Journal of Public Economics 139 (2016) 40–52

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Cognitive performance in competitive environments: Evidence from a natural experiment☆

Julio González-Díaz[a], Ignacio Palacios-Huerta[b,c,*]

[a]Universidad de Santiago de Compostela, Spain
[b]London School of Economics, United Kingdom
[c]Ikerbasque Foundation at UPV/EHU, Spain

ARTICLE INFO

Article history:
Received 2 October 2014
Received in revised form 4 May 2016
Accepted 5 May 2016
Available online 12 May 2016

Keywords:
Cognitive performance
Competition
Natural experiments

ABSTRACT

Competitive situations that involve cognitive performance are widespread in labor markets, schools, and organizations, including test taking, competition for promotion in firms, and others. This paper studies cognitive performance in a high-stakes competitive environment. The analysis takes advantage of a natural experiment that randomly allocates different emotional states across professional subjects competing in a cognitive task. The setting is a chess match where two players play an even number of chess games against each other alternating the color of the pieces. White pieces confer an advantage for winning a chess game and who starts the match with these pieces is randomly decided. The theoretical analysis shows that in this setting there is no rational reason why winning frequencies should be better than 50-50 in favor of the player drawing the white pieces in the first game. Yet, we find that observed frequencies are about 60-40. Differences in performance are also stronger when the competing subjects are more similar in cognitive skills. We conclude that the evidence is consistent with the hypothesis that psychological elements affect cognitive performance in the face of experience, competition, and high stakes.

1. Introduction

In recent years, economists have paid considerable attention to the relationship between perceptions and reasoning, and to emotions such as loss aversion, reference points, disappointment and others. There is evidence that these and other behavioral effects are in fact important for explaining a wide range of economic and social behavior.[1] Despite their potential importance, however, little is known about the relevance of these effects on cognitive performance. Do they exist? If so, do they persist in the face of experience, competition, and high stakes? These are the questions we study in this paper.

Understanding cognition is important. Numerous studies establish that measured cognitive ability is a strong predictor of occupational attainment, wages, and a range of social behaviors in adults, and several studies document its importance in predicting the schooling performance of children and adolescents.[2] An emerging body of literature also finds that "psychic" costs explain a range of economic and social behavior (see, e.g., Carneiro et al. (2003), Carneiro and Heckman (2003), Cunha et al. (2010), Heckman et al. (2006a)). Besides social and economic outcomes, recent research shows that cognitive ability is also important for financial market outcomes.[3] Thus, numerous settings represent competitive situations that involve cognitive performance (e.g., test taking, student competition in schools, competitions for promotion in certain firms and organizations, and others), and understanding the relationship between cognitive performance

☆ We thank Luis Cabral, David De Meza, Leontxo García, Diego González-Díaz, Robert Östling, Pedro Rey-Biel, Yona Rubinstein, David Str ömberg, Richard Thaler, Oscar Volij, two anonymous referees, and participants in seminars at LSE, Bristol, Munich, Pisa, Stockholm IIES, Erasmus, Ben Gurion, Collegio Carlo Alberto, IMEBE and the Chess Grand Slam Final Bilbao-Shanghai for useful comments. Financial supports from the Spanish Ministerio de Economía y Competitividad and FEDER (projects MTM2011-27731-C03, MTM2014-60191-JIN, ECO2012-31626, ECO2015-66027-P), from Xunta de Galicia (project INCITE09-207-064-PR), and from the Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco (IT-869-13) are gratefully acknowledged.

* Corresponding author at: Department of Management, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom.
E-mail addresses: julio.gonzalez@usc.es (J. González-Díaz), i.palacioshuerta@gmail.com (I. Palacios-Huerta).

[1] Camerer (2003), Rabin (1998) and DellaVigna (2009) provide excellent surveys.

[2] See, for instance, Neal and Johnson (1996), McArdle et al. (2009), and other references therein. Heckman et al. (2006b) review this literature and present an analysis of the effects of both cognitive and noncognitive skills on wages. They show that a model with one latent cognitive skill and one latent noncognitive skill explains a large array of diverse behaviors including schooling, work experience, occupational choice, and participation in various adolescent risky behaviors.

[3] See, for instance, Agarwal and Mazumder (2013), Bertrand and Morse (2011), Gerardi et al. (2010), and Cole and Shastry (2009).

http://dx.doi.org/10.1016/j.jpubeco.2016.05.001
0047-2727/

and psychological effects is an important question in the literature on human capital, schooling, behavioral economics and others.

This paper contributes to these strands of economics literature by studying the impact of psychological differences on cognitive performance in a competitive environment. The analysis benefits from the opportunity provided by a randomized natural experiment that, in effect, exogenously assigns different emotional states across subjects. Similar natural experiments to the one we study have been used to examine the role of psychological effects when subjects perform non-cognitive tasks, and this paper extends the analysis to study their impact on the performance in cognitive tasks. As such, and to the best of our knowledge, it represents the first study that evaluates the causal link from behavioral effects to cognitive performance in a competitive setting taking advantage of a natural experiment.

The randomized experiment comes from professional sports. Important elements of human behavior are starkly observable in these settings. As Rosen and Sanderson (2001) indicate, "if one of the attractions of sports is to see occasionally universal aspects of the human struggle in stark and dramatic forms, their attraction to economists is to illustrate universal economic principles in interesting and tractable ways." Thus, not surprisingly, a number of prominent findings in economics have been documented for the first time studying sports settings. For instance, without attempting to be exhaustive, Ehrenberg and Bognanno (1990) investigate incentive effects in golf tournaments, Szymanski (2000) studies discrimination using soccer data, Garicano et al. (2005) study social pressure as a determinant of corruption in a soccer setting, and Bhaskar (2009) and Romer (2006) analyze optimal decision-making using cricket and football data respectively.

Much like these sports settings, ours represents a valuable opportunity for studying an open question in the literature for a number of reasons:

First, the situation involves a tractable number of subjects (just two) competing at a game that is considered the ultimate cognitive sport (chess). The game they play has complete information and involves no chance elements. The game is strictly competitive or zero-sum. Pure conflict situations in which one person's gain is always identical to another's loss involve no potential elements of cooperation. As such they represent the cleanest possible context to study competitive behavior. Subjects compete in the same setting and under identical circumstances and, as we will see in the next section, the only difference is the randomly determined order in which they complete a task.

Second, and most importantly, we take advantage of existing results in the literature (to be discussed below) that show that the *order* of competition generates differences in emotional states. Using the same type of randomly assigned treatment and control of these emotional states we extend existing research to the study of performance on *cognitive* tasks in a competitive environment.[4]

Third, the setting involves professional subjects who are characterized by the highest degree of cognitive skills at the specific competitive task they perform as professionals (playing chess). Thus, we can study if biases exist in the face of experience, competition and high stakes. This is also important because existing research has found that individuals with higher cognitive ability demonstrate *fewer* and *less extreme* cognitive biases that may lead to suboptimal behavior.[5]

Fourth, direct measures of cognitive abilities are often lacking in the literature and can be measured only indirectly (through their correlation with other variables). The setting in this paper provides a highly precise measure of the cognitive ability of the players at the task they perform. In particular, subjects have a rating according to what is called the "ELO rating method" (see Section 4), and this rating estimates quite precisely the probability that one player will outperform the other at the cognitive task. This is a valuable advantage of the empirical setting.

Finally, the study concerns high-stakes decisions that subjects are familiar with, that really affect them, to which they are used, and that take place in their own real-life environment. In this sense, it involves a set of useful characteristics in terms of stakes, familiarity and nature of the environment. And from the perspective of observing and measuring behavior, a comprehensive dataset is available where choices, outcomes, and treatments are cleanly measured.

From the theoretical point of view, we also develop rational and behavioral models of optimal play to interpret the empirical evidence. Importantly, these models will contain a contribution to the game theoretical literature on repeated interactions and to the literature on multi-battle contests. In our setting, a match consists in the repeated play of a given stage game but, differently from standard repeated games, the total payoff that players obtain may not be a sum or an average of the payoffs in each period. The existing literature has studied the case of binary outcomes: in each stage game one player wins and the other loses (see Walker et al. (2011)), but we are aware of no study with more than two outcomes. The presence of a third outcome (in our context, win, lose, and tie) brings in the issue of how to chose risk during the match, which we incorporate into the formal frameworks. This represents a novel aspect with respect to the literature on multi-battle contests in which strategic risk taking is not a choice variable (e.g., Konrad and Kovenock (2009)).

The rest of the paper is structured as follows. Section 2 describes the natural experiment and a brief literature review. Section 3 develops formal rational and behavioral models of the task the subjects undertake. The models allow us to identify the conditions under which we may be able to conclude, using the average treatment effects from the natural experiment, whether behavioral elements have an impact on cognitive performance. Section 4 describes the data. Section 5 presents the main empirical evidence, and Section 6 concludes.

