Decreasing Differences in Expert Advice: Evidence from Chess Players *

Elias Bouacida[†] Renaud Foucart[‡] Maya Jalloul[§]

May 28, 2024

Abstract

We study the impact of external advice on the relative performance of chess players. We asked players in chess tournaments to evaluate positions in past games and allowed them to revise their evaluation following advice from a high or a low ability player. While high-quality advice has the potential to act as a "great equalizer," reducing the difference between high and low ability players, this is not what happens in our experiment. This is in part because our subjects ignore too much of the advice they receive. It is also because low ability players pay a higher premium than high ability ones by following their initial idea instead of high-quality advice.

Keywords: decreasing differences, expert, advice, chess, control
JEL-Codes: C78, C91, C93, D91, J24, O33

*We thank Jamal El Chamieh and Omar El Deeb for their support in organizing the experiment, and Alexander Matros and David Rietzke for taking the time to solve our positions. The paper also greatly benefited from the comments of Ro'ee Levy and Paola Manzini, as well as from those of participants to seminars and workshop in Lancaster and at the 1st Doctoriales ASFEE.

[†]Université Paris 8, elias.bouacida@univ-paris8.fr
[‡]Lancaster University Management School, r.foucart@lancaster.ac.uk
[§]Lancaster University Management School, m.jalloul@lancaster.ac.uk

1 Introduction

Does offering high quality advice help reduce the productivity gap between high and low ability workers? Mechanically, the answer is yes: the potential benefit from being able to rely on outside advice is higher if your own ability is lower, a property known as *decreasing differences*.

In a lab-in-the-field experiment with chess players participating in tournaments in Lebanon, we however find little evidence of decreasing differences. Our subjects reveal such a high preference for following their first idea and ignoring additional information that they forego a large share of the potential gains from the advice. Lower ability players end up paying the highest premium from ignoring good advice.

We partnered with a local academy to run our incentivized experiment alongside chess tournaments in several locations in Lebanon in the Summer of 2023. The main task our subjects had to perform was to evaluate the *pawn advantage* of 20 chess positions – a measure of which player is better positioned to win the game, and by how much. For each position, we first asked our subjects to make their own evaluation, by choosing one of four possible answers. We then provided them with the evaluation of an external adviser for the same position, and asked our subjects to evaluate it again.

One of our advisers is an International Master, among the top 6,000 players in the world, and better rated than all of our subjects. He provided correct advice for 75% of the positions. The second adviser is an everyday chess player with no formal rating, placing him at the bottom of our subjects, and offering correct advice only for 15% of the positions. In one treatment, we disclosed the rating of both advisers, but only told our subjects the advice came from one of them with equal probability. In the other treatment, we also informed the subjects of which expert the advice came from.

We defined as "high ability" subjects with an official chess rating in the top half of our sample, and as "low ability" those in the bottom half. Before receiving the high quality advice, high ability subjects had a rate of correct answers of 41.2%, and low ability ones of 32.9%. After receiving high quality advice, the rate increased to 50.8% (+9.6pp) for high ability subjects, and to 42.5% (+9.6pp) for low ability ones. In contrast, high ability subjects offered high quality advice could have increased their rate of correct answers by 22.0pp on average and low ability subjects could have increased it by 30pp by following advice when it was (in expectation) beneficial to do so. The absence of decreasing differences in expert advice in our experiment is therefore largely explained by the significantly higher premium paid by low-ability subjects ignoring good advice.

The fact that most ways of matching expertise involves decreasing differences has been shown it theory by Chade and Eeckhout (2018): the marginal impact of the quality of advice is decreasing in the ability of the person who receives it. Empirical evidence of decreasing differences in matching abilities in the labour market include production in garment factories (Hamilton *et al.*, 2003; Adhvaryu *et al.*, 2020) and student coursework in universities (Fischer *et al.*, 2023). A study of the US labour market shows that, in general, lower ability workers benefit more from being part of a team with a high ability partner (Herkenhoff *et al.*, 2024).

More recently, research on Artificial Intelligence (AI) has shown the potential for decreasing differences on the performance of lawyers (Choi and Schwarcz, 2023), programmers (Peng *et al.*, 2023), writers (Noy and Zhang, 2023), customer support (Brynjolfsson *et al.*, 2023), and consultants (Dell'Acqua *et al.*, 2023) in routine tasks. This evidence contrasts with the few studies looking at advice for the tasks in which subjects may perceive themselves as experts: Agarwal *et al.* (2023) finds that radiologists often fail to incorporate uncertain advice optimally, and Otis *et al.* (2023) that, among Kenyan entrepreneurs, advice increases the performance of high performers but actually hurts low performers.

The decision to ignore one's own signal and follow the advice of others is typically studied in economics in the context of information cascades (Anderson and Holt, 1997; Kübler and Weizsäcker, 2004), and there is evidence that subjects often like to bet on themselves even when it is optimal not to do so (Weizsäcker, 2010), and put a lower weight on information discovered by others (Conlon *et al.*, 2022). In psychology, a large literature studies how subjects tend to give a sub optimal weight on advice in their decision-making (Bailey *et al.*, 2022; Bonaccio and Dalal, 2006). This result is also linked to the idea of preference for decision rights or control premium (Bartling *et al.*, 2014; Owens *et al.*, 2014) ; and the "illusion of control" (Langer, 1975; Sloof and von Siemens, 2017) where subjects are overconfident when they make the decision themselves.

Our paper contributes to the scientific literature on decreasing differences, advice and control discussed above. The main novelty of our research is to explicitly measure the potential for decreasing differences in a context where our subjects have a certain level of expertise on their topic and have to confront that expertise with a possibly better external advice. Indeed, chess players who signed up to participate to an official tournament are arguably closer to professionals with knowledge of their job facing advice than participants in a lab experiment completing tasks for which they have no particular reason to feel qualified.

We also contribute to the literature on control and advice by providing results from a non-WEIRD (White, Educated, Industrialized, Rich, and Democratic) sample (Henrich *et al.*, 2010), as our subjects live in a Middle-Eastern country in the midst of a banking and political crisis. Finally, this paper is part of a literature using chess players to study human decisions, such as strategic behaviour in sequential games (Levitt *et al.*, 2011), gender differences in risk-taking (Gerdes and Gränsmark, 2010), social norms and the gender gap (Dilmaghani, 2021), or the role of superstars (Bilen and Matros, 2023).

The rest of the paper is organised as follows. In Section 2, we describe our experimental protocol and procedures. We present the results in Section 3 and conclude in Section 4.

2 The experiment

We ran the experiment during the Summer of 2023 in several cities in Lebanon, alongside tournaments organised by a local chess academy.[1] Our subjects were regular participants in tournaments. We describe their self-reported demographic characteristics in Table 3 in Appendix B. All subjects received the experimental material written both in English and Arabic.

