

Anticipating Peer Ranking Causes Hormonal Adaptations that Benefit Cognitive Performance

Carsten K.W. De Dreu^{1,2,*}, Klarita Gërxhani^{3,*} & Arthur Schram^{2,4,*}

¹Institute of Psychology, Leiden University (PB 9555, NL-2300 RB Leiden).

² Amsterdam School of Economics (PO Box 15867, NL-1001 NJ Amsterdam).

³ Department of Political and Social Sciences, European University Institute, Florence.

⁴ Department of Economics, European University Institute, Florence.

* All authors contributed equally to this paper. Corresponding Author: K. Gërxhani (klarita.gerxhani@eui.eu).

Author Acknowledgment and Contribution

This work was supported by the Behavioral Economics Research Priority Grant to AS and a Spinoza Award from the Netherlands Science Foundation to CKWDD (NWO SPI-57-242).

Author Contributions: CKWDD, KG and AS designed the study, analyzed the data, discussed results, and wrote the paper. AS and KG implemented the study. We thank Michael Giffin and Eliska Prochazkova for their assistance with data collection. The authors declare no conflict of interest.

Abstract

Performance ranking is common across a range of professional and recreational domains. Even when it has no economic consequences but does order people in terms of their social standing, anticipating such performance ranking may impact how people feel and perform. We examined this possibility by asking human subjects to execute a simple cognitive task while anticipating their performance being ranked by an outside evaluator. We measured baseline and post-performance levels of testosterone and cortisol. We find that (i) anticipating performance ranking reduces testosterone and increases cortisol; (ii) both these hormonal responses benefit cognitive performance; which explains why (iii) anticipation of being ranked by a peer increases cognitive performance.

Anticipating Peer Ranking Causes Hormonal Adaptations that Benefit Cognitive Performance

People often want to know where they stand relative to others. They compete to estimate relative strength, curiously scrutinize the performance of others, and become upset when their social standing is undermined or otherwise below their expectations. Indeed, the drive to enhance and protect one's standing within groups and organizations is at the core of social cognition and behavior (Festinger, 1954; Jost, Banaj & Nosek, 2004; Taylor & Lobel, 1989). It is also seen in other social animals who show-off their strength, battle for high standing and suffer when falling behind (Sapolsky, 2005, 2017; Twenge, et al., 2012). In fact, in human groups and societies social comparisons are continuously and routinely made and are an intricate and institutionalized part of life. Whether at work, in the classroom, or when recreating, humans anticipate being compared and ranked for their performance.

While omnipresent and seemingly unavoidable, whether and how the anticipation of being ranked affects performance is poorly understood. In earlier work the anticipation of performance ranking invariably coincided with economic incentives, thus clouding interpretation (Anderson, Ertac, Gneezy, List & Maximiano, 2013; Buser, Dreber, & Mollerstrom, 2017; Buckert et al., 2015). Some studies suggest that anticipating performance ranking may deteriorate cognitive performance. For example, preparing for a competitive (versus cooperative) interaction enhances cognitive rigidity and reduces creative thinking (Chen, Williamson & Zhou, 2012; De Dreu & Nijstad, 2008; Gneezy & Rustichini, 2000; Pashler, Johnson & Ruthruff, 2001; Staw, Sandelands & Dutton, 1981), and being observed by others increases stress reactions that slows down task performance (Kelsey, Blascovitch et al., 2000; Schmader, Johns, & Forbes, 2008). Other, more recent, studies suggest that anticipating performance ranking may improve cognitive

effort and performance (Schram, Brandts, & Gërkhani, 2019; Brandts, Gërkhani, & Schram, 2020). For example, relative to working alone, competitive incentives can be arousing (Harrison et al., 2001), and some level of socially-induced stress can promote cognitive performance (Oei et al., 2006). However, how and why socioeconomic incentives, and their absence, influence cognitive effort and performance remains poorly understood (Radl and Miller, this issue). Accordingly, our first goal here was to examine whether the anticipation of being ranked influences cognitive performance in pure absence of economic incentives to outperform others. Our second goal was to explore the possible mechanisms evoked by the anticipation of being ranked, focusing on adaptations in two key hormones – testosterone and cortisol. Testosterone belongs to the hypothalamic–pituitary–gonadal axis (HPG axis) and has a well-documented role in the regulation of social status and dominance-related behavior (Sapolsky, 2005; Cooke, Kavussanu, McIntyre & Ring, 2014; Eisenegger & Fehr, 2011; Newman, Sellers & Joseph, 2005). The possible effects of performance ranking on (changes in) testosterone release have not been examined. There is, however, considerable research on changes in testosterone during or while anticipating competition (e.g., Bateup, Booth, Shirtcliff, & Granger, 2002; Mehta & Josephs, 2006), showing that testosterone associates with increased risk-taking (Apicella, Dreber & Mollerstrom, 2014; Coates, Gurnell, & Rustichini, 2009; Cueva, Roberts, Spencer et al., 2015; Mazur & Booth, 1998; Stanton, Liening & Schultheiss, 2011), and with more heuristic thought processes and reduced cognitive performance (Goucki & Kimuram 1991). However, the precise linkage between (changes in) testosterone and cognitive performance remains poorly understood (Hua, Hildreth & Pelak, 2016). Hua and colleagues (2016) conclude from their systematic review that the variability in results may be due to different methods for testosterone measurement and disparate measures of cognitive function used, but nevertheless find promising associations between cognition and testosterone supplementation in men with low testosterone

levels (also see Thilers, MacDonald, & Herlitz, 2006). Whereas we expected that changes in testosterone accompany the effects of performance ranking on cognitive performance, it remains to be seen whether such associations are particularly strong for men compared to women.

