciej Wilamowski
Faculty of Economic Sciences
University of Warsaw
e-mail: mwilamowski@wne.uw.edu.pl

Abstract

In this project we sought to contribute to extant literature on overconfidence by identifying it in a large, heterogeneous sample making familiar, repeated choices in a natural environment which provides direct feedback. In Study 1 we elicited predictions of own finishing time among participants of the 2012 Warsaw Marathon. Their prediction errors turned out to be very highly correlated with the change in pace over the course of the run. In Study 2 we thus took this change in pace as a proxy for self-confidence and used existing field data of around one million participants. Both studies indicate that males as well as youngest and oldest participants tend to be more confident. In Study 2 we are also able to investigate national and cultural dimensions, confirming previously reported findings of relative overconfidence in Asians and providing some novel results, i.a. that relatively conservative societies tend to be more self-confident.

Keywords:
overconfidence, performance forecasts, gender differences, age effects, national culture

JEL:
C93, D01, Z1

Acknowedgements

We thank Patryk Cebartowski of Adidas and Dorota Kostrzewska of Goldbach Interactive for granting access to data that made Study 1 possible. We have benefitted from comments made by participants of seminars in Warsaw and Konstanz. Katarzyna Pfeifer and Emil Leśkiewicz provided research assistance. All errors remain ours.

For which of you, intending to build a tower, sitteth not down first, and counteth the cost, whether he have sufficient to finish it? Lest haply, after he hath laid the foundation, and is not able to finish it, all that behold it begin to mock him, Saying, This man began to build, and was not able to finish.
Luke 14, 28-30

Introduction

Suppose you have to assign an important and difficult task to one of your two employees: Mrs Lotte de Jong, 36 year old, coming from the Netherlands or Mr Vivaan Gupta, a 66 year old Indian. A priori they seem to be equally competent and they both honestly believe they can successfully complete it before deadline. Whom should you nominate? Is one or both of them likely to be in error, overestimating their actual ability?

Research in psychology credibly shows that in some situations most people do have a tendency to make such a mistake. This has been proposed to have substantial consequences in several life domains. It may result in excessive taking of chances in a range of situations – in finance, education, sports etc. It may lead to suboptimal project planning in organizations. As (Bénabou, Tirole 2002) assert, "From car accidents, failed dot.com firms, and day trading to the space shuttle disaster and lost wars, the costs of overconfidence are plain for all to see".[1]

But is systematic overconfidence (OC) real? The term has been applied to many phenomena (Moore, Healy 2008)[2] which tend to be quite difficult to pin down in the field. Consider the well-known result that substantial majority think they are above-average drivers (Svenson 1981) which clearly cannot be correct if "driving ability" is roughly symmetrically distributed. What is this ability exactly, though, and how can it be identified and compared? It is likely that individuals who mostly drive on highways will define good drivers as those who can drive safely at, say 100+ mph, and will typically, correctly, judge that their experience makes them better than average. Those who spend most of the time at the wheel in cities will primarily think about the ability to parallel-park etc. Is it very surprising that un-incentivized, imprecise questions are interpreted in an ever so slightly self-serving way?

The alternative approach is to use a laboratory setting, in which tasks can be well-defined and predictions of own performance—directly verified (Clark, Friesen 2009). However, this typically comes at a cost—most of these studies use small, homogenous, self-selected samples of student subjects making choices in highly artificial settings they encounter for the first time in their lives.

[1] To be sure, there is a positive side to reasonably high self-confidence, which may directly boost performance (Feather 1966) and satisfaction with the task.

[2] In particular, one can speak of individual's overestimation of own performance with respect to some objective benchmark. Then, there is "overplacement" with respect to others' performance (aka "better than average effect"). Both of these are types of unrealistically positive self-evaluations and may be associated with illusion of control and unsubstantiated optimism in general. Another sort of OC concerns one's knowledge and has been labeled miscalibration and "overprecision"; it may manifest itself i. a. in setting excessively narrow confidence intervals given pre-determined confidence level and overestimating the number of values that were covered by such intervals (miscalibration in frequency judgment). It remains subject to scrutiny to what extent these various biases are correlated (Hilton, Regner et al. 2011). In this study we focus on the first type of overconfidence.

Again, the problem of providing proper financial motivation to judge own abilities correctly is not always solved, partly because reward schemes affect incentives to perform, not only to predict performance. Issues such as biased choice of tasks and experimenter effect are also often raised in this respect.

One negative consequence of this state of affairs is that our knowledge of demographic and cultural correlates of overconfidence is quite limited, making choices such as the one between Mrs. De Jong and Mr. Gupta harder. Perhaps the only well-established effect is that males are found to be, on average, more overconfident than females; it is also often suggested that East Asians are more overconfident than Westerners. Sadly, even for these effects we may be unsure to what extent observed results are merely artifacts associated with different self-presentation styles.[3]

In this project we looked at incentivized predictions and proxies thereof in a very large, diverse, international sample of long-distance runners. Thereby we took advantage of the fact that amateur distance running is a highly popular sport in many countries, engaging individuals of both genders, all age groups and social strata. This is not meant to say that runners form a representative sample: in particular, according to the National Runners Surveys, runners (in the US at any rate) are on average substantially better educated and richer than the general population. This group might thus provide some insight such important subpopulations as managers and executives.

In our Study 1 participants of the Warsaw Marathon were asked to predict their own finishing time prior to the race. This task was framed as a contest encouraging correct forecasts. This setting appears highly suitable to studying one of the key types of overconfidence, namely overestimation of own actual performance. Indeed, we have a well-defined, precise, individual performance measure (own time). The subjects are used to making and disclosing similar predictions and have ample opportunity to learn and correct for biases, if any. Running time is also a very focal variable, many a runner actually treating improvement of their personal best, year's best etc. as priority in a typical running event.[4] The fact that running is a massively popular sport with several major races taking place yearly means that we can access very large, heterogeneous samples. In particular, the key contribution of our study is that we are able to estimate the link between demographic variables (notably age, which has received much less scrutiny in the literature than gender) and the tendency to fall prey to the bias. On top of that, we can also address the controversial unskilled and unaware hypothesis (Kruger, Dunning 1999) that (only) low-performers systematically overestimate their abilities.

[3] Indeed, willingness to be perceived by others as skilled, self-confident (or, conversely, self-restrained) etc. can substantially affect results, especially given that most studies give no incentive to tell the truth. Related to this, (Krawczyk 2012) found that gender OC gap diminished when subjects were offered monetary rewards for correct predictions. This result suggests that gender effects observed in non-incentivized studies may partly be due to relatively modest self-presentation style that the society perceives as appropriate for females.

[4] It may be important to note that there is only handful of prizes in a typical event that are well beyond the reach of amateurs, so finishing time is only relevant from the viewpoint of individual satisfaction with own performance.

In our Study 2 we make use of the observation of a strikingly strong relationship between prediction error and change in pace over the course of the run observed in our Study 1 sample: overly confident individuals tend to start too fast, in that they slow down much more than those with pessimistic or reasonable expectations (and much more than would be optimal). Midpoint split-times (which are available for hundreds of thousands of marathon participants worldwide) can thus provide a proxy for OC, an identification strategy that we employ to further investigate effects of gender and age and additionally nationality and ethnic background. We can thereby link overconfidence to culturally-shaped variables such as religious beliefs and political views.

In Studies 1 and 2 we confirm previous results of higher overconfidence in males and (in Study 2)—in Asians. On top of that, however, in Study 2 we deliver several new results, in particular being able to link OC with both very young and old age, as well as conservative, masculine cultures.

Literature review

Mechanics of overconfidence

Two major approaches to the question why people are overconfident can be found in the literature. The first concerns possible *functions* of such a bias. *Prima facie*, overconfidence represents a puzzle – how could it prove beneficial to hold systematically biased view of oneself, which is likely to occasionally result in a spectacular failure? Bénabou and Tirole (2002) provide one answer, arguing that OC may be adaptive as a motivational factor when time inconsistency is present. In other words, it may be useful in making us work hard when we should but do not want to. These authors plausibly assume that skill and effort are typically complements. Thus more positive view of oneself leads to greater effort. To the extent that insufficient effort would otherwise arise due to present bias (aka quasi-hyperbolic discounting), OC may indeed be adaptive.

The second approach is to focus on psychological antecedents and correlates of OC. One important factor is the self-serving attribution bias (SSAB) (Miller, Ross 1975)–the ego-protecting tendency to attribute failures to external conditions, such as luck or actions of others, and successes to internal conditions such as own skill and hard work. For example, Gervais and Odean (2001) showed that the traders in their sample were becoming overconfident by self-servingly updating their beliefs. Importantly, in view of our sample, studies also found such effects for athletes, although they generally seem to display internal locus of control (Scanlan, Passer 1980). The SSAB, or the "need to view [oneself] positively" (Heine, Lehman et al. 1999 p. 766) is ubiquitous: in their meta-analysis of 266 studies, (Mezulis, Abramson et al. 2004) found a huge mean effect size of $d = 0.96$. They also concluded that SSAB tends to decline slightly with age, especially in women (while there was no main effect of gender). Interestingly, Asians seemed to be less susceptible to the bias than Westerners.
While neuroscientific understanding of the SSAB is far from perfect, Krusemark et al. (2008) associated non-self-serving attributions with activity in the medial prefrontal cortex, an area believed to be responsible for cognitive control. These findings suggest that SSAB comes naturally and overcoming it is a matter of conscious cognitive effort (Coalson 2014).

Demographic effects

Turning to the question *who* is prone to OC, several studies showed that males tend to be more self-confident than females, at least for tasks that are considered ``masculine'' (Beyer, Bowden 1997, Lundeberg, Fox et al. 1994, Lenney 1977). This in particular is true for sports: a meta-analysis (Lirgg 1991) revealed a substantial gender effect. The mere fact that a substantial majority of participants in the events we are considering are male suffices to conclude that running is likely to be a male thing. On the other hand, running provides immediate, direct feedback. That might reduce bias and gender gap (Lenney 1977), although Beyer (2002) finds the opposite.

Much less is known of the impact of age on confidence, although population aging makes research on changes in decision making over lifespan highly relevant. The natural hypothesis is that wisdom, presumably acquired over long time, lies partly in the knowledge of one's limitations, as in "I know that I know nothing" ascribed to the archetypical sage of the ancient world, Socrates. It would also seem that some typical sins of the young age, such as driving too fast and under influence of alcohol, have something to do with excessive confidence in one's abilities. On the other hand, the elderly may anchor too much on the superior physical and mental abilities they once had, which would also result in overconfidence.