We recruited subjects before the tournament through the academy and paid for their registration (around $5) as a participation fee. The experiment took part in a separate room at times where our subjects were not competing. Each subject was randomly allocated either to a treatment with or without information on the adviser. Subjects received tasks booklets and answer sheets upon being seated. There were two rounds of tasks, each corresponding to evaluating ten positions.

A position is a description at a given point of a game of the positions of the pieces on the board (Figure 1). Positions are evaluated using the notion of *pawn advantage*, a measure of which player (White or Black) is better placed to win the game. We chose 20 positions from past games of chess using the Chessbase Mega database 2023, and picked half of them with a pawn advantage of 0.7 (a slight advantage) and the other half with 2.4 (a large advantage), either for Black (-2.4 and -0.7) or for White (0.7 and 2.4).[2] The

[1]The exact dates are August 15, August 20, September 2, and September 17, 2023. The pre-registration is available at: https://aspredicted.org/124_MSY. In line with the pre-registration, we stopped recruiting participants when we reached 100 subjects, so that we recruited a total of 103 subjects. Our total sample is however $n = 102$ as, in line with our pre-registration, we removed observations for which no choice were made and one of our subjects did not write anything in the second part of the answer sheet. The project has received IRB approval from Lancaster University.

[2]Chess players are in general reluctant to translate pawn advantages into winning probabilities, one reason being that there are not two but three possible outcomes in chess: a win, a loss, or a draw. According to one measure however (suggested by Sune Fischer and Radu Pannan based on 405,460 past games), a pawn advantage of 0.7 corresponds to a 60% probability of win and of 2.4 to a 80% probability of win - counting a draw as half a win. In our selection of positions, we followed this statistical regularity: of the games with a pawn advantage of ± 2.4, 7 ended with a win for the advantaged player, 2 with a draw, and 1 with a loss ; of the games with a pawn advantage of ± 0.7, 3 ended with a win for the

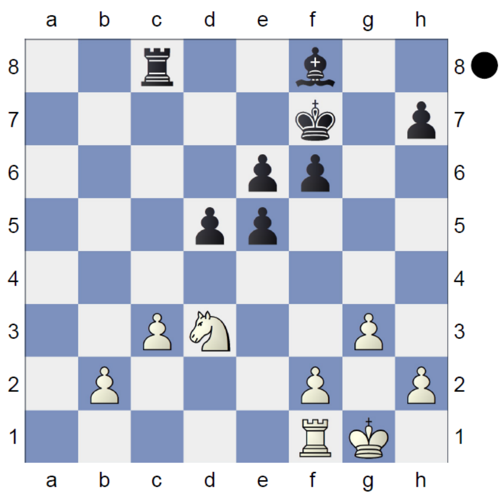


Figure 1: A position in a game of chess, as shown to our subjects.

task was to identify the correct evaluation out of the four possible ones (in Figure 1, the correct answer is -0.7). Evaluating positions is a standard exercise familiar to any chess player advanced enough to participate in an official tournament. As contemporary chess engines converge towards almost identical pawn advantages, there is no ambiguity as to which of the four evaluations is the correct answer.

In each round, subjects were given 8 minutes to complete the first part of the answer sheet with ten evaluations. Then, they were provided with the evaluations of one of our two advisers for the same positions. They were given 4 minutes to look back at their answers, compare with the advice, and complete the second part of the answer sheet with their possibly updated evaluations.

In the known adviser condition, we told subjects that the answers we gave them were coming from "a player with a rating of 2335" (H-adviser) for one of the rounds of ten evaluations, and from "an unrated player, who plays regularly for fun" (L-adviser) for the other one. In the unknown adviser condition, we told them in both rounds that "With equal probability, the player has a rating of 2335, or it is an unrated player, who plays regularly for fun". The rating refers to the *Elo rating,* the standard measure of chess performance.[3]

advantaged player, 6 with a draw, and 1 with a loss.

[3]The Elo Rating is a system created by Arpad Elo to compute the relative skill level of a player. When two players play against each other in a tournament registered with the international chess federation FIDE, the winner gains Elo points, and the loser loses points. The number of points gained and lost depends on the difference in ratings and on the expected outcome. Any player with a rating strictly lower than 1000 is considered as unrated by the FIDE (and in our sample). As a rule of thumb, a difference of 100 points in the Elo rating means that the best rated player is expected to win 5 out of 8 games. While Elo is an imperfect measure of ability, it is taken seriously by players.

After solving the two rounds of evaluations, subjects completed a short demographic questionnaire as well as questions about their stated preference for control (5 questions borrowed from Burger and Cooper, 1979). We provide the experimental material in the Online Appendix. All the sessions were administered by one of the co-authors of this study (Maya Jalloul), who read the experimental material aloud and ensured no one could cheat.

On top of the participation fee, we picked one of the 40 evaluations of each subject at random (20 evaluations before advice, and 20 after) and paid a variable amount of \$10 if the answer was correct.[4]

In line with our pre-registration, we divided our sample of $n = 102$ into two groups of equal size, based on their rating, and removed questions for which subjects did not answer.[5] The average Elo rating of our rated players is 1490, while our best subject is in the range 2100-2200. It should thus be clear to all our subjects that our high ability adviser with a rating higher than 2300 is more likely than them to correctly evaluate a position. It should also be clear that our low ability adviser is not of strictly higher ability than any of our subjects who all take part in a registered tournament. We plot the Elo distribution of our subjects in Figure 3 in Appendix B.

3 Results

We plot on Figure 2 the average share of correct answers in the low (l) and high (h) ability groups, before observing advice, and after observing low quality (L) and high quality (H) advice. This figure pools both the treatment where the type of the adviser is known and the treatment where it is not.

Before observing advice, h subjects evaluated 41.8% of the positions correctly, while the result for l subjects is 31.2%. Our H adviser provided 75% of correct answers, and our L adviser only 15%, less than the expected rate of someone answering at random. For this reason, the share of correct answers drops slightly, to 38.4% and 26.1% respectively after observing L-advice.