Cortisol belongs to the hypothalamic-pituitary-adrenal (HPA) axis. Cortisol has a well-documented role in stress-regulation (Sapolsky, 2005; Dickerson & Kemeny, 2004). Although the possible effects of performance ranking on (changes in) cortisol release have not been examined either, a negative correlation between status and cortisol levels has recently been observed (Sherman and Mehta 2020). Moreover, stress-induced cortisol responses can augment cognitive control and analytical performance (Chrousos, 2009; Goldfarb et al., 2016; Schwabe, Joels, Roozendaal, Wolf & Oitzl, 2012). For example, cortisol has been shown to predict specific performance factors, such as slowing down after making a mistake (Tops & Boksem, 2011) and making fewer decision-making errors under threat (Akinola & Mendes, 2012) and reduces risk-taking (Cueve et al., 2015). Changes in cortisol may thus accompany the effects of performance ranking on cognitive performance.

Although earlier work on the relationship between (changes in) testosterone and cortisol during cognitive performance returned somewhat mixed results, it also suggests that (i) reductions in testosterone during task execution may be associated with increases in cognitive performance, and (ii) increases in cortisol during task execution may be associated with increases in cognitive performance. If true, it would follow that (iii) performance-enhancing effects of the anticipation of being ranked by a peer should be due to reduced levels of testosterone and increased levels of cortisol. To examine these possibilities, we assessed changes in salivary testosterone and cortisol between the beginning of the experiment and upon completion of the cognitive performance task (and controlled for initial hormonal balance, gender, health and lifestyle (Fig. 1ABC).

Materials and Method

Participants

The experiment was run in June 2015 at the CREED laboratory of the University of Amsterdam. An earlier study on performance ranking and cognitive performance, using the same experimental set-up as used here, used sample sizes of 18 per treatment (Schram et al., 2019). Using the results from that study we expected a medium effect size between $d = 0.5$ (conservative) and $d = 0.8$ (liberal) for the main comparison. G-Power 3.1 (Faul et al., 2007), with $\alpha = 0.05$ and $1 - \beta = 0.80$, yielded a required sample size of 53 and 21 for each treatment. These sample sizes are calculated for simple t-tests. Instead, we will apply permutation t-tests, which have much higher power and allow for valid inference for as few as eight observations per category (Moir 1998, Schram et al. 2019). We will base most of our (mean-comparison) tests on categories of 47-48 subjects; for others some of the categories have $N = 24$. These numbers are well above the sample sizes typically used in studies tracking neurohormonal responses in cognitive performance tasks. For all our tests, we will report the number of observations they are based on.

In total, we recruited 124 participants, mainly undergraduate students at the University of Amsterdam. All of the more than 2000 potential participants in the CREED subject pool received an invitation to voluntarily sign up and participation was on a first-come, first-serve basis. We organized eight sessions: four sessions with 13 and four with 18, for a total of 124 participants. 28 of these were passive C-players as described below and one participant (a type B-player, see below) decided to leave during the experiment. This left us with 95 (active) participants for our data analysis. Unfortunately, data from one session of 12 participants have become unusable due to a corrupted data set, thus leaving 83 participants to be used for our statistical analyses.

- Insert Figure 1 About Here -

Experimental Procedures and Treatments

An outline of the experimental procedure is given in Fig 1A. In each session, before entering the laboratory, each participant is randomly assigned to one of three player types, denoted by A, B and C. Only types A and B enter the laboratory and do the computerized cognitive task described below. For this reason, we call them ‘active participants’. The instructions for both A- and B-players emphasize that this task has been shown to correlate positively with success in professional life. Participants were informed that evidence of this claim would be provided after the experiment if so desired. For this purpose, copies of Koedel and Tyhurst (2012) were available. The experimental software for this task was developed in Delphi at the Center for Research in Experimental Economics and political Decision making (CREED) by Jos Theelen. The software is available from the authors upon request.

C-players are taken to separate rooms and act as peers in a passive observatory way, which is why we call them ‘inactive participants’. In every session there are six A-players and six B-players (Fig. *SI*). Depending on the treatment (see below), there are either six or one C-player(s). Active participants fill out a questionnaire which assessed their current and past medical history, use of medication (including contraception), food intake (including alcohol, caffeine, nicotine), type and duration of physical activity in the past 24 hours, length and body weight, gender and age.

