Recognition of expertise in chess players

ROBERT I. REYNOLDS
Yeshiva University, New York City

What kinds of information facilitate the identification of expert performance? How well can chess players of different performance levels identify the level of players who produced a chess position, or the moves leading to it? Fifteen chess players with U.S. Chess Federation (USCF) ratings from 1300 to 2210 judged six unfamiliar chess positions taken from games between players with USCF ratings from 1400 to 2600. The moves immediately preceding the starting position were successively revealed, with rating and confidence estimations made at each move. Estimation error decreased as a function of number of moves revealed (p < .001). Higher rated players consistently made lower estimation errors (p < .01). Judges at all personal levels were more accurate about positions arising between players close in rating to themselves. A self-reference heuristic is proposed in which estimation of expertise is made relative to the judge's own projected performance.

How do we recognize expertise? What kinds of information facilitate the identification of expert performance? Can we recognize the expert production of someone who has superior technique or a different style? For many reasons I have chosen the problem domain of chess to investigate expert performance and its recognition. To develop the skill of a master chess player generally requires at least 10 years of study and competition. Chess players attempt to improve by studying the games of stronger players, much as students in other disciplines study examples of superior performance. The implicit rationale for such study is that the qualities that distinguish the stronger performer from the student will become apparent and be assimilated. Chess also provides a highly reliable rating system which distinguishes players on the basis of past performance. This rating system is, in fact, the most precise predictor of performance of any complex or highly skilled task.

An important theoretical question in the study of expert performance is whether chess masters are distinguished by the positions they recognize and create, the moves they make, or their search for the

AMERICAN JOURNAL OF PSYCHOLOGY
Fall 1992, Vol. 105, No. 3, pp. 409–415

best move. In Chase and Simon's (1973a) theory, "the most important processes underlying chess mastery are those immediate visual-perceptual processes rather than the subsequent logical-deductive thinking processes" (p. 215). Chase and Simon (1973b) suggest that masters have superior chess-playing ability because they can recognize a greater number of patterns of related pieces. According to them, the chess player's ability grows as specific patterns are stored in long-term memory.

There are many objections to this pattern recognition theory, including the demonstration by Charness (1981) that memory scores may vary when skill is held constant and by Holding and Reynolds (1982) that skill may vary when memory remains constant. Dynamic factors in search appear to be at least as important as pattern recognition. For example, the type of search heuristic that subjects use is superior to their recall scores in discriminating level of performance (Reynolds, 1982). More recently, Holding (1989) has shown that chess skill depends on the efficiency of operations on the search tree. The stronger players improve the accuracy of their evaluations as they progress through a game tree, whereas weaker players do not.

Studies on the "anchoring and adjustment" heuristic (Tversky & Kahneman, 1982) show that problem solvers at all levels of expertise make their initial judgments based on an often idiosyncratic point of reference, and adjust that decision in response to subsequent information. For example, subjects' initial predictions of behavior are often based on what they themselves would do in the situation (Davis, Hoch, & Ragsdale, 1986). Problem solvers at all levels of expertise typically fail to modify their initial decision sufficiently. In some circumstances, experts are poorer at making necessary adjustments to their initial judgments than are subjects with fewer expectations. For example, psychotherapists tend to "anchor" their clinical diagnosis to the symptoms presented at the beginning of a protocol (Friedlander & Stockman, 1983), though such distortions are not found in undergraduate students unaware of prototypical conditions that the symptoms may represent (Friedlander & Phillips, 1984). On the other hand, Grandmaster and Master chess players are more likely than lower rated players to modify their initial search tree in response to their evaluations of subsequent moves (Reynolds, 1991).

In the present experiment, I explore how well players of different performance levels can identify the ratings of those who produced a chess position or the moves leading to a chess position. It is hoped that this will help to tease apart the relative contributions of pattern recognition and search factors. It should be expected that seeing chess moves will facilitate the making of accurate judgments. This is consistent with an "ecological" approach (e.g., Gibson 1979; Neisser, 1976) which suggests that recognition is accurate to the extent that one observes the stimulus in action. Based on studies such as that of Davis et al. (1986), we may further hypothesize that subjects will estimate playing strength by comparing the demonstrated moves with those that they would have made were it their own game. If this hypothesis is true, subjects should be superior at estimating ratings of positions taken from games between players most like themselves in playing strength. We should also expect that subjects will tend to identify as coming from a different caliber of chess player the positions or moves not resembling their own.

EXPERIMENT

METHOD

Stimuli

Fifteen chess players were recruited from chess clubs in New York City. The United States Chess Federation rating system (Elo, 1978) provides a scale with a mean at 1500 and a standard deviation of 200 points. Every 200 points corresponds to nominal categories such as class "B," "A," "Expert," "Master," and "Senior Master." The 15 subjects ranged in performance level from 1300 to 2210 points, that is, from beginning tournament player (class "D") to Master. They were 17 to 65 years old and had 3 to 40 years of playing experience.