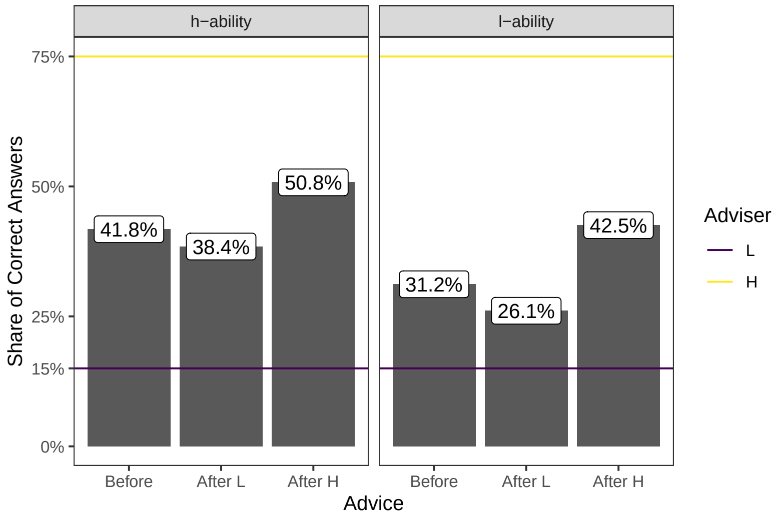
After observing high quality advice, the share of correct answers increases, but only slightly more for low-ability subjects (+11.3pp) than for high-ability subjects (+9pp).

[4]Given the difficult banking situation in Lebanon and the fact that some of our subjects were minor, we did not pay subjects directly in cash but with monetary vouchers for subsequent tournaments or other spending on the day. We only knew the subject number, and not their identity. We communicated a list of payments and subject numbers to the organizing chess academy, who then processed the payments based on a list they made allocating participant numbers to individuals.

[5]As 54 subjects had a formal Elo rating, the results are almost identical when considering a dichotomy rated/unrated instead.

As a consequence, we do not find any statistical evidence of decreasing differences in expert advice in our pre-registered tests. In the treatment with unknown adviser, the average share of correct answers when matching l subjects and L advice and h subjects and H advice (Positive Assortative Matching, PAM) is 35.2%. With Negative Assortative Matching of l subjects to H advice and h subjects to L advice (NAM), it is 38.1%. The p-value of the two-sided sample test of equal proportion between NAM and PAM is 0.384. In the treatment with known advice, the difference is even smaller (42.6% versus 41.3%, p=0.711). Pooling both treatments, the difference remains non-significant (p=0.365).[6] Our alternative pre-registered test in which we use the share of correct answers pre-advice instead of after L-advice also fails to capture statistically significant evidence of decreasing differences (see Table 4 in Appendix C).

Figure 2: Share of correct answers by subject type, before observing advice, and after observing Low and High quality advice.



Notes: pooling the treatment where the adviser type is known and the treatment where he is unknown to our subject.

Overall, subjects tend to ignore advice: around three-quarters of the choices are unchanged after observing advice (we report in Table 7 in Appendix D the figure for each treatment and type). Unchanged answers can however be for several reasons, one of them being that if a subject's answer is identical to the adviser's there is no reason to modify

[6]While we cannot rule out that some decreasing differences exist, our sample size should have been sufficient to identify any large effect. Indeed, pooling both treatments, with our realized sample size of 1,020 observations, the proportion with PAM is 38.4%. We would have been able to detect a significant difference ($p < 0.05$) with a power of 0.8 with a NAM proportion of 44.5%, while the observed one is 40.5% (see Vasilaky and Brock (2020) for explanations on why we computed the minimal detectable effect).

it. In Table 6 in Appendix D, we show that subjects keep their answers at least 93% of the time when they agree with the adviser.

Table 1 show what happens when the answers of participants differ from the advisers'. Participants mostly, and correctly, ignore advice from our L-adviser. In line with Schultze *et al.* (2017) who shows that some subjects feel the need to incorporate even useless advice; 9.5% of our high ability subjects update their evaluation following advice from an adviser they should expect to be worse than them. Low ability subjects ignore advice from unknown advisers 64.3% of the time – as compared to 79.7% for the high ability one – and ignore (46.8%) or move further away (1.8%) from our H-adviser roughly half of the time, only slightly less than our high ability subjects. In line with the findings of Alysandratos *et al.* (2020) on economic experts, we find no evidence that our subjects are able to distinguish good from bad advice when they do not know the identity of the adviser.

Table 1: How do subjects react to the advice received, when they disagree with it? (in percentages)

			Among those who disagree before[2]			
Type	Treatment	Disagree[1]	Keep	Follow	Closer	Further
l	Know H	66.5	46.8	39.2	12.3	1.8
l	Know L	72.2	78.8	14.0	7.3	0.0
l	Unknown	70.6	64.3	27.3	6.3	2.2
h	Know H	58.2	52.5	42.5	2.5	2.5
h	Know L	73.2	83.9	9.5	4.0	2.5
h	Unknown	67.0	79.7	15.3	3.3	1.7

We remove from this table the missing answers because we have no distance from the answer for them. We therefore slightly underestimate the disagreement percentage before receiving the advice.

[1] Percentage of different pre-advice answers with the adviser.

[2] Percentage of kept or changed answer (following, getting closer, or further away from advice) conditional on pre-advice answer being different from adviser's.

To see how much payment subjects left on the table by ignoring advice, we compare their choices in our experiment with two "heuristics" of always following or ignoring some type of advice.

Our first "Probability" heuristic corresponds to the first best choice of our subjects if

Table 2: Difference (in percentage points) between the average share of correct answers of h and l type subjects having received H-advice, following the probability heuristic and in the experiment.

	Types		
Treatment	l	h	P-value[1]
Unknown	32.4	23.0	0.056
Known	27.7	21.1	0.139
All	30.0	22.0	0.015

Premia for l and h are given in percentage points.

[1] P-value of the two-sided two sample t-test of equal premium between the h and l Elo types.

they were aware of their probability of answering correctly (approximated by their share of correct answers) as well as the probability of the advisers.[7]

For each subject, we pick these heuristics and see how many correct answers they could have achieved by following them. While imperfect (and not part of our pre-registration), this method gives us an illustrative idea of the potential of advice and its role as a great equalizer. Subjects could have done even better if they were able to identify the questions for which they are particularly confident to have a correct answer for instance.

We measure in Table 2 the premium subjects are paying in order to ignore high-quality advice, defined as the difference (in percentage points) in the share of correct answers post H-advice if they followed our heuristics and in the experiment. Across treatments, our low ability subjects would have a 30 percentage points higher share of correct answers after H-advice following the Probabilistic heuristics than they did in the experiment, as compared to 22.0pp for our high ability subjects. The difference is statistically significant. We can thus conclude that, be it because of overconfidence or intrinsic preference for keeping their original answer, our low-ability subjects ended up paying a higher premium than high-ability ones for ignoring good advice. We show in Appendix D.2 two other heuristics that show similar results.