2. The natural experiment

In a *chess match*, two players play an even number of *chess games*, typically about 6 to 10 games, against each other. Games are generally played one per day, with one or two rest days during the duration of the match. The basic procedure establishes that the two players *alternate* the colors of the pieces with which they play. In the first game, one player plays with the white pieces and the other with the black pieces. In the second game, the colors are reversed, and so on. Who plays with the white pieces in the first game is *randomly* determined, and this is the only procedural difference between the two players. According to the rules of FIDE (the Fédération Internationale d'Échecs, the world governing body of chess), the order is decided randomly under the supervision of a referee. This random draw of colors, which is typically conducted publicly during the opening ceremony of the match, requires that the player who wins the draw will play the first game with the white pieces. Therefore, the fact that players have no choice of order or color of the pieces makes it an ideal randomized experiment for empirically establishing causality.

The explicit randomization mechanism used to determine which player begins with the white pieces in a sequence of games where both players have exactly the *same* opportunities to play the *same* number of games with the *same* colors, have the *same* stakes, are

[4] As is well known, a randomized experiment is a powerful methodology not often available in the social sciences that ensures that the conditions for causal inference are satisfied (Manski, 1995). There is also a related literature suggesting that providing relative performance information (a consequence of the order of competition in our setting) affects performance (Azmat and Iriberri, 2010).

[5] See, for instance, the recent results in Gill and Prowse (forthcoming). Also, Benjamin et al. (2013) and Frederick (2005) report similar findings for high school and college students, respectively, using different measures of intelligence and cognitive ability.

in the *same* setting and where all other circumstances are identical, suggests that we should expect both players to have, ceteris paribus, exactly the *same* probability of winning the match. That is, absent behavioral effects associated with the order of colors, there is no rational reason why observed winning frequencies should be different from 50-50. [6] Yet, we find that this is not the case. As anticipation of the results, what we observe instead is that winning probabilities are about 60-40 in favor of the player who plays with the white pieces in the first and in all the odd games of the match.

As will be discussed in more detail later, playing with the white pieces is advantageous to win a chess game. This means that, ceteris paribus, the player playing with the white pieces in the odd games of the match is randomly allocated a greater likelihood to be *leading* during the course of the match. Conversely, his opponent, who plays with the white pieces in the even games of the match, is more likely to be *lagging.* Hence, this natural experiment shares the same basic design used recently in the literature to study the relevance of emotional or psychological states in understanding the behavior of subjects performing *non-cognitive* tasks in competitive environments. In particular, Apesteguia and Palacios-Huerta (2010), Genakos and Pagliero (2012), Pope and Schweitzer (2011), and Genakos et al. (2015) provide strong evidence for these effects from weightlifting, golf, penalty kicks in soccer, and diving competitions, respectively. Our study, therefore, extends existing research to the area of *cognitive performance* in a competitive environment using the same type of randomly determined asymmetry in emotional states. With respect to the term "emotional states," Sokol-Hessner et al. (2009) document how loss aversion is a basic hedonic property of our reaction to losing. In particular, they combine physiological measurements of arousal and various cognitive strategies to study how differences in arousal to losses relative to gains correlate with behavioral loss aversion. It is in this sense that we refer throughout the paper to the random determination of the order of play (the advantage of playing with the white pieces in the even games) as effectively randomizing *emotional* states.[7]

Finally, it seems appropriate to quote a reflection by Osborne and Rubinstein (1994) p.6) who were the first to identify the research potential of this specific natural setting (chess) to contribute to our understanding of bounded rationality, including the relationship between cognitive abilities and behavioral effects (italics added):

"When we talk in real life about games we often focus on the asymmetry between individuals in their abilities. For example, some players may have a clearer perception of a situation or have a greater ability to analyze it. These differences, which are so critical in real life, are missing from game theory in its current form. To illustrate the consequences of this fact, consider the game of chess. In an actual play of chess the players may differ in their knowledge of the legal moves and in their analytical abilities. In contrast, when chess is modeled using current game theory it is assumed that the players' knowledge of the rules is perfect and their ability to analyze it ideal. Results we prove […] imply that chess is a trivial game for "rational" players: an algorithm exists that can be used to "solve" the game. This algorithm defines a pair of strategies, one for each player, that leads to an "equilibrium" outcome with the property that a player who follows this strategy can be sure that the outcome will be at least as good as the equilibrium outcome no matter what strategy the other player uses. The existence of such strategies (first proven by Zermelo in 1913) suggests that chess is uninteresting because it has only one possible outcome. Nevertheless, chess remains a very popular and interesting game. Its equilibrium outcome is yet to be calculated; currently it is impossible to do so using the algorithm. Even if White, for example, is shown one day to have a winning strategy, it may *not* be possible for a human being to implement that strategy. Thus, while the abstract model of chess allows us to deduce a significant fact about the game, at the same time it omits the *most important determinant* of the outcome of an actual play of chess: the players' "abilities." Modeling asymmetries in abilities and in *perceptions* of a situation by different players is a fascinating challenge for future research, which models of "bounded rationality" have begun to tackle."

To the best of our knowledge, no previous research has taken the opportunity that this setting provides to study these aspects.

3. Rational and behavioral models of a match

A chess match is a nontrivial setting in which it is not possible to attribute differences in performance, if any, without first understanding what is the role that rational and behavioral elements may play in behavior. So, what is the role that these elements play in a chess match? Under what conditions may we conclude that psychological or rational effects have an impact on cognitive performance? In this section we provide a formal analysis to address these questions.

Recall that the randomly determined color of pieces generates one very specific type of asymmetry between the players: as playing with white pieces confers a strategic advantage in a chess game,[8] the random draw of colors means that players who begin playing with the white pieces are randomly given a greater opportunity to *lead*in the match and, conversely, those with the black pieces are given a greater opportunity to *lag* in the match.

We start with a canonical model (Subsection 3.1), which we then develop to include rational and psychological elements (Subsections 3.2 and 3.3). For the sake of exposition, we use chess terminology. Needless to say, the analysis also applies to other settings with repeated interactions in which the stage games have three possible outcomes.

3.1. The canonical model

Consider a *chess game* between two identical players: white and black. Let $W > 0$ denote the probability that the player with the white pieces (white) wins and $L > 0$ the probability that the player with the black pieces (black) wins. We assume that $W + L < 1$, so $1 - W - L > 0$ is the probability that the game ends in a draw. In chess it is strategically advantageous to play with the white pieces, which means that $W > L$. As just noted, empirically W is about 0.28–0.30 and L about 0.17–0.18.

A *canonical chess match* consists of T chess games, where T is an even number. In game 1, Player 1 plays with the white pieces and Player 2 with the black ones. In subsequent games the colors are alternated. Since a chess match is a constant-sum game, then, without loss of generality, we can assume that the utilities for each of the players are 1 if winning the match, 0 if losing, and 0.5 if they tie.[9]

[6] In Section 3 we qualify this statement.

[7] Other references with neurological and physiological evidence that support our use of these terms include Bechara et al. (1997), Schaefer et al. (2002), Ochesner et al. (2004), and Kermer et al. (2006)

[8] In the sample of matches we will study, 30% of the *games* were won by the players with the white pieces and 17% by the player with the black pieces; the rest are draws. In more than 165,000 chess games in the *Chessbase* dataset presented in the next section, which includes our data from chess matches, the win rates are 28 and 18%, respectively, when both players have an ELO rating above 2500(typically Grandmasters, the highest title a player can achieve).

[9] Since both players are identical, we can represent their preferences by the same utility functions. Then, without loss of generality, these utilities can be normalized so that the utility of a win is 1 and the utility of a loss is 0. Hence, since we are in a constant-sum game, the utility of a tie has to be 0.5.

Since both players are completely symmetric in a canonical chess match, the following result is straightforward:

Proposition 1. *The expected payoff in a canonical chess match is 0.5 for both players.*

3.2. Rational models

Preliminaries. A chess match is a dynamic tournament in which, in principle, players may not have the same effort conditions during the match and/or may choose the amount of risk to take depending on their leading/lagging state in the score. Typically, in the empirical studies in the literature that study non-cognitive performance in competitions, the task is effortless and risk either does not play any role (when the outcome of the task is binary, e.g., score or not) or can be cleanly taken into account.[10] In a chess game, however, strategic risk taking matters since there are three possible outcomes (either player may win the game or they may tie).

In what follows we try to understand the role that effort and risk may play in our empirical setting:

a. EFFORT. With respect to the idea that players can exert different effort during the match depending on the score, the design by FIDE of the typical chess match intends to ensure that all the games in the match are played under *identical* conditions and, in particular, that players have sufficient time to fully recover from the effort they exert: no more than one game is played each day and rest days are scattered during the duration of the match to ensure that players can play every single game in perfect physical conditions and can always exert the maximum cognitive effort. This characteristic allows us to abstract from modeling effort as a choice variable.[11]
b. RISK. A more important consideration is the fact that players may choose the risk they take during the match depending on the score. The role of strategic risk taking is, in general, not trivial and requires a formal analysis, which we provide below. Interestingly, the analysis shows that strategic risk taking is not neutral: it favors the player who starts the match playing with the *black* pieces. That is, absent behavioral effects this player should win significantly *more* often a chess match. The basic intuition for this result is the following. Lagging in the score may induce a player to choose to lower his expected performance by taking risks that he would otherwise not take in exchange for a greater probability to win a game and catch up in the score. Hence, the possibility of taking more risks and having more variable outcomes (e.g., increasing the chance of both winning *and* losing in exchange for a lower chance of tying) is an instrument at the disposal of the lagging player. This instrument, if anything, could help counteract any potential disadvantages given by the random determination of the colors. Clearly, the leading player can also tailor the risk he takes to the advantage that he has in the match and play more conservative strategies. However, no matter how conservative the leading player is, the lagging player can *always* drive the game into a win–lose lottery where the probability of winning is greater than if he had not chosen to take the additional risk.

