Bühren, Christoph; Frank, Björn

Working Paper

Chess players' performance beyond 64 squares: A case study on the limitations of cognitive abilities transfer

MAGKS Joint Discussion Paper Series in Economics, No. 19-2010

Provided in Cooperation with:
Faculty of Business Administration and Economics, University of Marburg

MAGKS

Aachen · Siegen · Gießen · Marburg · Kassel · Göttingen

Joint Discussion Paper Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 19-2010

Christoph Bühren and Björn Frank

Chess players' performance beyond 64 squares: A case study on the limitations of cognitive abilities transfer

Coordination: Bernd Hayo • Philipps-University Marburg
Faculty of Business Administration and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Chess players' performance beyond 64 squares: A case study on the limitations of cognitive abilities transfer*

by Christoph Bühren♦ and Björn Frank✉, University of Kassel

Abstract

In a beauty contest experiment with over 6,000 chess players, ranked from amateur to world class, we found that Grandmasters act very similar to other humans. This even holds true when they play exclusively against players of approximately their own strength. In line with psychological research on chess players' thinking, we argue that they are not more rational in a game theoretic sense per se. Their skills are rather specific for their game.

Keywords: chess, beauty contest, cognitive transfer

JEL C93, C72, D03

---

* We are indebted to Chessbase for making our experiment possible and to the chess players who, apart from taking part in our experiment, provided valuable comments, as did Joel Sobel and participants of the GfeW 2009 conference, a workshop in Kassel, the Hohenheimer Oberseminar in Esbjerg 2010, especially Jürgen Zerth.

♦ c.buehren@uni-kassel.de

✉ frank@uni-kassel.de, University of Kassel, IVWL, Nora-Platiel-Straße 4, 34127 Kassel, Germany,
☎ ++49-561-804-7187

1. Introduction

It is often a good idea to ask experts for help in solving certain problems. The trick is to find the right expert. Many auctions have optimal strategies and equilibria that game theorists are educated to find, and hiring game theorists as consultants has helped governmental agencies or firms in a number of important auctions (e.g., Milgrom, 2004). But who is an expert in game theory, apart from game theorists? This is an important question, for two reasons. First, if we find that people, at least certain people, behave according to game theoretic predictions where others do not, this would help to understand to which extent and in which cases game theory can be used for predictive purposes, rather than just as a normative theory. Second, it is important to understand the extent to which we can expect experts in one field to transfer their expertise into other fields. Many hiring decisions are based not on the current knowledge and training of the prospective employee, but on the extent to which his or her expertise can be transferred into new areas. Probably Goldman Sachs had a hypothesis on this question when they hired chess Grandmaster Luke McShane as a trader. Gerdes and Gränsmark (2010) motivate their study on differences between female and male chess players (with respect to risk-taking) with the claim "that findings based on chess can be transferred to other professions that are characterized by a high level of expertise" (p. 5).

However, a look at a larger range of chess players' biographies does not reveal a clear complementary talent: among (current or former) top chess Grandmasters, we find a successful entrepreneur (Miguel Najdorf), a Harvard economist (Kenneth Rogoff), a papyrologist (Robert Hübner), a former member of Norway's national soccer team (Simen Agdestein), a world class pianist (Mark Taimanov), a paranoid antisemite (Bobby Fischer), a taxi driver (Nicolas Rossolimo), a psychoanalyst (Reuben Fine) and Soviet dissidents like Viktor Korchnoi. Do they have anything in common except for their ability to play chess?

Until recently there was no evidence supporting this point for the transfer of chess playing abilities to other strategic situations. In this paper, we provide evidence on the extent to which chess players' expertise can be transferred. The task we confronted them with was the beauty contest. In section 2, we discuss how our work relates to two streams of research: first, economists' experiments on cognitive transfer, and second, psychologists' research on chess players' thinking. Section 3 reports our experimental findings, suggesting that a chess grandmaster is not necessarily a beauty contest grandmaster. Section 4 concludes.

2. Related Literature

2.1. Experiments on Cognitive Transfer

It is not straightforward how the "size" or "difficulty" of a transfer should be operationalized, let alone measured. However, one important aspect to be considered is causation. Presume that we find that abilities in tasks X and Y correlate positively, then this correlation might be due to causation or not. Presume that we deal with a group of people with a lot of past experience in task (or occupation) Y. For the first time in their lives, we confront them with task X and find that they are particularly good at X. It is natural to presume that being good at Y causes being good at X, but this is not necessarily true. First, there might be a common cause for good performance in X and Y, and second, there might even be reverse causation: People who (surprisingly or not) turn out to be good at task X might for some reason have been more likely to survive in occupation Y. These different reasons behind observed correlations help us to organize experimental evidence on cognitive transfer.

Unclear direction of the cognitive transfer

List and Haigh (2005) study the Allais paradox with students and with professional floor traders and report that the latter violate the independence axiom slightly less often than the students. This is an interesting result, yet we do not know whether it is due to a common cause, like a certain sense of rationality, or whether trading provides training that eventually reduces violations of the independence axiom, or whether violating the independence axiom reduces the likelihood of being sufficiently successful as a trader. Another study with unclear causation is that of Potters and van Winden (2000), who compare the behavior of students and of professional lobbyists in an experiment that reconstructs certain aspects of lobbying, finding that the lobbyists are more rational in the game-theoretic sense and earn more money.[1] Finally, Alevy et al. (2007) show that professional traders are "less Bayesian" in information cascades games, but perform better in other dimensions, whereas professional traders are subject to more, rather than less, myopic loss aversion than students according to Haigh and List (2005). These latter results serve to warn us that potentially not only cognitive abilities,