[7]We approximate a subject i's probability of evaluating a position correctly p_i by their share of correct answers pre-advice, and, similarly, the probability for experts to do so q_L and q_H, with $\bar{q} = \frac{q_L+q_H}{2}$ the probability for unknown advice. This "first-best" way of incorporating advice follows a simple decision rule: if $p_i > q_H$, ignore all advice ; if $p_i \in (\bar{q}, q_H)$, only follow the known advice of H ; if $p_i \in (q_L, \bar{q})$, follow all advice, known or unknown, except for the advice of L ; and if $p_i < q_L$, follow all advice.

Finally, we construct an index of the stated preference for control by aggregating the answers to our questions borrowed from Burger and Cooper (1979). We find that stated preference for control is correlated with the probability of a subject keeping their answers, after controlling for subject and answer characteristics (see Table 8 in Appendix D). We also confirm our main results in a regression (Table 5 in Appendix C). Following our pre-registration, we use both our binary definition of low and high-ability subjects, and a continuous measure based on the Elo rating, and run regressions for the known and unknown adviser treatment. As expected, a high ability adviser typically benefits subjects, and the better rated subjects are more likely to evaluate positions correctly. The interaction term between the adviser type and the rating of our subjects gives an alternative measure of the existence of decreasing differences. As in the main tests with two categories of subjects, it is not significant. We control for individual characteristics in Table 5 in Appendix D.

4 Conclusions

Digitalization and the development of Artifical Intelligence promise to give broad access to high quality specialist advice. In theory, one of the main consequences of this evolution is a compression in the distribution of productivity, reducing the difference between the best and worst performers. However, the literature on advice and preference for control tells us that subjects may simply not take up this advice.

In this paper, we used a sample of subjects with specialist knowledge in their topic in a natural setting – chess players evaluating chess positions during a chess tournament – to learn more about the "great equalizer" potential of advice. While we find evidence that improving the quality of advice could benefit low ability players more, most of the potential benefit of advice is wasted by subjects choosing to keep their initial evaluation. This preference for following their own expertise hurts low ability subjects the most, as they had the most to gain. The fact that low-ability subjects are also those paying the highest premium to follow their initial evaluation is consistent with the idea that the most able subjects are also the most able to follow advice. It could also be the case that ability in chess is not exogenous, and that the best rated players are precisely those who are able to listen to advice during their training.

Among the limitations of our paper is the fact that we do not distinguish between advice from humans and from computers. We did so because, since the landmark victory of chess engine Deep Blue versus the then world champion Garry Kasparov in 1997, chess players see algorithmic analysis of the games as the gold standard. This is precisely the reason why we could use the chess engines evaluation of the pawn advantage in our chosen positions as the unambiguously correct answer. Further studies of subject specialists such as our chess players would benefit from comparing computer-based advice and human one and see whether decreasing differences are more pronounced with the latter.

5 Bibliography

Adhvaryu, A., Bassi, V., Nyshadham, A. and Tamayo, J. A. (2020). *No line left behind: Assortative matching inside the firm*. Tech. rep., National Bureau of Economic Research.

Agarwal, N., Moehring, A., Rajpurkar, P. and Salz, T. (2023). *Combining human expertise with artificial intelligence: Experimental evidence from radiology*. Tech. rep., National Bureau of Economic Research.

Alysandratos, T., Boukouras, A., Georganas, S. and Maniadis, Z. (2020). The expert and the charlatan: an experimental study in economic advice. *Available at SSRN 3644653*.

Anderson, L. R. and Holt, C. A. (1997). Information cascades in the laboratory. *The American economic review*, pp. 847–862.

Bailey, P. E., Leon, T., Ebner, N. C., Moustafa, A. A. and Weidemann, G. (2022). A meta-analysis of the weight of advice in decision-making. *Current Psychology*, pp. 1–26.

Bartling, B., Fehr, E. and Herz, H. (2014). The intrinsic value of decision rights. *Econometrica*, **82** (6), 2005–2039.

Bilen, E. and Matros, A. (2023). The queen's gambit: Explaining the superstar effect using evidence from chess. *Journal of Economic Behavior & Organization*, **215**, 307–324.

Bonaccio, S. and Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational behavior and human decision processes*, **101** (2), 127–151.

Brynjolfsson, E., Li, D. and Raymond, L. R. (2023). *Generative AI at work*. Tech. rep., National Bureau of Economic Research.

Burger, J. M. and Cooper, H. M. (1979). The desirability of control. *Motivation and emotion*, **3**, 381–393.

Chade, H. and Eeckhout, J. (2018). Matching information. *Theoretical Economics*, **13** (1), 377–414.

Choi, J. H. and Schwarcz, D. (2023). Ai assistance in legal analysis: An empirical study. *Available at SSRN 4539836*.

Conlon, J. J., Mani, M., Rao, G., Ridley, M. W. and Schilbach, F. (2022). *Not learning from others*. Tech. rep., National Bureau of Economic Research.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F. and Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (24-013).

Dilmaghani, M. (2021). The gender gap in competitive chess across countries: Commanding queens in command economies. *Journal of Comparative Economics*, **49** (2), 425–441.

Fischer, M., Rilke, R. M. and Yurtoglu, B. B. (2023). When, and why, do teams benefit from self-selection? *Experimental Economics*, pp. 1–26.

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

Hamilton, B. H., Nickerson, J. A. and Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of political Economy*, **111** (3), 465–497.

Henrich, J., Heine, S. J. and Norenzayan, A. (2010). Most people are not weird. *Nature*, **466** (7302), 29–29.

Herkenhoff, K., Lise, J., Menzio, G. and Phillips, G. M. (2024). Production and learning in teams. *Econometrica*.

Kübler, D. and Weizsäcker, G. (2004). Limited depth of reasoning and failure of cascade formation in the laboratory. *The Review of Economic Studies*, **71** (2), 425–441.

Langer, E. J. (1975). The illusion of control. *Journal of personality and social psychology*, **32** (2), 311.

Levitt, S. D., List, J. A. and Sadoff, S. E. (2011). Checkmate: Exploring backward induction among chess players. *American Economic Review*, **101** (2), 975–990.

Noy, S. and Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Available at SSRN 4375283*.

Otis, N., Clarke, R. P., Delecourt, S., Holtz, D. and Koning, R. (2023). The uneven impact of generative ai on entrepreneurial performance. *Available at SSRN 4671369*.

Owens, D., Grossman, Z. and Fackler, R. (2014). The control premium: A preference for payoff autonomy. *American Economic Journal: Microeconomics*, **6** (4), 138–161.

Peng, S., Kalliamvakou, E., Cihon, P. and Demirer, M. (2023). The impact of ai on developer productivity: Evidence from github copilot. *arXiv preprint arXiv:2302.06590*.

Schultze, T., Mojzisch, A. and Schulz-Hardt, S. (2017). On the inability to ignore useless advice. *Experimental Psychology*.