This insight is not new; in fact, it is well known in the literature on the strategic choice of risk (variance and covariance) in dynamic competitive situations.[12] Yet, as indicated earlier, it has been studied neither in the game theoretical literature on repeated interactions nor in the literature on multi-battle contests. In terms of empirical implications, it means that if the player who starts the match playing with the *black* pieces exhibits significantly greater cognitive performance, then empirical evidence from average treatment effects alone will not allow us to conclude whether behavioral effects are present in the data. The reason being that his greater winning frequency may simply reflect the advantage that strategic risk taking confers.

Before formalizing the role of strategic risk taking we discuss an assumption specific to the empirical setting.

Assumption. In a typical chess game, the first mover advantage gives the player with the white pieces *at least as much* control over how "risky" the game will be. This is because he has at least as much control over the type of "opening" that will be played. Although there is not much discussion about this assumption in the chess community, chess is too complex to provide a theoretical foundation for it. We incorporate this asymmetry in the "technology" for risk taking in the models and provide two pieces of support for this assumption:

1. Experts' Assessment. It is not difficult to find statements from world elite players that support this assumption. For instance, former world champion Vladimir Kramnik (June 2011, interviewed after the Candidates Matches to qualify to challenge the reigning world champion, italics added) indicates: "My white games were all pretty complicated, tense and full of fight. I am responsible for my white games, and I was always trying to find a way to fight with white, even if I did not get an advantage. But *with black it is very difficult and incredibly risky to start avoiding drawish lines from the very beginning,*because it can easily just cost you a point in a very stupid way […] get a bad position, lose the game, lose the match and feel like an idiot? I didn't do it […]. It is a difficult decision which can easily backfire at this level."
2. Empirical Evidence. We know that because of strategic risk taking the expected performance of a player in a game should decrease when he is lagging. As a result, the assumption on the asymmetric technology for risk taking has the additional implication that it should decrease by a *greater* amount when the *black* player is lagging than when the white player is lagging. The empirical evidence is consistent with this implication: when lagging the observed decrease in performance is more than twice as large for black (about 16%) than for white (about 7%). [13]

[10] See, for example, Genakos and Pagliero (2012) and Apesteguia and Palacios-Huerta (2010).

[11] In related settings studied in the literature, effort is a choice variable. There is a body of literature that studies multi-battle contests in which players compete in a sequence of single component contests (battles) choosing effort (e.g., monetary expenditures). Importantly, in these settings, and different from ours, effort determines the size of the prize *both* in the component battles and in the overall battle. See for example Harris and Vickers (1987) in the context of a patent race, and Klumpp and Polborn's 2006 study of the dynamics of candidate performance and campaign expenditures in the US presidential primaries. Konrad and Kovenock (2009) characterize the unique subgame perfect equilibrium in these multi-battle contests when effort is a choice variable, but strategic risk taking is absent. Interestingly, in their setting, having effort as a choice variable is neutral in that it does not cause any deviation from 50-50 in the probability of winning the contest. For a survey of the theory of contests in sports see Szymanski (2003).

[12] See, for instance, Cabral (2002, 2003), Hvide (2002) and Hvide and Krinstiansen (2003).

[13] The expected performance is simply the number of points a player is expected to achieve in the current game. That is, his expected score can be computed as 1 × "prob. of winning" + 0.5 × "prob. of a draw" + 0 × "prob. of losing" . The table below shows the expected score for each player when lagging and when the match is tied for matches between players with ratings above 2500 (typically Grandmasters). Similar results are found for all other subsets we have examined, and for the whole sample of matches. Performance loss denotes the relative change in the expected score.

ELO ratings above 2500			
Expected score of white		Expected score of black	
Match tied	0.553	Match tied	0.447
White lagging	0.515	Black lagging	0.374
Performance loss	6.9%	Performance loss	16.3%

Models. We now present two models which incorporate into the canonical model the assumption that the "technology" for risk-return trade-off is at least as good for the player with the white pieces than for his opponent. Importantly, we will find that this assumption is sufficient (but not necessary) to show that the possibility of choosing the risk that is taken favors the player starting with the black pieces. The reason is that the main element driving the result is not this assumption but the "informational rent" of the player starting with the black pieces: Since risk taking increases the probability of winning a game at the cost of increasing by a larger amount the probability of losing, risk taking is especially useful for a lagging player. The intuition is again straightforward. Take a two-game match and suppose that the player starting with black has lost the first game. Then, in the second game, a draw is as bad as a loss and therefore he only cares about increasing his probability of winning. He will surely take risks *regardless* of whether or not white has more control over risk. Of course, the same would be true for the player starting with the white pieces if he had lost the first game, but this effect is less important since the probability of winning with the white pieces is greater than with the black pieces.

3.2.1. Model R1: only white controls risk

In this first model we assume that white has *all* the control over the risk involved in the game, an assumption that we relax in the following model by assuming that black also has some control over the risk. By taking a risky action, white can increase his probability of winning by $R_w > 0$ and his probability of losing by αR_w, with $\alpha > 1$.[14] Recall that in game 1 Player 1 plays with the white pieces and Player 2 with the black ones. Since white has all control over risk, Player 2 can guarantee for himself an expected payoff of at least 0.5 by mimicking in the even games the choices made by Player 1 in the odd games. We show below that Player 2 can in fact do strictly better. Intuitively, this is because he can benefit from the fact that he has more information when he has to make his choices concerning optimal risk-taking.

Proposition 2. *Consider a match consisting of $T = 2$ games in Model R1. Then, optimal play in this match leads to a higher expected payoff for Player 2 than for Player 1.*

Proof. See Appendix A.

Corollary 1. *In a match consisting of T games in Model R1, the expected payoff for Player 2 is greater than the expected payoff for Player 1.*

Proof. See Appendix A.

The proofs of these results show that the empirical fact that $W > L$ is not necessary. Since Player 2 is the only one who can choose risk in Period 2, he is also the first one who can make an informed choice of risk. This "informational rent" is enough to give him an edge in the match.

We show in the next model that if *both*players have some control over risk in *both* periods, then the fact that $W > L$ is crucial to prove that Player 2 has an advantage.

3.2.2. Model R2: both players control risk

In a given game both players can increase the probability of winning by taking a risky action. A risky action by white increases his probability of winning by $R_w > 0$ and his probability of losing by αR_w, with $\alpha > 1$. Similarly, a risky action by black increases his probability of winning by $R_b \geq 0$ and his probability of losing by αR_b.[15]Under the assumption that white has *at least as much* control over how "risky" a chess game is, we have $R_w \geq R_b$. We find below that this is a *sufficient* condition to obtain that Player 2 has an advantage in a match. Further, this condition is not necessary.

Note that, since $W > L$ and $\alpha > 1$, the following two conditions are satisfied when $R_w \geq R_b$:

C1: $R_b < \frac{W}{L} R_w$.
C2: $R_b < R_w + \frac{R_w(\alpha-1)(1-W-R_b)}{L+R_w}$.

We show next that these two conditions suffice to give Player 2 an advantage in a two-game match. The intuition is that, although both players can control risk in both periods, the possibility of choosing a risky strategy is particularly valuable when a player is lagging in the score. Since $W > L$, Player 2 is more likely to be lagging in the score than Player 1 and, hence, he is the one more likely to benefit from optimal risk taking in Period 2.

Proposition 3. *Consider a match with $T = 2$ games in Model R2. When C1 and C2 are satisfied, optimal play leads to a higher expected payoff for Player 2. A sufficient condition for this result is that $R_w \geq R_b$.*

Proof. See Appendix A.

This generalizes Proposition 2.[16] Although $R_w \geq R_b$ is sufficient for this result, it is clear from conditions C1 and C2 that it is not necessary as Player 2 will also have an advantage even in some cases where $R_w < R_b$. In other words, stating the result in terms of these two conditions is stronger than stating it in terms of $R_w \geq$ and R_b. And again the intuition is that the "informational rent" that Player 2 always has is independent of which player has greater control over risk in a game.

3.3. Behavioral models

We next extend the canonical and rational models to incorporate psychological elements. We try to adopt the simplest possible formulation that is both tractable and consistent with empirical evidence. As discussed earlier, empirical evidence from non-cognitive tasks supports the hypothesis that a gain/loss or leading/lagging asymmetry relative to a reference point has an impact on performance. This is the first aspect that we want to capture in the model. The literature offers various ways to formalize this idea. In particular, preferences with loss-aversion relative to a reference point have been widely adopted in both theoretical and empirical research, rationalizing a host of anomalies from labor supply, to consumer behavior and finance. The incorporation of these ingredients into economic theory dates back at least to Kahneman and Tversky (1979), and much of the early literature equated the reference point with an exogenous or history-dependent status quo. Recently, Köszegi and Rabin (2006, 2007) suggested an alternative approach of forward-looking, endogenous reference-point formation based on expectations. Here we take the simplest possible version and simply assume that two identical subjects perceive an even score in the match as their reference point and that their performance is a function of whether they are leading or lagging in the score. As will be clear below, other formulations are definitely possible. Yet, this one is tractable, proves

[14] Of course, we assume that $W + L + (1+\alpha)R_w < 1$.

