Decreasing Differences in Expert Advice: Evidence from Chess Players *

Elias Bouacida† Renaud Foucart‡ Maya Jalloul§

August 19, 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 lower-ability player. While high-quality advice has the potential to act as a "great equalizer," reducing the difference between high and lower-ability players, this is not what happens in our experiment. This is in part because lower-ability players pay a higher premium than higher-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, Danmo Lin and Paola Manzini, as well as from those of participants to seminars and workshop in Lancaster, Paris 8, the 1st Doctoriales ASFEE and 1st PBES. The pre-registration is available at https://aspredicted.org/124_MSY, the data is stored in https://doi.org/10.57745/UL502I, and the data analysis at https://decreasing-differences-ebouacida-92abd7c994ce472e25f3e62c7fbcaa.gitpages.huma-num.fr/

†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 higher and lower-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*.

Empirical evidence of decreasing differences in matching abilities in the labour market includes 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 finds that, in general, lower ability workers benefit more from being part of a team with a higher-ability partner (Herkenhoff *et al.*, 2024). Chade and Eeckhout (2018) show theoretically how this logic applies to the context of matching expertise.

In a lab-in-the-field experiment with chess players participating in tournaments in Lebanon, we find no significant evidence of decreasing differences in expert advice. 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 for ignoring good advice.

We partnered with a local chess academy to run our incentivized experiment alongside chess tournaments in several locations in Lebanon in the Summer of 2023. The main task 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 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 subjects to evaluate it again.

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

We define as "higher-ability" subjects with an official chess rating in the top half of the sample, and as "lower-ability" those in the bottom half. Before receiving the high-quality advice, higher-ability subjects had a correct answer rate of 41.2%, and lower-ability subjects had a rate of 32.9%.

After receiving high-quality advice, the rate increased to 50.8% (+9.6 percentage points) for higher-ability subjects, and to 42.5% (+9.6pp) for lower-ability ones. In contrast, higher-ability subjects offered high-quality advice could have increased their rate of correct answers by 22.0pp on average and lower-ability subjects could have increased it by 30.0pp by following advice when it was (in expectation) beneficial to do so. The absence of decreasing differences in expert advice in the experiment is therefore largely due to the significantly higher premium paid by lower-ability subjects when they ignore high-quality advice.

The main novelty of our research lies in explicitly measuring the potential for decreasing differences in a context involving expertise. Our subjects possess a significant level of expertise on their domain and must reconcile this with potentially superior external expertise, presented as advice. Indeed, chess players who signed up to participate in an official tournament are arguably more akin to professionals with domain knowledge receiving advice than to participants in a lab experiment completing tasks for which they have no particular reason to feel qualified.

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. 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.

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.

Finally, this paper follows a long tradition in the economic literature of using chess players to study human decision-making, including, among others, 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] 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 (the material in available in the Online Appendix).

We recruited subjects prior to the tournament through the academy and covered their registration (approximately $5) as a participation fee. The experiment was conducted in a separate room while subjects were not engaged in competitive play. Each subject was randomly assigned to either the treatment with or the treatment without information on the adviser. 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. Following 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, following the 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 advantaged player, 6 with a draw, and 1 with a loss.

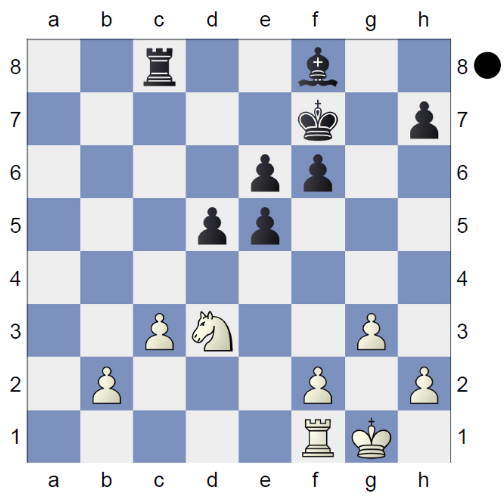


Figure 1: A position in a game of chess, as shown to our subjects.

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 the 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.[3]

In the *known adviser* treatment, we told subjects that the answers we gave them were coming from "a player with a rating of 2,335" (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 2,335, or it is an unrated player, who plays regularly for fun". The rating refers to the *Elo rating,* the standard measure of chess performance.[4]

[3]We chose this "independent-then-revise" procedure in order to be able to measure how advice improves the answers of our subjects. As compared to a procedure offering advice before answering, it is unclear whether it should lead to more or less advice taking (Rader *et al.*, 2015). On the one hand, answering first leads to an anchoring bias where subjects want to keep their original answer. On the other hand, subjects who have the same answer as the adviser keep it, while seeing advice immediately can lead to a push-away effect where subjects choose an answer just above or just below the advice.