Sloof, R. and von Siemens, F. A. (2017). Illusion of control and the pursuit of authority. *Experimental Economics*, **20** (3), 556–573.

Vasilaky, K. N. and Brock, J. M. (2020). Power(ful) guidelines for experimental economists. *Journal of the Economic Science Association*.

Weizsäcker, G. (2010). Do we follow others when we should? a simple test of rational expectations. *American Economic Review*, **100** (5), 2340–2360.

Appendix

A A simple theoretical framework

Consider two subjects l and h, with perfect information about their own probability of successfully solving a task p_i, $i \in \{l, h\}$, as well as the probability of the low L and high H ability advisers to do so $q_L < q_H$. In the case in which subjects do not know the identity of the adviser – but know both are equiprobable – we denote by $\bar{q} = \frac{q_L + q_H}{2}$ this probability.

Unless all subjects follow (or ignore) all types of advice, we should observe strictly *decreasing differences* if subjects correctly infer the probabilities and maximize their expected probability of finding the correct answer.

Define by $f(i, j)$ the probability that subject i solves a task correctly after observing advice j and assume that $q_L < p_l < p_h < q_H$. If subjects want to maximize their probability of success and know the identity of the adviser, $f(l, L) = p_l$, $f(h, L) = p_h$, and $f(l, H) = f(h, H) = q_h$.

It is easy to see that in that case, the function displays decreasing differences:

$$f(l, H) - f(l, L) > f(h, H) - f(h, L),$$

as the expression simplifies to $p_l < p_h$. This statement is equivalent to saying that Negative Assortative Matching (NAM) of subjects to advisers yields a higher expected share of correct answers than Positive Assortative Matching (PAM),

$$\frac{f(l, H) + f(h, L)}{2} > \frac{f(l, L) + f(h, H)}{2}.$$

The same result holds when considering the case of unknown advisers if $p_l < \bar{q} < p_h$, so that type l subjects follow all advice and type h do not follow any. In that case, $f(l, L) = q_L$, $f(h, L) = f(h, H) = p_h$, and $f(l, H) = q_H$. The condition for decreasing differences is then $q_H > q_L$, and the difference between NAM and PAM is higher than with known advisers. The reason is that a good adviser then not only helps more the low ability subjects, but it also protects them from following bad advice. Finally, if $\bar{q} \geq p_h$ or $\bar{q} \leq p_l$, the differences are constant and the probability of a correct answer in NAM is the same as in PAM. This result is trivial, as it simply states that if all subjects follow all advice, they also solve all problems with the same probability, and if they ignore all advice, the quality of advice has no influence on their success.

By the same logic, we can compare advice from H and no advice at all, where $f(i, 0) = p_i$ is the probability of the answer of subject i being correct before advice. With known adviser, the result is identical to the one above, as $f(i, L) = p_i$ for both types of subjects.

With unknown adviser, there are always decreasing differences unless all advice is ignored. If $\bar{q} < p_h$, the condition becomes $q_H > p_l$. If $\bar{q} \geq p_h$, it is $p_h > p_l$.

There are however two main biases and preferences that could influence our theoretical result of decreasing differences in the experiment. The first is that our subjects do not have full information on their probability of success and the one of their advisers. If lower performing subjects are also more overconfident than high ability ones, they may benefit relatively less from advice. The second is preference for following their initial idea: if lower ability subjects value more strongly keeping their first answer than high ability ones, they are less likely to follow advice for a given expected gain, decreasing the potential for advice to act as a great equalizer.

B Sample Description

Table 3 shows that most of our subjects are young men. Figure 3 that most of our subjects are rated, but the mode is not being rated. The proportion of unrated players means that the low Elo group is almost all made of unrated players.

Table 3: Demographic characteristics

Gender	
Female	10
Male	78
Undeclared	15
Age	
<18	34
18-29	42
≥30	14
Undeclared	13

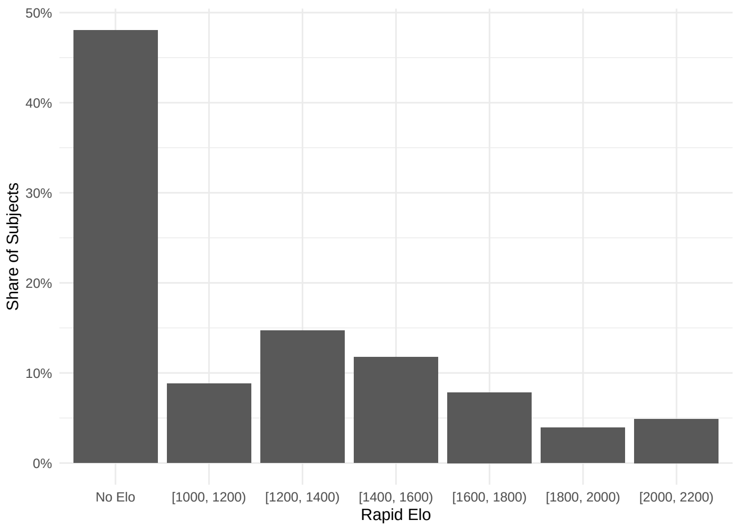


Figure 3: Distribution of Elo ratings among our subjects. The top rated adviser is above the upper limit.

C Main pre-registered test and regression analysis

In Table 4, we show the main pre-registered test. The difference in the share of correct answers after receiving advice with negative or positive assortative matching. Nothing is significant, showing that there is no decreasing differences effect. In Table 5 shows that the high quality advice has a positive effect on the share of correct answers, in both cases, but the effect is more pronounced in the Known Adviser case. The rating of the subject also has a positive effect, you we expect. Individual characteristics do not seem to matter much, outside of the ability of the subject.

Table 4: Main pre-registered test: comparing the share of correct answers post-advice under Negative Assortative Matching (NAM) and Positive Assortative Matching (PAM) of subjects to advisers.

Treatment	NAM	PAM	P-value[1]
Main test: H vs L-advice			
All	40.5%	38.4%	0.365
Unknown Adviser	38.1%	35.2%	0.384
Known Adviser	42.6%	41.3%	0.711
Alternative test: H vs No-advice			
All	42.5%	40.1%	0.301
Unknown Adviser	41.5%	36.9%	0.165
Known Adviser	43.3%	43.0%	0.951

[1] P-value of the two-sided two sample test of equal proportion between NAM and PAM.

D Additional Results

D.1 Keeping Answers

Table 7 shows that subjects correctly change their answers more after seeing H-advice. They fail to change them as often as they should however. It also shows that lower ability players change their answers more often that higher ability ones, which is expected. In Table 6, we show that most subjects keep their answers when they agree with the advisers'. Finally, in the regressions in Table 8, we show that higher Elo subjects tend to stick with their answers more often lower rated ones. When knowing the adviser, the baseline is that it is the L adviser, and subjects correctly keep their answer more often. On the other hand, as shown by the interaction terms between the H adviser and the Known Adviser, subject change significantly more often their answer when they know it is of good quality. Regression (2) shows that individual demographic characteristic do not play a significant role, outside of the ability of subjects.