At the beginning of a session, participants are told the rules of conduct. They are subsequently asked to open the closed envelop on their desk, which contains preliminary instructions. Participants read these privately, at their own pace. Instructions are reproduced in the *SI*. The experiment consists of two stages, as implemented in Schram et al. (2019). In Stage I, active participants (both A- and B-types) undertake an individual cognitive task. In Stage II (not

computerized), type A are required to report their result to a type C peer seated in a separate room (Fig S1). Instructions for Stage II are distributed after Stage I has been completed. Sessions lasted approximately 80 minutes, and active-participant earnings were performance based, with an average of €21.69 (including a €7 show-up fee) for active participants. Inactive participants receive a flat fee of €20 (including a €7 show-up fee).

We implemented three treatments. For all treatments, Stage I involves the real-effort cognitive task with the monetary payoff based on the individual's score (see below). Type B participants form a baseline-treatment group (B-T), and only perform this task. Before performing this same task in Stage I, type A participants know that in Stage II they have to report to a peer seated in a separate room (Fig 1A). The way they report their performance to a peer (C-type) depends on the treatment (see Fig. S1). In the 'No-Ranking' treatment (NR-T), each participant reports to a different peer and (truthfully) reads aloud the score, but not the rank. In this way, each C-player only hears the performance of a single A-player and can therefore not rank this performance against others'. In the 'Performance Ranking' treatment (PR-T), each participant individually and privately reports to the same peer and (truthfully) reads aloud his/her score as well as the ranking among the other participants in that treatment group. The peer in this treatment thus hears all A-players' performances and ranks, which allows the peer to compare them. The distinction between these two treatments uses the fact that ranking is by definition positional to isolate the mere effects of having to report one's result to a peer from the effects of performance ranking, i.e., being compared to others by a stranger (Heffetz & Frank, 2008).

In NR-T, after finishing the instructions, each type A participant is taken to a (distinct) C-player (each seated in a separate room) and reads aloud a text stating that s/he will return after the task to report her/his performance (see Section 6 of the S1). This is done to increase the anticipation of having to later report to the C-player. At the end of Stage II, each type A participant reports (one

at a time) to his or her C-player and reads aloud the own score of performance in Stage I task. Participants know their ranking but do not report this ranking to the C-player. In PR-T, each type A participant is also taken to a C-player after finishing the instructions, but this C-player is now the same peer for all. The type A participant reads aloud a text stating that s/he will return after the task to report her/his performance (see Section 6 of the *SI*). After performing the task in Stage I, each type A reports, one at a time, to the same C-player and reads aloud the own score and the rank amongst the type A participant from printed texts provided by the experimenter.

In both (NR-T and PR-T) treatments, C-players are not informed about the task, but are told that high scores indicate better performance than low scores. A-players know that the C-players do not know the task.

Cognitive Performance

In Stage I of the experiment, active participants are presented with a sequence of pairs of 10x10 matrices filled with two-digit numbers. These matrices appear at the bottom half of their computer monitor (Fig. S2). As described in Weber and Schram (2016), for each pair of matrices each participant has to individually find the highest number in the left matrix and the highest number in the right matrix and to calculate the sum of these two numbers. This sum must be entered in the window at the center-top of the monitor (Fig. S2). A correct answer yields one euro. We apply the piece-rate remuneration in all of our treatments, thus eliminating incentivized competition. After a number has been entered, two new matrices appear, regardless of whether the sum is correct or not.

The task continues for 15 minutes. Participants are informed that the numbers are ‘randomly generated’. We do not draw from a uniform distribution, however, because this would lead with a high probability to very high sums. Instead, we first draw for each cell a random

number between 40 and 99, say X. Subsequently, we drew a random number between 10 and X with equal probability.

Hormonal Measurement

Before the instructions of Stage I are distributed among the active participants, we collect their saliva samples using standard procedures (Bosch et al., 2011; Granger et al., 2004; Riad-Fahmy et al., 1987). To prevent abnormalities in salivary measures, participants were asked not to smoke, eat or drink anything except water, and to not brush their teeth two hours prior to participation. Participants provide saliva samples before they received instructions for Stage I and then again directly after completing Stage II. Subjects are handed a 25 ml sterile polypropylene tube, asked to swallow all saliva in their mouths, and then allow saliva to collect for 3 minutes, spitting once a minute. The first (baseline) sample is taken after a 25-minute habituation period and the second sample is taken after the cognitive task has been completed. Collected samples were put on ice immediately and within an hour stored at -20 degrees Celsius. Upon completion of the study, all samples were transported and analyzed for testosterone and cortisol at the University of Utrecht Medical Center.

Results

To start, we examined the simple relationship between experimental treatment, cognitive performance, and hormonal adaptations in two planned contrasts. The first compared baseline (B-T) against the two reporting treatments (NR-T and PR-T combined); this aims at studying whether the anticipation of having to report the score to a peer (which occurs in NR-T and PR-T), affects performance. The second compared performance ranking (PR-T) against the no-ranking treatments (NR-T and B-T combined); this aims to investigate whether performance is affected if

it can be compared by the peer to that of others (which occurs only in PR-T). For these comparisons, we applied permutation t-tests with 10,000 permutations (henceforth, PtT).