[4]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

After completing the two rounds of evaluations, subjects filled out a brief demographic questionnaire along with questions regarding their stated preference for control (5 questions borrowed from Burger and Cooper, 1979). The experimental material is available in the Online Appendix. All 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 $10 if the answer was correct.[5]

In accordance with our pre-registration, we divided the sample of 102 subjects into two equal-sized groups based on their rating. As 54 subjects had a formal Elo rating, the results remain nearly identical when considering a dichotomy of rated/unrated subjects instead. We classified questions for which subjects did not provide answers as incorrect. The average Elo rating of rated players is 1,490, with the highest rated subject falling within the range 2,100-2,200. Therefore, it should be clear to all subjects that the higher-ability adviser, with a rating of 2,335, is more likely than them to evaluate a position correctly than they are. Furthermore, it should be apparent that the lower-ability adviser does not possess strictly higher ability than any of our subjects, all of whom are participating in a registered tournament. The Elo distribution of subjects is illustrated in Figure 3 in Appendix B.

3 Results

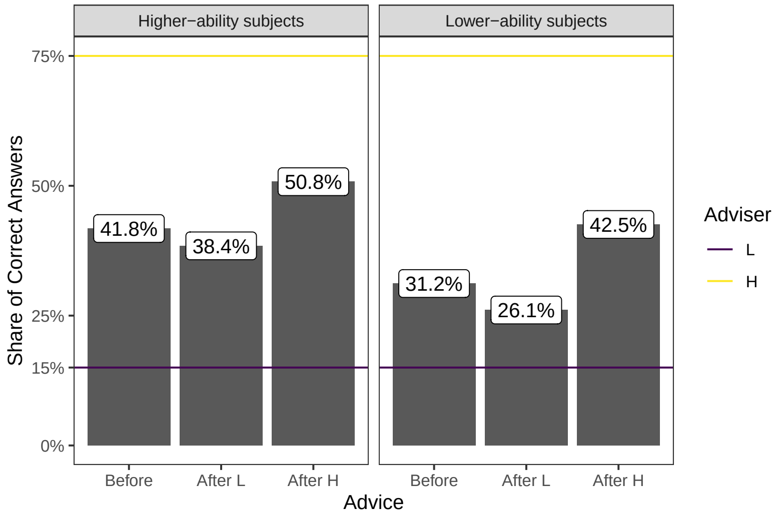
We show on Figure 2 the average share of correct answers in the lower (l) and higher (h) ability groups, before observing advice, and after observing low-quality (L) and high-quality (H) advice. This figure pools both treatments.

Before observing advice, higher-ability subjects evaluated 41.8% of the positions correctly, while the result for lower-ability subjects is 31.2%. Our higher-quality adviser

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 1,000 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.

[5] Given the difficult banking situation in Lebanon and the fact that some 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.

Figure 2: Share of correct answers by subject type, before observing advice, and after observing low and high-quality advice. Both treatments are pooled.



provided 75% of correct answers, and our lower-quality 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 low-quality advice.

Looking only at the questions for which subjects received high-quality advice, the rate of correct answers increases from 32.9% to 42.5% for lower-ability subjects, and from 41.2% to 50.8% for higher-ability subjects. The increase following high-quality advice is thus identical for both types of subjects, at 9.6pp. In contrast, if all subjects followed all high-quality advice, our lower-ability subjects would have seen their share of correct answers after receiving high-quality advice increase by 8.3pp more than higher-ability subjects.

As a consequence, we do not find any statistical evidence of decreasing differences in expert advice in our pre-registered tests. In the main test, we compare the average share of correct answers when matching lower-ability subjects and low-quality advice and higher-ability subjects and high-quality advice (Positive Assortative Matching, PAM) to Negative Assortative Matching of lower-ability subjects to high-quality advice and higher-ability subjects to low-quality advice (NAM). In the presence of decreasing differences, the gain from high-quality advice as compared to low-quality advice should be higher for lower-ability subjects than for higher-ability subjects, so that the share of correct answers with NAM should be higher than with PAM (see Appendix A for a formalization).