D.2 Alternative Heuristics

In Section 3, we showed one possible counterfactual heuristic of what subject could have done if they knew their average probability of being correct. Here, we propose two alternative heuristics.

First, we start with a very unrealistic "first-best" heuristic. For each evaluation, you

Table 5: Regression for the share of correct answers, with fixed effects at the position level. The two leftmost regression control for individual characteristics.

Adviser	Known	Unknown	Known	Unknown
H adviser	0.40 (0.022)	0.27 (0.049)	0.31458 (0.046)	0.345 (0.044)
	(0.159)	(0.128)	(0.148)	(0.160)
Elo[1]	0.31 (<0.001)	0.21 (0.018)	0.32638 (<0.001)	0.279 (0.023)
	(0.068)	(0.081)	(0.058)	(0.113)
H adviser×Elo[1]	-0.18 (0.129)	-0.12 (0.198)	-0.11242 (0.297)	-0.181 (0.161)
	(0.113)	(0.091)	(0.105)	(0.124)
Male			-0.08838 (0.027)	-0.095 (0.083)
			(0.037)	(0.052)
age>=30			-0.09838 (0.089)	-0.107 (0.100)
			(0.055)	(0.062)
age18-29			-0.00766 (0.815)	-0.037 (0.375)
			(0.032)	(0.040)
Control Index			0.00044 (0.989)	0.056 (0.173)
			(0.031)	(0.040)
Std.Errors	by: position	by: position	by: position	by: position
Num.Obs.	1080	960	960	800
R2	0.145	0.080	0.171	0.092
R2 Adj.	0.127	0.059	0.148	0.062

P-value in parentheses on the same line, standard deviation on the line below.

[1] Elo rating is divided by 1,000 to scale the coefficient and make the estimate more readable.

can change your choice to the adviser's one, considering that after the advice you know what position is correct. In essence, we are assuming that subject know what the correct answer is once they see the advice, but are constrained to using their pre-advice choice or the advice. In any other case, they keep their (wrong) answer.

Our second, "Elo" heuristic is a rule-of-thumb of only accepting advice from someone objectively better. For h subjects, it means only following known H-advice, while for l subjects, only ignoring known L-advice. This approach has the advantage of being simple and corresponds to information known ex-ante by the subjects, but it is simplistic and penalizes h subjects who never follow unknown advice, while it would be better to do so for some.

Table 6: Percentage of kept answers when subjects agree with the adviser.

			When Agree Before[2]	
Type	Treatment	Agree[1]	Keep	React
l	Know H	33.5	93.0	7.0
l	Know L	27.8	98.6	1.4
l	Unknown	29.4	94.0	6.0
h	Know H	41.8	96.5	3.5
h	Know L	26.8	95.9	4.1
h	Unknown	33.0	97.3	2.7

We remove from this table the missing answers because we have no distance from the answer for them. We therefore slightly overestimate the agreement percentage before receiving the advice.

[1] Percentage of identical pre-advice answers with the adviser.

[2] Percentage of kept or changed answer conditional on pre-advice answer being identical to the adviser's.

Table 7: Share of identical answers for l and h subjects after observing different types of advice.

Type	L-Advice	H-advice	Unknown Advice
l	80.4%	61.5%	69.4%
h	85.0%	69.6%	83.5%

Table 9 shows that no matter the heuristic chosen, l type subjects always pay a higher premium than h type subjects. The difference is not always significant, in particular in the first-best heuristic which by construction equalizes more than the other two.

Table 8: Regression for keeping the answer after receiving the advice, with fixed effects at the position level.

	(1)	(2)
Distance Correct[1]	-0.041 (0.002)	-0.0451 (0.002)
	(0.011)	(0.012)
H Adviser	-0.019 (0.544)	-0.0083 (0.806)
	(0.031)	(0.033)
Known Adviser	0.078 (0.003)	0.0473 (0.063)
	(0.023)	(0.024)
Elo[1]	0.177 (<0.001)	0.1702 (<0.001)
	(0.022)	(0.027)
H×Known Adviser	-0.168 (<0.001)	-0.1685 (<0.001)
	(0.038)	(0.038)
Control Index		0.0356 (0.085)
		(0.020)
Male		0.0456 (0.157)
		(0.031)
Age >=30		0.0440 (0.089)
		(0.025)
Age 18-29		0.0070 (0.701)
		(0.018)
Std.Errors	by: position	by: position
Num.Obs.	1999	1749
R2	0.072	0.080
R2 Adj.	0.060	0.065

Notes: Robust standard errors clustered at the position level. In parenthesis on the same line are the p-value, below the standard error.

(2) adds demographic controls but restrict the sample.

[1] Absolute distance from the correct answer in pawn advantage.

[2] Elo is divided by 1,000 to scale the coefficient and make it easier to understand.

Table 9: Difference (in percentage points) between the average share of correct answers of h and l type subjects having received H-advice, following our heuristics and in the experiment.

		Types		
Treatment	Heuristic	l	h	P-value[1]
Unknown	Probabilistic	32.4	23.0	0.056
Known	Probabilistic	27.7	21.1	0.139
All	Probabilistic	30.0	22.0	0.015
Unknown	Elo	36.8	-5.7	<0.001
Known	Elo	27.7	21.1	0.139
All	Elo	32.2	9.0	<0.001
Unknown	First-Best	43.2	37.8	0.240
Known	First-Best	34.6	28.9	0.168
All	First-Best	38.8	32.9	0.055

Premia paid for ignoring H-advice are given in percentage points.

[1] P-value of the two-sided two sample t-test of equal control premium between the h and l Elo types.

FOR ONLINE PUBLICATION ONLY: EXPERIMENTAL INSTRUCTIONS IN THE KNOWN ADVISOR TREATMENT

Response Sheet - ورقة الاجابة

Participant number – (ة)رقم المشترك:A1___________

ELO - التصنيف:

Your predictions - التوقعات

Round 1 – الجولة الاولى:

رقم الوضع Position Number	**Part 1 - الجزء الأول**				**Part 2 - الجزء الثاني**			
	-2.4	**-0.7**	**+0.7**	**+2.4**	**-2.4**	**-0.7**	**+0.7**	**+2.4**
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								

Round 2 – الجولة الثانية:

رقم الوضع Position Number	**Part 1 - الجزء الأول**				**Part 2 - الجزء الثاني**			
	-2.4	**-0.7**	**+0.7**	**+2.4**	**-2.4**	**-0.7**	**+0.7**	**+2.4**
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								

(Please complete both sides of the sheet)

الرجاء تعبئة جانبي الورقة

Additional Info – معلومات إضافية

Age - العمر:

Gender - الجنس:

How much do you agree with the following statements - ما مدى موافقتك على العبارات التالية ?