[1] Cooper (2006) finds that experienced managers are better able than students to overcome coordination failures. It is not clear, however, whether this finding is correctly classified as 'cognitive transfer', or whether it just confirms that managers are good in managing.

but also their transfer might be rather specific and should not be generalized without complementing evidence.[2]

Unambiguous direction of the cognitive transfer

Some artificial field experiments suggest that there can indeed be cognitive abilities transfers in the causal sense: being good at Y causes an increased likelihood of being good at X. The reason is that task X is unlikely to have an impact on task Y. The prime examples are studies on soccer players' use of mixed strategies in simple experimental games. Task Y is playing soccer. Soccer players might be good at mixing strategies, as it is important to randomize, i.e., not to perform in a predictable way, in penalty kicks. However, the reverse causality is unlikely to be at work here[3]: effectively mixing is only a small part of playing soccer, and certain players in a team can - and do - specialize in penalty kicks, hence not everyone has to be good at it.

There is convincing evidence that soccer players come close to randomizing optimally in penalty kicks (Chiappori et al., 2002; Palacios-Huerta, 2003; Azar and Bar-Eli, 2010). Evidence on the question whether this ability extends to the domain of other strategic situations is mixed in an irritating way. Palacios-Huerta and Volij (2008) find that professional soccer players, unlike students, play (nearly) optimal mixed strategies in a 2x2 and in a 4x4 zero-sum card game in the laboratory, even avoiding serial dependence of their strategy choices.[4] However, Levitt et al. (2010) failed to replicate this result with American soccer players.[5] They also, and more surprisingly, do not find that highly skilled poker players come close to the minimax predictions in a context that is unfamiliar to them.[6] Sports that require intellectual, rather than physical, training provide attractive subjects for the study of cognitive transfers, as the transfer itself is part of the training. Professionals in intellectual sports such as poker should be able to understand and explain what they are doing, in stark contrast to, say, soccer players. (Consequently, some of the latter fail terribly as trainers.) Hence it was quite reasonable that Palacios-Huerta and Volij (2009) hypothesize that

[2] This is also pointed out in Levitt et al. (2010), section IV.

[3] The caveats by Palacios-Huerta and Volij (2008, p.112) notwithstanding.

[4] But see Wooders (2010) for a critical reexamination of the data from Palacios-Huerta and Volij (2008).

[5] Palacios-Huerta and Volij (2008) use Spanish soccer players. Levitt et al. (2010), section IV, list a number of further differences in experimental conditions, but none of these would be expected to be the cause of the stark differences in the results.

[6] In the case of poker, where randomization is essential to success, a different result would have left us, again, with the question of causality being undecided. The result provided by Levitt et al. (2010), however, has an unambiguous interpretation: There was no transfer of cognitive abilities from poker to the simple zero-sum card games played in the lab.

professional chess players would play rationally in games other than chess. An example in line with their hypothesis is one of the best female chess players Almira Skripchenko, who started playing poker no earlier than in 2003, but recently got the "French Poker Award 2009".[7] Indeed, Palacios-Huerta and Volij (2009) report extremely uniform behavior of all 26 Grandmasters who participated in their centipede game experiments. All grandmasters in their sample

a) seemed to be aware of the subgame perfect Nash equilibrium,

b) seemed to presume that other chess players are highly likely to be aware of the logic of backward induction, and

c) were ignorant of the idea that starting by playing the non-equilibrium move "right" instead of "down" would update their opponent's belief about their rationality in the game-theoretic sense, which could make the opponent play "right" in turn.

Levitt et al. (2009) challenge this result in their replication: not a single Grandmaster out of 16 chooses the Nash equilibrium strategy at the first node. In contrast to these recent experiments, we chose one which allows a continuous strategy choice: the beauty contest.[8] In many ways, the beauty contest is more similar to chess than the centipede game: chess and the beauty contest are constant-sum games, which the centipede game is not. Furthermore, chess and the beauty contest are practically impossible to solve with backward induction and they both have an almost infinite number of different outcomes[9], while the centipede game, as played by Palacios-Huerta and Volij (2008), has 7 and is solvable by backward induction. On the other hand, chess is a two-person game, while we played the beauty contest with a large number of players. And while the expected rationality of opponent(s) is only relevant for chess players' decisions in rare instances[10], it is decisive in the beauty contest. Altogether, the beauty contest, just like the centipede game, requires a considerable cognitive transfer.

[7] Interviewed by the magazine "Card Player Europe" on October 1st 2009, she argued that "chess players are already formed to be good poker players, they possess fundamental qualities to perform, especially in tournaments – capacity to concentrate for a long time, analytical skills, calculating variations or probabilities, and patience". Further she believes she benefits from the "presumption of intelligence" that other players have about her. (http://www.cardplayer.com/cardplayer-magazines/65760-phil-ivey-6-10/articles/18741-generation-next-almira-skripchenko)

[8] While the first proper beauty contest experiment was conducted by Nagel (1995), the game was invented and actually used in a newspaper contest by Ledoux (1981, 1983); see Bühren et al. (2009) for a historical account of the birth of the beauty contest.

[9] For chess, Simon's (1972, p. 166) estimate is 10^{120}, while Ewerhart (2002) argues that it is indeed ∞.

[10] For example, when deciding whether to accept a draw offer in a slightly worse position.

But what kinds of chess players' skills are likely to be transferable to the beauty contest? For the centipede game, Palacios-Huerta and Volij (2009) claim to have found the answer: backward induction. The next subsection summarizes psychologists' research on this issue.