Systematic evidence seems scarce, even if one looks for findings concerning all types of overconfidence. One relevant study is Kovalchik et al. (2005) which found healthy elderly (average age 82) better calibrated (i.e. better able to estimate the probability that they are right) than young adults (average age 20). Conversely, Hansson et al. (2008) observed overconfidence (in interval judgment) increase over lifetime, an effect partly mediated by deteriorating cognitive abilities, see also Crawford and Stankov (1996). Likewise, in their study of Canadian investors Bhandari and Deaves (2006) found greater overconfidence (judged probability that one's answer to question testing financial knowledge is correct minus actual fraction of correct answers was higher) closer to retirement age. Bruine de Bruin et al. (2007) and Charness and Villeval (2007) reported rather mixed results concerning dependence between age and confidence and Dahl et al. (2009) found no effect on calibration in their sample of 60-93 year olds. Likewise, Chen et al. (2007) reported no impact of age on number of stocks or trading volume (both of which are linked to overconfidence) in Chinese investors.

Experience and skill effects

A variable that is intricately related to age is experience. While some research exists exploring its link with OC, it has often suffered from lack of relevant background data and difficulty in making time-pressured professionals respond to a sufficiently extensive survey in a controlled environment. These problems can be illustrated by Menkhoff et al. (2006). These authors found that experienced fund managers were more likely to deem their performance as "above average"; because Menkhoff and colleagues did not have meaningful controls for actual performance of their subjects, it is hard to conclude that these claims were indeed unsubstantiated. Likewise Menkhoff et al. (2013) found that more experienced investors tended to provide smaller confidence intervals for future values of stock market indices. Again, with just two intervals per subject it is impossible to blame it on greater overprecision.

Lambert et al. (2012) had 10 subjective probability judgment items allowing to identify miscalibration. They also measured the better than average effect and illusion of control. However, the fact that no significant difference was found between students and professionals (loan officers) could partly be due to the fact that only 20 representatives of the latter group could be recruited. The authors also had to use different items for the two groups.

McKenzie et al. (2008) did better in this respect, reaching 43 IT professionals to compare with their student sample. They also cleverly played with knowledge level – some questions concerned the IT and some the University of California San Diego, so that the students would become experts in the latter case, while professionals would be novices. However, the authors used different procedures – professionals filled in the survey on their own, so one cannot be sure they did not check the Web for correct answers. Results showed that experts provided narrower confidence intervals than novices, but this corresponded to their superior actual knowledge (midpoints were closer to the true value) so that hit rates were identical.

Glaser et al. (2010) recruited a sizable sample of investment bankers, who were found to be more overconfident in some tasks than student controls. Again, the study involved no incentives and was conducted online (so that subjects could e.g. easily cheat on general knowledge questions).

Overall, no clear and ultimately reliable picture of the impact of experience on overconfidence seems to emerge. The same is largely true of another important determinant of performance, i.e. skill. In a much-cited study Kruger and Dunning (1999) suggested that unskilled individuals suffer from a "dual burden" – not only do they perform poorly but additionally fail to recognize that. However, these findings have been strongly contested (Krueger and Mueller 2002, Burson et al. 2006) . In particular, it seems they may to a large extent be attributed to statistical regression to the mean.

Cross-cultural effects

In a series of studies Yates and colleagues (e.g. Yates, Lee et al. 1996) found that Asian, particularly Chinese subjects tend to show greater confidence in general knowledge questions than Westerners. These authors also provided some evidence that this overconfidence does not merely reflect a specific response style. Indeed in Yates et al. (1997) Asian subjects were willing to act upon their false beliefs (although stakes were miniscule). By contrast, Americans' judgments inferred from such choices were just as strongly overconfident as those of the Asians, suggesting that their relatively modest direct responses were shaped by self-presentation style.

Acker and Duck (2008) compared British and East Asian undergraduates in a different task: they asked them to predict their relative performance (and assess to what extent it would be driven by their skill vs. luck and other external conditions) in a stock-market game. These subjects were also asked to predict their own grades. Neither of the tasks was incentivized. Again, Asian students showed substantially greater overconfidence.

Very few studies went beyond the East Asia vs the West comparison. Lundeberg et al. (2000) asked students in five countries for confidence judgments concerning items in an exam they were taking. Palestinian students were strongly over-confident, while Taiwanese showed modesty when actually wrong (good ``discrimination''), with Israeli, the Dutch and the Americans in

between. Using similar methodology Puncochar et al. (1996) found Israeli students to be more overconfident than those in the United States.

In studies focusing on the SSAB, Americans were found to be more prone to it than Finns (Nurmi 1992) and Indians more than Canadians (Higgins, Bhatt 2001). The Japanese could be one of the very few cultures not affected by the bias (Heine, Lehman 1997), although (Aldridge, Islam 2012) found no difference between Japanese and Australian athletes.

We are not aware of a study with large, heterogeneous sample encompassing many nationalities, which could allow establishing which of many differences between countries might drive the difference.

Politics

Very few studies addressed the link between overconfidence and political opinions. One can speculate that voters overconfident about their own prospect earnings will tend to oppose income redistribution, because they do not expect to benefit from it. Another potential factor contributing to the association between conservatism and overconfidence may be religious beliefs. Indeed, religious individuals tend to be right-wing voters while their hope for divine interventions (especially benefitting the believers) may boost confidence. Jost et al. (2003) associated both such biases as unreasonable optimism and overconfidence and conservative political views with what they call *fair-market ideology*.

Recently Ortoleva and Snowberg (2013) used a sample of three thousand US voters to find how political views correlate with overconfidence. The latter was inferred from self-declared confidence about and actual accuracy of responders' estimates of current and future rates of inflation and unemployment. They found that increasing such overprecision by one standard deviation corresponded to a shift towards more conservative views by roughly 20% of a standard deviation. It was also associated with greater variance, so that particularly the extreme right tended to be predominantly overconfident.

Related studies using data from distance running and other sports

Allen et al. (2013) recently reported analyses using over nine million observations from several running events. They focused on goal setting, finding that marathon finishing times tend to bound at round numbers, suggesting aversion to failing to meet the target (being stronger than satisfaction about exceeding it by the same margin). This tendency seemed particularly strong for males. The importance of goals was also corroborated by Sackett et al. (2013), who found that marathoners asked to set a goals run almost seven minutes faster than the control group who received no such request.

Also relevant to our findings, there are some studies in sports literature suggesting that pacing may differ between genders in long-distance running (Allen and Dechow 2013). The debate as to what could be causing this effect seems to have been inconclusive so far, see Deaner et al. (2014) and studies cited therein. In particular, there are no physiological differences that could clearly account for the effect (prof. Jerzy Żołądź, Dep. of Physiology and Biochemistry, University School of Physical Education, Kraków, Poland – personal communication).

Gender differences in strategic choices may be easier to spot in other disciplines. In particular, Böheim and Lackner (2013) investigated the frequency with which male and female athletes choose to pass a height in high jump and pole vault events (i.e. wait until the bar is raised). Controlling for skill and stamina, passing a height is a relatively risky, self-confident and competitive choice in these sports. The authors find that females choose this option less often than males and less often than would have been optimal for their final results. There is also some recent evidence of relatively conservative female choices in chess (Gerdes, Gränsmark 2010) and tennis (Paserman 2007, Anbarci et al. 2014).

Study 1

In Study 1 participants of the 2012 Warsaw Marathon were asked to take part in a contest involving predicting their own finishing times. They were told that prizes would be awarded to the contestants whose actual finishing times are closest to the prediction. The winners would receive a quality T-shirt with their name and finishing time imprinted on it.

Note that under this reward scheme it is quite easy to maximize the probability of winning. One can simply make a prediction that is much greater (i.e. worse) than the time needed to complete the course and wait just before the finish line to cross it in the right moment. Less extreme strategy would involve deliberate slowing down in the latter stage of the run. This is a clear case of moral hazard, whereby the prizes go to those altering their behavior rather than those providing a good prediction.

On the other hand, the nature of the prize gives some incentives to make a much more ambitious prediction. Given the reward function and the ratio of the number of prizes to the number of participants it is easy to conclude that the finishing time printed on the shirt will be equal or almost equal to the predicted time. Therefore the prediction is also an indirect decision on what time will be printed on the shirt. It seems obvious that a shirt with good finishing time (whereby "good" will mean different things for different runners, depending on their age, gender, experience, skill, and ambition) is more attractive than one with a mediocre one.

It is impossible to identify and measure the two effects described above. The method is thus not incentive compatible but it does give some reasons to make a balanced prediction. Most importantly, it provides analogous incentives to different demographic groups, thus providing a basis for meaningful comparisons between them. Moreover, looking at the results one can easily obtain a rough indication of the prevalence of the "moral hazard" problem described above – if only few people finish right on time (or +/- a second or two), then we may be fairly sure that the issue is of minor importance.

Samples and procedures

Study 1 took place at the 2012 Warsaw Marathon. All 6800 runners were invited to take part in the contest sponsored by Adidas by providing their forecasts during two days before the race. A total of 345 participants[5], of whom 20% were women, took part in the contest and competed for 50 prizes. The age of participants ranged from 18 to 69, with a mean of 36.9 and standard deviation of 9.5.

Results

We analyzed the error of forecast in both absolute (AFE) and relative terms (RFE). Our measures are defined as follows:

$$\text{AFE} = \text{actual time} - \text{forecasted time}$$
$$\text{RFE} = (\text{actual time} - \text{forecasted time})/(\text{forecasted time})$$

Both measures take positive values when runner finished the race slower than he or she predicted. If the values are systematically higher in one group than another, we shall refer to it as showing more self-confidence. Interpreting positive values as overconfidence and negative as underconfidence is clearly not justified at individual level. Even at group level it is slightly problematic, as there may be factors (such as weather) that unexpectedly affect performance of all runners. Figure 1 shows the distribution of AFE. It has positive skew with mean (median) equal to 630 (358) and standard deviation of 1282.

The concerns that many participants give a non-ambitious forecast and slow down intentionally, thereby hitting their target time, appear unsubstantiated – there is no spike near zero forecast error. However there are obvious outliers in the data set; this is not surprising given that there was no direct penalty (other than a prize missing) even for the most absurd predictions. It seems natural to take values of equal to -2000 and 4000 as the cut-off points, because only isolated observations are beyond this interval. This procedure results in the removal of some 3% of all entries.