Cognitive performance is measured as the number of correct summations in the summation task. Fig 2A shows performance levels per treatment. Participants performed better when performance had to be reported rather than not (PtT; B-T versus not-B-T; $p=0.045$; $N_1=41$; $N_2=42$). For change in testosterone (Fig 2B), we likewise find a stronger decline in testosterone when performance had to be reported rather than not (PtT; B-T versus not-B-T; $p=0.004$; $N_1=41$; $N_2=42$). The testosterone change is also different when performance would be ranked rather than not (PtT; PR-T versus not-PR-T; $p=0.023$; $N_1=59$; $N_2=24$). For changes in cortisol (Fig 2C) no significant effects were found. Furthermore, Fig 2D shows that decreasing testosterone (but not cortisol) positively relates to cognitive performance (Pearson's correlation coefficient; $r = -0.24$, $p = 0.032$ for testosterone; $r = -0.07$, $p = 0.519$ for cortisol). Note that for the effects of hormonal changes on performance we follow the recent literature (e.g., Apiceli et al. 2014; Cueve et al. 2015; Stanton et al. 2011) and consider absolute changes in hormone levels. Nominally similar, but statistically slightly weaker, results are obtained if we consider relative changes instead.

-- Insert Figure 2 About Here --

To account for interdependencies between anticipating performance ranking, hormonal adaptations and cognitive performance, along with possible moderation by participant gender, health and lifestyle (see *SI*), we followed-up these initial analyses with a simultaneous equations model, formalized in eqs. (1):

$$\begin{cases} \Delta T = \alpha_0 + \alpha_1 G + \alpha_2 X + \alpha_3 X \cdot G + \alpha_4 Y + \alpha_5 Z + \varepsilon_1 \\ \Delta C = \beta_0 + \beta_1 G + \beta_2 X + \beta_3 X \cdot G + \beta_4 Y + \beta_5 Z + \varepsilon_2 \\ P = \gamma_0 + \gamma_1 G + \gamma_2 \Delta T + \gamma_3 \Delta T \cdot G + \gamma_4 \Delta C + \gamma_5 \Delta C \cdot G + \varepsilon_3 \end{cases} \quad (1)$$

The first two equations of (1) depict the relationships in Fig. 1B and the third equation depicts the relationship in Fig 1C. $\Delta T(\Delta C)$ denotes the change in testosterone (cortisol) level; G

denotes gender (with 0 = male; 1 = female); X denotes anticipated performance ranking (represented by three treatment groups: $B-T$, $NR-T$, and $PR-T$); Y denotes initial hormone levels and Z denotes a set of health and lifestyle variables¹; P denotes cognitive performance. The coefficients $\alpha_i, \beta_i, \gamma_i$ quantify the relationships indicated by arrows in Fig 1BC and will be estimated using our experimental data. Note that some of these constitute vectors of coefficients. For example, because anticipated performance ranking (X) comprises a vector of two elements (the treatment dummies $NR-T$ and $PR-T$, with $B-T$ serving as a benchmark), α_2 and β_2 each constitute a vector of two coefficients. Our use of the interaction terms ($X \cdot G, \Delta T \cdot G, \Delta C \cdot G$) allows the effects of anticipated performance ranking on hormonal change and the effects of hormonal change on performance to differ between men and women. Finally, the $\varepsilon_i, i = 1,2,3$ denote random disturbance terms. We cannot assume these to be independently distributed because of the simultaneous equations structure where the dependent variables of the first two equations enter as independent variables in the third.

To account for the correlations between independent variables and disturbances that follow from these simultaneous equations, we apply a 3SLS regression framework. The coefficients $\alpha_2, \alpha_3, \beta_2, \beta_3, \gamma_2, \gamma_3, \gamma_4, \gamma_5$ capture the main relationships of interest. A first estimation of the coefficients showed that we cannot reject the null hypotheses that $\gamma_3 = 0$ and $\gamma_5 = 0$. This means that there is no difference between men and women in how performance responds to hormonal adaptations. We therefore drop the interaction terms $\Delta T \cdot G, \Delta C \cdot G$ from the third equation in (1).

¹ The health and lifestyle variable includes the number of hours slept on the previous night; the Body Mass Index (BMI); a dummy indicating whether the participant is on medication; the average number of cigarettes smoked per day; minutes spent on sports in the past 24 hours; a dummy indicating sexual activity in past 24 hrs; the number of glasses of alcohol in the past 24 hours; the number of cups of coffee in the past 24 hrs; and (for women), a dummy variable indicating the use of an anti-conceptive pil.

Table 1 presents the coefficients estimated for the remaining relationships. Consistent with the results shown in Fig 2 we find that anticipating performance ranking (measured by the α_2 and β_2 coefficients for $PR - T$) predicts a reduction in testosterone, and an increase in cortisol, and that reductions in testosterone and increases in cortisol enhance cognitive performance (as measured by γ_2 and γ_4). We observe no moderation of gender on the effects of anticipated performance ranking on changes in testosterone (α_3 for PR-T). For changes in cortisol, the effects are observed only for men (for women, $\beta_2 + \beta_3 \approx 0$ for PR-T). Finally, the increase in cortisol observed for NR-T and PR-T (β_2) are of similar size and both significant for men. Similarly, for women, $\beta_2 + \beta_3$ is not significantly different for NR-T and PR-T ($p=0.245$). This means that for cortisol, but not for testosterone, simply reporting to a peer has the same effect as being ranked.