2.2. How Chess Players Think

In order to get an impression of how likely it is, ex ante, that chess players are able to transfer their abilities to other cognitive tasks, it is important to understand what actually goes on when they are sitting at the chessboard. Figure 1 displays perhaps the best researched chess position with regard to the cognitive processes of chess players. De Groot (1965) reports on thinking-aloud protocols from 18 players, including the then-world champion[11], his predecessor and four other Grandmasters. Newell and Simon (1965) provide one further protocol, with more extensive documentation and detailed transcripts as parts of a game tree.

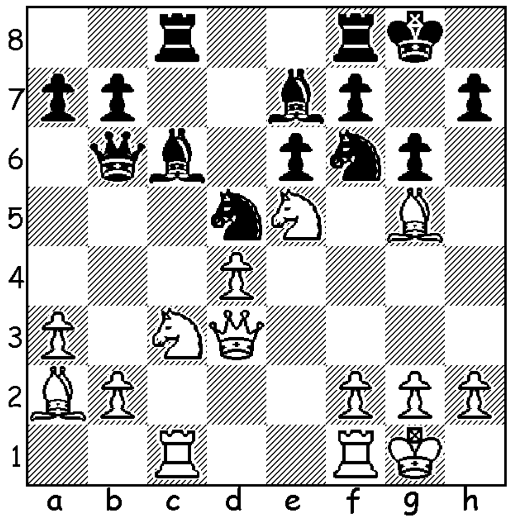


Fig. 1. Position from de Groot versus Scholtens April 10, 1936, after move 16

The position in figure 1 is typical and gives a good impression of the number of possible continuations at the beginning of the middlegame. White (to move) has 56 different moves, and Black's number of legal moves depends on what White does, though it is approximately 40. There are about 3,000,000 possible ways for the game to proceed in the next two moves (i.e., four ply in a game theoretic sense). What is more, restricting the calculations to only two moves makes no sense. Backward induction thus becomes plainly impossible when the game tree is almost infinitely large and the relevant end nodes are unknown; there are also no traces

[11] Alexander Alekhine; the investigations took place in 1938/1939 and 1943, they were first published in Dutch in de Groot (1946), but without the full protocols.

of backward induction in the protocols reported by de Groot (1965) and Newell and Simon (1965).[12]

With some knowledge of typical patterns, a reasonable aim emerges (i.e., forking the queen on b6 and the rook on f8), with a possible means to permit the knight on e5 to move there. Only strong players will see this possibility though. In this case, it is not backward induction that makes the difference but rather the ability to spot the relevant end node among millions of others that remain unobserved. It thus makes sense that Herbert Simon often used chess playing as an example of satisficing (e.g., Simon, 1955, 1972); players calculate forward and then stop when they find a reasonable path. Satisficing, not backward induction, represents the practical solution to an intractable game tree in chess.

The question of whether stronger players calculate more moves than weaker players, and how much farther they look ahead, remains subject to debate (de Groot, 1946, 1965; Campitelli and Gobet, 2004; Bilalić, et al., 2008). There is no doubt, however, that what really makes the difference is the stronger players' ability to "see" and calculate the relevant moves. By measuring players' eye movements, Charness et al. (2001) show that expert chess players fixate more on the relevant pieces than do players of intermediate skill. Klein et al. (1995) find that better chess players consider fewer potential moves from a chess position than do players with a medium skill rating, but those they do consider are more relevant, i.e. better, moves. This is possible because expert chess players have stored a lot of positional and tactical patterns or "chunks", which is known because they are much better than amateur players or beginners at reconstructing chess positions shown to them for only a few seconds - but only if it is a "realistic" position, not a random placement of the pieces (Chase and Simon, 1973; Gobet and Simon, 1996). In support of this interpretation, Amidzic et al. (2001) find in a magnetic imaging study that Grandmasters, compared with amateur chess players, exhibit markedly more brain activity in the frontal and parietal cortices, which indicates that they use their long-term memory.

This is in line with Kelly's (1985) research on the personalities of 2,209 chess players. The results of his questionnaires indicate that the latent variable "intuition" is very strongly related to chess playing strength. Kelly (1985, p. 284) concludes that "chess is much more of an intuitive than a thinking game, especially at master level."

[12] This is in line with the examples of thinking aloud by British Grandmaster Daniel King, recorded on a DVD (Fritztrainer power play 10, Hamburg, Chessbase 2009).

Psychologists' research has been nicely summed up by one participant of our experiment (on which we report in the next section), who offered the following brilliant analogy: "Have you ever looked for mushrooms with an expert mushroom searcher? Where you see only leaves and dirt, the mushroom searcher immediately spots the mushrooms: Would you say that, for this reason, the mushroom searcher is more intelligent than you?" To a beginner, a chess game is as messy as leaves and dirt, whereas strong chess players can spot the relevant aspects and use chess-specific patterns that they have stored in their minds. Nevertheless, playing tournament chess might provide chess players with a kind of training that has effects beyond the sphere of chess, effects which have been overlooked so far by psychologists. Together with the fact that recent field experiments with chess players led to mixed results, this motivates our own experimental investigation with a large number of chess players.

3. Chess Players in a Beauty Contest

3.1. All Against All

In June 2009, 6,112 chess players accepted our invitation posted on www.chessbase.de and www.chessbase.com to take part in an online experiment. They were asked to state a number (not necessarily an integer) between 0 and 100, the winning number being the one closest to two-thirds of the average. We did not tell them that this game is known as the beauty contest. The prize for the winner was a €200 Chessbase voucher, and those in second and third places received €100 and €50 vouchers, respectively. We used Chessbase vouchers instead of cash prizes to increase the credibility of our experiment. Furthermore, unlike cash prizes, Chessbase vouchers can be delivered internationally quite easily, and they are as good as money to chess players, considering the products and services offered by Chessbase.