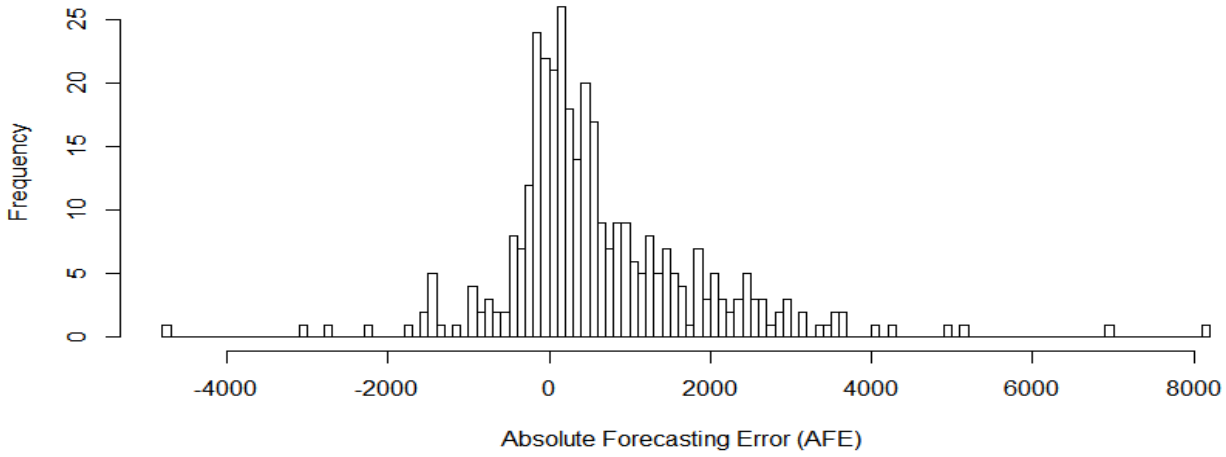

Figure 1. Distribution of AFE among runners in the Warsaw Marathon.

[5] The total number of the contest participants was equal to 650, but only for 345 of them it was possible to get additional information (age, gender, split time). It is difficult to tell how many participants were actually aware of the contest and made a conscious decision not to participate.

In Table 1 we show results of OLS regressions of OC measures on available demographics and split times (how many seconds it took the runner to cover the first 21.1 km/13.1 miles). In models (1) and (2) we take the AFE as the dependent variable, while it is the RFE in models (3) and (4). Models (1) and (3) are run on the entire sample, while outliers are removed in (2) and (4). Our results strongly indicate that male runners tend to be more confident than females. Absolute gender gap is approximately equal to 10 minutes and in relative terms it is around 4%. The effect of age on the forecasting error is weak and nonlinear, with middle-aged individuals (between 30 and 40, depending on the specification) seeming to show least confidence. The interpretation of the coefficient that stands next to the split time is less straightforward. For example, it is quite possible that the probability of an unforeseeable negative shock to performance differs between faster and slower runners. If such effects are relatively weak, it would seem that lower skill is associated with greater OC, as the coefficient next to split time is positive and statistically significant, even when RFE is used as a dependent variable (with outliers removed in specification 4). The findings are thus highly consistent across all four specifications.

Table 1. Effects of runners characteristics on forecasting error

Dependent variable	AFE		RFE	
Specification	(1)	(2)	(3)	(4)
male	624.197*** (175.181)	544.111*** (137.197)	0.10486** (0.102)	0.10369*** (0.101)
age	-71.154* (42.16)	-82.136* (34.101)	-0.10115** (0)	-0.10056* (0)
age^2	0.184 (0.154)	0.195* (0.143)	0.10001* (0)	0.10001* (0)
21 km split	0.121** (0.107)	0.123*** (0.105)	0.1000007 (0)	0.100001** (0)
const	106.165 (965.194)	194.131 (767.191)	0.11996* (0.11)	0.10547 (0.105)
outliers cut	NO	YES	NO	YES
R^2	0.10448	0.10784	0.10324	0.10545

Notes. N=345, std. errors in parentheses. Signif. codes: '***' – 0.1%; '**' – 1%; '*' – 5%

To set stage for Study 2 we checked whether a good proxy for OC could be found to be used in races, in which no data on forecasts is available. The change in pace is a natural candidate as mentioned before – overly confident individuals are likely to start too fast and be forced to slow down substantially. Just as with the forecasting error the change of pace can be expressed in absolute (Absolute Pace Change, APC) and relative terms (Relative Pace Change, RPC):

$$APC = \text{time at finish} - 2(\text{21 km split})$$
$$RPC = (\text{time at finish} - \text{21 km split})/(\text{21 km split})$$

Both measures increase when runners slow down in the second half of the race. Physiology literature suggests that optimal performance at marathon can be achieved with steady pace sustained for the whole duration of the race (Joyner 1991) which corresponds to APC being close to zero and RPC close to one. It is also safe to assume that runners are aware of that, as nearly any relevant book or website cautions the runner not to start too fast. A substantial fraction of significant slowdowns is thus expected to follow from overconfident judgment of own capabilities resulting in overly fast initial pace.

This is confirmed by correlations between our measures of forecast error and pace change in the 2012 Warsaw Marathon data shown in Table 2. They never drop below 0.68 in the general population. The correlations are somewhat weaker for women (which may be a random occurrence in view of the lower number of observations in this group) yet remain well above .5. We see these values as remarkably high, given that many unpredictable events, such as minor injuries, will weaken the link between our two dimensions of interest. Hence we conclude that the data strongly supports the use of pace change, measured by APC or RPC, as a proxy for forecasting error.

Table 2. The correlation between forecasting error and pace change variables

	AFE			RFE		
	All	Male	Female	All	Male	Female
APC	0.7137	0.7311	0.5756	0.6881	0.7066	0.5369
RPC	0.6965	0.7094	0.553	0.6977	0.7125	0.5294

Study 2

To obtain a large number of observations, in Study 2 we gathered data from three major marathon series: Chicago, New York and London. The marathons were chosen based on the following criteria: data availability, size, low altitude, flat course and international popularity. The most limiting factor was data availability, as for many marathons the split times are not readily available (although could in principle be obtained from the organizers). However, the three locations we include already provide a data set of more than one million observations, 38% of which are females. There are around 480 000 observations from Chicago, 430 000 from New York and 140 000 from London. We also have 85 000 additional observations from three smaller marathon series: Orlando, Houston and Hamburg. The oldest observations come from 1998 and the most recent from 2014, although time spans are different for each locations. Runners' age ranges from 14 to 80 years[6] with mean (median) equal to 39.6 (39) and standard deviation of 11. Cultural diversity can also be studied, as we have 69 nationalities for which there are more than 100 observations (and 56 nationalities with more than 200 observations).

Distribution of the dependent variables

Figure 2 shows 21km split[7] and final times for all the runners. Both have distinct spikes at round values; they are particularly pronounced in the latter (for example at 3h=10800s and 4h=14400s).[8] Other than that, they are close to symmetry, with mean equal to 7700s and 16300s for 21 km split and overall time respectively.

The resulting distributions of our two proxies of OC are shown in Fig. 3. It turns out that as many as 85% of participants run the second half of the course slower than the first (a so-called positive split). Both APC and RPC are asymmetrically distributed with positive skew. The mean (median) value of APC is 884 (719) seconds and 112 (110) for RPC.

[6] A handful of observations outside of the depicted range have been excluded as they are quite likely to be mistakes. This has no bearing on results.

[7] We use this most common term, although more precisely it is the time at 21 097.5 m, i.e. half the distance.

[8] As mentioned before, this effect has been described in more detail by (Allen, Dechow et al. 2013).

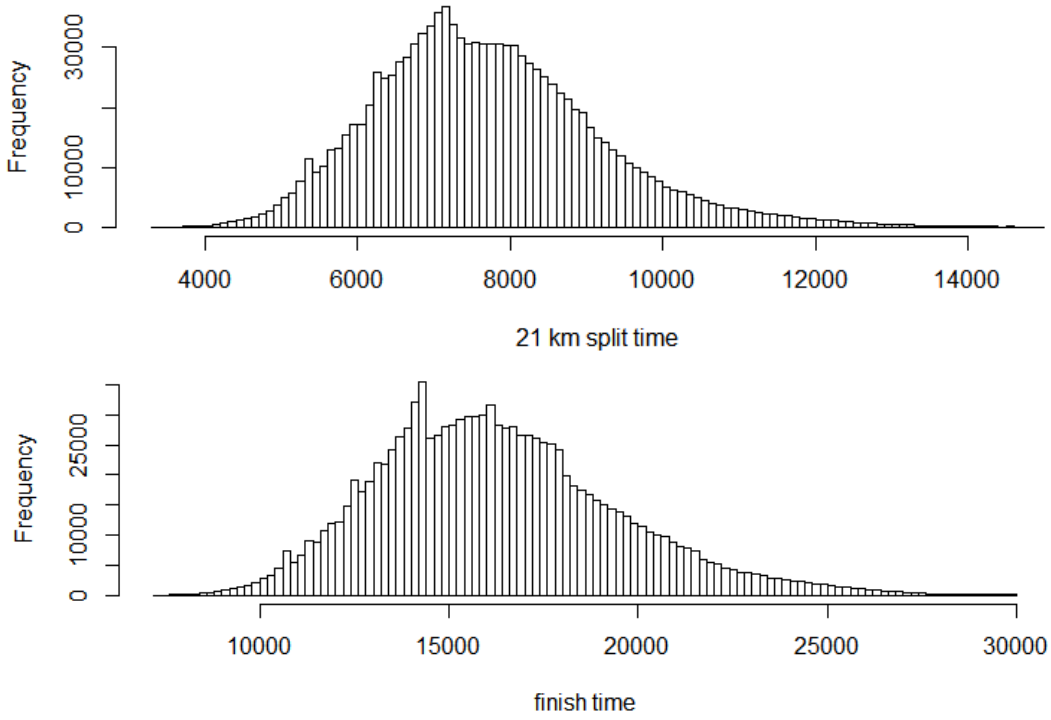

Figure 2. Distributions of 21km split and final times

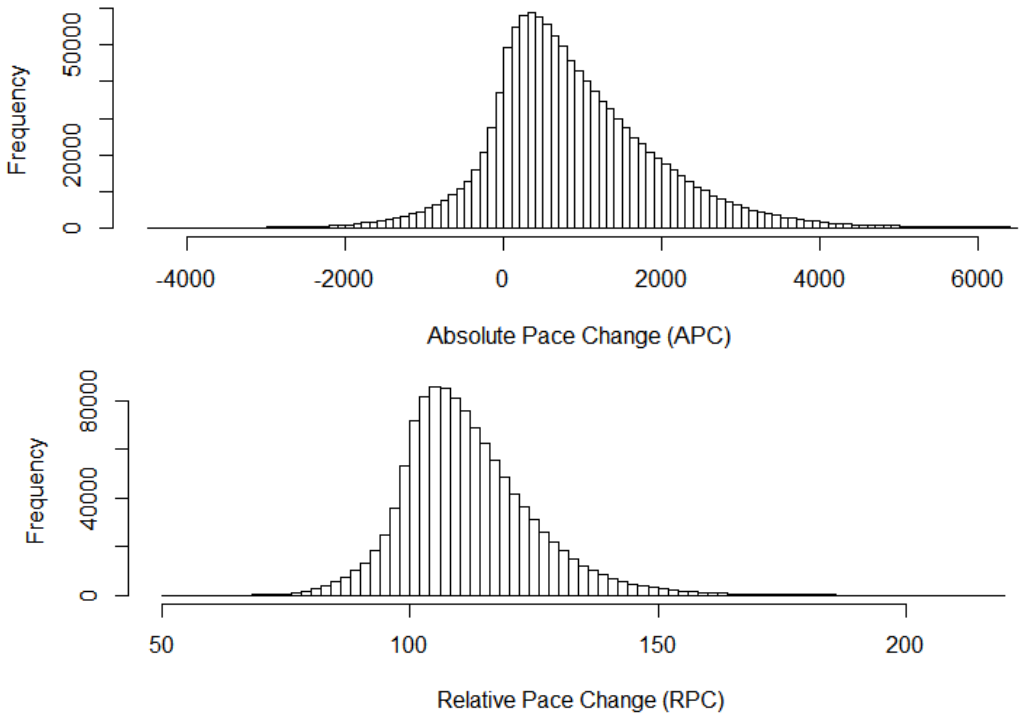

Figure 3. Distribution of the two measures of pace change.