We concluded our analyses by examining the model fit. We did a 100-fold repetition of out-of-sample estimation. For every repetition we randomly selected an estimation group E and a prediction group P . Every subject had a 50% chance of being in either. In every repetition we estimated the model on E and predicted the number of correct summations for each subject in P . In every repetition we then calculated for every subject in P a variable F , which is the predicted number of correct summations divided by the observed number. If the prediction for an individual is precise, then $F = 1$. A prediction that is 10% too high, for example, gives $F = 1.1$. For every repetition, we calculated the average F in P . We did so separately for each treatment (B-T, NR-T, PR-T). This gives 3 numbers per repetition and 300 in total. The observed values per treatment, ordered from low to high, show some over-estimation (most values are larger than 1; Fig 3A), and that 82% of the predictions are between -20% and $+20\%$ of the observed number. Across all repetitions, the average prediction is very accurate (Fig 3B).

Discussion and Conclusion

Our estimation results show, taken together, that anticipating performance ranking causes a decrease of 14.20 pmol/L of testosterone and (for men) an increase of 1.98 nmol/L of cortisol, which accounts for an estimated $(-0.12 * -14.20 =) 1.7$ and $(0.52 * 1.98 =) 1.0$ correct summations, respectively. Put differently, the hormonal adaptations to anticipated performance ranking associate with 2.7 (1.7) additional correct summations for men (women) which is almost 19% (almost 13%) of the mean number of correct summations that males (females) have when they perform the task in isolation. Thus, the anticipation of being ranked by a peer decreases testosterone and increases cortisol, and through this combination of reduced testosterone and increased cortisol cognitive performance can improve quite substantially. These findings are similar for males and females, but stronger for males, and remain whether or not we control for a range of life-style variables. Accordingly, we conclude that anticipating performance ranking enhances cognitive performance because such anticipation impacts hormonal adaptations in both testosterone and cortisol.

Others before us have observed effects of social comparison and performance ranking embedded in competitive incentive structures. Whereas competitive incentives can produce a surge of testosterone that in turn facilitates competitive performance, being observed has been associated with increased cortisol and, sometimes, reduced performance. Here we find a diametrically different pattern of results when competitive incentives are eliminated. Merely anticipating being ranked on performance reduces testosterone and increases cortisol, and increases cognitive performance. This suggests that the anticipation of performance ranking *per se* results in qualitatively different hormonal adaptations than does working under competitive incentives and when being observed by others. The effects we find are non-trivial, as under continued exposure to specific external pressures, such as social performance ranking, the

associated hormonal adaptations of the kind observed here can exert structural effects on brain and behavior (Sapolsky, 2005, 2017). As noted, performance ranking is omnipresent and an almost unavoidable fact of both professional and recreational life. Our results suggest that this formal and informal institutionalization of performance ranking can impact human cognitive performance. Even in the absence of economic incentives, merely anticipating performance ranking already increases cognitive performance when and probably because it increases cognitive effort, reduces overconfidence, and motivates people to go the extra mile.

References

- Akinola, M., & Mendes. W.B. (2012). Stress-induced cortisol facilitates threat-related decision making among police officers. *Behavioral Neuroscience*, *126*, 167-174.
- Anderson, S., Ertac, S., Gneezy, U., List, J.A., & Maximiano, S. (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economic Statistics*, *95*, 1438 – 1443.
- Apicella, C. L., Dreber, A. & Mollerstrom, J. (2014). Salivary testosterone change following monetary wins and losses predicts future financial risk-taking. *Psychoneuroendocrinology* *39*, 58–64.
- Bateup, H.S., Booth, A., Shirtcliff, E.A., & Granger, D.A. (2002). Testosterone, cortisol, and women’s competition. *Evolution and Human Behavior*, *23*, 181-192.
- Bosch, J.A., Veerman, E.C.I., de Geus, E.J. & Proctor, G.B. (2011). A-Amylase as a reliable and convenient measure of sympathetic activity: Don’t Start Salivating Just Yet!” *Psychoneuroendocrinology*, *36*, 449–53.
- Brandts, J., Gërkhani, K., & Schram, A. (2020). Are there Gender Differences in Status Ranking Aversion?. *Journal of Behavioral and Experimental Economics*, *84*, 101485.
- Buckert, M., Schwieren, Ch. Kudielka, B., & Fiebach, Ch. (2015). How stressful are economic competitions in the lab? *Journal of Economic Psychology*.
- Buser, T., Dreber, A., Mollerstrom, J. (2017). The impact of stress on tournament entry. *Experimental Economics*, *20*, 506-530.
- Chen, Williamson & Zhou. (2012). Reward system design and group creativity: An experimental investigation. *Accounting Review*, *87*, 1885 – 191.
- Chrousos, G.P. (2009). Stress and disorders of the stress system. *Nature Reviews Endocrinology*, *5*, 374-381.