The target number in our first round equaled 21.43, that is, two-thirds of the average guess of 32.15 and far from the Nash equilibrium of 0. Likewise, the first round in Rosemarie Nagel's (1995) first beauty contest experiment resulted in a target number of 24.49. Playing the beauty contest online produces very similar results (24.13 in Rubinstein, 2007).[13] The chess players' guesses fall within the range provided by other humans.

[13] Playing the game as a newspaper or magazine contest gives participants more time, and they generally think one step further ahead. The target number calculated from a magazine experiment by Selten and Nagel (1998) was 14.7; for Financial Times readers it equalled 12.6 (Thaler, 1997), and for readers of the Spanish newspaper Expansión it was 17.0 (Bosch-Domènech et al., 2002). However, Schou (2005) reports a target number of 21.6 when playing with 19,196 readers of the daily newspaper Politiken.

And what about the Grandmasters, of whom we have 28 in our sample? While the average guess in our complete sample was 32.15, the Grandmasters' average was slightly higher: 32.96! As the group of Grandmasters is a small subsample of our top-level players, their results only serve to give a first impression of our results. More generally, table 1 suggests no clear relationship between playing strength, measured using the Elo rating[14], and the number chosen in the beauty contest.[15] The OLS regressions[16] in table 2 confirm this impression. While the numbers chosen by better chess players are significantly lower, this relation is minuscule in its extent: On average, chess players guess one integer lower if they have a rating that is about 210 points higher (or even 310, if we control for "guessing effort", a variable which will be explained later).

Table 1
Number chosen in round 1 of the beauty contest: Summary statistics

r		n	Mean	Std. Dev.	Median	Min	Max
1	Rating < 1600[+]	2646	33.37	23.08	32	0	100
2	1600 ≤ Rating < 1800	664	32.01	21.68	30.0625	0	100
3	1800 ≤ Rating < 2000	1065	32.12	21.33	30	0	100
4	2000 ≤ Rating < 2200	1091	30.98	21.73	27.45	0	100
5	2200 ≤ Rating < 2400	551	29.15	20.15	25	0	100
6	2400 ≤ Rating *	95	30.56	22.47	22.8	0	100
	Grandmasters	28	32.96	25.03	26.175	0	100
	All	6112	32.15	22.22	29.6125	0	100

[+] Including unrated players

* Including Grandmasters

[14] Henceforward, "rating" refers to the international Elo rating (see Elo, 1978) if players have one. If they lack this rating, we use national ratings, such as DWZ in Germany, which are equally scaled. With some practice, amateurs can earn 1200 rating points quickly. The group with the lowest rating (≤ 1600) includes unrated participants, whose exact playing strength is unknown to us. However, if it corresponded to an Elo rating over 1600, the player would, in most cases, actually have a rating. Ratings above 2000 require intensive training, preferably at a young age. The world champion is rated approximately 2800. The difference between a Grandmaster and an International Master is about 200. The expected result of a player against someone with 200 Elo points more is 2.5 points out of a maximum of 10 in a 10-game match.

[15] However, the mean ranks of the rating groups listed in table 1 are not equal according a Kruskal-Wallis test ($p<0.01$).

[16] We also calculated two-limit Tobit models for the first and second rounds; the regression coefficients and marginal effects are nearly identical to the OLS regression. Thus, the OLS coefficients, which are easier to interpret, seem to be robust.

Table 2
Round 1 regression results

	Dependent variable: Chosen number			Dependent variable: Absolute difference between chosen number and target number		
Rating	-.0048*** (4.14)	-.0032* (1.99)		-.0031*** (3.69)	-.0017 (1.45)	
Guessing effort		-.5979*** (13.08)	-.6245*** (16.06)		-.4080*** (12.39)	-.4341*** (15.42)
Constant	40.84*** (18.33)	90.63*** (17.99)	87.53*** (24.45)	24.15*** (14.99)	57.24*** (15.77)	57.00*** (21.98)
n	4043	1739	2499	4043	1739	2499
Adj. R^2	.0040	.0931	.0933	.0031	.0841	.0866

t-statistics in parentheses *: significant at the 5% level; **: at the 1% level; ***: at the 0.1% level

An obvious objection to this result is that the strong chess players in our sample might have "seen" the theoretical solution, but presumed the average participant would make a less sophisticated guess. The better they are in chess, the better they might be in guessing other people's guesses, though this supposition sounds more applicable to poker than to chess. Anyway, as we report in table 2 (columns 4 - 6), better chess players are closer to the winning number, but the amount of difference is tiny; a rating that is 320 points higher brings the chess player one integer closer to the target number. If we control for guessing effort, this relation even becomes insignificant. Again, Grandmasters perform slightly below the average. For the whole sample, the mean absolute difference between the chosen number and the target number is 18.62, whereas for Grandmasters, it is 20.00.

Our sample includes only chess players, but their playing strength differs greatly: from the lowest level to world class (including a former world champion). While our beauty contest experiment does not allow a within-subject comparison of strategy choices across opponents, we asked the second round participants, ex post, to provide their guess what the chosen numbers in round 1 had been. They had knowledge of the *overall* target number from round 1 as well as the sizes of rating groups listed in table 1, and they guessed the target numbers within these groups. For every group, the person who offered the best guess received a €50 Chessbase voucher.