	Strongly disagree أعارض بشدة	Disagree أعارض	Neither agree nor disagree لا أوافق ولا أعارض	Agree أوافق	Strongly agree أوافق بشدة
I try to avoid situations where someone else tells me what to do. أحاول تجنب المواقف التي يقول لي فيها شخص آخر بما يجب القيام به.					
I prefer to be a leader rather than a follower. أفضل أن أكون قائدًا وليس تابعًا.					
I enjoy making my own decisions. أنا أستمتع باتخاذ قراراتي بنفسي.					
I would rather someone else took over the leadership role when I'm involved in a group project. أفضل أن يتولى شخص آخر الدور القيادي عندما أشارك في مشروع جماعي.					
There are many situations in which I would prefer only one choice rather than having to make a decision. هناك العديد من المواقف التي أفضل فيها خيارًا واحدًا فقط بدلاً من الاضطرار إلى اتخاذ قرار.					

A1i

Here are ten positions that occurred in real chess games which have been chosen from a dataset of previous games from the Mega Database 2023.

We will ask you to evaluate 20 games over two rounds: 1 and 2. We will pick one of your evaluations at random and you will receive a voucher of $10 if your answer was correct.

Please complete **Round 1, Part 1** of the Response Sheet by indicating for each game your best estimate of the pawn advantage, which can be +0.7, -0.7, +2.4, or -2.4. Please check the box corresponding to your choice (only one possible answer). Note that the positions have a pawn advantage of ±0.7 and one of ±2.4 with equal probability.

Once you have completed **Round 1, Part 1**, please wait for the experimenter to give you the next set of instructions.

You have a total of 8 minutes to complete this part.

فيما يلي عشرة أوضاع حصلت في جولات شطرنج حقيقية وقد تم اختيارها من مجموعة بيانات للألعاب السابقة من Mega Database 2023.

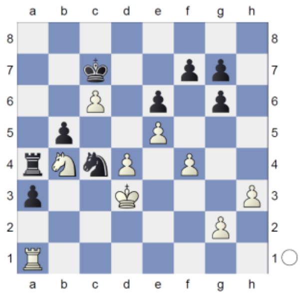
سنطلب منك تقييم 20 وضع على مرحلتين: الجولة الأولى والجولة الثانية. سوف نختار أحد تقييماتك بشكل عشوائي وستتلقى قسيمة بقيمة 10 دولارات إذا كانت إجابتك صحيحة.

يرجى تعبئة **الجولة 1، الجزء 1** من ورقة الإجابة بالإشارة إلى أفضل تقدير لديك لكل وضع حسب أفضلية ال pawn advantage والتي يمكن أن تكون 0.7+ أو 0.7- أو 2.4+ أو 2.4-. يرجى تحديد المربع المقابل لاختيارك (إجابة واحدة فقط ممكنة). ملاحظة: من المحتمل أن يكون الوضع مع أفضلية 0.7±، **أو** 2.4±، مع احتمالية متساوية.

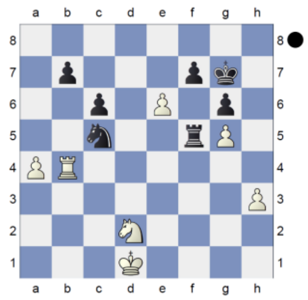
بمجرد الانتهاء من الجولة 1، الجزء 1، من فضلك انتظر أن يعطيك المشرف المجموعة التالية من التعليمات.

لديك 8 دقائق لإكمال هذا الجزء.

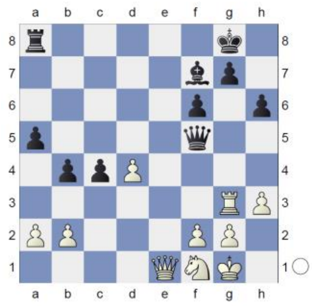
1



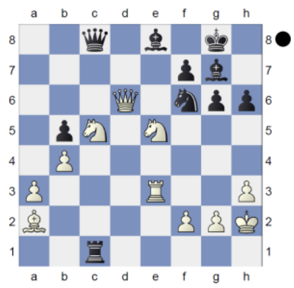
2



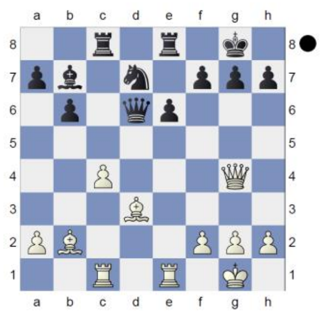
3



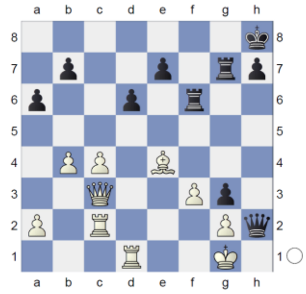
4



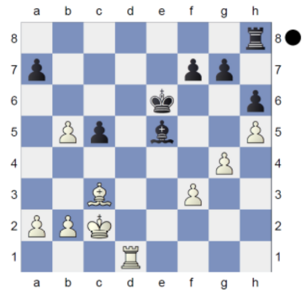
5



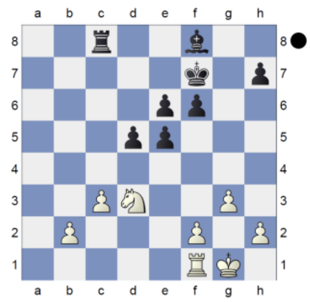
6



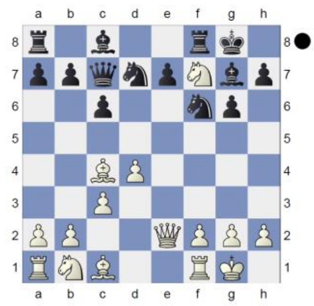
7



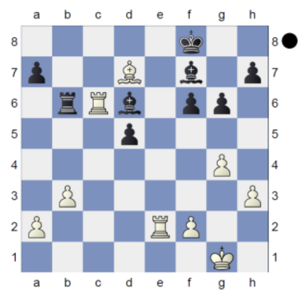
8



9



10



A1i

We will now provide you with some additional information about the ten positions.

We have asked **a player with a rating of 2335** to evaluate the ten games in the same conditions as you. You can find their prediction in the table below.