- Coates, JM, Gurnell, M., & Rustichini, A. (2009). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proceedings of the National Academy of Sciences USA*, *106*, 623-628.
- Cooke, A., Kavussanu, M., McIntyre, D., & Ring, C. (2014). Effects of competition on endurance performance and the underlying psychological and physiological mechanisms. *Biological Psychology*, *86*, 370 – 378.
- Cueva, C., Roberts, R., Spencer, T. *et al.* Cortisol and testosterone increase financial risk taking and may destabilize markets. *Sci Rep* *5*, 11206 (2015)
- De Dreu, C.K.W., & Nijstad, B.A. (2008). Conflict and creativity: Threat-rigidity or motivated focus? *Journal of Personality and Social Psychology*, *95*, 648-661.
- Dickerson, S.S., & Kemeny, M.E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*, 355-391.
- Eisenegger, C., & Fehr, E. (2011). The role of testosterone in social interaction. *Trends in Cognitive Science*, *15*, 263 – 274.
- Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M. & Fehr, E. (2010). “Prejudice and truth about the effect of testosterone on human bargaining behaviour.” *Nature*, *463*, 7279.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.
- Festinger L. (1954). A theory of social comparison. *Human Relations*, *7*, 117.
- Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics*, *115*, 791-810.
- Goldfarb, E.V., Froboese, M.I., Cools, R., & Phelps, E. (2016). Stress and cognitive flexibility: Cortisol increases are associated with enhanced updating but impaired switching. *Journal*

of Cognitive Neuroscience, 29, 14 – 24.

Gouchie, C., & Kimura, D. (1991). The relationship between testosterone levels and cognitive-ability patterns. *Psychoneuroendocrinology*, 16, 323 – 334.

Granger, D.A., Shirtcliff, E.A., Booth, A., Kivlighan, K.T, & Schwartz, E.B. (2004). The ‘Trouble’ with Salivary Testosterone. *Psychoneuroendocrinology*, 29, 1229–1240.

Harrison, L.K., Denning, S., Easton, H.L., Hall, J.C., Burns, V.E., Ring, C. & Carroll, D. (2001). The effects of competition and competitiveness on cardiovascular activity. *Psychophysiology*, 38, 601-606.

Heffetz, O., & Frank, R. (2008). Preferences for status: Evidence and economic implications. In J. Benhabib, M. Jackson, & A. Bisin (Eds.), *Handbook of Social Economics* (Vol. 1A, pp. 69–91). Amsterdam: Elsevier.

Hua, J.T., Hildreth, K.L., & Pelak, V.C. (2016). Effects of testosterone therapy on cognitive functioning in aging: A systematic review, *Cognitive and Behavioral Neurology*, 29, 122-138.

Jost, J.T., Banaji, M.R., & Nosek, B.A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, 25, 881 – 919.

Kelsey, R.M., Blascovich, J., Leitten, C.L., Schneider, T.R., Tomaka, J. & Wiens, S. (2000). Cardiovascular reactivity and adaptation to recurrent psychological stress: The moderating effects of evaluative observation. *Psychophysiology*, 37, 748 – 756.

Koedel, C. & E. Tyhurst (2012). Math skills and labor-market outcomes: Evidence from a resume-based field experiment. *Economics of Education Review* 31, 131-140.

Mazur, A., & Booth, A (1998). Testosterone and dominance in men. *Behavioral and Brain Sciences*, 21, 353+.

- Mehta, P.H., & Josephs, R.A. (2006). Testosterone change after losing predicts the decision to compete again. *Hormones and Behavior*, *50*, 684-692.
- Newman, M.L., Sellers, J.G., & Josephs, R.A. (2005). Testosterone, cognition, and social status. *Hormones and Behavior*, *47*, 205 – 211.
- Oei, N.Y.L., Everaerd, W.T.A.M., Elzinga, B.M., Van Well, S. & Bermond, B. (2006). *Stress*, *9*, 133 – 141.
- Pashler, H., Johnston, J.C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology*, *52*, 629-651.
- Riad-Fahmy, D., Read, G.F., Walker, R.F., Walker, S.M., & Griffiths, K. (1987). Determination of ovarian steroid hormone levels in saliva. An overview. *Journal of Reproductive Medicine*, *32*, 254–72.
- Sapolsky, R.M. (2005). The influence of social hierarchy of primate health. *Science*, *308*, 648 – 652.
- Sapolsky, R.M. (2017). *Behave: Humans at their best and worst*. New York, Penguin.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*, 336–356.
- Schram, A.J.H.C., Brandts, J., & Gërkhani, K. (2019). Social-Status Ranking: A Hidden Channel to Gender Inequality under Competition. *Experimental Economics*, *22*(2), 396-418.
- Schwabe, L., Joels, M., Roozendaal, B., Wolf, O.T., & Oitzl, M.S. (2012). Stress effects of memory: An update and integration. *Neuroscience and Biobehavioral Reviews*, *36*, 1740-1749.
- Sherman, G.D., & P.H. Mehta (2020). Stress, cortisol, and social hierarchy. *Current opinion in psychology* *33*, 227-232.