Asking for target numbers separately for players with different ratings (e.g., rated under 1600 or over 2400) means rubbing the research hypothesis under the participants' nose. Yet the expected impact of rating on the number chosen was low, but not as low as the actual impact, as shown in figure 2. Our Grandmasters did not presume this relationship existed, but their guesses were worse than the average guesses.

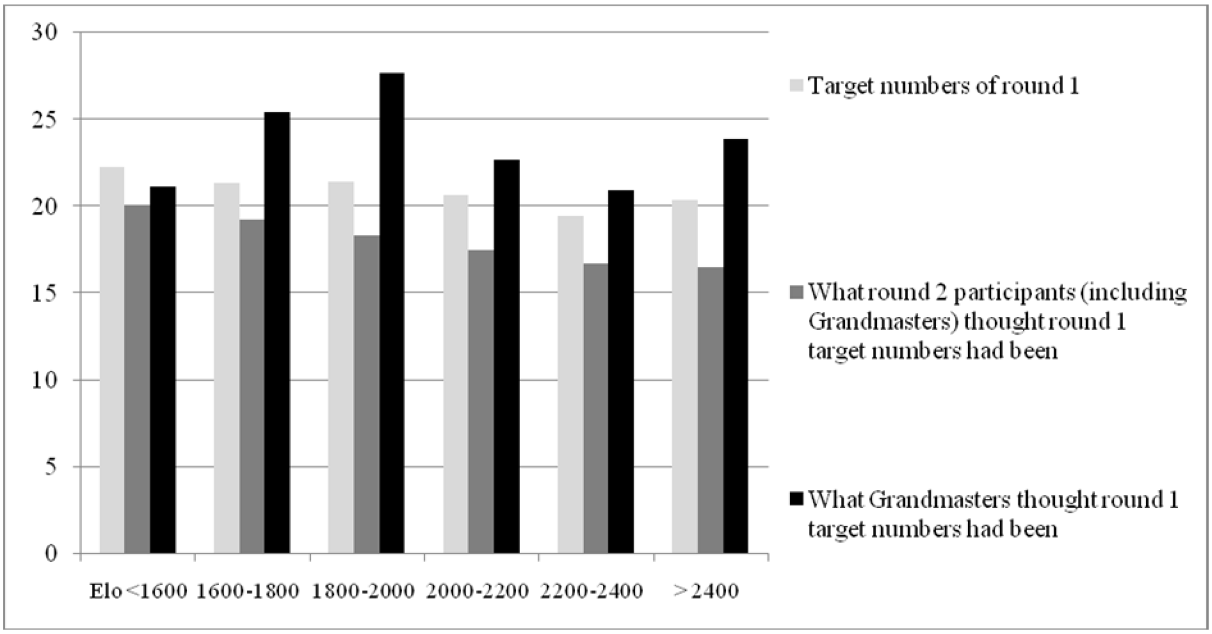


Fig. 2. Estimation of round 1 results by round 2 participants

These data pertaining to participants' guesses about the first-round results are useful for another purpose as well. Rubinstein (2007) has shown that thinking effort, proxied as the time taken to make a decision in an online beauty contest experiment, has a marked impact on the chosen number. We construct a different proxy and arrive at a similar result.

Specifically, if r denotes the rating group, n(r) is the number of people in the respective group, and G^r_i indicates subject i's guess for the respective target group, the weighted mean of the guesses should equal T, the actual overall target number for round 1:

$$T = \sum_{r=1}^{6} \frac{n(r)}{\sum_{r=1}^{6} n(r)} \cdot G_i^r$$

Because T and n(r) are public knowledge for all r, guesses that do not fulfill this condition are dominated. Guessing effort usually reduces the difference between the left- and right-hand sides of the above equation. Therefore, we define the guessing effort GE_i as:

$$GE_i = 100 - \left| T - \sum_{r=1}^{6} \frac{n(r)}{\sum_{r=1}^{6} n(r)} \cdot G_i^r \right| .$$

Less than 3% of our subjects simply stated T for every group. Because these answers could bias our proxy, we eliminate them from regressions that contain GE.

In Rubinstein's (2007) experiment, students make more reasonable guesses when they take more time to make their decision. Similarly, we find a highly significant and sizable impact of GE on the chosen number (see table 2). When we compare Rubinstein's and our distribution of guesses, which represent different depths of reasoning (table 3), we find that the chess players in our study are similar to the students in Rubinstein's study with regard to both the mean and the distribution of chosen numbers.

Table 3
Comparison with Rubinstein (2007)

Guess	Rubinstein (2007): One-period beauty contest	Chess players, round 1
51–100	20%	18%
50	16%	5%
35–49	11%	14%
33–34	11%	10%
23–32	10%	13%
22	4%	5%
16–21	6%	9%
14–15	2%	3%
2-13	9%	15%
0-1	11%	7%

Rubinstein (2007), who allowed only integers to be chosen, finds that students needed most time for answering 22, a number that is close to our winning number and is implied by a plausible expectation of others' degree of sophistication. In line with our previous results, we do not find that the likelihood of choosing 22 (in our case, a number between 21.5 and 22.5) depends strongly on the rating. Disregarding unrated players, those who chose 22 have an average of 1874, while the overall rating average is 1889.

3.2. Round 2 in One's Own League

For round 2 of our online beauty contest, we invited, via email, those first round participants who had agreed to take part in another round. In this round, they only played against players of their own rating group. For every group, €100 Chessbase vouchers provided incentives to win within that category.