Table 3. OLS regressions of the APC

Specification	1	2	3	4	5	6	7	8
male	277.03***	273.37***	287.97***	288.09***	327.25***	330.98***	289.29***	331.75***
	(19.92)	(12.99)	(22.07)	(20.36)	(20.6)	(18.67)	(20.75)	(18.97)
age 14-21	182.05***	173.02***	165.8***	165.71***	119.26***	116.09***	175.71***	119.38***
	(16.68)	(13.04)	(10.33)	(9.41)	(20.45)	(16.61)	(11.17)	(19.46)
age 21-30	72.98	43.7	29.24***	26.82***	33.34**	28.35**	57.31	30.9**
	(51.52)	(23.61)	(4.65)	(2.82)	(12.93)	(10.14)	(35.53)	(11.65)
age 41-50	192.18	120.87	59.88**	69.72***	35.25	46.04*	180.86	42.31*
	(129.13)	(68.69)	(21.48)	(21.09)	(21.51)	(20.44)	(105.03)	(20.66)
age 51-60	263.79*	214.92***	143.48***	161.99***	100.65***	120.85***	267.31**	113.93***
	(111.81)	(61.34)	(20.64)	(18.46)	(24.52)	(21.95)	(94.28)	(22.38)
age 61-70	342.16***	296.62***	232.56***	246.37***	171.93***	187.62***	349.47***	184.61***
	(90.34)	(46.12)	(10.3)	(9.85)	(15.48)	(15.17)	(77.06)	(15.04)
age 71-80	692.12***	522.22***	380.72***	391.48***	323.55***	337***	660.27***	333.85***
	(140.84)	(78.18)	(19.99)	(20.83)	(24.32)	(22.98)	(101.61)	(22.89)
age 14-21 (male)	32.31*	29.32	36.64***	35.85***	34.57***	31.67**	26.87	29.48**
	(13.21)	(17.23)	(9.01)	(10.45)	(9.55)	(10.75)	(15.71)	(10.36)
age 21-30 (male)	186.82*	71.35**	10.87	10.4	-20.86	-23.53	146.5***	-25.98
	(72.88)	(25.76)	(26.26)	(23.21)	(41.39)	(38.74)	(42.02)	(40.96)
age 41-50 (male)	-186.15	-110.05	-70.15*	-69.52*	-66.51*	-66.68*	-163.38	-66.15*
	(121.36)	(63.88)	(30.66)	(29.13)	(28.83)	(27.11)	(95.16)	(27.73)
age 51-60 (male)	-241.24*	-172.92**	-132.22***	-133.66***	-119.07***	-122.57***	-223.84*	-120.96***
	(112.33)	(55.98)	(31.09)	(27.32)	(32.59)	(28.45)	(87.08)	(29.86)
age 61-70 (male)	-224.59*	-162.06**	-124.87***	-127.68***	-111.59**	-116.27***	-211.27*	-114.62***
	(107)	(53.31)	(32.47)	(28.91)	(35.75)	(31.99)	(84.11)	(33.16)
age 71-80 (male)	-396.06**	-229.83*	-121.8*	-132.1*	-138.85*	-151.13**	-350.73**	-144.23**
	(150.28)	(91.46)	(59.87)	(57.47)	(53.99)	(49.77)	(109.37)	(50.41)
21kmSplit	0.5384***	0.5424***	0.5128***	0.5269***	0.4901***	0.5001***	0.5353***	0.494***
	(0.0506)	(0.0491)	(0.048)	(0.0485)	(0.0512)	(0.0506)	(0.0498)	(0.0522)
21kmSplitSQ	-29.18E-6***	-29.54E-6***	-27.11E-6***	-28.10E-6***	-24.45E-6***	-25.26E-6***	-29.28E-6***	-24.80E-6***
	(2.25E-6)	(1.94E-6)	(1.93E-6)	(1.80E-6)	(2.26E-6)	(2.06E-6)	(2.11E-6)	(2.20E-6)
tempC	27.78***	26.67***	27.33***	27.42***	---	---	27.57***	---
	(1.83)	(2.63)	(2.97)	(2.93)			(2.11)	
tempC^2	2.81***	3.42***	3.49***	3.48***	---	---	2.77***	---
	(0.82)	(0.36)	(0.23)	(0.23)			(0.72)	
male*tempC	8.93***	8.48***	8.22***	8.24***	---	---	8.78***	---
	(1.82)	(2.06)	(2.15)	(2.11)			(1.86)	
const	-1803.69***	-1831.75***	-1732.2***	-1778.05***	-1363.39***	-1359.25***	-1703.68***	-1403.17***
	(248.51)	(278.63)	(267.06)	(273.34)	(266.4)	(277.64)	(225.52)	(292.18)
nationality dummy	NO	YES	NO	YES	NO	YES	NO	NO
location dummies	NO	NO	YES	YES	NO	NO	NO	NO
year*location dummies	NO	NO	NO	NO	YES	YES	NO	YES
GH variables	NO	NO	NO	NO	NO	NO	YES	YES
R^2	0.075	0.093	0.096	0.101	0.181	0.186	0.082	0.183

Notes. male*tempC was standardized to have zero mean. so that the estimates for male are comparable across all specifications.
N=1150990 (spec. 1-6). 1145211 (spec. 7-8). std. errors clustered by nationality in parentheses. Signif. codes: '***' – 0.1%; '**' – 1%; '*' – 5%

Table 4. OLS regressions of the RPC

Specification	1	2	3	4	5	6	7	8
male	3.52*** (0.21)	3.47*** (0.14)	3.63*** (0.22)	3.63*** (0.2)	4.09*** (0.22)	4.14*** (0.19)	3.66*** (0.21)	4.14*** (0.19)
age 14-21	2.03*** (0.22)	1.92*** (0.17)	1.88*** (0.14)	1.88*** (0.13)	1.32*** (0.27)	1.28*** (0.22)	1.95*** (0.15)	1.32*** (0.26)
age 21-30	0.81 (0.59)	0.46 (0.25)	0.32*** (0.06)	0.29*** (0.02)	0.36* (0.15)	0.3** (0.11)	0.62 (0.4)	0.33* (0.13)
age 41-50	2.36 (1.54)	1.52 (0.82)	0.81** (0.28)	0.94*** (0.28)	0.51 (0.27)	0.65* (0.27)	2.22 (1.25)	0.6* (0.27)
age 51-60	3.33* (1.35)	2.76*** (0.75)	1.92*** (0.27)	2.16*** (0.26)	1.41*** (0.32)	1.67*** (0.3)	3.38** (1.13)	1.58*** (0.3)
age 61-70	4.31*** (1.12)	3.77*** (0.59)	3.01*** (0.16)	3.19*** (0.18)	2.29*** (0.24)	2.49*** (0.26)	4.4*** (0.96)	2.45*** (0.25)
age 71-80	8.24*** (1.6)	6.22*** (0.85)	4.58*** (0.21)	4.72*** (0.24)	3.89*** (0.26)	4.06*** (0.26)	7.86*** (1.12)	4.02*** (0.25)
age 14-21 (male)	1.02*** (0.18)	0.99*** (0.23)	1.07*** (0.13)	1.06*** (0.14)	1.09*** (0.13)	1.05*** (0.14)	0.97*** (0.21)	1.03*** (0.15)
age 21-30 (male)	2.4** (0.88)	1.03** (0.32)	0.34 (0.28)	0.33 (0.24)	-0.03 (0.46)	-0.07 (0.42)	1.92*** (0.51)	-0.1 (0.45)
age 41-50 (male)	-2.32 (1.45)	-1.4 (0.76)	-0.95* (0.38)	-0.94* (0.37)	-0.9* (0.36)	-0.9** (0.34)	-2.04 (1.13)	-0.89* (0.35)
age 51-60 (male)	-3.06* (1.33)	-2.24*** (0.65)	-1.77*** (0.38)	-1.79*** (0.33)	-1.62*** (0.39)	-1.67*** (0.34)	-2.85** (1.02)	-1.64*** (0.36)
age 61-70 (male)	-2.88* (1.25)	-2.13*** (0.61)	-1.7*** (0.39)	-1.74*** (0.34)	-1.55*** (0.42)	-1.61*** (0.38)	-2.71** (0.97)	-1.58*** (0.39)
age 71-80 (male)	-4.95** (1.59)	-2.96*** (0.88)	-1.73** (0.56)	-1.85*** (0.52)	-1.92*** (0.51)	-2.07*** (0.44)	-4.39*** (1.09)	-1.98*** (0.46)
21kmSplit	37.36E-4*** (7.41E-4)	37.62E-4*** (7.42E-4)	34.15E-4*** (7.36E-4)	35.53E-4*** (7.61E-4)	31.63E-4*** (7.56E-4)	32.54E-4*** (7.68E-4)	36.60E-4*** (7.34E-4)	31.68E-4*** (7.79E-4)
21kmSplitSQ	-25.94E-8*** (3.63E-8)	-26.27E-8*** (3.49E-8)	-23.40E-8*** (3.42E-8)	-24.45E-8*** (3.51E-8)	-20.40E-8*** (3.70E-8)	-21.19E-8*** (3.71E-8)	-25.84E-8*** (3.55E-8)	-20.60E-8*** (3.78E-8)
tempC	0.356*** (0.017)	0.342*** (0.026)	0.35*** (0.031)	0.351*** (0.03)	---	---	0.353*** (0.02)	---
tempC^2	0.035*** (0.01)	0.042*** (0.004)	0.045*** (0.003)	0.044*** (0.003)	---	---	0.034*** (0.009)	---
male*tempC	0.16*** (0.02)	0.15*** (0.03)	0.15*** (0.03)	0.15*** (0.03)	---	---	0.16*** (0.02)	---
const	94.62*** (3.34)	94.4*** (3.73)	95.38*** (3.64)	94.98*** (3.75)	100.08*** (3.63)	100.31*** (3.76)	95.83*** (3.11)	99.67*** (3.94)
nationality dummy	NO	YES	NO	YES	NO	YES	NO	NO
location dummies	NO	NO	YES	YES	NO	NO	NO	NO
year*location dummies	NO	NO	NO	NO	YES	YES	NO	YES
GH variables	NO	NO	NO	NO	NO	NO	YES	YES
R^2	0.074	0.090	0.092	0.097	0.167	0.172	0.081	0.170

Notes. male*tempC was standardized to have zero mean, so that the estimates for male are comparable across all specifications.
N=1150990 (spec. 1-6), 1145211 (spec. 7-8), std. errors clustered by nationality in parentheses. Signif. codes: '***' – 0.1%; '**' – 1%; '*' – 5%

Estimation results: individual effects

We investigate the determinants of our dependent variables in an Ordinary Least Squares model with clustered standard errors by runners` nationalities, see Table 3 for APC and Table 4 for RPC. All of our specifications include the key individual variables – gender, age as well as 21km split (and 21km split squared). In subsequent specifications we additionally account for fixed effects of each nationality (2) and marathon location (3) or both (4).[9] Specifications (1) through (4) also include mean temperature during the race (and temperature squared) as well as its interaction with the male gender. Further specifications have fixed effects for each specific run (year-location dummies), with (6) or without (5) nationality effects. Models (7) and (8) resemble (2) and (5) respectively, except that effects of Hofstede cultural dimensions are estimated instead of national fixed effects – we will come back to this analysis in the next subsection.