[15] Obviously, it has to be the case that $W + R_w + \alpha R_b + L + R_b + \alpha R_w \leq 1$.

[16] Obviously, when $R_b = 0$, conditions C1 and C2 are trivially satisfied and this result reduces to Proposition 2 in Model R1.

convenient from a formal perspective, and captures the basic insights of more general formulations.[17]

Let k denote the difference between the number of games won and the number of games lost by the player who plays with the white pieces; that is, k is positive when he is leading and negative when he is lagging. Recall that the natural experiment randomizes the identity of the players more likely to be leading and lagging in the score. Now, the probability that the player with the white pieces wins the current game is $W + e(k)$ and the probability that he loses is $L - e(k)$, where $e(\cdot)$ is an increasing function that captures the behavioral element with $e(0) = 0$. Winning generates elation whereas losing generates disappointment and discouragement. As we are in a zero-sum game, $e(\cdot)$ captures this asymmetry and its impact on relative performance. We further assume that $-e(-1) \geq e(1)$, which means that the decrease in performance when lagging by one game for the player with the white pieces is at least as large as the increase in performance when leading by one game.[18] Clearly, $e(\cdot)$ must also be such that $W + e(k) \leq 1$, and $L - e(k) \geq 0$. This formulation captures in a parsimonious manner the basic ingredient of the loss aversion effect typically considered in more general specifications in the literature.[19] Players have no control over this effect; in other words it is not under volitional control.[20]

3.3.1. Model B1: no strategic risk taking

We next prove that when risk taking is not a choice variable Player 1 has an advantage. For intuition note that, since $W > L$, Player 2 is more likely to start game 2 lagging in the score ($k = -1$) than leading in the score ($k = 1$), and that the properties of $e(\cdot)$ imply that this effect will negatively affect his performance in game 2.

Proposition 4. *Consider a match consisting of $T = 2$ in Model B1. If $e(-1) < 0$, then the expected payoff of Player 1 is higher than the expected payoff of Player 2.*

Proof. See Appendix A.

The proof of this result shows that the edge that the psychological effect gives to Player 1 increases with the edge he has in the first game, that is with $W - L$: Playing the first game with the white pieces makes his opponent more likely to lag in the match and thus more likely to be subject to the decrease in performance caused by $e(-1)$. We show below that the result extends to matches of arbitrary length.

Proposition 5. *Consider a match consisting of T games in Model B1. If $e(-1) < 0$, then the expected payoff for Player 1 is greater than the expected payoff for Player 2.*

Proof. See Appendix A.

3.3.2. Model B2: strategic risk taking

Thus far we have seen that strategic risk taking favors Player 2 (Section 3.2) while psychological effects $e(\cdot)$ favor Player 1. Next, we study the model in which both effects are considered simultaneously and the extent to which they can be compared. We do this under the assumptions on risk of Model R1 (white has all control over risk), not only because the resulting model is more tractable but also because it is the one where the effect of strategic risk taking is strongest.[21] To further facilitate the comparison, we fix a special form for the $e(\cdot)$ function.[22] We assume that there is $\lambda > 0$ such that $e(k)$ is essentially of the form $k\lambda$; that is, the magnitude of the effect is proportional to k.[23] This also means that $-e(-k) = e(k)$. So the leading/lagging state impacts performance in the same manner for the two types of pieces.

The following result, whose proof builds upon the proof of Proposition 5, says that the behavioral effect can be larger than the effects of strategic risk taking.

Proposition 6. *Consider a match consisting of T games in Model B2 with $\lambda > 0$. If R_w is small enough relative to λ, then the expected payoff for Player 1 is greater than the expected payoff for Player 2 (for all $\alpha \geq 1$).*

Proof. See Appendix A.

It may also be interesting to obtain some sufficient condition for the above result connecting λ and R_w. This is what we do next.

Proposition 7. *Consider a match consisting of $T = 2$ games with $0 < \lambda < M$. Then, if $\lambda > \frac{WR_w}{W-L}$, optimal play in this match leads to a higher expected payoff for Player 1 than for Player 2.*

Proof. See Appendix A.

It is interesting to note that the parameter α does not play any role in this condition on λ,[24] and that this condition is quite intuitive.[25] Finally, we note that a thorough numerical analysis

[17] Current research in economic theory is trying to understand how to empirically distinguish among different models of reference dependence that share similar formulations but specify different processes of reference point formation (see, e.g., Masatlioglu and Raymond (2014)). This is not at all trivial. In fact, providing separation between competing accounts of reference-dependence is empirically difficult, often pushing the limits of experimental feasibility. See Sprenger (2015) for a novel experimental design that successfully accomplish this separation. Distinguishing among competing models, however, is beyond what can be studied in our empirical setting.

[18] "Disappointment," "elation" and "discouragement" are terms used both in the economics literature (e.g., Gill and Prowse (2012) and Abeler et al. (2011) in contexts of effort provision) and in the psychology literature.

[19] Goldman and Rao (2014) provide supporting evidence for these ingredients from basketball. Using data from hundreds of thousands of plays, they find that among NBA players: (1) expectations do not influence the reference point, which appears remarkably stable around zero, and (2) when trailing, players perform worse on focus-intensive effortless tasks (they shoot free throws with lower accuracy), a finding also in Mertel (2011). Also in the context of basketball but at the level of teams, which may exert effort and choose risk, Berger and Pope (2011) find evidence where lagging by a little at half time can lead to winning at the end of the match.

[20] An intriguing theoretical innovation is the possibility of incorporating a conscious *choice* of anticipation (how to mentally prepare) as a mechanism through which reference points are formed as beliefs (see Sarver, 2014).

[21] Of course, a similar analysis is possible under the assumptions of Model R2.

[22] We have studied several variations of the function $e(\cdot)$, obtaining comparable results.

[23] Formally, in order to ensure that we have well-defined probabilities for the three possible outcomes of a chess game we need that, for all k, $W+R_w+e(k) \leq 1$, $L-e(k) \geq 0$, and $L+\alpha R_w+e(k) \leq 1$. So, we simply let $M = min\{1-W-R_w, L, 1-L-\alpha R_w\}$ and define

$$e(k) = \begin{cases} \min\{k\lambda, M\} & \text{if } k \geq 0 \\ -\min\{-k\lambda, M\} & \text{if } k < 0. \end{cases}$$

[24] The reason is that this condition is computed when Player 1 plays safe in game 1 and recall that $\alpha > 1$ implies that Player 2 only plays risky when he is lagging. In such a case, Player 2 is indifferent between a draw and a loss so he only cares about his probability of winning, which increases by R_w regardless of the value of α.

[25] The greater R_w is, the greater the psychological effect has to be to overcome the effect of strategic risk taking. And as $W - L$ represents how much more likely Player 1 is to be leading rather than lagging after game one, then, greater values of $W - L$ mean that the psychological effect will come into play more often in favor of Player 1 in game two, and this leads to smaller threshold values on λ. Another sufficient condition on λ can also be obtained by studying his expected payoff from playing risky.

suggests that a similar condition holds for matches of arbitrary length T.[26]

SUMMARY. We conclude from the theoretical analysis that rational effects operate in favor of Player 2 whereas behavioral effects suggested by the existing literature operate in favor of Player 1. Thus, evidence from average treatment effects will be consistent with the hypothesis that psychological effects are a relevant determinant of cognitive performance if the player starting with the white pieces (Player 1) significantly outperforms his opponent (Player 2). Likewise, the evidence will support the hypothesis that rational strategic risk taking is a significant determinant of observed performance if the player starting with the white pieces (Player 1) is outperformed by his opponent (Player 2).

4. Data

The dataset comes from *Chessbase*'s megabases, which are the most comprehensive databases in chess. They have detailed data on about 5 million games beginning in the XVIth century. We study *all* the matches during the period 1970–2010, namely about 511 matches with about 3000 chess games. We select these four decades since 1970 is the year when FIDE adopted the ELO rating system; that is, the year after which records on the cognitive ability of the players at this task, as measured by this rating, exist. The dataset is comprehensive as it includes *all* matches classified as such in Chessbase's megabases that exactly fit the randomized experiment described in Section 2.[27]

A valuable characteristic of the data set is that we have a reliable measure of the cognitive ability of the players performing the task. Players have a rating according to the ELO rating method, and the difference between two players' ELO ratings is functionally related to an estimate of the probability that one of the players will beat the other should they play a chess game. More precisely, a player's ELO rating is represented by a number that increases or decreases based upon the outcome of games between rated players. After every game, the winning player takes points from the losing player, and this number of points depends monotonically on the rating difference between the two players.[28] Nowadays, the top 10 players in the world typically have an ELO rating between 2770 and 2850 points, the top 100 players a rating above 2650, and players with a rating above 2500 points are professionals who have the title of Grandmaster, which is the highest title that a player can achieve.

Table 1 provides a description of the dataset and some pretreatment characteristics which, as expected, are not significantly different across the players. This is also confirmed when the dataset is split into different subsamples (see Table B1 in Appendix B).