As we noted previously, all respondents received information about the target number from round 1. Therefore, the average guess should decline in the subsequent round, like in all other multi-period beauty contests before. One should expect a strong negative correlation between beauty contest numbers and Elo ratings in the second round, because good chess players know that they are playing against only good players. That is, the belief that game theoretic rationality correlates with chess playing strength should result in low numbers for groups with high Elo ratings. Indeed, as for round 1, the descriptive statistics for round 2 (table 4) suggest a negative correlation[17], but again the size of the effect is small. Furthermore, it might be partially due to the smaller group size for stronger players.[18]

Table 4
Beauty contest round 2, summary statistics

r		n	Mean	Std. Dev.	Median	Min	Max
1	Rating < 1600+	897	26.86	18.89	21.4769	0	100
2	1600 ≤ Rating < 1800	324	27.26	19.13	21.745	0	99.999
3	1800 ≤ Rating < 2000	486	25.75	18.51	21.4145	0	99
4	2000 ≤ Rating < 2200	483	24.14	18.64	18.332	0	100
5	2200 ≤ Rating < 2400	241	23.73	18.94	16.874	0	100
6	2400 ≤ Rating *	50	19.11	15.89	14.381	0	69
	Grandmasters	13	20.33	17.84	15.55	0	64.01
	All	2481	25.70	18.79	20.3785	0	100

+ Including unrated players

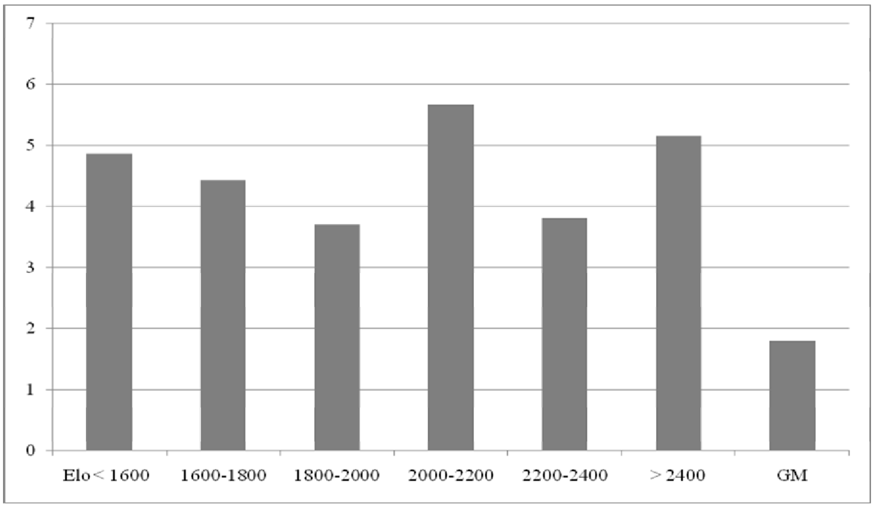
* Including Grandmasters

We consider a further dimension of cognitive ability with our beauty contest, namely, information processing. In figure 3, we depict the differences between the first- and second-round guesses. Before round 2, players received two pieces of information: the average

[17] Mean ranks of the rating groups (with respect to number chosen) are not equal according to a Kruskal-Wallis test (p<0.001).

[18] We are indebted to Oliver Kirchkamp and Karim Sadrieh for alerting us to this effect: As the guesser's number is also taken into account, the weakly dominating strategy is not (2/3)·100, but approximately 66.644 for a group size of 1000, and slightly less, namely about 66.216, for a group of 50, see Nagel (1999), p.109.

number for all players, and notification of the approximately equal strength of their competitors. If stronger chess players think more steps ahead in the beauty contest and presume that other strong players do so as well, then these pieces of information should lead to a greater difference between the second- and first-round guesses among stronger players. As we show in figure 3, however, no such pattern emerges. Starting with comparable means in round 1 (table 1), the decrease in numbers in round 2 is highest, on average, for players with an Elo rating between 2000 and 2200 and lowest for Grandmasters.



Note: This figure is based not on a direct comparison of means from tables 1 and 4 but rather on the difference of the round 2 and round 1 numbers for those who took part in both rounds

Fig. 3. Average difference between first and second round

If we regress the chosen number and measures of performance in round 2 on guessing effort (GE) and rating (table 5), we achieve results similar to those from round 1. The guesses of the better chess players in round 2 are significantly lower and nearer the target number, but the extent of difference is still not very great. In the second round, chess players guessed one integer lower if their ratings were approximately 130 points higher (or 170 points higher if we control for GE). They come one integer nearer to the target number if their rating is 240 or even 330 points higher, depending on specification. The impact of the GE variable in our second round, unsurprisingly, is higher than that in round 1, because we used data from the round 2 participants to calculate this attribute. Nevertheless, these differences in the GE coefficients are not notable, which indicates a certain consistency in the decision-making processes across different periods. Chess players with a higher GE submit lower numbers, fall farther below their round 1 guesses, and come closer to the target number of round 2.