- Stanton, S. J., Liening, S. H. & Schultheiss, O. C. (2011). Testosterone is positively associated with risk taking in the Iowa Gambling Task. *Hormones and behavior* 59, 252–256.
- Staw B.M., Sandelands, L.E., & Dutton, J.E. (1981). Threat-rigidity effects in organizational behavior—A multi-level analysis. *Administrative Science Quarterly*, 26, 501-524.
- Taylor, S.E., & Lobel, M. (1989). Social-comparison activity under threat – Downward evaluation and upward contacts. *Psychological Review*, 96, 569 – 575.
- Thilers, P.P., MacDonald, S.W.S., & Herlitz, A. (2006). The association between endogenous free testosterone and cognitive performance: A population-based study in 35-90 year old men and women. *Psychoneuroendocrinology*, 31, 565-576.
- Tops, M., & Boksem, M.A.S. (2011). Cortisol involvement in mechanisms of behavioral inhibition. *Psychophysiology*, 48, 723-732.
- Twenge, J.M., Gentile, B., DeWall, C.N., Ma, D., Lacefield, K., & Schurtz, D.R. (2009) Birth cohort increases in psychopathology among young Americans, 1938-2007. A cross-temporal meta-analysis of the MMPI. *Clinical Psychology Review*, 30, 145 – 154.
- Weber, M., & Schram, A.J.H.C. (2016). The non-equivalence of labour market taxes: A real-effort experiment. *Economic Journal*, 127, 2187-2215.

Table 1 Coefficients of model (1) were estimated with 3SLS (N=83). The full regression model includes a gender dummy (0 = male; 1 = female), health and lifestyle variables, and benchmark levels of testosterone and cortisol. A full overview of the estimates is in (Supplementary Materials, 27). Results for NR-T and PR-T are relative to the benchmark of conducting the task without reporting performance (B-T). (**/***) indicates that the coefficient concerned is statistically significantly at the 10%-(5%/1%-) level (Wald tests).

Relationship in Fig. 1	Coefficient in Eq. (1)	Estimated coefficient	Standard error
Anticipated performance ranking affects change in testosterone	α_2 (NR-T)	-8.33	9.58
	α_2 (PR-T)	-14.20	8.13*
Moderation of being female on performance ranking effect on testosterone	α_3 (NR-T)	-8.34	15.41
	α_3 (PR-T)	-0.32	14.84
Anticipated performance ranking affects change in cortisol	β_2 (NR-T)	2.49	1.22**
	β_2 (PR-T)	1.98	1.03*
Moderation of being female on performance ranking effect on cortisol	β_3 (NR-T)	-5.00	1.96**
	β_3 (PR-T)	-2.43	1.89
Effect of change in testosterone on performance	γ_2	-0.12	0.04***
Effect of change in cortisol on performance	γ_4	0.52	0.26**
N = 83			

Fig. 1 Overview of the Experiment and Model Summary. **(A)**. Timeline of the experimental procedures and measurements. **(B/C)** Hypothesized relations between anticipated performance ranking and cognitive performance through hormonal adaptations. Symbols represent the relationships defined in eqs (1). **(B)** Anticipated performance ranking predicts changes in testosterone and cortisol, controlling for gender, health, and lifestyle (27). **(C)** Changes in testosterone and cortisol predict cognitive performance, possibly moderated by participant gender.