Just as we did in Study 1, we find a strong and robust gender effect. On average, men slow down in the second half of the run more than women do. This effect persists across all specifications and both dependent variables. In absolute terms it ranges between 4.5 and 5.5 minutes for an average runner and in relative terms between 3 and 4 pct. points. These effects are substantial, given that overall mean APC and RPC (for women) are 13.2 minutes and 110% resp.

The non-linear and gender-specific effect of age is represented in Figure 4. We find that oldest participants are most confident, with APC (RPC) up to some 400 s (5%) higher than in younger individuals, depending on the specification. Yet youngest participants are also overconfident, resulting in a U-shaped pattern, another confirmation of Study 1 results. The lowest point on the resulting parabola depends on the model and the gender. In our preferred specifications including year/place dummies the minimum is at the age of 35 for women and 45 for men (model 6). Additionally the U-shape is flatter for men than for women. Average effects of APC and RPC for men are very similar in all age groups from 21 to 60 years old. The variability of APC and RPC is higher for women and statistically significant even between age groups 21-30 and 31-40. With these estimates the average difference of relative pace between a 35 year old and a 75 year old woman is around 3 minutes (1.5 pct. points). It is also interesting to note that the gender difference is strongest for youngest individuals – adolescent males are unusually self-confident. By contrast, the gender gap is very modest in the elderly.

[9] Although all the marathons have similar characteristics (flat course, low altitude, etc.) they still differ in some ways like the layout of the course or average humidity. When we control for these differences with marathon-specific dummies (models (3) and (4)) we see that on average the dependent variable is almost 6 minutes higher in London then in Chicago. By contrast, there is no statistically significant difference of runners pace between New York and Chicago. Overall, judging from R2 values, these dummies play a very significant role in explaining variance of the independent variables.

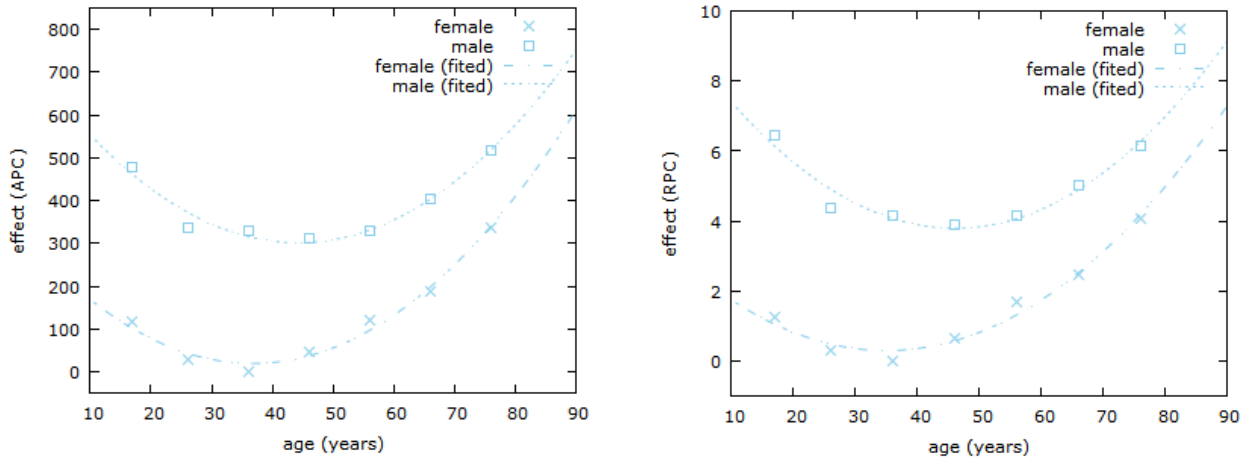

Figure 4. Age effects for APC and RPC in specification 6.

As for the relationship between our general measure of performance – the 21 km split – and pace change we find it to take an inverted U-shape with maximum value for around 10 000 seconds (94th centile of 21 km split time variable) when APC is the dependent variable and 7 600 seconds (50th centile) for RPC, see Figure 5. In absolute terms the 21 km split is responsible for up to 13 minutes of pace change variability and around 2.5 pct. points in relative terms[10]. Thus in absolute terms slower runners change their pace more than more expert runners. The effect is similar for relative pace change as the value of RPC is on average the highest for runners who finish first half of the course in around two hours. One must note that a brisk walker can have a 21 km split time of 9000 seconds. Therefore apparent lower overconfidence in slowest runners must be treated with some caution – they could just hardly slow down any more.

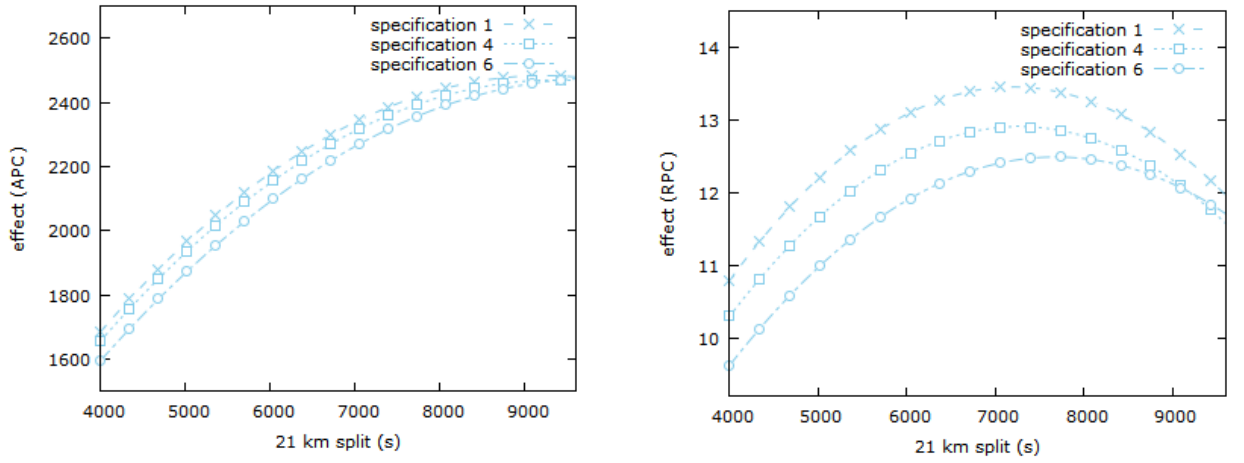

Figure 5. Twenty-one km split times effects for APC and RPC in specifications 1, 3 and 6.

[10] In what follows we continue to refer to APC estimates in minutes and RPC estimates in pct. points.

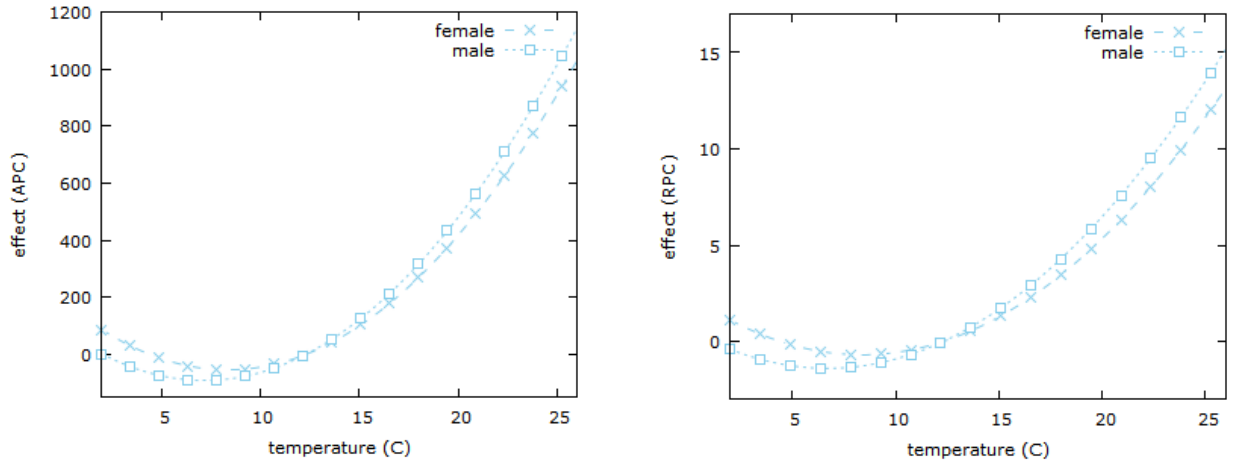

Figure 6. Temperature effects for males and females for APC and RPC in specification 6.

We turn now to the impact of weather conditions. It turns out that the way in which temperature affects runners differs between the sexes. Figure 6 shows that both genders run relatively slower in the latter part of the run for temperatures above 10 degrees centigrade or so. What is noteworthy though, this effect is stronger for males: with every centigrade men slow down by additional 9 seconds (0.16 pct. point) compared to women, a strikingly consistent estimate across all specifications. This difference widens the gender gap by additional 3 minutes (3.84 pct. points) comparing the highest against the lowest temperature in our sample (25 vs. 7 centigrade).

Interestingly, this pattern is essentially identical to the way in which temperature affects overall performance of amateur runners: overly low temperatures make little difference, but overly high ones are strongly detrimental and the effect might be slightly stronger for males (El Helou, Tafflet et al. 2012). Together, these observations suggests that runners, especially male runners, underestimate how adverse condition will affect their performance. Indeed, the absolute value of the negative correlation of pace with temperature is nearly twice as half in the latter half than in the former (ca. 18% vs. ca. 9%).

Study 2: National and cultural effects

We turn now to the role of nation-level, cultural variables in explaining the relative pace of runners. When we add nationality dummies to the model (taking the USA represented by the largest number of runners as the base group), then the values of coefficients range from -5,05% to 4.35%, see Figure 7. These effects are jointly significant at 0,0001% and 56 out of 68 of them are individually significant at 0.1%.