Table 1
Dataset and Pretreatment Characteristics

Panel A: type and length of matches

	Matches	Games
	N	N
Type of match		
World championship	83	689
Non-world championship	428	2181
Elite	133	953
Non-elite	378	1917
Length of match		
Number of rounds < 6	214	658
Number of rounds 6–8	203	1231
Number of rounds > 8	94	981
All	511	2870

Panel B: pretreatment Characteristics of Players

	Player 1	Player 2	Difference
ELO rating	2507.66	2510.18	−2.52
	(164.01)	(160.78)	(105.82)
Age	30.82	31.52	−0.69
	(10.83)	(11.14)	(15.67)
ELO criterion	0.475	0.524	−0.049
	(0.493)	(0.493)	(0.986)
Age criterion	0.478	0.521	−0.043
	(0.487)	(0.487)	(0.973)

Notes: The variables *ELO criterion* and *Age criterion* indicate the proportion of times that the player shows a better entry in the criterion, *ELO rating* and *Age* respectively, at the time of the match (= 1 if higher, = 0 if lower, = 0.5 if same). Standard deviations in parentheses.

5. Empirical evidence

5.1. Evidence for professionals

In order to have an initial sense of the data, we begin by studying all matches where players have an ELO rating above 2500 and the rating difference between the players is no greater than 100 points. We choose this subset because these players are professionals, the stakes in the matches they play are high, and matches with a difference in ratings above 100 ELO points are quite uneven, as the strong player is expected to win with a very high probability regardless of other factors. This sample concerns 197 matches with a total of 1317 chess games.

As one would expect from the random treatment, the average quality of the players that begin with the white pieces (mean = 2620.1, std. deviation=62.1) and the average quality of the players that begin with the black pieces (mean = 2616.7, std. deviation = 57.4) are statistically identical (p-value = 0.56). If the order had no effect on the outcome of a match, the proportions of matches won by the two players should be statistically identical. Yet, we find that there is a significant and quantitatively important difference: the player who begins playing the first game with the white pieces wins 57.4% of the time, a proportion that is statistically different from 50% at the five percent significance level (p-value = 0.046).[29] The analysis in the previous section indicates that this average treatment effect arising from the randomly determined difference in the order of play is consistent with the hypothesis that psychological effects

[26] It seems natural that the same condition that is sufficient to give Player 1 an advantage in a two-game match is sufficient for longer matches as well. However, we have not been able to obtain an analytic expression. The main challenge comes from the fact that in a match of length T the best reply of a player (whether to play risky or safe) at each possible situation depends on all the parameters of the game and also on the current score. Yet, to study this conjecture, we have simulated matches for over more than a million random parameter configurations constrained by $\lambda > \frac{WR_W}{W-L}$ (and for each parameter configuration we solved for match lengths going from $T = 2$ to $T = 32$). In each and every one of these instances, the expected payoff of Player 1 was greater than the expected payoff of Player 2.

[27] As such, it does not include matches played versus a computer, matches without a perfect alternation of colors, or matches where there is an incumbent who wins in case of a tie (such as various current and past World Championship final matches). We also exclude two observations of matches in which various games were played in the same day without the standard resting time between games.

[28] In case of a draw, the lower rated player also gains a few points from the higher rated player. For a more detailed explanation of the ELO rating system we refer to Chapter B.02 in the FIDE Handbook (https://www.fide.com/fide/handbook.html) in World Chess Federation (2104).

[29] Using Graham's (2015) TL estimator to correct for the non-independence caused by having certain players playing more than one match in the sample the p-value is 0.037. This estimator is also used in the regression specifications. Further, in the subset of matches where players played in just one match, where non-independence is obviously not an issue, the proportions are maintained around 60-40.

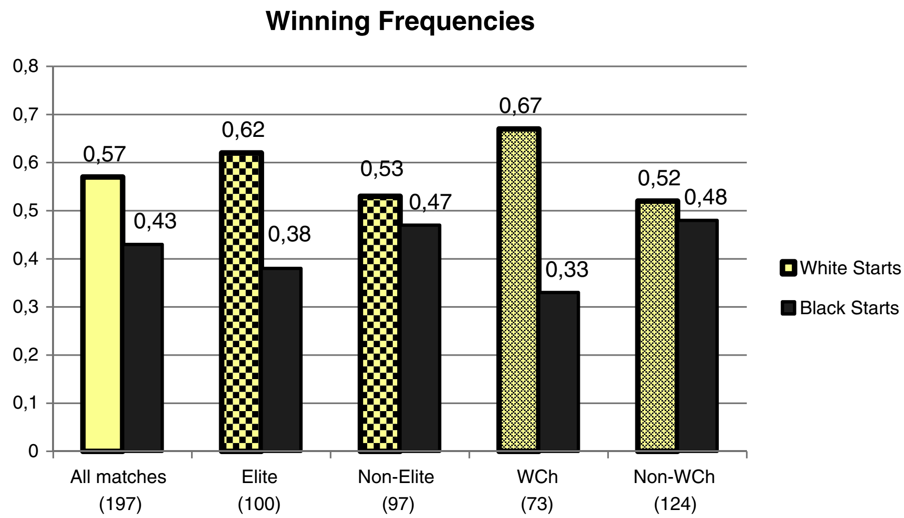

Fig. 1. The number of matches in each category are reported in parentheses. The p-values of the proportions Chi-square tests are 0.04 (All matches), 0.02 (Elite matches), 0.68 (Non-Elite matches), 0.004 (World Championship matches) and 0.79 (Non-World Championship matches).

resulting from the consequences of the playing order are a significant determinant of cognitive performance.[30]

In Fig. 1 we split these data into "Elite" versus "Non-Elite" matches, and in matches for the World Championship versus other matches. "Elite" matches are those played by players with an ELO rating above 2600, and World Championship matches are matches belonging to the World Championship cycles organized by FIDE. These are two intuitive ways of selecting arguably more important matches, where the stakes are even higher, and players are more skilled and have a deeper preparation.

We find that for Elite matches winning frequencies are 62-38 and for World Championship matches 67-33. These frequencies are statistically different from 50-50 at standard significance levels (for Elite matches p -value = 0.021, for World Championship matches p-value = 0.005). Thus, the magnitude and significance of the effects increase when considering Elite and World Championship matches.

5.2. Regression results

This subsection first reports the complete set of results for the three different samples of professionals studied earlier (professionals with a rating above 2500, and Elite and World Championship matches) with the same maximum rating difference. In each case we consider two different specifications (Table 2).

Not surprisingly, the results confirm the previous evidence: the effect of starting the match playing with the white pieces is positive and strongly significant in each of the regressions, typically with p-values below 0.05 and even below 0.01. Further, the impact becomes greater in magnitude and statistically more significant in the more important matches (Elite and World Championship). As expected, the difference in ELO ratings between the players also has a positive and significant impact in the probability of winning a match in every regression specification.

In Table 3 we report the main results. In the first two columns we consider *all* the matches in Chessbase's megabase *regardless*of the ELO level of the players and with no limit on the difference in ELO ratings. These are the most general specifications. In the next three columns we report the most complete specification for three minimum ELO levels as well (corresponding with 2200, 2400 and 2600 ratings).

The results continue to confirm the strongly significant effect of starting the match playing with the white pieces. Similarly, and not surprisingly, the difference in ELO ratings continues to have a positive and significant impact on the probability of winning a match. The same results arise in columns three to five for the various minimum ELO levels considered. A central result is that they are particularly strong, in terms of size and significance, at the highest level.[31] Together with the evidence from Elite and World Championship matches in the previous table, it indicates that the biases are strongest in the most important (and mentally more stressful) matches. Thus, the clear patterns we observe are consistent with the interpretation that increased stakes amplify the differences in cognitive performance associated with the effect we document.[32]

Finally, we also note that a small percentage of matches (not included in the sample) ended up tied. The same findings are obtained in the corresponding ordered probit regressions with the three outcomes (win, loss, tie) when these matches are included. These results are reported in Appendix B.

5.3. Additional testable implication

We next take further advantage of the opportunity provided by the fact that we have a reliable measure of the cognitive ability of the players to study the following testable prediction: given the undoubted role that other factors may play in determining the winner of a chess match, it should be the case that the effects of beginning with the white or black pieces significantly contribute to determining the outcome of a match *only* in relatively symmetric

[30] In the raw data, when the first game ends in a draw the winning probability (frequency) becomes higher for the player who started with the black pieces (45-55). Also, the likelihood of winning the match is higher when white wins the first game (87-13) and when black wins the first game (17-83). However, once the match begins and the first game is played we do not have the effect of randomization anymore as subsequent play is endogenous to the outcome of the first game. A random effects dynamic panel data model with lagged dependent variables and unobserved heterogeneity would then be needed to obtain unbiased and consistent estimates of the different correlation effects of the final score with interim scores.

[31] In the regression specifications of Tables 2 and 3 , the interaction of "ELO difference" with "Starting with White Pieces" is not significant and has no significant impact on the coefficient estimates of the rest of variables.

[32] Consistent with these findings, Goldman and Rao (2014) find that the accuracy of NBA players in focus-intensive effortless "free throws" is lower not only when trailing but also in the NBA playoffs and when the games are nationally televised.

Table 2
Probit Regressions for Winners of a Chess Match.