In the treatment with unknown adviser, the share of correct answers with PAM is 35.2%, and it is 38.1% with NAM. The p-value of the two-sided two sample t-test of equal proportion between NAM and PAM is 0.384. With known adviser, the difference is even smaller (42.6% versus 41.3%, $p = 0.711$). Pooling both treatments, the difference remains non-significant ($p = 0.365$). While we cannot rule out that some decreasing differences exist, our sample size should have been sufficient to identify any large effect. Following Vasilaky and Brock (2020), we look at the minimal detectable effect. 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%. In our alternative pre-registered test, we use the share of correct answers pre-advice instead of after lower-quality adviser. We also fail to capture statistically significant evidence of decreasing differences (see Table 4 in Appendix C).

Overall, subjects tend to ignore advice: around three-quarters of the choices are unchanged after observing advice (we report in Table 6 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 it. In Table 7 in Appendix D, we show that subjects keep their answers at least 93% of the time when they agree with 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]			
Subjects	Treatment	Disagree[1]	Keep	Follow	Closer	Further
Lower-ability	Know H	66.5	46.8	39.2	12.3	1.8
	Know L	72.2	78.8	14.0	7.3	0.0
	Unknown	70.6	64.3	27.3	6.3	2.2
Higher-ability	Know H	58.2	52.5	42.5	2.5	2.5
	Know L	73.2	83.9	9.5	4.0	2.5
	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.

Table 1 show what happens when the answers of subjects differ from the advisers'. Lower-ability subjects ignore advice from unknown advisers 64.3% of the time – as compared to 79.7% for the higher-ability one – and ignore (46.8%) or move further away (1.8%) from the higher-quality adviser roughly half of the time, only slightly less than higher-ability subjects. In line with the findings of Alysandratos *et al.* (2020) on economic experts, we find no evidence that subjects are able to distinguish good from bad advice when they do not know the identity of the adviser.

Participants also mostly, and correctly, ignore low-quality advice. In line with Schultze *et al.* (2017) some subjects however feel the need to incorporate even useless advice; 9.5% of our higher-ability subjects update their evaluation following advice from an adviser they should expect to be worse than them (the figure is 14.0% for lower-ability subjects).

We confirm our main results through regression analysis, as detailed in Table 5 in Appendix C. Following our pre-registration plan, we employ both our binary classification of lower- and higher-ability subjects, and a continuous measure based on the Elo rating, conducting regressions for both the known and unknown adviser treatments. Consistent with expectations, the higher-ability adviser typically benefits subjects, and better-rated subjects are more likely to accurately evaluate positions. The interaction term between advice quality and subjects' rating provides an alternative measure of the existence of decreasing differences, which, as in the main tests using two subjects categories, is not significant. We control for individual characteristics in Table 5 in Appendix D.

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

Our first "Probability" heuristic corresponds to the optimal choice of our subjects if they were aware of their probability of answering correctly (approximated by their share of correct answers) as well as the probability of the advisers.[6] This is for instance the simple heuristic payoff-maximizing subjects would choose after a large number of trials and errors.

For each subject, we apply this heuristic to determine the number of correct answers that could have been achieved by following it. Table 2 quantifies the premium paid by subjects for ignoring high-quality advice, defined as the difference (in percentage points) in the share of correct answers after receiving high-quality advice if they had followed our heuristic compared to their performance in the experiment. Across treatments, lower-ability subjects would have achieved a 30.0 percentage points higher share of correct answers after high-quality advice by following the heuristic, compared to a 22.0pp in-

[6] 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 "probabilistic" 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.

Table 2: Difference (in percentage points) between the average share of correct answers of higher-ability and lower-ability subjects having received high-quality advice, following the probability heuristic and in the experiment.

	Subjects		
Treatment	Lower-ability	Higher-ability	P-value[1]
Unknown adviser	32.4	23.0	0.056
Known Adviser	27.7	21.1	0.139
All	30.0	22.0	0.015

Premia for lower- and higher-ability are given in percentage points.

[1] P-value of the two-sided two sample t-test of equal premium between the lower- and higher-ability subjects.

crease for higher-ability subjects. The difference is statistically significant. Therefore, we conclude that lower-ability subjects pay a higher premium than higher-ability subjects for ignoring high-quality advice, potentially due to overconfidence or a preference for maintaining their original answer (i.e., anchoring bias, status quo bias or preference for control).