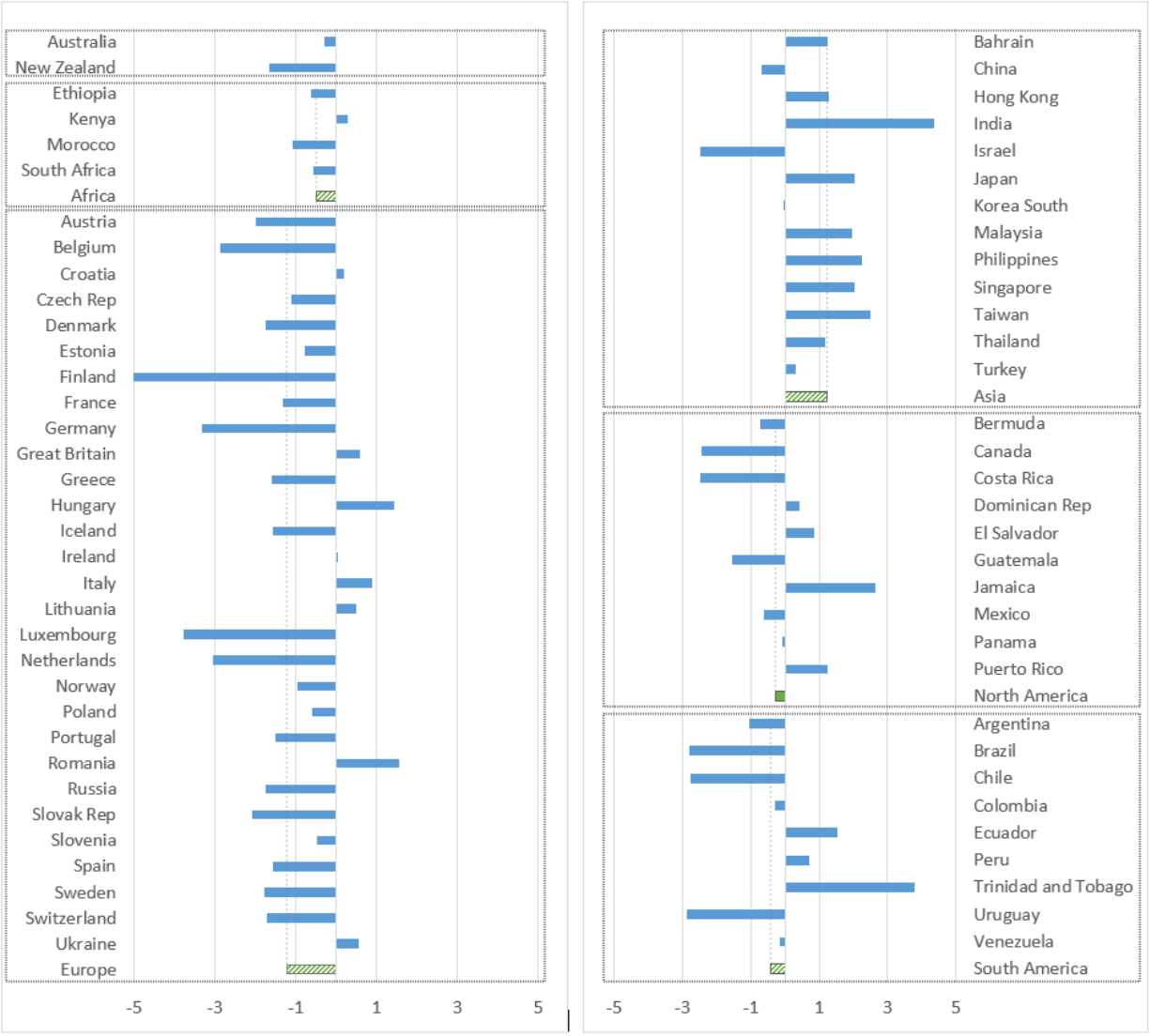

Figure 7. National effects for RPC.

Our results confirm the previously reported finding that (East) Asian cultures promote OC. Indeed, with the exception of China and South Korea, all Central and East Asian dummies are positive, large and significant, while other continents appear more diverse.

The estimates associated with specific nationalities do not immediately allow us to understand what cultural characteristics could be linked to high or low confidence. For this reason, models (7) and (8) include interpretably country-specific variables, rather than individual indicators.

Recognizing their importance in the literature on cultural differences (and esp. cross-cultural management) we rely on Geert Hofstede's (henceforth GH) cultural dimensions in this respect. There are four of them in the original formulation of the framework (Hofstede 1983).

Individualism-collectivism corresponds to the degree to which "people in a country prefer to act as individuals rather than as members of groups" (Hofstede, 1994, p. 6). Members of individualistic societies are expected to take care of themselves and their household members only, whereas collectivistic societies have a dense network of dependencies and require strong loyalty towards the ingroup. Europeans and North Americans tend to show individualistic tendencies, whereas societies of South America and Asia are typically collectivistic. Some research on the link between this dimension and overconfidence recently emerged. (Chui, Titman et al. 2010) found that individualism is associated with profitability of momentum trading strategies, which could stem from investors' overconfidence (Daniel, Hirshleifer et al. 1998). (Ferris, Jayaraman et al. 2013) reported that overconfident CEOs tended to originate from countries high on individualism. We therefore expect a positive relationship here.

The dimension of **masculinity-femininity** reflects which gender's values are dominant. Members of masculine societies tend to be assertive and strive to acquire possessions, whereas feminine societies care more about "friendly atmosphere", security and cooperation. Given substantial evidence linking OC with the male gender, one can expect members of more masculine cultures to be relatively self-confident.

Power distance is defined as "the extent to which a society accepts the fact that power in institutions and organizations is distributed unequally" (Hofstede, 1980). This dimension thus represents acceptance for hierarchy; Northern European nations tend to score very low on this dimension, whereas Latin Americans score very high. To the extent that individual mistakes are less likely to cause dramatic effects in horizontal organizations, high confidence may be associated with high power distance.

The last of Hofstede's original dimensions, that of **uncertainty avoidance** corresponds to "the extent to which a society feels threatened by uncertain and ambiguous situations and tries to avoid these situations by providing greater career stability, establishing more formal rules, not tolerating deviant ideas and behaviors, and believing in absolute truths and the attainment of expertise" (Hofstede, 1980). One might expect that self-confidence diminishes the need for uncertainty avoidance, so that the two will correlate negatively.

Hofstede and collaborators later added two further dimensions to the framework. **Long vs. short term orientation** represents the time perspective that members of each society typically take. (Ferris, Jayaraman et al. 2013) found that CEOs whose cultural environment is long-term oriented tend to be less overconfident, so a negative relationship could be anticipated.

The final dimension, **Indulgence vs. restraint** informs whether a society "allows relatively free gratification of basic and natural human drives". English-speaking countries and Latin America tend to be high on indulgence, whereas most of Asia appears to be restrained. Even more than in the case of the previous dimension, we do not find very compelling reasons to expect any particular effect here.

We used the data reported in (Hofstede 2010). Because we found strong correlations between two pairs out of GH variables (power distance with individualism/collectivism and long/short term orientation with indulgence/restraint), we would always include just one of each pair. To check for robustness, we tested all four combinations of variables. The coefficients are reported in Table 5. It is notable that with the GH variables the R^2 of our model rises almost half as much as when 68 nationality dummies are used (e.g. comparing models 1, 7 and 2).

Table 5. Estimates for GH variables for RPC

Specification	8	9	10	11
Power Distance	0.0586** (0.0217)	---	0.0371* (0.018)	---
Individualism	---	0.0043 (0.0109)	---	0.0079 (0.0099)
Masculinity	0.0454*** (0.0096)	0.044*** (0.0109)	0.0443*** (0.0106)	0.0423*** (0.011)
Uncertainty Avoidance	-0.0634*** (0.016)	-0.0187 (0.0124)	-0.0365** (0.0117)	-0.004 (0.0144)
Long-Term Orientation	---	---	-0.0125 (0.0122)	-0.0207 (0.016)
Indulgence	-0.0199 (0.0119)	-0.0034 (0.019)	---	---

Notes. N=1145211, std. errors clustered by nationality in parentheses.
Signif. codes: '***' – 0.1%; '**' – 1%; '*' – 5%

Our hypothesis is strongly confirmed for masculinity, for which large, robust, and highly significant positive effect is found. We take it as another piece of evidence that overconfidence is a male trait. Power distance is associated with more and uncertainty avoidance with less confidence as expected, although these effects are not always strongly significant. We find no evidence in any specification for the anticipated link between individualism and overconfidence, nor any impact of the last two variables, for which little effect was expected in the first place.

For an additional insight, we have considered various items from the World Values Survey (WVS), which provides a comprehensive, internationally comparable data set. Again, numerous items were highly correlated, making it difficult to include them simultaneously in the model. Table 6 shows the estimates for variables individually appended to specification (5) for the RPC.

Table 6. Estimates for WVS variables for RPC (model 5)

Religion is not important	-1.3639* (1.1949)
I am not happy at all	-1.2235 (1.6029)
My health is very poor	-0.6025 (0.3684)
I have great deal of choice over my life	0.2048 (0.2647)
It is not like me to think up new ideas and be creative	2.0395 (0.4918)
It is not like me to think that being rich is important	-0.4816 (0.3911)
It is not like me to think that being successful is important	0.9111 (0.3109)
It is not like me to think that adventure and taking risk is important	-2.3472 (0.3749)
We need larger income differences as incentives for individual effort	1.4896*** (0.4553)
Government ownership of business and industry should be increased	-0.6985** (0.4568)
People should take more responsibility to provide for themselves	0.5147 (0.7754)
Competition is harmful. It brings out the worst in people	-0.8171* (0.152)
I practically never attend religious events.	-0.745* (0.3119)
Claiming government benefits to which you are not entitled is always justifiable	0 (0.2066)
Avoiding a fare on public transport is always justifiable	-0.0528 (0.2626)
Cheating on taxes if you have a chance is always justifiable	-0.9284* (0.259)
Someone accepting a bribe in the course of their duties is always justifiable	-0.4652 (0.4248)
Homosexuality is always justifiable	-0.7536*** (1.2719)
Prostitution is always justifiable	-1.2147*** (-2.0121)
Abortion is always justifiable	-0.3943 (-0.4085)
Divorce is always justifiable	-0.8863*** (-0.188)
Euthanasia is always justifiable	-0.3021 (0.3372)
Suicide is always justifiable	-0.6657 (1.7608)
For a man to beat his wife is always justifiable	0.4353 (-0.4321)

Notes. The statements in the table are given in an abbreviated of original WVS questions for readability.
Higher values of a variable indicate more agreement towards the statement.
N= 1116830, std. errors clustered by nationality in parentheses. Signif. codes: '***' – 0.1%; '**' – 1%; '*' – 5%

Adding the WVS variables gives a new perspective on cultural correlates of overconfidence. Negative coefficients for questions concerning religious beliefs and practices indicate that religiosity may be associated with higher OC. Entries for questions about acceptance of homosexuality, divorce and prostitution point in the same direction, yet are much higher and more significant, suggesting that it may be primarily general conservatism that drives OC. This is corroborated by the large positive coefficient for acceptance for high income differences and negative for government ownership. Some dimensions that could a priori be plausibly linked to overconfidence, such as overall happiness, turn out insignificant.