	Matches with a minimum ELO: 2500		Elite matches minimum ELO: 2600		World championship matches	
Intercept	−0.170	−0.170	−0.300	−0.300	−0.391	−0.390
	(16.68)	(17.58)	(23.10)	(24.61)	(29.27)	(31.43)
Starts with white pieces	0.340**	0.340**	0.598***	0.600***	0.782***	0.780***
	(0.133)	(0.128)	(0.190)	(0.188)	(0.224)	(0.228)
Rounds	0.000	0.000	0.000	0.000	−0.000	0.000
	(0.020)	(0.020)	(0.030)	(0.030)	(0.038)	(0.039)
Year	0.000	0.000	0.000	0.000	0.000	0.000
	(0.008)	(0.008)	(0.012)	(0.012)	(0.016)	(0.017)
ELO difference	0.012***	0.012***	0.013***	0.013***	0.011***	0.011***
	(0.001)	(0.001)	(0.002)	(0.0028)	(0.003)	(0.003)
ELO points	0.000	0.000	0.000	0.000	−0.000	0.000
	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)
Age difference		−0.006		−0.000		−0.005
		(0.006)		(0.010)		(0.012)
Age		0.000		0.000		0.000
		(0.009)		(0.014)		(0.015)
Maximum ELO difference=100	Yes	Yes	Yes	Yes	Yes	Yes
N (matches)	197	197	100	100	73	73
Log-Likelihood	−234.12	−233.24	−114.35	−114.34	−81.99	−81.82
Akaike information criterion	480.24	482.48	240.69	244.69	175.97	179.63

Notes: *** denotes significant at the 1 percent significance level, ** at the 5 percent level, and * at the 10 percent level. The independent variables are the following: "Starts with white pieces" is a dummy variable that equals 1 if the player started playing the first game in the match with the white pieces; "Rounds" is the number of rounds in the match; "Year" is the year the match takes place; "ELO difference" is the difference in ELO points with respect to the opponent; "Maximum ELO difference" is the maximum ELO difference between the players in the sample; "Player's ELO" is the ELO rating of the player; "Age difference" is the difference in age with respect to the opponent, and "Age" is the age of the player. All the variables involving the rating and the age of the players are at the time the match takes place. The number of matches in which the difference in ELO levels is above 100 is 6 (World Championships), 16 (Elite, ELO>2600), 39 (ELO>2500), 68 (ELO>2400), 127 (ELO>2200), and 142 (complete sample). The number of tied matches are 4 (World Championships), 17 (Elite, ELO>2600), 33 (ELO>2500), 49 (ELO>2400), 67 (ELO>2200), and 73 (complete sample). Standard errors in parentheses (Graham (2015)).

matches. That is, in any of the formal models considered in the previous section, the more similar in cognitive strength the two players are, the greater the effects should be, and differences in the ability of the players should attenuate the differences in performance observed in the natural experiment. In other words, the order of colors should presumably tip the balance only when other factors are relatively similar, and this effect should steadily decrease as players are more different in their cognitive skills. We study this implication in Fig. 2 which reports the evidence for the sample of matches studied in Subsection 5.1.

Matches are sorted by the difference in ELO ratings between the players, and then divided into quartiles from more similar players (quartile 1) to less similar players (quartile 4). We find that the p-values of the proportions Chi-square tests are 0.02 (quartile 1), 0.44 (quartile 2), 0.88 (quartile 3) and 0.88 (quartile 4). Further, reassuringly, when ELO differences are larger than 100 (not reported in the

Table 3
Probit Regressions for Winners of a Chess Match

	Complete sample		Minimum ELO		
			2200	2400	2600
Intercept	−0.101	−0.100	−0.120	−0.130	−0.280
	(11.50)	(11.66)	(12.18)	(14.62)	(24.32)
Starts with white pieces	0.202**	0.200**	0.240**	0.260**	0.570***
	(0.090)	(0.090)	(0.096)	(0.109)	(0.188)
Rounds	−0.000	0.000	0.000	0.000	0.000
	(0.014)	(0.014)	(0.016)	(0.018)	(0.039)
Year	−0.000	0.000	0.000	0.000	0.000
	(0.005)	(0.005)	(0.006)	(0.007)	(0.012)
ELO difference	0.008***	0.009***	0.010***	0.012***	0.015***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)
ELO points	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)
Age difference		−0.011**	−0.012***	−0.015***	−0.001
		(0.004)	(0.004)	(0.005)	(0.010)
Age		0.000	0.000	0.000	0.000
		(0.006)	(0.006)	(0.007)	(0.014)
Maximum ELO difference	None	None	None	None	None
N (matches)	438	438	414	315	116
Log-Likelihood	−465.62	−459.23	−427.65	−327.54	−116.24
Akaike information criterion	943.24	934.45	871.29	671.08	248.48

Notes: *** denotes significant at the 1 percent significance level, ** at the 5 percent level, and * at the 10 percent level. The independent variables are the following: "Starts with white pieces" is a dummy variable that equals 1 if the player started playing the first game in the match with the white pieces; "Rounds" is the number of rounds in the match; "Year" is the year the match takes place; "ELO difference" is the difference in ELO points with respect to the opponent; "Maximum ELO difference" is the maximum ELO difference between the players in the sample; "Player's ELO" is the ELO rating of the player; "Age difference" is the difference in age with respect to the opponent, and "Age" is the age of the player. All the variables involving the rating and the age of the players are at the time the match takes place. The number of tied matches are 4 (World Championships), 17 (Elite, ELO>2600), 33 (ELO>2500), 49 (ELO>2400), and 67 (ELO>2200). Standard errors in parentheses (Graham (2015)).

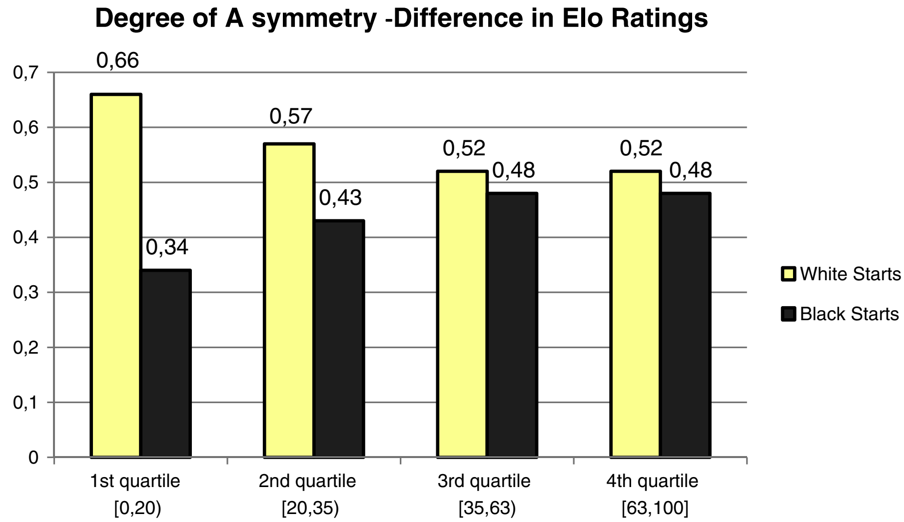

Fig. 2. Matches are sorted by the difference in ELO ratings into four intervals (quartiles). The endpoint of the first interval is chosen so that it cuts off the lowest 25% of the sorted ELO differences. Similarly, the other endpoints cut off the 50%, 75% and 100%, respectively. The intervals are reported below each quartile. The number of matches in each quartile is 61, 42, 46 and 48 respectively. The differences in the number of matches across the intervals come from the discreetness of the ELO ratings, which until the late 1999 moved in increments of 5 points.

figure), there is no significant effect (p-value=0.7423). Thus consistent with the basic hypothesis, *only* in matches between players of similar cognitive ability are there significant differences in performance between the players. And as predicted the size of the effect increases when players become more similar in cognitive skills.[33]

6. Concluding remarks

Understanding all aspects of "competition" is central to economics, and understanding the effects of cognitive and noncognitive abilities is important not only in economics but in areas ranging from cognitive psychology to neuroscience. Competitive situations that involve cognitive performance are widespread in labor markets, education, and organizations, including test taking, student competition in schools, competition for promotion in firms, and numerous other settings. This paper contributes to the theoretical and empirical literature on dynamic competitive situations, which shows that incorporating behavioral elements arising from the *state* of the competition may offer significant insights about human behavior that otherwise would be lost. First, we have developed rational and behavioral models that incorporate this ingredient. Second, in terms of empirics, the literature has found that these emotional states are important for explaining the behavior of professional subjects performing non-cognitive tasks in sports such as golf, soccer, basketball and weightlifting. Our findings show that they are also important in competitive cognitive tasks.[34]

We hope these results will stimulate further research. We have studied the impact of randomly allocated emotional differences on cognitive decision making in a competitive situation involving high stakes, sophisticated players, and elaborate decision processes. Previous research has found that individuals with higher cognitive ability tend to exhibit fewer and less extreme cognitive biases that may lead to suboptimal behavior. Thus, an open question for future research is the extent to which these effects are important in other parts of the distribution of cognitive abilities, in tasks and settings with lower stakes, and even among the poor (Mani et al., 2013). This may have relevant public policy implications.