Table 5

Round 2 regression results

	Dependent variable: Chosen number in round 2			Dependent variable: Absolute difference between chosen number and target number in round 2			Dependent variable: Difference between first and second round guess		
Rating	-.0075*** (-4.95)	-.0060*** (-4.22)		-.0042*** (-3.35)	-.0030* (-2.52)		.0022 (1.37)	.0018 (1.14)	
Guessing effort		-.6938*** (-16.87)	-.7503*** (-22.61)		-.6001*** (-17.71)	-.6558*** (-24.32)		.1107* (2.39)	.1161** (3.01)
Constant	39.69*** (13.44)	100.28*** (22.31)	94.13*** (30.79)	21.35*** (8.70)	73.82*** (20.01)	73.31*** (29.51)	.22 (0.07)	-8.66 (-1.83)	-5.83 (-1.64)
N	1638	1558	2358	1638	1558	2358	1631	1554	2296
Adj. R^2	.0141	.1682	.1779	.0062	.1741	.2003	.0005	.0036	.0040

t-statistics in parentheses; *: significant at the 5%-level; **: significant at the 1%-level; ***: significant at the 0.1%-level

4. Conclusion: Human Behavior in Chess Players

On ne joue pas aux échecs avec un bon cœur
(Nicolas-Sébastien de Chamfort)

Our experiment provided the potential for cognitive abilities transfer from chess playing skills to game-theoretic rational behavior in a beauty contest. Two rounds with slightly different rules (all against all vs. in one's own league) and two different performance measures (chosen number and distance to target number) should have been able to reveal any abilities transfers. The results of our study, however, rather illustrate the boundaries of cognitive abilities transfer across different contexts (in line with Loewenstein, 1999, p. F28).

As argued in section 2.2, good chess players are supposed to have a fine intuition. In our beauty contest, guesses of better chess players failed to outperform.[19] Further, chess playing skills are not strongly correlated to more rational choices in our experiment.[20] Better chess players do not look more steps/moves ahead in the beauty contest. This also holds true in the second round where the assumption of common rationality can even be rejected for a small group of professional chess players with an Elo-Rating above 2400.[21]

The results of our beauty contest are in line with those of Levitt et al. (2009) but in contrast to the findings of Palacios-Huerta and Volij (2009). Whereas Palacios-Huerta and Volij (2009) claim that chess players are brilliant in backward induction, Levitt et al. (2009) do not question the causal connection between chess playing and backward induction skills, but show that there is still no relation between the likelihood of conforming to the game-theoretic prediction in the centipede game and performance in a backward induction solvable constant sum game. In line with psychological research, we argued in section 2.2 that backward induction is pretty useless for practical chess. The Nash-behavior of the Grandmasters in Palacios-Huerta and Volij (2009) could be the consequence of another characteristic of chess players: they are highly competitive, especially in two person games. That is, they might have focused on beating their fellows instead of maximizing their income.[22] In this sense, the beauty contest is different from the centipede game and from chess because it is typically a multi-person game.

[19] cf. tables 2 and 5 as well as figure 2
[20] cf. tables 1 and 2
[21] cf. table 4 and 5
[22] The replication by Levitt, List, and Sadoff (2009) deviates from the design by Palacios-Huerta and Volij (2009) in that the former let their participants play more than once, reinforcing motives like reciprocity.

Considering the results of our beauty contest experiment, we propose that it is hard to conclude that chess players are beings of supernatural rationality. (Note that recent experimental evidence due to Burnham et al., 2009, and Rydval et al., 2009, did show that cognitive abilities correlate with performance in guessing games.[23]) A conclusion that tentatively hints how our results could go along with previous findings on chess players in the centipede game would be this: On the one hand, "intrapersonal spillovers" (Fennell, 2009, p. 96) from chess to game-theoretic understanding are negligible, hence typically strong chess players cannot be expected to see rational solutions where others do not. However, where both strong chess players and other subjects see the rational solution, chess players might be more likely to behave accordingly (i.e., to choose subgame perfect strategies) *under certain circumstances*. These circumstances were certainly missing in our beauty contest, where social preferences are irrelevant. Future research could shed more light on this hypothesis through within subject comparison of different experimental tasks, maybe not only with chess playing subjects, but also with, say, poker players or with professionals who succeed in both mind sports.

[23] Burnham et al. (2009) measure cognitive ability directly through IQ test-like questions and show that subjects with higher test scores come closer to the Nash equilibrium in a beauty contest experiment. Rydval et al. (2009) measure various process-oriented cognitive characteristics and find that the likelihood of dominance violation in dominance solvable guessing games is negatively related with short-term memory, intrinsic motivation to engage in cognitively demanding tasks and premeditation, i.e., the propensity to think through the task carefully.

References

Alevy, J.E., Haigh, M.S., List, J.A., 2007. Information cascades: evidence from a field experiment with financial market professionals. Journal of Finance 62, 151-180.

Amidzic, O., Riehle, H.J., Fehr, T., Wienbruch, C., and Elbert, T., 2001. Pattern of focal γ-bursts in chess players. Nature 412, 603.

Azar, O.H., Bar-Eli, M., 2010. Do soccer players play the mixed-strategy Nash equilibrium? forthcoming in Applied Economics.

Bilalić, M., McLeod, P., Gobet, F., 2008. Expert and 'novice' problem solving strategies in chess: sixty years of citing de Groot (1946). Thinking & Reasoning 14, 395–408.

Bosch-Domenech, A., Montalvo, J.G., Nagel, R., Satorra, A., 2002. One, two, (three), infinity...: newspaper and lab beauty-contest games. American Economic Review 92, 1687–1701.

Bühren, C., Frank, B., Nagel, R., 2009. A historical note on the beauty contest. Mimeo. https://cms.uni-kassel.de/unicms/fileadmin/groups/w_030516/BC/A_historical_note_on_the_Beauty_Contest.pdf

Burnham, T., Cesarini, D., Johannesson, M., Lichtenstein, P. Wallace, B., 2009. Higher cognitive ability is associated with lower entries in a *p*-beauty contest, Journal of Economic Behavior and Organization 72, 171–175.

Camerer, C., 2003. Behavioral game theory, experiments in strategic Interaction. Princeton University Press, Princeton.