This method is obviously imperfect and was not part of our pre-registration. However, it provides an illustrative idea of the potential impact of advice and its role as a great equalizer. In Appendix D.2, we present two other heuristics that yield similar results: a "first best" heuristic in which subjects select the correct advice when available for each question, and an "Elo" heuristic in which they only follow the advice of an objectively superior adviser.

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 the stated preference for control is correlated with the probability of a subject keeping their answers, after controlling for subject and positions characteristics (see Table 8 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 domain 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 lower-ability subjects 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 lower-ability subjects the most, as they had the most to gain. The fact that lower-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*.

Rader, C. A., Soll, J. B. and Larrick, R. P. (2015). Pushing away from representative advice: Advice taking, anchoring, and adjustment. *Organizational Behavior and Human Decision Processes*, **130**, 26–43.

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 lower-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 higher-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 higher-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 indicates that the majority of our subjects are young men. Figure 3 illustrates that while most of our subjects are rated, the most common category is being unrated. The number of unrated players means that the lower-ability group is mainly composed of unrated players.

Table 3: Demographic characteristics

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

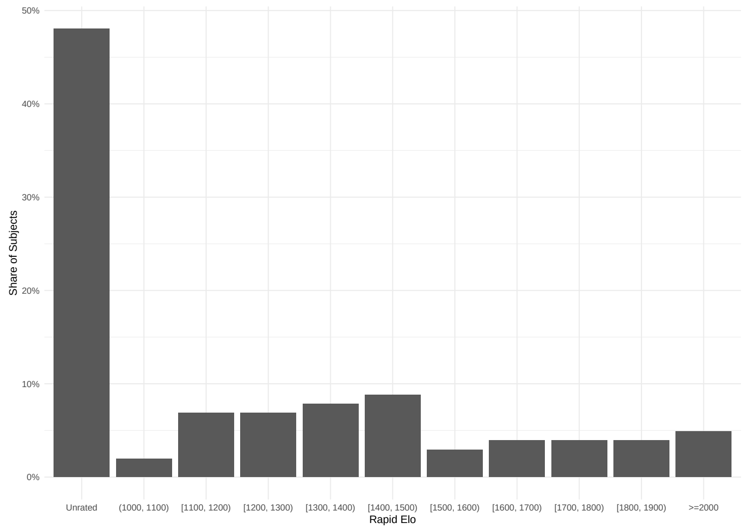


Figure 3: Distribution of Elo ratings among our subjects. The higher-ability adviser is rated above the upper limit.

C Main pre-registered test and regression analysis

In Table 4, we show the main pre-registered tests. The difference in the share of correct answers after receiving advice with Negative or Positive Assortative Matching. We do not find any significant evidence of decreasing differences in expert advice. We report the main regression analysis in Table 5. It shows that high-quality advice has a positive effect on the share of correct answers, in both cases. Consistent with expectations, the effect is more pronounced in the Known Adviser treatment. Also consistent with expectation, higher rated subjects have a higher rate of correct answers. Individual characteristics do not seem to matter much, outside of the relative ability of subjects.

Table 4: Main pre-registered tests: comparing the share of correct answers after receiving 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 t-test of equal proportion between NAM and PAM.

D Additional Results

D.1 Keeping Answers

Table 6 shows that subjects correctly change their answers more after seeing high-quality 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, at least in part because their answers are more likely to be different from the high-quality advice. In Table 7, we show that most subjects keep their answers when they agree with the advisers'. Finally, the regressions in Table 8 show that subjects with a higher Elo rating tend to stick with their answers more often than lower rated ones. When knowing the adviser, the baseline is the lower-quality advice, and subjects correctly keep their answer more often. As shown by the interaction terms between the higher-quality advice and the Known Adviser, subjects 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 presented a counterfactual heuristic of what subject could have achieved if they were aware of their average probability of being correct. Here, we propose two alternative heuristics.

Table 5: Regression for the rate 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 $\geq$ 30			-0.09838 (0.089)	-0.107 (0.100)
			(0.055)	(0.062)
18 $\leq$ Age $\leq$ 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
R^2	0.145	0.080	0.171	0.092
Adj. R^2	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.

Table 6: Share of identical answers for lower- and higher-ability subjects after observing different qualities of advice.

Subject	Low-quality advice	High-quality advice	Unknown advice
Lower-ability	80.4%	61.5%	69.4%
Higher-ability	85.0%	69.6%	83.5%

First, we start with a highly unrealistic “first-best” heuristic. For each evaluation, subjects can switch their choice to align with the adviser’s recommendation, assuming that they know the correct position after receiving the advice. Essentially, this assumes

Table 7: Percentage of kept answers when subjects agree with the adviser.

			When Agree Before[2]	
Subjects	Treatment	Agree[1]	Keep	React
Lower-ability	Know H	33.5	93.0	7.0
	Know L	27.8	98.6	1.4
	Unknown	29.4	94.0	6.0
Higher-ability	Know H	41.8	96.5	3.5
	Know L	26.8	95.9	4.1
	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.

that subjects recognize the correct answer once they see the advice but are limited to choosing between their initial decision and the adviser's suggestion.

Our second heuristic, the "Elo" heuristic, is a rule-of-thumb that involves accepting advice only from someone objectively better. For higher-ability subjects, this means following only known higher-quality advice, while for lower-ability subjects, it means ignoring only known lower-quality advice. This approach is straightforward and relies on information available to the subjects beforehand. However, it is simplistic and disadvantages higher-ability subjects who might benefit from following some unknown advice.

Table 9 demonstrates that regardless of the heuristic applied, lower-ability subjects always pay a higher premium than higher-ability subjects. The difference is not always statistically significant, particularly with the first-best heuristic which inherently equalizes more than the other two heuristics.

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 lower- and higher-ability subjects having received high-quality advice, following our heuristics and in the experiment.

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

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

[1] P-value of the two-sided two sample t-test of equal premium between the lower- and higher-ability subjects.

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

Response Sheet - ورقة الاجابة

Participant number – رقم المشترك(ة):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.

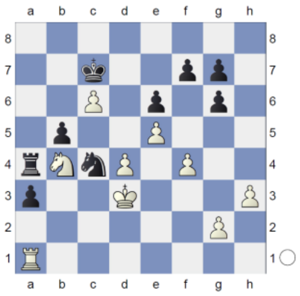
سنطلب منك تقييم 20 وضع على مرحلتين: الجولة الأولى والجولة الثانية. سوف نختار أحد تقييماتك بشكل عشوائي وستتلقى قسيمة بقيمة 10 دولارات إذا كانت إجابتك صحيحة.

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.

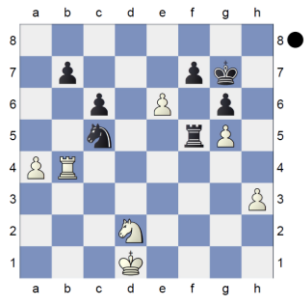
بمجرد الانتهاء من الجولة 1، الجزء 1، من فضلك انتظر أن يعطيك المشرف المجموعة التالية من التعليمات.

لديك 8 دقائق لإكمال هذا الجزء.