A second open question concerns models of reference points. Current developments in the theoretical and experimental literature are trying to understand how to distinguish among various models of reference dependence, and how these models relate to other models of non-expected utility theory which rely on different psychological intuitions.[35] This represents an important and necessary contribution for continued progress. The results in this paper suggest that the study of cognitive performance, both in individual decision-making and in game theoretical situations, may represent a fruitful area for future research to help us learn about processes of reference point formation.

Appendix A

A.1. Proof of Proposition 2

Proof. We solve the game backwards. We start in Period 2, where Player 2 plays with the white pieces. If Player 2 lost the first game, he can only get a positive payoff by winning the second game, so he chooses the risky action. If the first game was a draw then, since the utility for a tie in the match is 0.5, Player 2 plays safe. That is, he does not want to transfer probability from the probability of drawing to the probability of winning if this entails transferring an even greater probability from the probability of drawing to the probability of losing (recall that $\alpha > 1$). If Player 2 won the first game he just wants to minimize the probability of losing the second game, so he plays safe.

[33] This result may also contribute to explaining the previous findings that the effect is stronger for Elite and World Championship matches. These are matches where players have a deeper preparation and hence, conditional on their rating, their deeper preparation may make them more similar in cognitive skills during their matches.

[34] In schools, for instance, providing students with relative performance information (indicative of the state of the competition) has an impact on future performance (Azmat and Iriberri (2010)).

[35] See Masatlioglu and Raymond (2014), Sprenger (2015), and Sarver (2014).

We look now at Period 1. We find that the expected payoff for Player 1 is less than 0.5 regardless of his action. The expected payoff for Player 1 if he chooses the safe action is:

$$W\left(1-\frac{W+R_w}{2}\right)+(1-W-L)\left(L+\frac{1-W-L}{2}\right)+\frac{L^2}{2},$$

and this reduces to $\frac{1}{2}(1-WR_w)$, which is less than 0.5. His expected payoff when choosing the risky action is:

$$(W+R_w)\left(1-\frac{W+R_w}{2}\right)+(1-W-L-(1+\alpha)R_w)\left(L+\frac{1-W-L}{2}\right)+\frac{(L+\alpha R_w)L}{2},$$

and this reduces to $\frac{1}{2}\left(1-R_w^2-R_w(\alpha(1-W)-(1-W-L))\right)$, which is also less than 0.5 (recall that $\alpha>1$). ■

A.2. Proof of Corollary 1

Proof. We already know that Player 2 can get an expected payoff of 0.5 simply by mimicking the actions of Player 1 in the previous game. Assume now that he always follows the mimicking strategy except in game T in the following case: if the score is tied at the end of game $T-2$, then he plays risky in game T if he lost game $T-1$ and plays safe otherwise. Proposition 2 implies that this is a profitable deviation from the mimicking strategy and, therefore, its expected payoff for Player 2 is greater than 0.5. ■

A.3. Proof of Proposition 3

Proof. We solve the game backwards starting with Period 2. We distinguish three cases.

Case. Suppose Player 1 won game 1. Then, in game 2, Player 1 is indifferent between a victory and a draw, so he plays safe. Player 2 is indifferent between a loss and a draw, so he plays risky. The expected utility for Player 1 in this case is:

$$u_1=\frac{1}{2}(W+R_w)+1(1-(W+R_w)).$$

Case. Suppose Player 1 lost game 1. Following the above reasoning, Player 1 plays risky in game 2 and Player 2 plays safe. The expected utility for Player 1 in this case is:

$$u_2=\frac{1}{2}(L+R_b).$$

Case. Suppose that game 1 was a draw. Then, in game 2, since the utility for a tie in the match is 0.5, both players play safe. That is, they do not want to transfer probability from the probability of drawing to the probability of winning if it entails transferring an even greater probability from the probability of drawing to the probability of losing (recall that $\alpha>1$). The expected utility for Player 1 in this case is:

$$u_3=1L+\frac{1}{2}(1-(W+L)).$$

We move now to Period 1. We can compute the expected utilities for each combination of strategy profiles of the players in period 1. Let u_{SR} denote the expected utility of Player 1 when he plays safe and Player 2 plays risky and similarly the other utilities. Then:

$$\begin{aligned}
u_{SS}&=Wu_1+Lu_2+(1-W-L)u_3\\
u_{RR}&=(W+R_w+\alpha R_b)u_1+(L+\alpha R_w+R_b)u_2\\
&\quad+(1-(W+L+(1+\alpha)(R_w+R_b)))u_3\\
u_{RS}&=(W+R_w)u_1+(L+\alpha R_w)u_2+(1-((W+R_w)+(L+\alpha R_w)))u_3\\
u_{SR}&=(W+\alpha R_b)u_1+(L+R_b)u_2+(1-((W+\alpha R_b)+(L+R_b)))u_3.
\end{aligned}$$

We now show that, given Period 2's optimal behavior, Player 2 can ensure for himself an expected utility higher than 0.5 by playing safe in period 1. Then, whatever Player 2's optimal choice is in period 1, it also gives him an expected payoff higher than 0.5. We compute the utilities of Player 1 with his two actions in period 1 when Player 2 is playing safe. First, u_{SS} reduces to $\frac{1}{2}(1+R_bL-R_wW)$, which is less than 0.5 when C1 is satisfied. Second, u_{RS} reduces to

$$u_{RS}=\frac{1}{2}\left(1+L(R_b-R_w)-R_w^2+R_w(1+(\alpha-1)W+\alpha(-1+R_b))\right),$$

which after some algebra can be rewritten as:

$$u_{RS}=\frac{1}{2}\left(1-(L+R_w)(R_w-R_b)-R_w((\alpha-1)(1-W-R_b))\right),$$

which is less than 0.5 when C2 is satisfied. ■

A.4. Proof of Proposition 4

Proof. The expected payoff of Player 1 is given by

$$W\left(1-\frac{W+e(-1)}{2}\right)+(1-W-L)\left(L+\frac{1-W-L}{2}\right)+L\frac{L-e(1)}{2},$$

which reduces to $\frac{1}{2}+\frac{-e(-1)}{2}W-\frac{e(1)}{2}L\geq\frac{1}{2}+\frac{-e(-1)}{2}(W-L)$ where the inequality follows from the assumption $-e(-1)\geq e(1)$. Since $W-L>0$ and $-e(-1)>0$, the expected payoff of Player 1 is greater than 0.5. Note that rather than assuming $-e(-1)\geq e(1)$, we could instead have made the weaker assumption that $-e(-1)W>e(1)L$. ■

A.5. Proof of Proposition 5

Proof. A T-game match is composed of many two-game mini-matches. Suppose that a two-game mini-match is about to start and let k denote the difference between the number of games won and the number of games lost by Player 1. We denote by $P_1(k,s)$ the probability that Player 1 is leading/lagging by $k+s$ games after the mini-match.

Following the above notation, $P_1(0,s)$ denotes the probability that Player 1 is leading/lagging by s games after the first two games of the match. Then, $P_1(0,2)=W(L+e(1))$, $P_1(0,1)=(1-W-L)(W+L)$, $P_1(0,0)=(1-W-L)^2+L(L-e(1)+W(W-e(1))$, $P_1(0,-1)=(1-W-L)(W+L)$, and $P_1(0,-2)=L(W+e(1))$. In particular, $P_1(0,2)-P_1(0,-2)=(W-L)e(1)>0$ and $P_1(0,1)=P_1(0,-1)$.

The above probabilities show that Player 1 has a higher probability of being better-off than Player 2 after the first two games. What we show next it is also easier for Player 1 to make a comeback than it is for Player 2. More precisely, we show that, for each $k\in\mathbb{N}$ and each $s\in\{2,1,0,-1,-2\}$, $\sum_s^2P_1(k,s)\geq\sum_s^2P_1(-k,-s)$. In words, the probability that Player 1 improves his score by at least s

points when the current score is k is greater than the probability that it is for Player 2 to improve by at least s points when the score is $-k$. Establishing this result independently of k and s, combined with $P_1(0,2)-P_1(0,-2)>0$ and $P_1(0,1)=P_1(0,-1)$, delivers the result.