Looking back at your own evaluation in **Round 1, Part 1** on the Response Sheet, please complete **Round 1, Part 2**. You are free to change or keep your previous predictions based on the information on this sheet.

سنزودك الآن ببعض المعلومات الإضافية حول الاوضاع العشر.

لقد طلبنا من **لاعب(ة) تصنيفه 2335** أن يقيم الاوضاع العشر في نفس ظروفك. يمكنك الاطلاع على توقعاتهم في الجدول أدناه.

بالنظر إلى تقديرك في **الجولة 1**، **الجزء 1** في ورقة الإجابة، يرجى إكمال **الجولة 1**، **الجزء 2**. لك مطلق الحرية في تغيير توقعاتك السابقة أو الاحتفاظ بها بناءً على المعلومات الواردة في هذه الورقة.

رقم الوضع Position Number	التفوق Pawn advantage
1	-2.4
2	-0.7
3	-0.7
4	+2.4
5	+0.7
6	-0.7
7	+2.4
8	-0.7
9	-0.7
10	+2.4

You have a total of <u>4 minutes</u> to complete this part.

لديك 4 دقائق لإكمال هذا الجزء.

A1ii

Now, we will repeat the previous exercise with a new set of ten positions.

Please complete **Round 2, Part 1** of the Response Sheet. This is the same procedure as for **Round 1.**

Once you have completed **Round 2, Part 1**, please wait for the experimenter to give you the next set of instructions.

الآن، سنكرر التمرين السابق بمجموعة جديدة من عشر أوضاع.

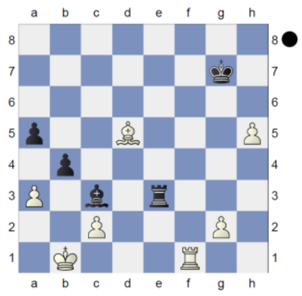
يرجى تعبئة **الجولة الثانية، الجزء 1** من ورقة الإجابة. هذا هو نفس الإجراء المتبع في الجولة الأولى.

بمجرد الانتهاء من الجولة 2، الجزء 1، من فضلك انتظر أن يعطيك المشرف المجموعة التالية من التعليمات.

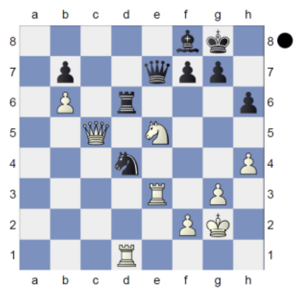
You have a total of <u>*8 minutes*</u> *to complete this part.*

لديك 8 دقائق لإكمال هذا الجزء.

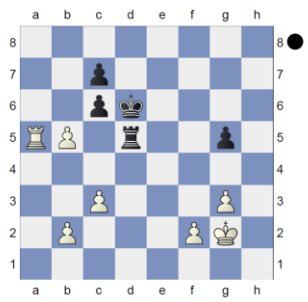
11



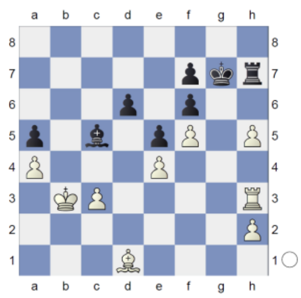
12



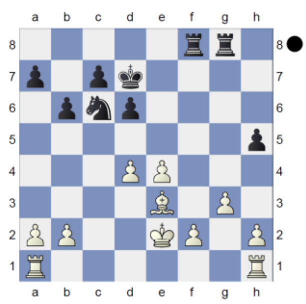
13



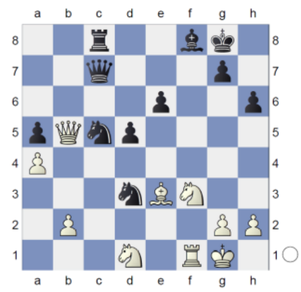
14



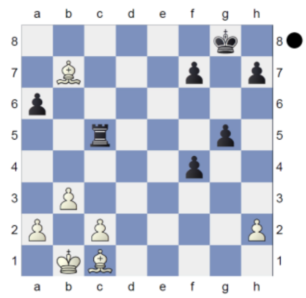
15



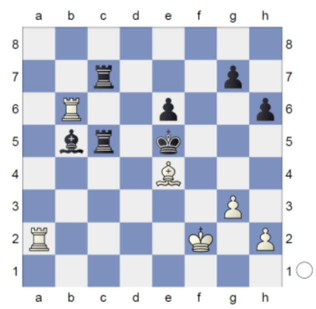
16



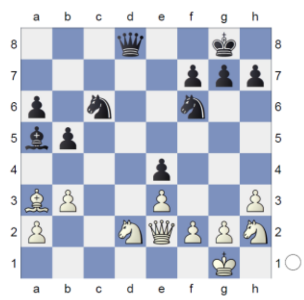
17



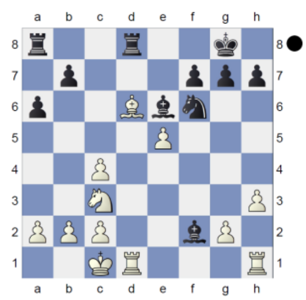
18



19



20



A1ii

We will now provide you with some additional information about the ten positions.

We have asked **an unrated player**, who plays regularly for fun, to evaluate the ten positions in the same conditions as you. You can find their predictions in the table below.

Looking back at your own evaluation in **Round 2, Part 1** on the Response Sheet, please complete **Round 2, Part 2**. You are free to change or keep your previous predictions based on the information on this sheet and to look at the prediction sheet.

سنزودك الآن ببعض المعلومات الإضافية حول الاوضاع العشر.

لقد طلبنا من **لاعب(ة)غير مصنف، يلعب بانتظام من أجل التسلية**، أن يقيم الاوضاع العشر في نفس ظروفك. يمكنك الاطلاع على توقعاتهم في الجدول أدناه.

بالنظر إلى تقديرك في **الجولة 2، الجزء 1** في ورقة الإجابة، يرجى إكمال **الجولة 2، الجزء 2**. لك مطلق الحرية في تغيير توقعاتك السابقة أو الاحتفاظ بها بناءً على المعلومات الواردة في هذه الورقة.

رقم الوضع Position Number	التفوق Pawn advantage
11	-2.4
12	+0.7
13	+2.4
14	-0.7
15	+0.7
16	-2.4
17	-0.7
18	-2.4
19	-0.7
20	-0.7

You have a total of 4 minutes to complete this part.

لديك 4 دقائق لإكمال هذا الجزء.

When this is over, please complete the personal information questions at the back of the response sheet.

عندما تنتهي من هذا الجزء، يرجى إكمال أسئلة المعلومات الشخصية في الجزء الخلفي من ورقة الإجابة.