The data we have for American participants (of US-based events) allows further investigation. Indeed, we can almost always tell their state of origin. That amounts to nearly 700'000 observations from all the states (with 40 states represented by more than one thousand runners each). Thanks to that it is possible to estimate coefficients for specific state dummies, taking the runners from the rest of the world as the base group. We used Model 6, except that we replaced the dummies for nations with dummies for states and we added a variable corresponding to the (logarithm of the) distance between the marathon location and runner's home state [11]. Standardized state dummies are represented in Figure 8.

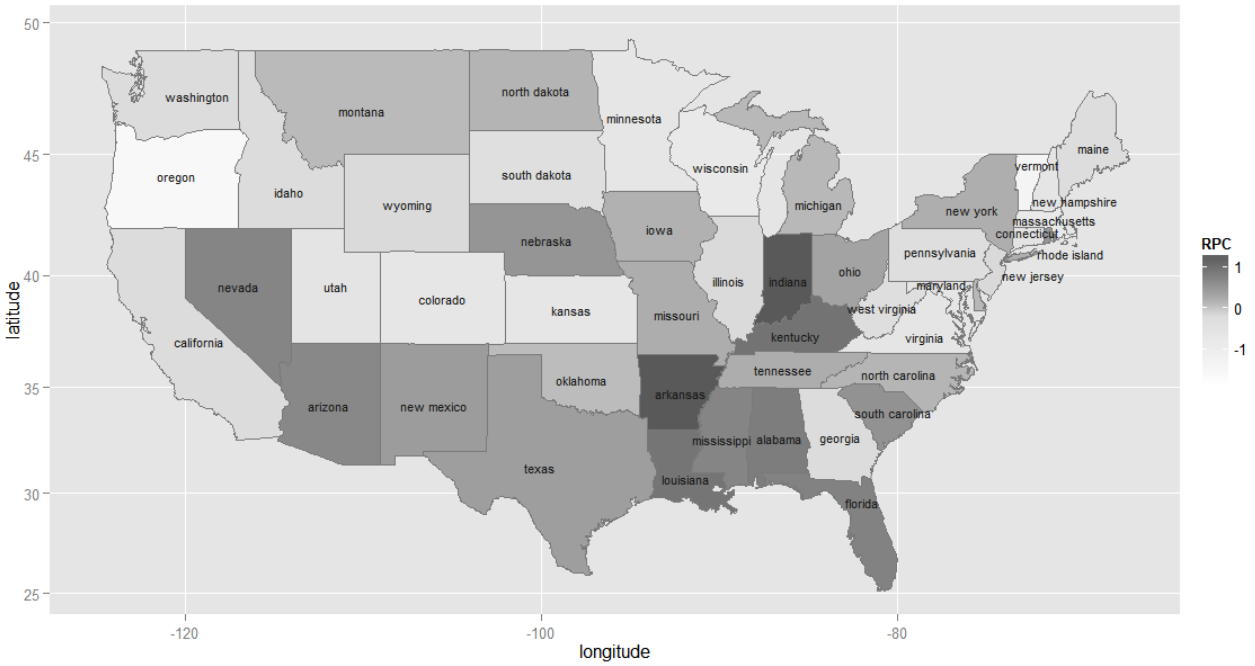

Figure 8. State effects for runners from the USA for RPC.

Notes: to visualize the differences between states, the values were standardized by subtraction of average coefficient of all states. The map is available in color at http://coin.wne.uw.edu.pl/mwilamowski/research/oc/oc_map_us_color.png

In what we find a striking congruity with our WVS results observed in the international sample, south-eastern "bible belt" states such as Arkansas, Louisiana and Alabama appear to be relatively "overconfident", whereas more liberal New England and West Coast states – much less so.[12]

[11] For simplicity we assumed for the purpose of distance calculations that all runners lived in the state's largest city or were split equally between the largest cities if they were comparable in size. The specifics of this procedure do not seem to make much difference.

[12] To be sure, the pattern is not "perfect", as particularly manifested by high OC of Indiana and low OC of Georgia.

To verify this observation more formally, we have calculated the correlation between state dummies and their various characteristics, see table 7.[13] Again, the results are consistent with the findings from WVS data: religiousness and conservative views strongly correlate with overconfidence. In particular, the RPC correlates with the share of non-religious citizens at 0.52 and with the share of citizens with liberal views at -0.49. It can be noted that figures for political preferences, albeit going in the expected direction, are substantially lower. By contrast, OC appears to be strongly negatively related to general indicators of development, such as higher education rate, GDP or HDI. The figure for firm failure rate which could be expected to be highly positive in view of (Camerer, Lovallo 1999), turns out to be low and insignificant.

Table 7. Correlations between state characteristics and effects for RPC and APC

	Pearson correlation		Kendall correlation	
Variable	RPC	APC	RPC	APC
Republican lean (%)	0.2	0.18	0.07	0.06
Democrat lean (%)	-0.22	-0.21	-0.11	-0.1
Conservative views (%)	0.38	0.34	0.22*	0.19*
Moderate views (%)	-0.02	-0.04	-0.05	-0.04
Liberal views (%)	-0.49	-0.44	-0.27**	-0.23*
Very religious (%)	0.44	0.36	0.28**	0.22*
Moderately religious (%)	0.29	0.28	0.17	0.17
Nonreligious (%)	-0.51	-0.43	-0.33***	-0.26**
Protestant (%)	0.31	0.27	0.23*	0.21*
Catholic (%)	-0.07	-0.03	-0.08	-0.04
No religious preference (%)	-0.52	-0.45	-0.31**	-0.25**
Economic Confidence Index	-0.37	-0.37	-0.2*	-0.19
Job Creation Index	-0.09	-0.11	-0.04	-0.06
Payroll to population (%)	-0.48	-0.48	-0.33***	-0.33***
Underemployed (%)	0.3	0.29	0.24*	0.23*
Government workers (%)	-0.13	-0.1	-0.02	0
Overall Well-Being	-0.47	-0.42	-0.38***	-0.33***
Diabetes (%)	0.56	0.53	0.39***	0.36***
Obese (%)	0.46	0.42	0.28**	0.25*
Exercise frequently (%)	-0.26	-0.19	-0.18	-0.14
Uninsured (%)	0.43	0.4	0.33***	0.31**
Firm failure rate	0.12	0.15	0.02	0.03
GDP (BEA, 2013)	-0.48	-0.47	-0.28**	-0.25*
Share of higher education (USCB, 2009)	-0.67	-0.64	-0.47***	-0.44***
Human Development Index (MoA, 2010)	-0.55	-0.51	-0.41***	-0.38***

Notes. All variables but last four come from Gallup state of states for year 2013
Variable describing firm failure rate was computed using the data describing firm characteristics available at census.gov.

[13] We do not present here the (qualitatively very similar) results of alternative approach of direct estimation using state-specific dummies because of strong imbalance in the number of observations.

Discussion

Admittedly, pace change is only an indirect measure of overconfidence. However, it appears to be a good proxy. First, it is very highly correlated with prediction errors as reported before. Second, there is a very natural mechanism linking the two, in which overconfident beliefs lead to setting overly high initial speed. Third, we confirm previously existing results for differences between various demographic groups, notably high OC in males and (East) Asians.

Still, alternative psychological reasons for the observed differences in pace changes could be suggested. One concerns risk taking. A relatively ambitious initial pace could be a conscious strategy giving some hope for a very good final time, at the cost of a high risk of a much poorer result. There is indeed some evidence that males may be more risk-seeking (Byrnes, Miller et al. 1999) and ambitious (Costantini 1990). However, we observe that the pace change is actually lowest in the fastest (i.e. expert) runners, consistent with the idea that it results from a bias that goes away with experience, given sufficient feedback.[14] Instead, if it was a conscious strategy, it could perhaps only escalate as one's personal best improves.

Another potential explanation could be that runners are affected by others' behavior. If many participants in the specific starting zone tend to start relatively fast, the others could be tempted to join in, even if it contradicts their planned strategy. If anything, we would expect such liability to influence to be more prevalent among women (Eagly 1978), which would in any case only strengthen our finding of gender gap.

Summary and conclusions

We have used two quite different methods to identify overconfidence in amateur runners. They turned out to bring strikingly consistent results – strong gender effect and weaker, U-shaped pattern for age. The results also seem to be highly robust across econometric specifications within each method. Furthermore, the direction and sizes of these effects seem quite plausible in view of pre-existing literature. Importantly, they are non-negligable. For example, according to our estimates for age, gender, India and the Netherlands, Mr Gupta from the opening example is expected to be nearly one standard deviation more confident than Mrs de Jong, making her actually much more promising for any task, for which the two happen to signal roughly equal competence.

[14] We have also run some preliminary analyses suggesting that APC and RPC go down on subsequent performances of the same individual. This effect will be addressed in a companion paper.

References

ACKER, D. and DUCK, N.W., 2008. Cross-cultural overconfidence and biased self-attribution. *The Journal of Socio-Economics,* **37**(5), pp. 1815-1824.

ALDRIDGE, L.J. and ISLAM, M.R., 2012. Cultural differences in athlete attributions for success and failure: The sports pages revisited. *International Journal of Psychology,* **47**(1), pp. 67-75.

ALLEN, E.J. and DECHOW, P.M., 2013. The'Rationality'of the Long Distance Runner: Prospect Theory and the Marathon. *Available at SSRN 2342396,* .

ALLEN, E.J., DECHOW, P.M., POPE, D.G. and WU, G., 2013. Reference-Dependent Preferences: Evidence from Marathon Runners.

ANBARCI, N., LEE, J. and ULKER, A., 2014. Win at All Costs or Lose Gracefully in High-Stakes Competition? Gender Differences in Professional Tennis. *Journal of Sports Economics,* , pp. 1527002514531788.

BÉNABOU, R. and TIROLE, J., 2002. Self-confidence and personal motivation. *Quarterly Journal of Economics,* , pp. 871-915.

BEYER, S., 2002. The effects of gender, dysphoria, and performance feedback on the accuracy of self-evaluations. *Sex Roles,* **47**(9-10), pp. 453-464.

BEYER, S. and BOWDEN, E.M., 1997. Gender Differences in Seff-Perceptions: Convergent Evidence from Three Measures of Accuracy and Bias. *Personality and Social Psychology Bulletin,* **23**(2), pp. 157-172.