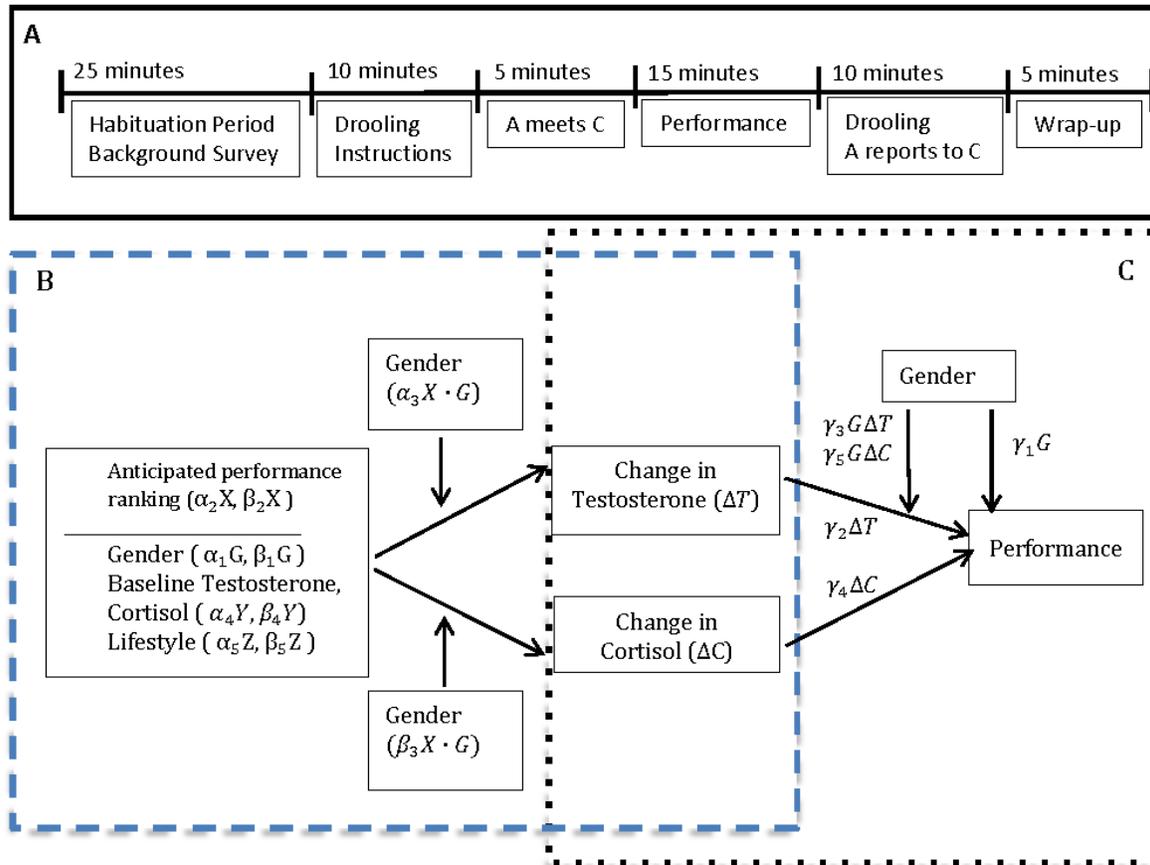


Fig. 2

Responses to Anticipated Performance Ranking. **(A)** Cognitive performance (number of correct responses) as a function of treatment (shown Mean \pm SE). **(B)** Changes in testosterone ($\mu\text{mol/L}$) from pre- to post-task (shown Mean \pm SE). **(C)** Changes in cortisol (nmol/L) from pre- to post-task (shown Mean \pm SE). **(D)** Correlations between change in testosterone (shown on left y-axis) and cortisol (shown on right y-axis) and cognitive performance. Green diamonds (red triangles) show individual pairings of cognitive performance and change in testosterone (cortisol). Green (red dashed) line shows best fitting linear regression of change in testosterone (cortisol) on cognitive performance, with $\text{coefficient} = -1.52, p = 0.032, N = 83$; and $\text{coefficient} = -0.06, p = 0.519, N = 83$, respectively.

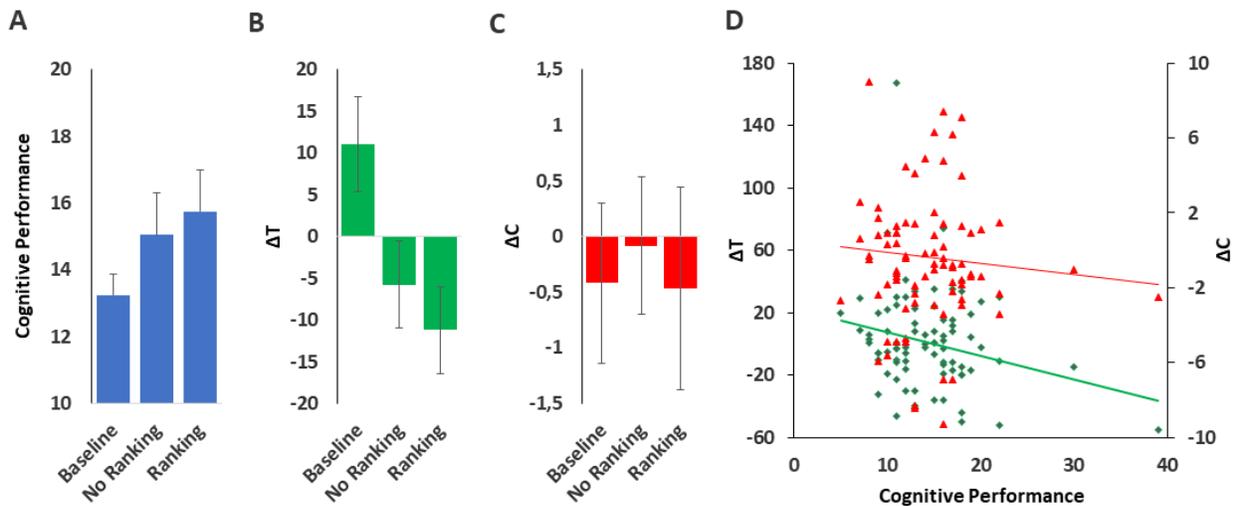


Fig 3 Predicted and Observed Performance. (A) Observed values per treatment, ordered from low to high. Perfect prediction is achieved when $F = 1.0$. Observed cases within the dashed area are predicted with 80% accuracy. (B). Out-of-sample predictions, showing the average performance for 100 subsamples of 50% of the data as estimated by the model (light blue) on the remaining 50% compared to their observed performance (dark blue).