The difference $P_1(k,2)-P_1(-k,-2)$ reduces to $(W-L)(e(k+1)-e(k))$ which is non-negative, since $e(\cdot)$ is an increasing function. Now, $P_1(k,1) = P_1(-k,-1)$ and, hence, $P_1(k,2)+P_1(k,1)-P_1(-k,-2)-P_1(-k,-1)$ reduces again to $(W-L)(e(k+1)-e(k))$. $P_1(k,0)-P_1(-k,0) = (W-L)(2e(k)-e(k+1)-e(k-1)$ and so $P_1(k,2)+P_1(k,1)+P_1(k,0)-P_1(-k,-2)-P_1(-k,-1)-P_1(-k,0)$ equals $(W-L)(e(k)-e(k-1))$ which is also non-negative. The next inequality follows from the fact that $P_1(k,-1)=P_1(-k,1)$ and the final one is immediate since $\sum_{s=-2}^{2} P_1(k,s) = \sum_{s=-2}^{2} P_1(-k,-s) = 1$. ■

A.6. Proof of Proposition 6

Proof. Note that Model B1 is a particular case of Model B2 in which $R_w = 0$. We could now replicate the probabilistic analysis in the proof of Proposition 5 and obtain new expressions for the probabilities computed there. Importantly, all of them would be continuous in R_w. Therefore, since the case $R_w = 0$ leads to a strictly greater expected payoff for Player 1, the same will also be true if R_w is small enough. For the observation regarding α we just need to note that if a value of R_w works for $\alpha = 1$, it will also work for greater α (which means that risk taking is more costly). ■

A.7. Proof of Proposition 7

Proof. By repeating the arguments in the proof of Proposition 2, we can get that the expected payoff for Player 1 if he chooses the safe action is:

$$W\left(1-\frac{W+R_w-\lambda}{2}\right)+(1-W-L)\left(L+\frac{1-W-L}{2}\right)+\frac{L(L-\lambda)}{2}.$$

This reduces to $\frac{1}{2}(1-WR_w+\lambda(W-L))$, which is more than 0.5 if $\lambda > \frac{WR_w}{W-L}$. ■

Appendix B. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jpubeco.2016.05.001

References

Abeler, J., Falk, A., Goette, L., Huffman, D., 2011. Reference points and effort provision. Am. Econ. Rev. 101, 470–492.
Agarwal, S., Mazumder, B., 2013. Cognitive abilities and household decision making. Am. Econ. J. Appl. Econ. 5, 193–207.
Apesteguia, A., Palacios-Huerta, I., 2010. Psychological pressure in competitive environments: evidence from a randomized natural experiment. Am. Econ. Rev. 100, 2548–2564.
Azmat, G., Iriberri, N., 2010. The importance of relative performance feedback information: evidence from a natural experiment using high school students. J. Public Econ. 94, 435–452.
Bechara, A., Damasio, H., Tranel, D., Damasio, A.R., 1997. Deciding advantageously before knowing the advantageous strategy. Science 275, 1293–1295.
Benjamin, D., Brown, S.A., Shapiro, J., 2013. Who is 'behavioral'? Cognitive ability and anomalous preferences. J. Eur. Econ. Assoc. 111, 1231–1255.
Berger, J., Pope, D., 2011. Can losing lead to winning? Manag. Sci. 57, 817–827.
Bertrand, M., Morse, A., 2011. Information disclosure, cognitive biases, and payday borrowing. J. Financ. 66, 1865–1893.
Bhaskar, V., 2009. Rational adversaries? Evidence from randomised trials in one day cricket. Econ. J. 119, 1–23.
Cabral, L., 2002. Increasing dominance with no efficiency effect. J. Econ. Theory 102, 471–479.
Cabral, L., 2003. R&d competition when firms choose variance. J. Econ. Manag. Strateg. 12, 139–150.
Camerer, C.F., 2003. Behavioral Game Theory. Princeton University Press, Princeton, NJ.
Carneiro, P., Hansen, K., Heckman, J.J., 2003. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. Int. Econ. Rev. 44, 361–422.
Carneiro, P., Heckman, J.J., Heckman, J.J., 2003. Human capital policy In: Krueger, A.B., Friedman, B. (Eds.), Inequality in America: What Role for Human Capital Policies?. MIT Press, Cambridge, MA.
Cole, S., Shastry, G.K., 2009. Smart Money: The Effect of Education, Cognitive Ability, and Financial Literacy on Financial Market Participation. Working Paper, 2009, Harvard Business School.
Cunha, F., Heckman, J.J., Schennach, S.M., 2010. Estimating the technology of cognitive and non-cognitive skill formation. Econometrica 78, 883–931.
DellaVigna, S., 2009. Psychology and economics: evidence from the field. J. Econ. Lit. 47, 315–372.
Ehrenberg, R., Bognanno, M., 1990. Do tournaments have incentive effects? J. Polit. Econ. 98, 1307–1324.
Frederick, S., 2005. Cognitive reflection and decision making. J. Econ. Perspect. 19, 25–42.
Garicano, L., Palacios-Huerta, I., Prendergast, C., 2005. Favoritism under social pressure. Rev. Econ. Stat. 87, 208–216.
Genakos, C., Pagliero, M., 2012. Interim rank, risk taking, and performance in dynamic tournaments. J. Polit. Econ. 120, 782–813.
Genakos, C., Pagliero, M., Garbio, E., 2015. When pressure sinks performance: evidence from diving competitions. Econ. Lett. 132, 5–8.
Gerardi, K., Goette, L., Meier, S., 2010. Financial Literacy and Subprime Mortgage Delinquency: Evidence from a Survey Matched to Administrative Data. Working Paper.
Gill, D., Prowse, V., 2012. A structural analysis of disappointment aversion in a real effort competition. Am. Econ. Rev. 102, 469–503.
Gill, D., Prowse, V., 2015. Cognitive, ability, character skills, and learning to play equilibrium: a level-k analysis. J. Polit. Econo. forthcoming.
Goldman, M., Rao, J.M., 2014. Loss aversion around a fixed reference point in highly experienced agents, UC san diego mimeo.
Graham, B.S., 2015. An econometric model of link formation with degree heterogeneity, UC-berkeley mimeo.
Harris, C., Vickers, J., 1987. Racing with uncertainty. Rev. Econ. Stud. 54, 1–21.
Heckman, J.J., Lochner, L.J., Todd, P.E., 2006. Earnings equations and rates of return: the mincer equation and beyond. In: Hanushek, E.A., Welch, F. (Eds.), Handbook of the Economics of Education. 1. pp. 307–458.
Heckman, J.J., Stixrud, J., Urzua, S., 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. J. Labor Econ. 24, 411–482.
Hvide, H.K., 2002. Tournament rewards and risk taking. J. Labor Econ. 20, 877–898.
Hvide, H.K., Krinstiansen, E.G., 2003. Risk taking in selection contests. Games Econ. Behav. 42, 172–179.
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47, 263–291.
Kermer, D.A., Driver-Linn, E., Wilson, T.D., Gilbert, D.T., 2006. Loss aversion is an affective forecasting error. Psychol. Sci. 17, 649–653.
Klumpp, T., Polborn, M.K., 2006. Primaries and the new hampshire effect. J. Public Econ. 90, 1073–1114.
Konrad, K.A., Kovenock, D., 2009. Multi-battle contests. Games Econ. Behav. 66, 256–274.
Köszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. Q. J. Econ. Rev. 121, 1133–1165.
Köszegi, B., Rabin, M., 2007. Reference-dependent risk attitudes. Am. Econ. Rev. 97, 1047–1073.
Mani, A., Mullainathan, S., Shafir, E., Zhao, J., 2013. Poverty impedes cognitive function. Science 341, 976–980.
Manski, C.F., 1995. Identification Problems in the Social Sciences. Harvard University Press, Cambridge, MA.
Masatlioglu, Y., Raymond, C., 2014. A behavioral analysis of stochastic reference dependence, university of michigan and oxford university, mimeo.
McArdle, J.J., Smith, J.P., Willis, R., 2009. Cognition and economic outcomes in the health and retirement study. Master Thesis. NBER Working Paper, 15266.
Mertel, M., 2011. Under Pressure: Evidence from Repeated Actions by Professional Sportsmen. CEMFI, Madrid.
Neal, D.A., Johnson, W.R., 1996. The role of pre-market factors in Black–White wage differences. J. Polit. Econ. 15, 869–898.
Ochesner, K.N., Ray, R.D., Cooper, J.C., Robertson, E.R., Chopra, S., Gabrieli, J.D.E., Gross, J.J., 2004. For Better or worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. Neuroimage 23, 483–499.
Osborne, M.J., Rubinstein, A., 1994. A Course in Game Theory. MIT Press, Cambridge, MA.
Pope, D., Schweitzer, M., 2011. Is tiger woods loss averse? Persistent bias in the face of experience, competition, and high stakes. Am. Econ. Rev. 101, 129–157.
Rabin, M., 1998. Psychology and economics. J. Econ. Lit. 36, 11–46.
Romer, D., 2006. Do firms maximize? Evidence from professional football. J. Polit. Econ. 114, 340–365.
Rosen, S., Sanderson, A., 2001. Labour markets in professional sports. Econ. J. 111, 47–68.
Sarver, T., 2014. Optimal reference points and anticipation, duke university, mimeo.
Schaefer, S.M., Jackson, D.C., Davidson, R.J., Aguirre, G.K., Kimberg, D.Y., Thompson-Schill, S.L., 2002. Modulation of amygdalar activity by the conscious regulation of negative emotion. J. Cogn. Neurosci. 14, 913–921.

Sokol-Hessner, P., Hsu, M., Curley, N.G., Delgado, M.R., Camerer, C.F., Phelps, E.A., 2009. Thinking like a trader selectively reduces individuals' loss aversion. Proc. Natl. Acad. Sci. 106, 5035–5040.
Sprenger, C., 2015. An endowment effect for risk: experimental tests of stochastic reference points. J. Polit. Econ. 123, 1456–1499.
Szymanski, S., 2000. A market test of discrimination in the english soccer leagues. J. Polit. Econ. 108, 590–603.
Szymanski, S., 2003. The economic design of sporting contests. J. Econ. Lit. 41, 1137–1187.
Walker, M., Wooders, J., Amir, R., 2011. Equilibrium play in matches: binary Markov games. Games Econ. Behav. 71, 487–502.
World Chess Federation, 2014. FIDE Rating Regulations. FIDE Handbook (Chapter B.02 July https://www.fide.com/fide/handbook.html