BHANDARI, G. and DEAVES, R., 2006. The demographics of overconfidence. *The Journal of Behavioral Finance,* **7**(1), pp. 5-11.

BÖHEIM, R. and LACKNER, M., 2013. Gender and Competition: Evidence from Jumping Competitions.

BRUINE DE BRUIN, W., PARKER, A.M. and FISCHHOFF, B., 2007. Individual differences in adult decision-making competence. *Journal of personality and social psychology,* **92**(5), pp. 938.

BURSON, K.A., LARRICK, R.P. and KLAYMAN, J., 2006. Skilled or unskilled, but still unaware of it: how perceptions of difficulty drive miscalibration in relative comparisons. *Journal of personality and social psychology,* **90**(1), pp. 60.

BYRNES, J.P., MILLER, D.C. and SCHAFER, W.D., 1999. Gender differences in risk taking: A meta-analysis. *Psychological bulletin,* **125**(3), pp. 367.

CAMERER, C. and LOVALLO, D., 1999. Overconfidence and excess entry: An experimental approach. *American economic review,* , pp. 306-318.

CHARNESS, G. and VILLEVAL, M., 2007. Cooperation, competition, and risk attitudes: An intergenerational field and laboratory experiment.

CHEN, G., KIM, K.A., NOFSINGER, J.R. and RUI, O.M., 2007. Trading performance, disposition effect, overconfidence, representativeness bias, and experience of emerging market investors. *Journal of Behavioral Decision Making,* **20**(4), pp. 425-451.

CHUI, A.C., TITMAN, S. and WEI, K.J., 2010. Individualism and momentum around the world. *The Journal of Finance,* **65**(1), pp. 361-392.

CLARK, J. and FRIESEN, L., 2009. Overconfidence in Forecasts of Own Performance: An Experimental Study*. *The Economic Journal,* **119**(534), pp. 229-251.

COALSON, D., 2014. *Self-Serving Bias: A Review of Research on Variability and Outcomes,* .

COSTANTINI, E., 1990. Political women and political ambition: Closing the gender gap. *American Journal of Political Science,* , pp. 741-770.

CRAWFORD, J.D. and STANKOV, L., 1996. Age differences in the realism of confidence judgements: A calibration study using tests of fluid and crystallized intelligence. *Learning and Individual Differences,* **8**(2), pp. 83-103.

DAHL, M., ALLWOOD, C.M. and HAGBERG, B., 2009. The realism in older people's confidence judgments of answers to general knowledge questions. *Psychology and aging,* **24**(1), pp. 234.

DANIEL, K., HIRSHLEIFER, D. and SUBRAHMANYAM, A., 1998. Investor psychology and security market under-and overreactions. *the Journal of Finance,* **53**(6), pp. 1839-1885.

DEANER, R.O., CARTER, R.E., JOYNER, M.J. and HUNTER, S.K., 2014. Men are More Likely than Women to Slow in the Marathon. *Medicine and science in sports and exercise,* .

EAGLY, A.H., 1978. Sex differences in influenceability. *Psychological bulletin,* **85**(1), pp. 86.

EL HELOU, N., TAFFLET, M., BERTHELOT, G., TOLAINI, J., MARC, A., GUILLAUME, M., HAUSSWIRTH, C. and TOUSSAINT, J., 2012. Impact of environmental parameters on marathon running performance. *PLoS One,* **7**(5), pp. e37407.

FERRIS, S.P., JAYARAMAN, N. and SABHERWAL, S., 2013. CEO overconfidence and international merger and acquisition activity. *Journal of Financial and Quantitative Analysis,* **48**(01), pp. 137-164.

GALLUP, 2010. State of the states. Retrieved from http://www.gallup.com/poll/125066/State-States.aspx

BEA, 2013 Retrieved from http://www.bea.gov/newsreleases/regional/gdp_state/gsp_newsrelease.htm

GERDES, C. and GRÄNSMARK, P., 2010. Strategic behavior across gender: A comparison of female and male expert chess players. *Labour Economics,* **17**(5), pp. 766-775.

GERVAIS, S. and ODEAN, T., 2001. Learning to be overconfident. *Review of Financial Studies,* **14**(1), pp. 1-27.

GLASER, M., LANGER, T. and WEBER, M., 2010. Overconfidence of professionals and lay men: individual differences within and between tasks? .

HANSSON, P., RÖNNLUND, M., JUSLIN, P. and NILSSON, L., 2008. Adult age differences in the realism of confidence judgments: Overconfidence, format dependence, and cognitive predictors. *Psychology and aging,* **23**(3), pp. 531.

HEINE, S.J. and LEHMAN, D.R., 1997. The cultural construction of self-enhancement: an examination of group-serving biases. *Journal of personality and social psychology,* **72**(6), pp. 1268.

HEINE, S.J., LEHMAN, D.R., MARKUS, H.R. and KITAYAMA, S., 1999. Is there a universal need for positive self-regard? *Psychological review,* **106**(4), pp. 766.

HIGGINS, N. and BHATT, G., 2001. Culture moderates the self-serving bias: Etic and emic features of causal attributions in India and in Canada. *Social Behavior and Personality: an international journal,* **29**(1), pp. 49-61.

HILTON, D., REGNER, I., CABANTOUS, L., CHARALAMBIDES, L. and VAUTIER, S., 2011. Do positive illusions predict overconfidence in judgment? A test using interval production and probability evaluation measures of miscalibration. *Journal of Behavioral Decision Making,* **24**(2), pp. 117-139.

HOFSTEDE, G., 1980. Motivation, leadership, and organization: do American theories apply abroad?. *Organizational dynamics*, *9*(1), 42-63.

HOFSTEDE, G., 2010. *Cultures and organizations (3rd ed.).* London: McGraw-Hill.

HOFSTEDE, G., 1983. National cultures in four dimensions: A research-based theory of cultural differences among nations. *International Studies of Management & Organization,* , pp. 46-74.

JOST, J.T., BLOUNT, S., PFEFFER, J. and HUNYADY, G., 2003. Fair market ideology: Its cognitive-motivational underpinnings. *Research in Organizational Behavior,* **25**, pp. 53-91.

JOYNER, M.J., 1991. Modeling: optimal marathon performance on the basis of physiological factors. *J Appl Physiol,* **70**(2), pp. 683-687.

KOVALCHIK, S., CAMERER, C.F., GRETHER, D.M., PLOTT, C.R. and ALLMAN, J.M., 2005. Aging and decision making: A comparison between neurologically healthy elderly and young individuals. *Journal of Economic Behavior & Organization,* **58**(1), pp. 79-94.

KRAWCZYK, M., 2012. Incentives and timing in relative performance judgments. A field experiment. *Journal of Economic Psychology,* .

KRUEGER, J. and MUELLER, R.A., 2002. Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of personality and social psychology,* **82**(2), pp. 180.

KRUGER, J. and DUNNING, D., 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology,* **77**(6), pp. 1121.

KRUSEMARK, E.A., KEITH CAMPBELL, W. and CLEMENTZ, B.A., 2008. Attributions, deception, and event related potentials: An investigation of the self-serving bias. *Psychophysiology,* **45**(4), pp. 511-515.

LAMBERT, J., BESSIÈRE, V. and N'GOALA, G., 2012. Does expertise influence the impact of overconfidence on judgment, valuation and investment decision? *Journal of Economic Psychology,* **33**(6), pp. 1115-1128.

LENNEY, E., 1977. Women's self-confidence in achievement settings. *Psychological bulletin,* **84**(1), pp. 1.

LIRGG, C.D., 1991. Gender differences in self-confidence in physical activity: A meta-analysis of recent studies. *Journal of Sport & Exercise Psychology,* .

LUNDEBERG, M.A., FOX, P.W., BROWN, A.C. and ELBEDOUR, S., 2000. Cultural influences on confidence: Country and gender. *Journal of educational psychology,* **92**(1), pp. 152.

LUNDEBERG, M.A., FOX, P.W. and PUŃĆOHAŔ, J., 1994. Highly confident but wrong: gender differences and similarities in confidence judgments. *Journal of educational psychology,* **86**(1), pp. 114.

MCKENZIE, C.R., LIERSCH, M.J. and YANIV, I., 2008. Overconfidence in interval estimates: What does expertise buy you? *Organizational behavior and human decision processes,* **107**(2), pp. 179-191.

Measure of America (MoA), 2010 A Project of the Social Science Research Council. Retrieved from http://www.measureofamerica.org/maps/

MENKHOFF, L., SCHMELING, M. and SCHMIDT, U., 2013. Overconfidence, experience, and professionalism: An experimental study. *Journal of Economic Behavior & Organization,* **86**, pp. 92-101.

MENKHOFF, L., SCHMIDT, U. and BROZYNSKI, T., 2006. The impact of experience on risk taking, overconfidence, and herding of fund managers: Complementary survey evidence. *European Economic Review,* **50**(7), pp. 1753-1766.

MEZULIS, A.H., ABRAMSON, L.Y., HYDE, J.S. and HANKIN, B.L., 2004. Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological bulletin,* **130**(5), pp. 711.

MILLER, D.T. and ROSS, M., 1975. Self-serving biases in the attribution of causality: Fact or fiction? *Psychological bulletin,* **82**(2), pp. 213.

MOORE, D.A. and HEALY, P.J., 2008. The trouble with overconfidence. *Psychological review,* **115**(2), pp. 502.

NURMI, J., 1992. Cross-cultural differences in self-serving bias: Responses to the attributional style questionnaire by American and Finnish students. *The Journal of social psychology,* **132**(1), pp. 69-76.

ORTOLEVA, P. and SNOWBERG, E., 2013. *Overconfidence in Political Behavior,* .

PASERMAN, M.D., 2007. *Gender differences in performance in competitive environments: evidence from professional tennis players,* .

PUNCOCHAR, J., FOX, P., FRITZ, W. and ELBEDOUR, S., 1996. Highly confident, but wrong, revisited overseas, *annual meeting of the American Educational Research Association, New York* 1996.

SACKETT, A.M., WU, G., WHITE, R.J. and MARKLE, A.B., 2013. Harnessing optimism: How eliciting goals improves performance. *unpublished paper,* .

SCANLAN, T.K. and PASSER, M.W., 1980. Self-serving biases in the competitive sport setting: An attributional dilemma. *Journal of Sport Psychology,* **2**(2),.

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

U.S. Census Bureau (USCB); 2009 Educational Attainment by State. Retrieved from http://www.census.gov/compendia/statab/2012/tables/12s0233.xlsYATES, J.F., LEE, J. and BUSH, J.G., 1997. General knowledge overconfidence: cross-national variations, response style, and “reality”. *Organizational behavior and human decision processes,* **70**(2), pp. 87-94.

YATES, J.F., LEE, J. and SHINOTSUKA, H., 1996. Beliefs about overconfidence, including its cross-national variation. *Organizational behavior and human decision processes,* **65**(2), pp. 138-147.

Faculty of Economic Sciences
University of Warsaw
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl