Intuitive Choices in Chess Option-Generation are Task and Expertise Dependent: A Conceptual Replication

Markus Raab[1], Sofie Klenk[2], Zoltán Medvegy[3], Ivanna Korobeinikova[4], Lesia Korobeinikova[4,5], Georgiy Korobeynikov[4,5], Elisa Weinand[1], and Tamás Sterbenz[3]

[1]German Sport University Cologne, Germany; [2]University of Salzburg, Austria; [3]Hungarian University of Sports Science, Hungary; [4]National University of Ukraine on Physical Education and Sport, Ukraine; [5]Uzbek State University of Physical Culture and Sports, Uzbekistan

Correspondence: Markus Raab, raab@dshs-koeln.de

Abstract

The aim of this study was the conceptual replication of a study assessing when experts trust their intuition. Medvegy et al. (2022) found that experts generate options based on option validity that matches the current situation, generate only a few options, and often pick the first one generated. Results vary based on expertise assessed by Elo ratings. We tested whether these results hold when comparing individuals with a slightly lower, but still expert, level of expertise. Therefore, we conceptually replicated the exact design and stimuli and asked chess players for their intuitive choice, further candidate moves, and the best choice. We predicted that the findings in experts would replicate and be due to situational factors, such as when the task prompts more strategic rather than tactical choices in chess. Results indicate effects of both expertise and task characteristics, indicating a broader account of when individuals close to expert level trust their intuition. However, unlike the original study, lower-rated players did not show a clear quality advantage of the first generated option, suggesting that the Take-the-First heuristic becomes reliably effective only at higher levels of expertise. At the same time, intuitive choices were not worse than deliberate ones, and players were especially vulnerable in tactical situations under time pressure. Conclusions are drawn regarding when to trust intuition.

Keywords

chess, intuition, Take-the-First, option generation, fast and frugal heuristic

Where does intuition come from?

"Intuition is driven by your massive experience in chess, with structures and plans that exist in each position. As soon as I see a position that I know, my feeling tells me what to do. I do not need to think about what I want to achieve": Vincent Keymer (2023), grand chess master, from an interview in *Die Zeit*, a German national weekly newspaper (translation by the first author).

Introduction

In situations that are ambiguous and uncertain, decisions need to be made frequently and multiple ways to move forward need to be generated. The theorizing of decision processes and how options are generated has received attention in studies of individual decisions in contexts such as chess (Blanch et al., 2020; Klein et al., 1995), sports (Hepler & Feltz, 2012;

Johnson & Raab, 2003), and areas of business and management (e.g., Butler & Scherer, 1997). In some of the experimental situations, participants had to generate options and then select from them, as opposed to being given a set of choice alternatives (Johnson & Raab, 2003). Whether choosing among many options improves performance (more-is-better effect) or diminishes performance (less-is-more effect) is heavily debated and may depend on the person, the task, and situational factors (for a meta-analytical account, see Chernev et al. 2015; Scheibehenne et al., 2010; for theoretical accounts, see Gigerenzer, 2008; Gigerenzer & Gaissmaier, 2011; Gobet, 1997, 2018; Raab & Gigerenzer, 2015; for empirical investigations, see Del Missier, Visentini, & Mäntylä, 2015; Gobet & Simon, 1996; Klatt et al., 2019; Medvegy et al., 2022; Moxley et al., 2012; Nordgren & Dijksterhuis, 2009).

In a meta-analysis, asking "if there can ever be too many options," Scheibehenne et al. (2010) presented a forest plot of 50 studies showing how the effect of the number of available options affects choices. Importantly, even for the same task, more options sometimes produced more desired or appropriate choices, but in other experiments, a lower number of options was beneficial. A limitation of most of these studies was that participants were often not systematically assessed on their expertise and experience with that task, or how much time they had to make a decision, such that both the person and the situation may add to the explanation. It seems quite likely that persons with high expertise may rely on fewer options for a satisfying decision, especially when time is limited, such as in many sport situations or when the chess clock is running down. A second limitation is that in previous research on testing how the number of options affects performance, neither expertise nor the nature of the choice could always be operationalized consistently. Chess, in contrast, is a task in which humans or artificial machines can evaluate the quality of a move, and for which player expertise is well-defined (Kasparov, 2007).

To address the limits of previous work, we tested expertise-dependent effects of the less-is-

more vs. more-is-better claims in chess decisions with a conceptual replication of the study by Medvegy et al. (2022). Any conceptual replication study risks that the specific sampling of participants or experimental procedures may produce different results. Whether these differences indicate that effects are not robust or that the specific conceptual variations of design choices are meaningful are hard to judge based on isolated study pairs. Our motivation to conceptually replicate the specific study of Medvegy et al. (2022) was two-fold. First, the expertise of chess players in their sample was very high – grandmasters and continental tournament players – which may have skewed their choices toward generating only a few options per situation. Including players with lower expertise allows us to test if there is a specific threshold of expertise needed to replicate the less-is-more effect in chess. Second, we aimed to use the self-reported limitations of the study from Medvegy et al. (2022), such as the small sample size, the limited variance in expertise in the sample, and specific methodological decisions, and turn them into an empirical – conceptually optimized – test. For instance, relying on Elo scores alone to assess chess expertise (Elo, 1978) may be limiting if time pressure or different situations require more or fewer options to be generated. Therefore, we added Bullet, Blitz, Rapid and Classical ratings; otherwise, we kept the design as close as possible to identical to the original.

The original study used an option-generation paradigm to test the Take-the-First heuristic (Johnson & Raab, 2003), which predicts a less-is-more effect for experts, whereby picking the first generated option and generating only a few options leads to better decisions. This heuristic is especially relevant when humans can rely on task experience and are under time pressure (Belling et al., 2015; Musculus et al., 2019). In Medvegy et al. (2022), chess players of the European Universities Championship were shown pictures of positions (see Figure 1 below) and were asked to make a fast, intuitive decision about their next move, then generate further appropriate moves. After that, they were asked to choose the best move from those generated

and evaluate the quality of this move on a Likert scale. The presented positions varied in whether they required tactical or strategic decisions and whether time pressure was applied or not, allowing for testing potentially contradictory findings across tasks and conditions.

Their results showed that players with the highest Elo scores (Elo, 1978) performed better according to expert ratings and chess engines. The difference between levels of expertise was more pronounced in tactical decisions than in strategic decisions. The categorization of a decision as tactical or strategic was based on grandmasters' evaluations of chess positions. Tactical decisions in chess tend to be concrete, whereas strategic decisions leave more freedom for long-term planning. A quote from former world champion Kasparov (2007, p. 41) makes this transparent: "While strategy is abstract and based on long-term goals, tactics are concrete and based on finding the best move right now. Tactics are conditional and opportunistic, all about threat and defense. If you don't immediately exploit a tactical opportunity the game will almost certainly turn against you".

Further results of the original study indicate that decisions of experts and advanced players improve when no time pressure is applied compared to time pressure conditions. However, for the highest level of chess masters, applying time pressure did not change the high quality of moves. Importantly, for testing the Take-the-First heuristic, the mean number of generated options was about 3 ($M = 3.11$), consistent with the ill-defined nature of the task. This procedure is well established in problem-solving and creativity research that enables ill-defined tasks and allows for differentiation between better or worse solutions (Johnson & Raab, 2003). Players often chose their first generated option as the best move, and move quality decreased from the first to the second to the third option, supporting intuition-based accounts of chess expertise (e.g., Chassy, Lahaye, & Gobet, 2023).

The limits of the study by Medvegy et al. (2022) include a small ($N = 34$) and selective sample using a between-subject design (three groups of about 10 players with an Elo score above 2000). In addition, move quality seemed to differ between grandmasters' evaluations and different chess engines. For the conceptual replication, we tried to overcome these limitations by testing a larger sample ($N = 34$ in the original study compared to $N = 92$ in the current study). We also chose an expertise level with a mean below an Elo rating of 2000 and a higher variance of expertise, allowing us to test whether a specific expertise threshold exists that separates a less-is-more and a more-is-better effect. Such a strategy has been recently advocated in a special issue on chess in the *Journal of Expertise* arguing that future studies should use a wider range of expertise and manipulate exposure time (e.g. Chassy et al., 2023). Thus, we conceptually replicated the study by Medvegy et al. (2022) with a sample with more variance of expertise that still allowed us to differentiate expertise. In the study by Medvegy et al. (2022), the participants were divided into groups depending on their expertise: an Elo rating under 2,000 ($M = 1,659$, $SD = 284$), over 2,000 ($M = 2,227$, $SD = 105$) and grandmasters over 2,399 ($M = 2,506$, $SD = 71$) whereas we used expertise as a continuous variable.

In addition, we advanced the expertise differences to be sensitive to the manipulations of tactical/strategic and time pressure/no time pressure by using as argued above different ratings: the Bullet online rating, the average Blitz online rating, the average Rapid online rating, the average Classical rating, and the average official Elo FIDE rating (International Chess Federation or World Chess Federation, commonly referred to by its French acronym FIDE). Whereas the average scores (at 6th of January 2024) of rapid rating (641), blitz rating (640) and bullet rating (602) are comparable, online rating may have slightly higher scores than the classical ratings given online additional support (e.g. AI, chess-engines) cannot be controlled for. The variance of expertise, however, required us to rely on comparable scores between participants and accept that, if not for all participants, we have scores for each rating, as not all players play fast and slow chess game tournaments. However, we assumed that this diversity of expertise in preferences for fast

or slow games may be more sensitive to the specific task manipulations such as type of position and time-pressure compared to the average official FIDE rating. At the same time, comparing expertise based on different ratings provides insights into how stable expertise effects are when studies rely on a single metric – such as the classic FIDE Elo rating – for all kind of tasks and situations. Our findings contribute to the ongoing discussion of domain-general versus domain-specific effects in expertise (see a meta-analysis of Kalen et al., 2021).

Taken together, the literature suggests that the debate between less-is-more and more-is-better (Chernev, Böckenholt, & Goodman, 2015) must be addressed empirically and systematically when assessing expertise effects. We therefore used chess as a test bed that allowed us to vary person, task, and situation factors, measuring choice outcomes and decision processes with established paradigms (Johnson & Raab, 2003) and to conceptually replicate previous research (Medvegy et al., 2022).

Hypotheses

As the present study used the same theoretical framework and specified performance and expertise evaluations, we assumed the same effect direction as in the original research by Medvegy et al. (2022). This refers to those players with comparable Elo ratings, serving as a confirmatory analysis. Exploratory analyses focused on whether there is a specific threshold of expertise that would allow us to separate a less-is-more effect (e.g., fewer options generated result in better choices) and an opposite more-is-better effect.

In sum, we predicted that option generation and resulting choices would follow the Take-the-First heuristic (Johnson & Raab, 2003; Raab & Johnson, 2007), such that chess players generate a few options to choose from and pick the first choice they generated, as previously demonstrated in chess and other expert domains (e.g. Gobet & Simon, 1996; Medvegy et al., 2022) Operationally, this means that we expected (a) higher move quality for the first generated option compared to later ones, consistent with prior findings on sequential option generation in experts (Johnson & Raab, 2003; Chassy et al., 2023), (b) a higher likelihood that the first generated option is also selected as the final best choice, and (c) fewer generated options overall with increasing expertise. However, given that option generation is differently developed across the tested variance of expertise, we expect a less-is-more effect mainly in higher expertise and potentially a more-is-better effect in lower expertise. Thus, lower expertise should be associated with generating more options, showing smaller differences in move quality between the first and later options and relying less on the first generated move, in line with theoretical accounts suggesting that intuitive heuristics become effective only at higher levels of domain-specific expertise (Gobet, 1997; Raab & Gigerenzer, 2015; Medvegy et al., 2022). This rationale is based on the assumption that lower expertise results in less trust in the first choice, and thus, the differences in quality of choices between the first generated option and the final decision are moderated by expertise. In addition to heuristic use, we conceptually replicated whether expertise interacts with the task effects of strategic vs. tactical decisions and time pressure vs. no time pressure, as in the original study. Specifically, we expected stronger Take-the-First patterns in tactical situations and under time pressure, reflected in fewer generated moves and smaller differences between first and final choices.

Methods

The design used expertise as a continuous predictor and tested in a regression to what extent expertise explains chess choices. The task variations (tactical vs. strategic and time pressure vs. no time pressure) were used as additional moderators. Chess choices were used as criteria, based on the study by Medvegy et al. (2022), including the average number of options, decision quality based on the position of the option generated, and the number of first-generated options that were ultimately selected as best.

Participants

Participants were recruited from the German Chess Association email list. We deliberately restricted our sample to the lower end of the study by Medvegy et al. (2022) with an average Elo rating of below 2000, to examine the threshold of expertise necessary for the reliable use of the Take-the-First heuristic. G*Power was used to define a sample size for testing chess choices. We used the supplementary data by Medvegy et al. (2022) to calculate the effect size ($f = .31$) for the main effects of expertise and predictions for the Take-the-First heuristic on the number of generated options. This yielded a recommended sample size of 82 participants for a conservative a priori ANOVA (LeBel et al., 2017). We recruited 92 players to compensate for potential drop-outs.

A total of 92 chess players took part in the study (see Table 1), of whom 84 (91.3%) were male and 8 (8.7%) were female. Their average age was $M = 31.67$ years, 95% CI [28.42, 34.93], $SD = 15.71$, with ages ranging from 9 to 72 years. The performance of players was based on chess performance scores across different game types: (I) Bullet, in which each player has one to two minutes for the whole game; (ii) Blitz, which ranges from three to five minutes per player. (iii) Rapid, which gives each player 20 to 30 minutes for a game; and (iv) Classical, which is most used in tournaments, and where each player has about one to three hours for one game (depending on age and with additional time settings such as a 30-second increment per move).

If participants had an official FIDE rating, they could enter that, too ($N = 70$). Ethical approval was obtained from the local university (# 123/2021), and informed consent was obtained from all participants. For chess players under 18, parental consent was collected. Table 1 shows demographic characteristics and rating distributions.

Table 1. Demographic characteristics of the sample ($N = 92$)

Variable	Percentages and Mean Values (with Standard Deviations)
Gender	
Male	91.3 %
Female	8.7 %
Age	31.67 (15.71)
Expertise Ratings	
Bullet Rating ($N = 70$)	2018.60 (367.17)
Blitz Rating ($N = 87$)	2121.38 (294.38)
Rapid Rating ($N = 66$)	2089.58 (252.91)
Classical Rating ($N = 41$)	2006.61 (236.19)

Materials and Chess Task

We used two tactical and two strategic positions as displayed in Figure 1. The positions were selected and categorized by two grandmasters who additionally work as chess coaches. The positions were previously used in the study of Medvegy et al. (2022) and taken for reasons of comparability. It can be argued that the ambiguity of the positions, a prerequisite for option-generation paradigms in problem-solving and creativity research, varies depending on the number of strong possible moves in those positions. For instance, in Figure 1, position 2 has six good alternative choices, whereas other positions have fewer alternatives. Likewise, it can be argued that position 4 in Figure 1 could be assigned to both a tactical and a strategic choice; thus, participants' choices may depend on their individual mindset. However, for a conceptual replication, we needed to ensure a standard for comparison, and participants were not informed about the type of position, how they should generate options, or how many they should generate, in order not to influence option generation and to keep instructions identical to the original study. As in the original study, the tactical positions were presented first, followed by the strategic positions, and participants were instructed to treat each position as independent.
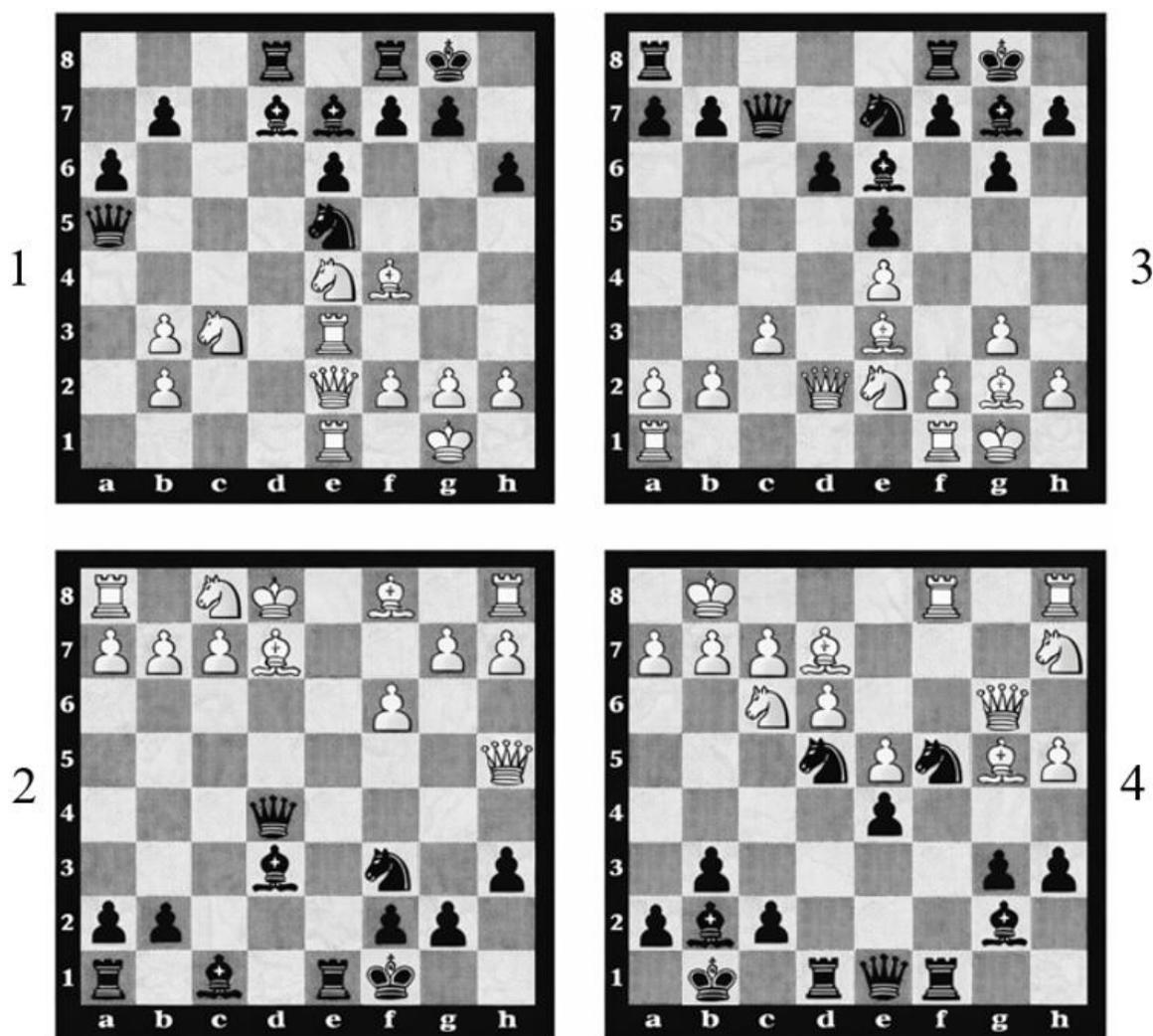
Figure 1. Chess positions presented to participants. White to move in positions 1 and 3; Black to move in positions 2 and 4.

Note. Tactical positions are on the left, strategic positions are on the right. Based on the study by Medvegy et al. (2022), the best move for position 1 is 1. b4 (pawn on square b3 moves to b4), and for position 2 is 1. Bb4 (bishop from f8 to b4), for position 3 is 1. Rfc1 (rook from f1 to c1) and for position 4 is 1. g5 (pawn from g6 to g5).

Procedure

The study was programmed and conducted using LimeSurvey (LimeSurvey GmbH, 2003) and was available in both German and English. Participants first had to fill out some demographic data, as well as all their chess ratings they held. After a detailed instruction that emphasized solving the positions without any human or artificial assistance, participants were provided with the four chess positions described above.

For each of the positions, we followed the procedure of Medvegy et al. (2022) in which

players first had 15 seconds to choose the best move. Then, they had 45 seconds to enter additional candidate moves, and after that, an additional four minutes to calculate the consequences of some of the options they had chosen by simulating possible opponents' responses and deciding on a final best move. Subsequently, they were given a Likert scale for a subjective evaluation of their best move. After finishing this procedure for the first position, they proceeded to the next chess position (as defined in order by Figure 1). We replicated the same order as in the original study. Given the

instruction that positions should be treated independently, no sequence effects are assumed or any fatigue effects, given the low number of trials and a task that participants enjoy daily.

Measures

To test the use of the Take-the-First heuristic, we assessed several dependent variables:

1. **Number of alternative moves generated**
    For each position, we recorded how many candidate moves participants entered during the option-generation phase (45 seconds). This measure reflects the quantity of generated options and serves as the key indicator of less-is-more versus more-is-better effects.

2. **First-best choice consistency**
    We coded whether the first generated move was also selected as the final best move after the calculation phase. In decision research that is often called dynamical inconsistency (Johnson & Raab, 2003). This served as an indicator of reliance on the first option and reflects the core prediction of the Take-the-First heuristic.

3. **Number of optimal moves generated**
    Based on expert ratings and Stockfish evaluations, we coded how often participants generated a move that was classified as optimal for the given position.

4. **Quality of moves**
    The quality of the first, second, and third generated options was evaluated using expert ratings (for comparison, we used the same solutions as Medvegy et al., 2022) and the Stockfish chess engine, including the NNUE (neural network extension) evaluation, as done previously in the study by Medvegy et al. (2022). As in the original study, in the conceptual replication, both the chess engine and the participants indicated that the positions are indeed ill-defined, in the sense that multiple good options can be generated, but that these options differ in move quality.

5. **Subjective evaluations of the final (best) move**
    For subjective move evaluation, we used four items ranging on a scale from -4 to 4,

on which participants had to assess the purpose of their best move. The four items were: (1) Whether their preferred move was tactical or strategic, (2) Whether the aim was short-term or long-term, (3) Whether the thinking was intuitive or deliberate (the phrase "well thought through" was used to avoid linguistic barriers, as the word "intuitive" is far more common than "deliberate"), and (4) Whether the aim of the move is rather defensive or active.

This part of the survey was important because strong moves can result from different intentions. Considering both the type and the intention of a move allowed us to differentiate levels of expertise and validate our manipulations. For instance, a move can be tactically motivated but still have a long-term objective: positioning the rook and queen on the same file, with pieces in between, prepares a long-term attack along that file while simultaneously addressing a short-term threat, such as an attack on the moved queen. This also constitutes an example of a tactical defensive move; a simpler example would be avoiding checkmate. Conversely, strategic moves can also serve a short-term purpose—for instance, preventing a pawn from breaking through by occupying the relevant square with one's piece. Whether such a move is active or defensive depends on the prior position of the piece. Furthermore, the rating scales may reveal metacognitive evaluations of the positions. According to Blanch et al. (2020), expert chess players adapt their search strategies depending on the type of position, whereas less experienced players tend not to do so.

Expertise differentiation was based on different player ratings that reflect their strength in chess, such as the official FIDE rating or online ratings like Blitz or Bullet scores for fast-paced games. It must be noted that online game ratings tend to be significantly more inflated than official ratings such as FIDE. For example, a strong player with an official FIDE rating of 2,300 Elo might have an online rating of 2,800 if

they excel at Bullet games—a rating that would correspond to world-class level if it were official. Consequently, tendencies across online and official ratings are comparable, though the absolute rating values themselves are not directly equivalent. The online ratings were taken from the website *chess.com*, which is the world's largest chess server with over 200 million registered users, of whom approximately 20 million are active.

Statistical Analyses

The statistical analyses were conducted using RStudio 2023.06.0 (Posit team, 2023). To test the hypotheses, we estimated a series of linear regression models in which chess decision variables served as dependent variables and expertise was entered as a continuous predictor. Based on our hypotheses, we included the following dependent variables (a) the number of generated options, (b) move quality of the first generated option, and (c) first-best choice consistency. Expertise was operationalized using the different rating formats (Bullet, Blitz, Rapid, Classical, and FIDE), which were entered separately to account for task-specific expertise effects.

The general form of the regression model was $Y_i = \beta_0 + \beta_1(\text{Expertise}_i) + \varepsilon_i$, where Y_i represents the respective decision outcome for participant i, β_0 the intercept, β_1 the effect of expertise, and ε_i the error term. For analyses including task characteristics, tactical versus strategic position and time pressure versus no time pressure were entered as additional predictors or tested using condition-wise comparisons, consistent with Medvegy et al. (2022).

Model fit was evaluated using the coefficient of determination (R^2 and adjusted R^2), and overall model significance was assessed using F-tests. Regression coefficients were interpreted based on standardized beta weights and corresponding p-values.

To test whether the first-generated option differed in quality from later options, we conducted repeated-measures ANOVAs with the serial position of the generated move (first,

second, third, fourth) as a within-subject factor. This analysis addressed the core prediction of the Take-the-First heuristic. Task-specific and situational effects (tactical vs. strategic positions; time pressure vs. no time pressure) were examined using paired-sample t-tests, consistent with the original study. Additional exploratory analyses assessed potential interactions between expertise and task characteristics.

Together, these analyses directly map onto our hypotheses regarding (a) reliance on the first generated option, (b) expertise-dependent differences in option generation and move quality, and (c) the impact of situational factors such as time pressure.

Results
Chess moves

We investigated whether chess players with an Elo rating below 2,000 rely on the Take-the-First heuristic. The results indicate that they do: on average, they generated $M = 2.2$ options and selected their first generated move as the best option in 37.3% of the cases. However, in contrast to the findings of the original study, the quality of the generated moves did not differ substantially between the first, second, third, and fourth options (see Figure 2).

A direct comparison with the findings of Medvegy et al. (2022) indicates that the move evaluations in our sample were generally lower. This difference appears to be driven by the lower average expertise in the present study. When comparing only our highest-rated participants with the least experienced group in Medvegy et al. (2022), the overall pattern of results is more similar, although the absolute evaluation values remain slightly lower. This suggests that the relationship between expertise and the quality of first and subsequently generated moves replicates in direction, but not fully in magnitude.

Participants also self-evaluated the quality of their chosen moves. On average, their intuitive choice was not the only viable option, as reflected by the number of different moves generated and the variation in selected best

moves across participants. In position 1, a total of 14 different best moves were suggested; in position 2, 19; in position 3, 11; and in position 4, 17. This variability indicates that the task elicited a wide range of plausible, high-quality options, rather than biasing participants toward

selecting their first move as the best. These counts were aggregated across both time pressure and no time pressure conditions. For future replication studies, see details in Table S1 in supplemental materials at the end of this article
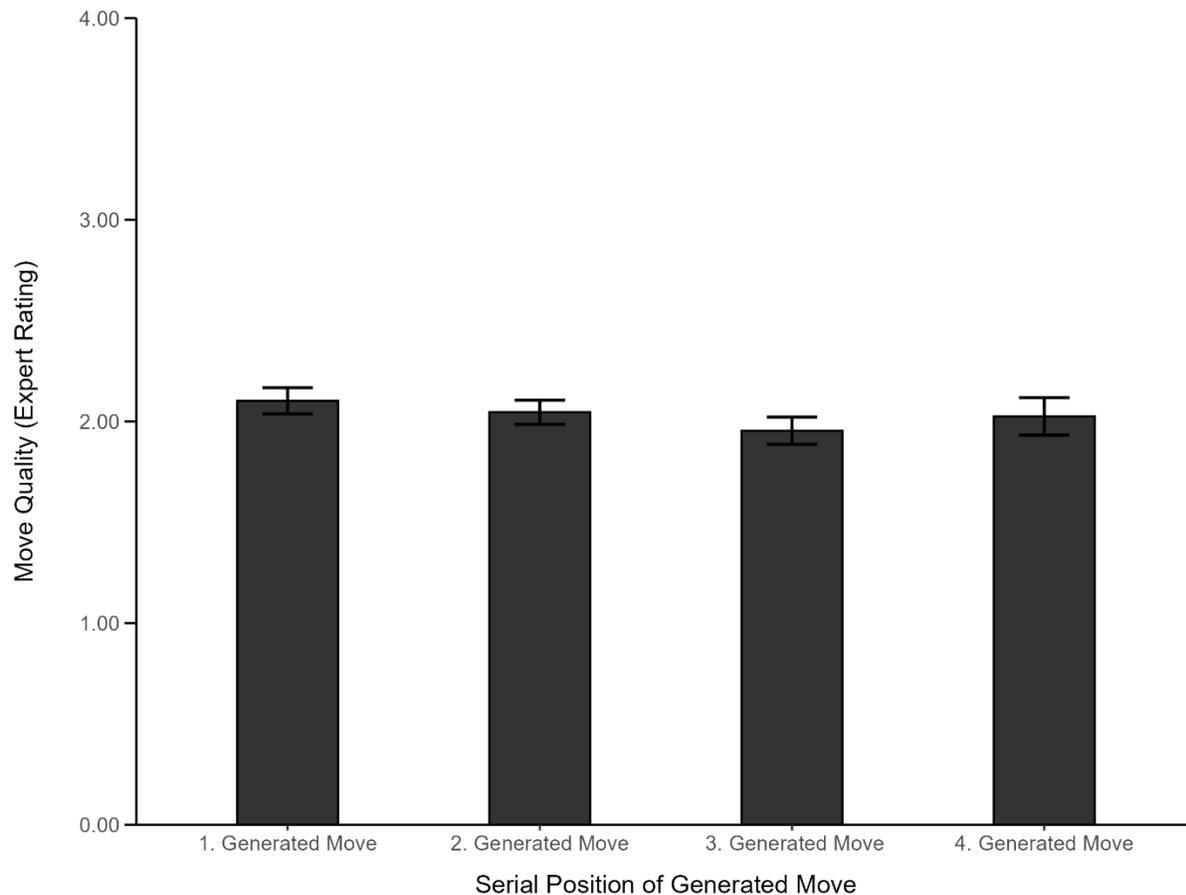
Figure 2. Mean expert ratings of the first four moves generated under time pressure with standard error bars

Note. The moves are shown in the order in which participants generated them: the first move was produced during the initial 15-second intuitive phase, and the subsequent moves were generated during the following 45-second option-generation phase. Only the first four generated options were analyzed because many participants did not produce more than four candidate moves.

Descriptive Data

An important argument for including multiple ratings in the conceptual replication is that different ratings may reflect task-specific performance relative to our factors, such that Blitz ratings may be more sensitive to time pressure conditions, fast tactical choices, or the first option generated. Thus, it is important to evaluate correlations of the ratings themselves. Table 2 (next page) shows correlations between the online ratings and FIDE ratings. The latter

correlates significantly with Bullet ($r = .218$) and Blitz ratings ($r = .421$). With Rapid and Classical, the correlations are non-significant. These results are not surprising, as the ratings of non-experts can differ substantially. For example, a player can be very good in Rapid games but have a low FIDE rating. When testing predictions for the conceptual replication, we nonetheless observed correlations with the different ratings to assess task-specific performance factors, such as our manipulations of time pressure or tactical versus strategic positions.

Table 2. Coherence of ratings among participants

Ratings	Bullet	Blitz	Rapid	Classical
Blitz	.56**[.41, .71]	—		
Rapid	.22*[.05, .47]	.19[.17, .59]	—	
Classical	.05[-.37, .04]	-.14[-.33, .1]	.45**[-.28, .16]	
FIDE	.22*[.02, 41]	.42**[.22, .66]	.05[-.14, .27]	-.08[-.27, .13]

** the correlation is significant on a two-side .01 level; * the correlation is significant on a two-side .05 level

Hypotheses: Expertise and Option Generation

The correlations between expertise ratings and the number of candidate moves are summarized in Table 3. Bullet and Blitz ratings significantly correlated with the number of generated options across all four positions, with stronger associations in tactical positions. Rapid, Classical, and FIDE ratings also correlated with option generation, particularly in the two tactical positions. These results are consistent with the conceptual with the conceptual replication and indicate that expertise predicts the number of options generated across positions. In summary, the ratings appear to reflect both the quality and quantity of choices across all tested positions. Given that correlations are sensitive to the specific sample, we do not place full emphasis on the size of the correlations. However, the significant directions predicted based on the study we conceptually replicated provide reliable information from an independent sample.

Table 3. Correlations between expertise ratings and the number of candidate moves generated across all positions [95% confidence intervals in brackets].

Expertise Rating	Position 1 (Tactical)	Position 2 (Tactical)	Position 3 (Strategic)	Position 4 (Strategic)
Bullet	.44** [.26, .6]	.34** [.11, .53]	.34** [.10, .53]	.28* [.04, .47]
Blitz	.35** [.18, .51]	.33** [.15, .49]	.23* [.03, .42]	.25* [.06, .42]
Rapid	.39** [.2, .54]	.42** [.27, .57]	.14 (ns)	.28* [.09, .45]
Classical	.36* [.11, .58]	.46** [.23, .65]	ns	ns
FIDE	.25* [.07, .41]	.30** [.14, .46]	ns	ns

** the correlation is significant on a two-side .01 level; * the correlation is significant on a two-side .05 level; ns = non significant

Hypotheses: Expertise and Task-Dependent Option Generation and Quality of Choice

Take-the-First Heuristic

The first generated choice often outperforms subsequently generated options in an option-generation paradigm (Raab & Johnson, 2007). In this study, we conceptually replicated the findings of Medvegy et al. (2022) to investigate whether this effect also holds for participants with lower Elo ratings than those in the original study (see Figure 2). A repeated-measures ANOVA revealed no significant effect of the serial position of generated moves on move quality as rated by experts, $F(3,$ 992) = 1.73, p = .159. In contrast to Medvegy et al. (2022), we found that participants with lower Elo ratings did not show a clear advantage of the first generated option over later ones. Although descriptively, the first option received higher ratings than the later ones, this effect was not statistically significant despite the larger sample size. This supports the notion that the advantages of Take-the-First heuristic and intuitive decision-making become evident only at higher levels of expertise. Regardless of descriptive trends or statistical significance, we found that participants' ratings were correlated with move quality. For

instance, the Blitz rating significantly correlated with move quality in position 1, $r = .356**$, 95% BCa [.134, .542], and position 3, $r = .486**$, 95% BCa [.280, .644], which may indicate that Blitz performance is related to making good initial choices. However, since the official FIDE rating also significantly correlated with position 1, $r = .414**$, 95% BCa [.252, .567], and position 3, $r = .316*$, 95% BCa [.093, .516], task-specific effects cannot be ruled out. Moreover, because FIDE and Blitz ratings themselves are correlated, they are not fully independent (see Table 2).

We used simple t-tests to evaluate situational and task-specific effects (see Table 4). Significant differences emerged, particularly between time pressure conditions in tactical and strategic moves. This effect was present across multiple dependent variables, including the number of generated moves, the number of first generated moves that were ultimately selected as best, the number of best-generated moves based on expert ratings, the quality of the first move, and the subjective evaluations.

Table 4. Mean values of dependent variables (standard deviation in brackets) and t-test results for comparison of tactical and strategic positions

Variable	All Positions	Tactical Positions	Strategic Positions	t	p	d [95% CI]
Number of Alternative Moves Generated						
Time Pressure	1.89 (1.40)	1.74 (1.44)	2.03 (1.34)	1.99	.048	0.21 [0.00, 0.41]
No Time Pressure	0.53 (1.02)	0.55 (1.09)	0.51 (0.95)	0.41	.684	-0.04 [-0.25, 0.16]
Number of Identical First Moves under Time Pressure vs. Number of Optimal Moves Generated without Time Pressure	0.38 (0.49)	0.28 (0.45)	0.47 (0.5)	3.76	< .001	0.40 [0.19, 0.61]
Time Pressure	0.07 (0.26)	0.14 (0.35)	0.01 (0.07)	5.16	< .001	-0.54 [-0.75, -0.33]
No Time Pressure	0.13 (0.34)	0.21 (0.41)	0.05 (0.21)	4.68	< .001	-0.50 [-0.71, -0.29]
Quality of First Move under Time Pressure						
Expert Rating	2.32 (1.38)	2.22 (1.68)	2.41 (0.99)	1.29	.2	0.13 [-0.07, 0.34]
Stockfish Rating	-1.14 (2.18)	-1.80 (2.61)	-0.51 (1.41)	5.66	< .001	0.62 [0.40, 0.84]
Quality of First Move without Time Pressure						
Expert Rating	2.93 (1.29)	3.11 (1.52)	2.75 (0.98)	2.59	.01	-0.28 [-0.49, -0.07]
Stockfish Rating	-0.60 (1.82)	-1.05 (2.27)	-0.14 (1.03)	4.82	< .001	0.52 [0.30, 0.73]
Subjective Evaluation of the Final Move						
Tactical vs. Strategic	-0.13 (2.94)	-1.23 (2.84)	0.97 (2.61)	7.75	< .001	0.81 [0.59, 1.02]
Short-term vs. Long-term	-0.19 (2.72)	-1.27 (2.56)	0.89 (2.45)	8.29	< .001	0.86 [0.65, 1.08]
Intuitive vs. Deliberate	-0.67 (2.48)	-0.92 (2.5)	-0.43 (2.44)	1.9	.059	0.20 [-0.01, 0.40]
Defensive vs. Active	1.90 (2.09)	2.66 (1.73)	1.14 (2.14)	7.53	< .001	-0.79 [-1.00, -0.57]

Differences in Move Quality Dependent on Time Condition and Task

In the study by Medvegy et al. (2022), differences in move quality were interpreted as task-specific effects, as less-is-more effects should result in better move quality under time pressure or when tactical decisions are required (see Figure 3). However, in our data, move quality in tactical decisions *without* time pressure was significantly higher than in strategic decisions (see Table 4). Under time pressure, the first moves were better in strategic decisions, but only when move quality was evaluated using Stockfish ratings, not when using expert ratings as in Medvegy et al. (2022; see Table 4).

When we additionally calculated quality differences by subtracting the move quality under time pressure from the move quality with sufficient time, no significant effects emerged. This indicates that the findings of good first choices in the given task cannot be attributed solely to time pressure or tactical demands (see supplemental material, Table S2, for additional analyses). A further set of correlations between the factors indicates that the behavior of participants is correlated (see Table 5, next page).
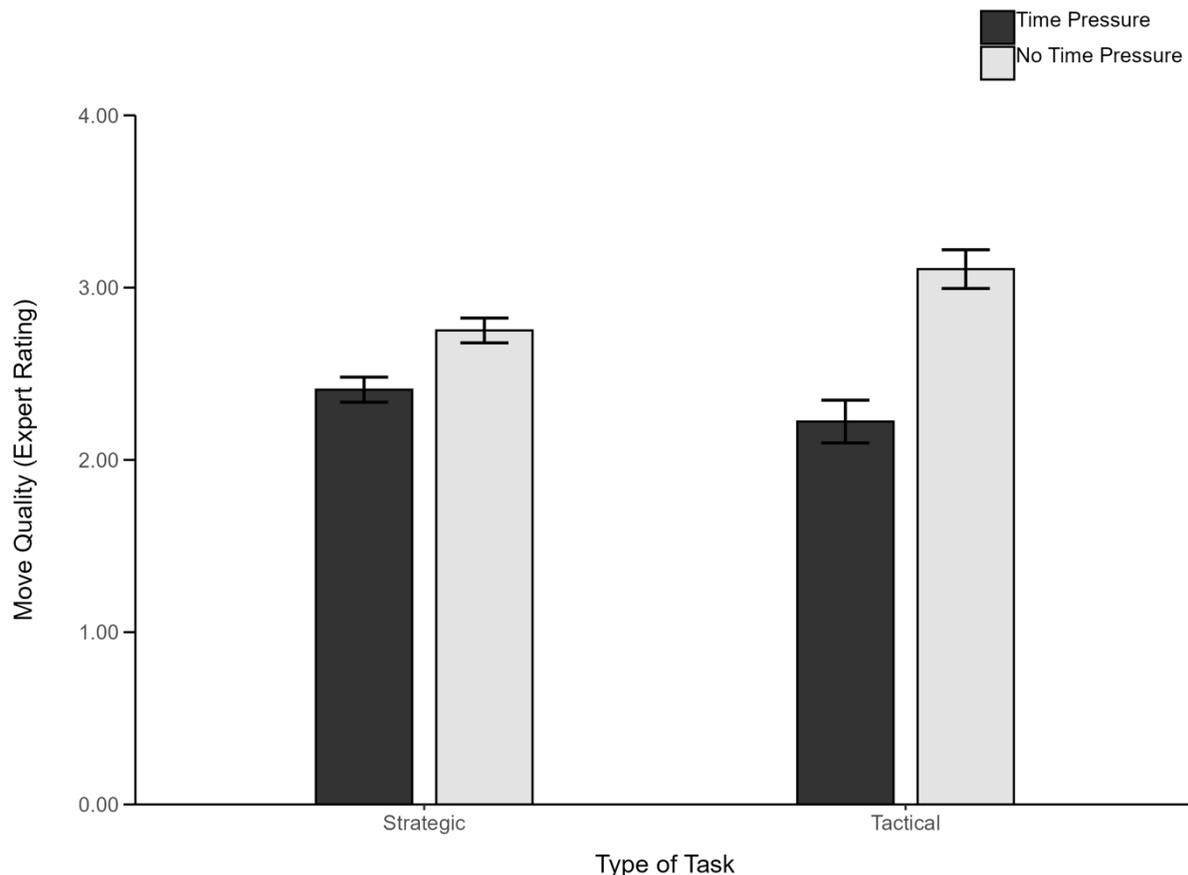
Figure 3. Descriptive values for two-way interaction effects of task and time condition for move quality with standard error bars

Table 5. Correlations between subjectively evaluated dimensions with *p*-values

Variable	1. Tactical vs. Strategic	2. Short-term vs. Long-term	3. Intuitive vs. Deliberate	4. Defensive vs. Active
1		.70 [.64, .75] *p* < .001	.37 [.27, .45] *p* < .001	-.37 [-.45, -.28] *p* < .001
2			.33 [.24, .42] *p* < .001	-.35[-.43, -.25] *p* < .001
3				-.07 [-.17, .04] *p* = .198

Note. Values in square brackets indicate the 95% confidence interval for each correlation.

In Table 5, tactical and strategic as well as short-term or long-term choices correlate highly as the constructs are related, as introduced above. Importantly, the choice mode of intuitive and deliberative choices is related to short-term and long-term goals as indicated by the quote of chess players. Not surprisingly, the only non-significant relation is between the decision mode (intuitive, deliberative) and the strategy of playing defensively or actively as this seems to be independent.

Discussion

The goal of the current study was to examine how decision makers leverage their intuition in strategic and tactical situations by conceptually replicating the study by Medvegy et al. (2022). We used the replication strategy to address some limitations and provide additional choice measures and expertise differentiations. We compared the decision-making process of chess players in tactical and strategic game situations, both with and without time pressure. The main novelty of our study is that we replicated findings in an independent, larger sample by assessing intuitive choices among players with an Elo rating below 2000 and with different expertise ratings depending on the speed of the games. In the following, we discuss our findings.

The conceptual replication provides new insights into different expert rating procedures in chess and into the relationship between tactical and strategic moves. When comparing our results directly with those of Medvegy et al. (2022), it becomes evident that the overall quality evaluations in our sample were lower. This is expected given that our participants' expertise levels were substantially below those of the original sample. However, when restricting the comparison to the most expert players in our sample and the least expert group in Medvegy et al. (2022), the patterns regarding the serial position effect and reliance on the first move appear broadly consistent, suggesting that the Take-the-First mechanism generalizes across adjacent expertise ranges. We substantiate this by the following reasoning.

First, the correlation coefficients between expertise ratings and the number or quality of candidate moves were largely consistent and appeared to be independent of the specific rating category. In addition, ratings from the same person across different formats—Classical and fast games—were significantly correlated. We conclude that the method of assessing expertise, whether through different chess experts or rating classifications, may not be as critical as previously assumed. Within the online ratings, both fast and slow time controls showed significant correlations with each other. The official rating was mostly correlated with Bullet and Blitz ratings. This may be due to the fact that many players prefer fast games online, which leads them to play more games in these formats – making their ratings in these categories more reliable.

Second, the correlations between the subjectively evaluated dimensions tactical – strategic and short-term – long-term were high across all positions. A possible explanation is that chess players do not evaluate positions purely tactically or strategically, but rather consider their interdependencies—for example, pursuing a short-term tactical advantage as part of a broader strategic plan. Contrary to our predictions, the correlations between the dimensions tactical – strategical and intuitive – deliberate were also positive rather than negative.

Why do we need such a conceptual replication? An important reason for the replication was to investigate whether the effect could be observed in a larger sample with lower expertise compared to the original study. The results indicate that the quality of moves—whether generated as the first, second, third, or fourth option—did not differ as much as in highly expert players such as grandmasters. This finding supports the idea that Take-the-First decisions require a high level of expertise, as has also been shown in other domains (Raab & Johnson, 2007). However, based on our hypotheses, we can assume that players affiliated with a national chess association—given their Elo ratings—might still demonstrate a less-is-more effect and show that first moves are better than subsequently generated options when moves are produced sequentially.

Importantly, tactical and strategic choices—both in terms of the number of generated moves and their quality—differ depending on time pressure. This provides an important leverage point for designing valid paradigms to assess decision-making in chess. Another novel finding is that the stronger the player (as indicated by their Elo rating), the smaller the difference in move quality between conditions with and without time pressure. This pattern was observed in our sample with lower Elo ratings and stands in contrast to the grandmaster and master samples in Medvegy et al. (2022). Overall, these results may stimulate further discussion on the nature of expertise in chess. Our findings challenge the idea of generalizable expertise by highlighting context specificity:

some participants may be more familiar with particular positions or structures—e.g., due to their preferred opening repertoire (see also Campitelli & Gobet, 2004; Campitelli, Gobet & Bilalić, 2014).

Limitations and Future Research

What counts as a good move? A strength—and at the same time a weakness—of understanding expert choices in real and complex situations lies in the assessment of choice quality, as has been evident since the early research in chess (de Groot, 1965). It is a strength because expert choices exhibit high external validity. However, it is also a weakness, as tasks like chess make it difficult to assess the quality of moves—especially when made by grandmasters—without knowing the reasoning behind them, such as whether a move serves a short-term or long-term purpose. Our study showed that comparing participants' moves to computer-generated moves, to the evaluations in Medvegy et al. (2022), or to the original grandmasters' choices in the source games leads to different benchmarks for what counts as a "best" move.

How to test expert chess players? Defining experts broadly as general chess players—i.e., domain-specific experts—may not be sufficiently sensitive when task designs vary, for example, between tactical and strategic situations or between Blitz and classical formats. Our approach of using different domain-specific but task-sensitive ratings offers a promising path forward for capturing the nuances of expertise across varying contexts. The ambiguity of positions – a prerequisite for option-generation paradigms in problem-solving and creativity research – varies depending on the number of strong possible moves in a position. For example, some positions offer multiple good alternatives, whereas others provide fewer; some can be classified as both tactical and strategic. Such ambiguity may contribute to differences in participants' choices and interacts with individual mindsets, as seen in our Methods section and we conclude that testing expert chess players may include multiple of those situational and task conditions to be representative of their expertise.

Moreover, moves should be directly conducted and saved on the board in future studies. We consulted a computer scientist specializing in chess programming to explore technical improvements. A novel observation was that some participants added evaluations to their variations (such as +=, !, etc., commonly used in annotated chess games) or verbally commented on their ideas. Such data could be incorporated in follow-up studies to systematically examine pattern recognition processes via eye-tracking or metacognitive thinking through chess notation in lab experiments.

How to sample chess players? Differences in the sampling procedure, however, may require additional considerations in future research. For instance, some participants mentioned in the debriefing that the time allotted was too short to adequately calculate and enter variations. This might be due to the fact that, on average, the chess players in our sample had a lower Elo rating than those in the original study by Medvegy et al. (2022). Future studies should control more rigorously for expertise, experience, and age differences by applying inclusion and exclusion criteria in combination with the specific ratings and choice quality assessments we introduced to advance the current state of the art.

When do experts trust their intuition? Our findings suggest that a specific level of expertise is needed to rely on intuition and thus the less-is-more effects, as reflected in the Take-the-First heuristic, is beneficial for performance. Naturally, not all players fully trust their intuition and thus interindividual differences in preferences for intuitive versus deliberative decision-making could be assessed in future work (cf. Betsch, 2004, for a relevant scale). Beyond characteristics of the person, uncertainty of choices, the number of possible and equally appropriate solutions in a position may explain differences between the first generated move and the final chosen one. Anecdotally, it was often reported—especially by stronger players—that participants felt uncertain about their choices. This is an intriguing aspect, as it is known that players tend to opt for forcing moves when unsure how

to proceed. This tendency is supported by the most frequently chosen moves in our study and could be further investigated in future research.

Conclusions

In our study, we examined how decision makers leverage their intuition in both strategic and tactical situations. The results suggest that at lower levels of expertise, intuition is not as effective as it is for grandmasters, and players require more time to calculate variations. However, we found that experts' intuitive choices were not worse than their decisions after careful deliberation which becomes important when time is limited. We also analyzed decisions separately for strategic and tactical positions. According to our findings, lower-skilled chess players are more vulnerable in tactical situations than in strategic ones, especially under time pressure. Extending previous research on athletes (Hepler & Feltz, 2012; Johnson & Raab, 2003; Raab & Johnson, 2007) to chess players, we reached a similar conclusion: an expert's first thought—the first generated option (Take-the-First heuristic)—is, in most cases, a good choice.

Acknowledgements

Authors' Declarations

in accordance with the Ethical Principles of the *Journal of Expertise.*

The authors declare that data can be provided on request.

ORCID iDs

Markus Raab
https://orcid.org/0000-0001-6546-1666

Zoltán Medvegy
https://orcid.org/0000-0001-8187-0451

Ivanna Korobeinikova
https://orcid.org/0000-0003-1662-8006

Lesia Korobeinikova
https://orcid.org/0000-0001-8648-316X

Georgiy Korobeynikov
https://orcid.org/0000-0002-1097-4787

Elisa Weinand
https://orcid.org/0009-0006-5005-4973

References

Belling, P. K., Suss, J., & Ward, P. (2015). Advancing theory and application of cognitive research in sport: Using representative tasks to explain and predict skilled anticipation, decision-making, and option-generation behavior. *Psychology of Sport and Exercise, 16*, 45–59. https://doi.org/10.1016/j.psychsport.2014.08.001

Betsch, C. (2004). Präferenz für Intuition und Deliberation (PID): Inventar zur Erfassung von affekt- und kognitionsbasiertem Entscheiden [Preference for Intuition and Deliberation (PID): An Inventory for Assessing Affect- and Cognition-Based Decision-Making]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 25*(4), 179–197. https://doi.org/10.1024/0170-1789.25.4.179

Blanch, A., Ayats, A., & Cornado, M. P. (2020). Slow and fast chess performance across three expert levels. *Psychology of Sport and Exercise, 50,* Article 101749. https://doi.org/10.1016/j.psychsport.2020.101749

Butler, A. B., & Scherer, L. L. (1997). The effects of elicitation aids, knowledge, and problem content on option quantity and quality. *Organizational Behavior and Human Decision Processes*, 72, 184–202. https://doi.org/10.1006/obhd.1997.2737

Campitelli, G., & Gobet, F. (2004). Adaptive expert decision making: Skilled chess players search more and deeper *ICGA Journal, 27(4.),* 209–216.

Campitelli, G., Gobet, F., & Bilalić, M. (2014). Cognitive processes and development of chess genius: An integrative approach. In D. K. Simonton (Ed.), *The Handbook of Genius* (pp. 350-374). New York: Wiley-Blackwell.

Chassy, P., Lahaye, R., Didierjean, A. *et al.* (2023). Intuition in chess: a study with world-class players. *Psychological Research* 87, 2380–2389. https://doi.org/10.1007/s00426-023-01823-x

Chassy, P., Lahaye, R., & Gobet, F. (2023). Templates but not emotions facilitate the information flow between long-term and working memory: A Sternberg study with chess experts. *Journal of Expertise, 6,* 37–55.

Chernev, A., Böckenholt, U., &, Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis, *Journal of Consumer Psychology, 25*(2), 333–358. https://doi.org/10.1016/j.jcps.2014.08.002.

de Groot, A. D. (1965). *Thought and choice in chess*. Mouton. (Original work published 1946)

Del Missier, F., Visentini, M., & Mäntylä, T. (2015). Option generation in decision making: Ideation beyond memory retrieval. *Frontiers in Psychology*, *5*, Article 1584. https://doi.org/10.3389/fpsyg.2014.01584

Elo, A. (1978). *The rating of chess players, past and present*. Batsford.

Gigerenzer, G. (2008). *Gut feelings: Short cuts to better decision making*. Penguin Books.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gobet, F. (1997). A pattern-recognition theory of search in expert problem solving. *Thinking & Reasoning*, *3*(4), 291–313. https://doi.org/10.1080/135467897394301

Gobet, F. (2018). *The psychology of chess.* New York: Routledge. https://doi.org/10.4324/9781315441887

Gobet, F., & Simon, H. A. (1996). Recall of random and distorted chess positions: Implications for the theory of expertise. *Memory & Cognition, 24,* 493–503. https://doi.org/10.3758/BF03200937

Hepler, T. J., & Feltz, D. L. (2012). Take the first heuristic, self-efficacy, and decision-making in sport. *Journal of Experimental Psychology: Applied, 18*(2), 154–161. https://doi.org/10.1037/a0027807

Johnson, J. G., & Raab, M. (2003). Take the first: Option-generation and resulting choice. *Organizational Behavior and Human Decision Processes, 91*, 215–229. https://doi.org/10.1016/S0749-5978(03)00027-X

Kalén, A., Bisagno, E., Musculus, L., Raab, M., Pérez-Ferreirós, A., Williams, A. M., Araújo, D., Lindwall, M., & Ivarsson, A. (2021). The role of domain-specific and domain-general cognitive functions and skills in sports performance: A meta-analysis. *Psychological Bulletin, 147*(12), 1290–1308. https://doi.org/10.1037/bul0000355

Kasparov, G. (2007). *How life imitates chess.* W. Heinemann.

Klatt, S., Noël, B., Musculus, L., Werner, K., Laborde, S., Lopes, M. C., Greco, P. J., Memmert, D., & Raab, M. (2019). Creative and Intuitive Decision-Making Processes: A Comparison of Brazilian and German Soccer Coaches and Players. *Research Quarterly for Exercise and Sport, 90*(4), 651–665. https://doi.org/10.1080/02701367.2019.1642994

Klein, G., Wolf, S., Militello, L., & Zsambok, C. (1995). Characteristics of skilled option generation in chess. *Organizational Behavior and Human Decision Processes, 62*(1), 63–69. https://doi.org/10.1006/obhd.1995.1031

Keymer, V. (2023, January 5). [Interview]. *Die Zeit*, p. 26.

LeBel, E. P., Campbell, L., & Loving, T. J. (2017). Benefits of open and high-powered research outweigh costs. *Journal of Personality and Social Psychology*, *113*(2), 230–243. https://doi.org/10.1037/pspi0000049

LimeSurvey GmbH, Hamburg, Germany. URL http://www.limesurvey.org

Medvegy, Z., Raab, M., Tóth, K., Csurilla, G., & Sterbenz, T. (2022). When do expert decision makers trust their intuition? *Applied Cognitive Psychology*, 36(4), 748–757. https://doi.org/10.1002/acp.3958

Moxley, J. H., Ericsson, K. A., Charness, N., & Krampe, R. T. (2012). The role of intuition and deliberative thinking in experts' superior tactical decision-making. *Cognition, 124,* 72–78.

Musculus, L., Ruggeri, A., Raab, M., & Lobinger, B (2019). A developmental perspective on option generation and selection. *Developmental Psychology*, *55*(4), 745–753. https://doi.org/10.1037/dev0000665

Nordgren, L., & Dijksterhuis, A. (2009). The devil is in the deliberation: Thinking too much reduces preference consistency. *Journal of Consumer Research*, *36*, 39–46. https://doi.org/10.1086/596306

Posit team. (2023). RStudio: Integrated Development Environment for R [Software]. Posit Software, PBC. http://www.posit.co/

Raab, M., & Gigerenzer, G. (2015). The power of simplicity: A fast-and-frugal heuristics approach to performance science. *Frontiers in Psychology, 6,* Article 1672. https://doi.org/10.3389/fpsyg.2015.01672

Raab, M., & Johnson, J. G. (2007). Expertise-based differences in search and option generation strategies. *Journal of Experimental Psychology: Applied, 13,* 158–170. https://doi.org/10.1037/1076-898X.13.3.158

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, *37*(3), 409–425. https://doi.org/10.1086/651235

Supplement

Table S1. Best moves for each position as proposed by Stockfish 13+, including move quality evaluation by the engine.

| Position 1		Position 2		Position 3		Position 4	
Move	Eval.	Move	Eval.	Move	Eval.	Move	Eval.
b4	1.6 *[5]*	Bd7	1.4	Rfc1	0.5 *[5]*	g5	1.4 *[5]*
Bxe5	-3.8	Bb4	0.8 *[5]*	b3	0 *[2]*	**a5**	0.1 *[3]*
Qh5	-1.1 *[4]*	**Bc5**	1.0 *[4]*	c4	0 *[4]*	Nxf2	0.9 *[4]*
Rg3	-2.9 *[3]*	Bd6	1.4	**Bh6**	-0.9 *[3]*	Qd7	0 *[3]*
Nd6	-4.0 *[2]*	Nd4	0 *[3]*	Rfd1	-0.2 *[3]*	Qc7	0
Nf6+	-3.4 *[2]*	Nb4	-3.1	Bg5	-0.2 *[3]*	h5	0.4 *[3]*
Nc5	-3.1	Bf5	0.8	Bf3	-0.8	Bh6	0
Nd2	-3.9	Kb8	1.0	h4	-0.1	Rc7	0
Ng3	-5.4	Bd5	0.3 *[2]*	f4	-0.2 *[2]*	Ncd6	-1.2 *[2]*
Bg3	-4.0	Re8	0.8 *[2]*	Rad1	-0.4 *[2]*	Ned6	-0.2 *[2]*
Bxh6	-4.2	Rd1+	-6.6	h3	-0.4	b5	-0.3 *[2]*
Qd3	-9.9	Bb3	-5.1			Ncd2	-6.2
Rd1	-4.5	Bh3	-4.4			Ned2	-5.3
Ng5	-5.9	Ba3	-3.0			Bf8	-0.5
		Qxe2	-7.0			f5	-0.1
		Bg4	-5.3			Ng5	0
		b5	-3.9			Na5	-5.6
		Rd4	-6.0				
		Qd5	-0.1				

Note. Most frequently chosen moves are marked in bold. The first row displays the strongest moves according to the engine. Pawn has no abbreviation and just the space moving to is used. K for king, Q for queen, R for rook, B for bishop, and N for knight. The multiplication sign (×) is used when an opponent's figure is captured. If the opponent's king is in check, the symbol "+" is used. The move evaluation by Medvegy et al. (2022) is included in [brackets]. See more on abbreviations for chess pieces and position on the board: https://en.wikipedia.org/wiki/Algebraic_notation_(chess). Note that the Stockfish score is expressed as a positive number for white and a negative number for black. The higher the score, the greater the advantage.

The strongest moves are congruent with the study of Medvegy et al. (2022). Furthermore, the engine rates alternative options differently, which indicates why choice accuracy in ambiguous situations requires both the use of human and artificial experts. In the second position, Medvegy et al. (2022) also rated h5 and g5 with 2 points, in the third position Rac1 with 2 points, and in the fourth position Qf6 with 3 points. Im the current sample, no one chose one of these as their best move; therefore, they are not included in Table S1.

Table S2. Summary of linear regression results for quality of first moves dependent on time condition

Dependent Variable	Predictor	Coefficient (β)	Standard Error (SE)	t	p	R^2	Adj. R^2
Time Pressure Expert Rating	Intercept	-0.3993	1.3241	0.30	.764	0.05	0.005
	Bullet Rating	-0.0003	0.0006	0.42	.674		
	Blitz Rating	0.0008	0.0013	0.64	.523		
	Rapid Rating	0.0003	0.0009	0.29	.776		
	Classical Rating	0.0006	0.0010	0.57	.571		
	FIDE Rating	-0.0001	0.0008	0.15	.880		
Time Pressure Stockfish Rating	Intercept	-3.6427	2.3744	1.53	.128	0.05	0.003
	Bullet Rating	0.0003	0.0010	0.28	.777		
	Blitz Rating	0.0027	0.0022	1.20	.234		
	Rapid Rating	-0.0005	0.0016	0.32	.750		
	Classical Rating	-0.0009	0.0018	0.50	.616		
	FIDE Rating	-0.0005	0.0014	0.36	.722		
No Time Pressure Expert Rating	Intercept	0.3561	1.4887	0.24	.812	0.07	0.026
	Bullet Rating	-0.0012	0.0007	1.91	.060		
	Blitz Rating	0.0008	0.0014	0.57	.570		
	Rapid Rating	0.0004	0.0010	0.44	.658		
	Classical Rating	0.0023	0.0011	2.17	.032 *		
	FIDE Rating	-0.0011	0.0009	1.30	.198		
No Time Pressure Stockfish Rating	Intercept	-2.7970	1.9065	1.47	.146	0.05	0.003
	Bullet Rating	-0.0018	0.0009	2.21	.030 *		
	Blitz Rating	0.0021	0.0018	1.15	.254		
	Rapid Rating	0.0012	0.0012	0.99	.326		
	Classical Rating	0.0004	0.0014	0.32	.751		
	FIDE Rating	0.0009	0.0011	-0.85	.397		

* The t-value is significant on a two-side .05 level.
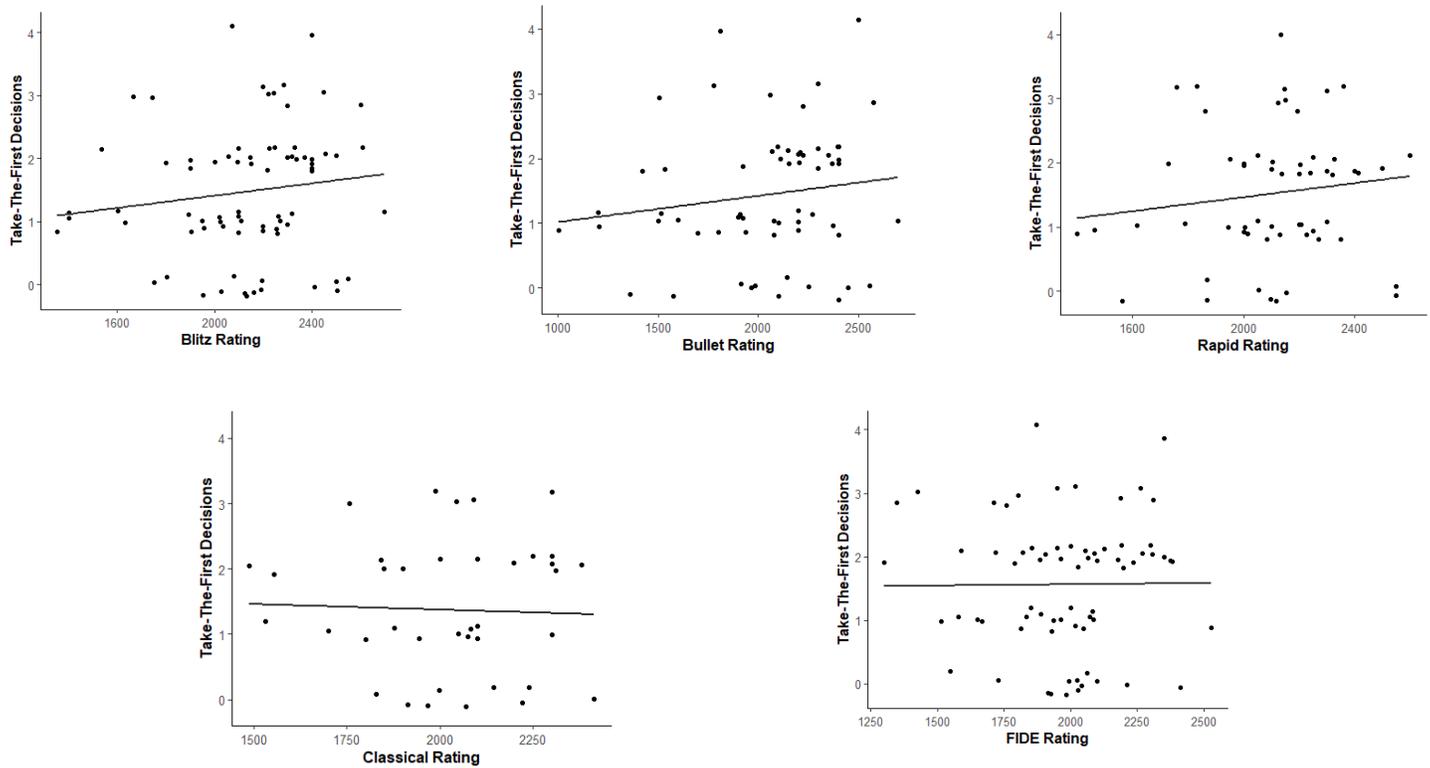
Figure S1. Correlations between Take-the-First Decisions and different ratings. Note that the fast games (Bullet, Blitz and Rapid ratings) indicate a positive relation to Take-the-First decisions whereas for classical and FIDE this relation does not exist.