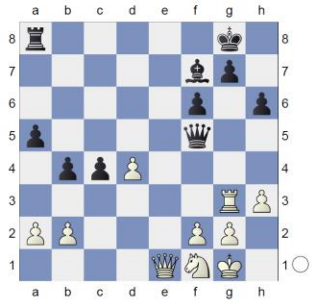
1



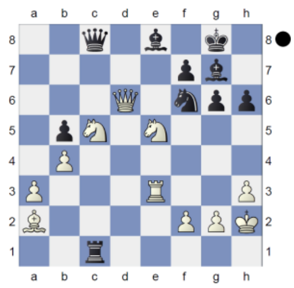
2



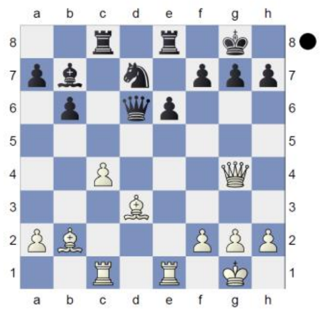
3



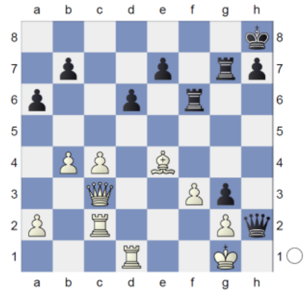
4



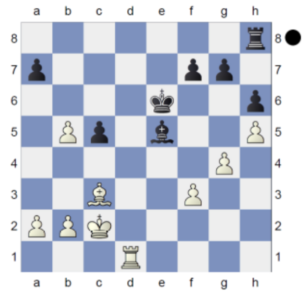
5



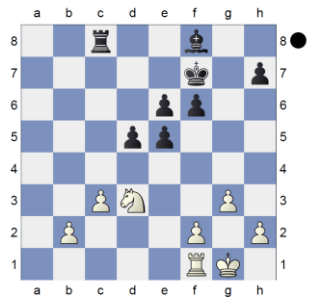
6



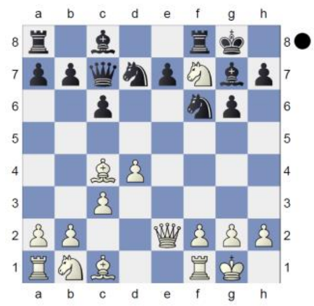
7



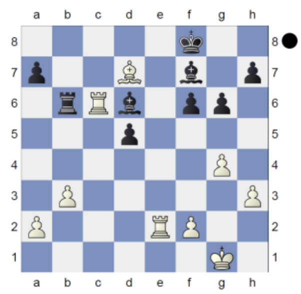
8



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 4 minutes 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.

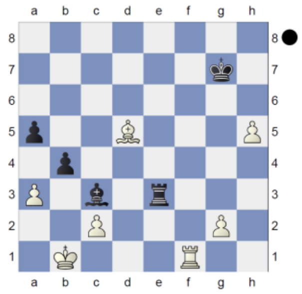
يرجى تعبئة **الجولة الثانية، الجزء 1** من ورقة الإجابة. هذا هو نفس الإجراء المتبع في الجولة الأولى.

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

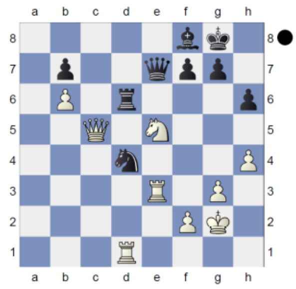
You have a total of 8 minutes to complete this part.

لديك 8 دقائق لإكمال هذا الجزء.

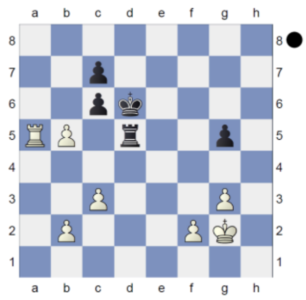
11



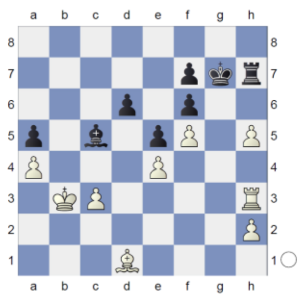
12



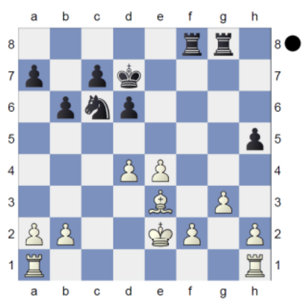
13



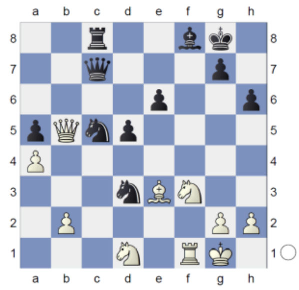
14



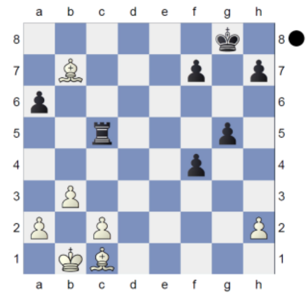
15



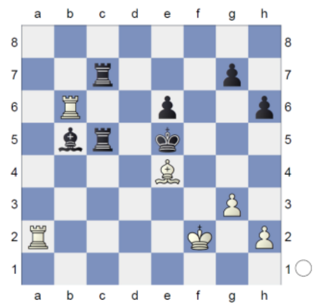
16



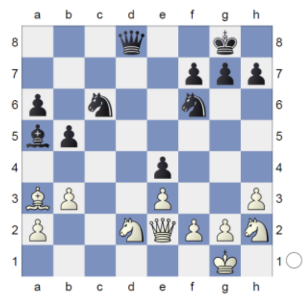
17



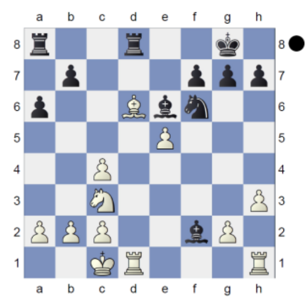
18



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.