Campitelli, G., Gobet, F., 2004. Adaptive expert decision making: skilled chess players search more and deeper. ICGA Journal 27, 209-216.

Charness, N., Reingold, E.M., Pomplun, M., Stampe, D.M., 2001. The perceptual aspect of skilled performance in chess: evidence from eye movements. Memory and Cognition 29, 1146-1152.

Chase, W.G., Simon, H.A., 1973. Perception in chess. Cognitive Psychology 4, 55-81.

Chiappori, P.-A., Levitt, S., Groseclose, T., 2002. Testing mixed strategy equilibria when players are heterogeneous: the case of penalty kicks in soccer. American Economic Review 92, 1138-1151.

Cooper, D.J., 2006. Are experienced managers experts at overcoming coordination failure? Advances in Economic Analysis and Policy 6, 1-50.

de Groot, A.D., 1946. Het denken van den schaaker. Noord-Holland, Amsterdam.

de Groot, A.D., 1965. Thought and choice in chess. Mouton, The Hague.

Elo, A.E., 1978. The rating of chessplayers, past and present. Arco, New York.

Ewerhart, C., 2000. Chess-like games are dominance solvable in at most two steps. Games and Economic Behavior 33, 41-47.

Ewerhart, C., 2002. Backward induction and the game-theoretic analysis of chess. Games and Economic Behavior 39, 206-214.

Fennell, L.A., 2009. Willpower and legal policy. Annual Review of Law and Social Science 5, 91–113.

Gerdes, C., Gränsmark P., 2010. Strategic behavior across gender: a comparison of female and male expert chess players. Mimeo, IZA DP No. 4793.

Gobet, F., Simon, H.A., 1996. Recall of random and distorted positions: implications for the theory of expertise. Memory and Cognition 24, 493-503.

Haigh, M., List, J.A., 2005. Do professional traders exhibit myopic loss aversion? An experimental analysis. Journal of Finance 60, 523-534.

Kelly, E.J., 1985. The personality of chessplayers. Journal of Personality Assessment 49, 282-284.

Klein, G., Wolf, S., Miltello, L., Zsambock, C., 1995. Characteristics of Skilled Generation in Chess. Organizational Behavior and Human Decision Processes 62, 63-69.

Ledoux, A., 1981. Concours résultats complets. Les victimes se sont plu à jouer le 14 d'atout. Jeux et Stratégie 2, 10-11.

Ledoux, A., 1983. Concours J&S 83. On a retrouvé la stanta barbara! Jeux et Stratégie 4, 12-15.

Levitt, S., List, J., Reiley, D.H., 2010. What happens in the field stays in the field: professionals do not play minimax in laboratory experiments. forthcoming in Econometrica.

Levitt, S., List, J., Sadoff, S., 2009. Checkmate: exploring backward induction among chess players. NBER Working Paper No. 15610.

List, J.A., Haigh, M.S., 2005. A simple test of expected utility theory using professional traders. Proceedings of the National Academy of Science 102, 945-948.

Loewenstein, G., 1999. Experimental economics from the vantage-point of behavioural economics. Economic Journal 109, F25-F34.

Milgrom, P., 2004. Putting auction theory to work. Cambridge University Press, Cambridge.

Nagel, R., 1995. Unraveling in guessing games: an experimental study. American Economic Review 85, 1313-1326.

Nagel, R., 1999. A survey on experimental beauty contest games: bounded rationality and learning. In: Budescu, D.V., Erev, I., Zwick, R. (Eds.), Games and human behavior: essays in honor of Amnon Rapoport. Erlbaum, New Jersey, pp. 105-142.

Newell, A. Simon, H.A., 1965. An example of human chess play in the light of chess playing programs. In: Wiener, N., Schadé, J.P. (Eds.), Progress in biocybernetics. Vol. 2, Elsevier, Amsterdam, pp. 19-75.

Palacios-Huerta, I., 2003. Professionals play minimax. Review of Economic Studies 70, 395-415.

Palacios-Huerta, I., Volij, O., 2008. Experientia docet: professionals play minimax in laboratory experiments. Econometrica 76, 71-115.

Palacios-Huerta, I., Volij O., 2009. Field centipedes. American Economic Review 99:,1619-1635.

Potters, J., van Winden F., 2000. Professionals and students in a lobbying experiment. Professional rules of conduct and subject surrogacy. Journal of Economic Behavior and Organization 43, 499-522.

Rubinstein, A., 2007. Instinctive and cognitive reasoning: a study of response times. Economic Journal 117, 1243–1259.

Rydval, O., Ortmann, A., Ostatnicky, M., 2009. Three very simple games and what it takes to solve them. Journal of Economic Behavior and Organization 72, 589-601.

Schou, A., 2005. Gæt-et-tal konkurrence afslører at vi er irrationelle. Politiken, September 22. http://politiken.dk/erhverv/article123939.ece, retrieved 9.9.2009.

Selten, R., Nagel, R., 1998. Das Zahlenwahlspiel-Ergebnisse und Hintergrund. Spektrum der Wissenschaft, February, 16-22.

Simon, H.A., 1955. A behavioral model of rational choice. Quarterly Journal of Economics 69, 99-118.

Simon, H.A., 1972. Theories of bounded rationality. In: McGuire C.B., Radner, R. (Eds.), Decision and Organization. North-Holland, Amsterdam, pp. 161-176.

Thaler, R.H., 1997. Giving markets a human dimension. Financial Times 6, June 16.

Wooders, J., 2010. Does experience teach? professionals and minimax play in the lab, forthcoming: Econometrica.