Procedure

Each subject was individually tested on positions from six different chess games. The order of presentation was randomized for each subject. The positions were arranged with standard tournament chess pieces and board. Subjects were instructed that the positions were between players of similar ratings, ranging from 1400 to 2600 points. After looking at a position for 2 min, subjects estimated the ratings of the contestants, and wrote the estimate on a score sheet; they indicated their confidence on a scale ranging from 0 (*no confidence*) to 10 (*complete confidence*). They were then shown the two half-moves immediately preceding the initial position, and they again provided rating and confidence estimates. This procedure was continued twice more, for a total of six half-moves preceding the initial position. Subjects were not informed as to their accuracy during the experiment. Total testing time for the six games was approximatey 1 hr per subject.

Materials

Games were selected so as to be (a) unfamiliar to the subjects, (b) between players of almost identical ratings (± 10 points), and (c) having equal material after Move 20, from which the critical positions were derived. The six test positions were taken from games between players with ratings of 1400, 1600, 1800, 2200, 2400, and 2600. Each was randomly selected from a pool of five positions, all of which satisfied the above criteria.

RESULTS

The estimation error for the initial positions, averaged across all subjects and games, was 278.9 points, or nearly 1.5 nominal categories. Mean estimation error decreased with each move revealed: 233.8 after the first move, 188.9 after the second, and 177.7 after the third. Because the position ratings were equally distributed across a range of 1200 points (1400–2600), the expected error from random guessing (with replacement) across this range was 600 points. If subjects assumed that the range of values positions could have varied continuously from 1200 to 0, then the expected error would be 300 points.

An analysis of variance (ANOVA) for repeated measures showed significant variance due to number of moves revealed, $F(3, 42) = 13.15$, $p < .001$. The Newman-Keuls post hoc test indicated significantly greater errors in response to the initial position than to the first move ($Q = 3.53$, $p < .05$), second move ($Q = 7.04$, $p < .01$), and third move ($Q = 7.91$, $p < .01$). There was also a significant difference between the first move and second move ($Q = 3.51$, $p < .05$), and the first move and third move ($Q = 4.38$, $p < .01$). There was no significant improvement in estimation error in response to the third move.

To what extent is accuracy related to the judge's rating? Ratings and error rates were negatively correlated for both the initial position and the move estimations. The correlations were all significant at $p < .01$: initial position (−.66, $N = 90$), first move (−.31, $N = 90$), second move (−.33, $N = 90$), third move (−.35, $N = 90$).

In Table 1, the subjects are divided into three groups of 5 similarly rated subjects. Players of all ratings improved after seeing moves. ANOVA for repeated measures was applied to Table 1: $F(2, 12) = 8.26$, $p < .01$, for the rating groups; $F(3, 36) = 4.19$, $p < .02$, for

Table 1. Mean estimation error at three rating categories

U.S. Chess Federation rating	Initial	Moves		
		1st	2d	3d
2000–2210	198.4	173.4	141.8	123.2
1650–1900	281.6	259.8	205.0	205.0
1300–1600	356.8	268.2	220.0	205.0

the moves variable. There was no interaction between moves and rating.

Confidence scores increased as a function of number of moves revealed. When averaged across subjects, the initial position yielded a mean score of 4.49, first move 5.57, second move 6.45, and third move 7.58. Confidence increased with degree of accuracy ($r = .32$, $N = 360$, $p < .01$), which, as indicated above, was related to number of moves presented. Confidence also increased with the judge's rating ($r = .14$, $N = 360$, $p = < .01$).

Evidence for the hypothesis that subjects are superior at estimating ratings of players most like themselves comes from correlations between error and rating difference between the judge and the judged. For all positions and moves, there were positive correlations between error and rating difference. These correlations were statistically significant for the initial position ($r = .63$, $N = 90$, $p < .01$), first move ($r = .24$, $N = 90$, $p < .05$), second move ($r = .37$, $N = 90$, $p < .01$), and third move ($r = .35$, $N = 90$, $p < .01$).

Evidence for the corollary hypothesis that subjects will identify positions or moves not resembling their own as coming from a different performance level is found in correlations between the judges' ratings and their mean estimations. There was a significant negative correlation in response to the initial position ($r = -.58$, $N = 90$, $p < .01$) and after the first move ($r = -.61$, $N = 90$, $p < .01$). This indicates that under conditions of little confidence and high error, judges tend to identify the participants as belonging to a different class from their own. The correlations in response to the second and third moves were very low ($-.03$ and $-.17$, respectively). The negative correlations between judges' ratings and their mean estimations indicate that subjects do not have a bias to respond with estimates close to their own ratings.

DISCUSSION

Subjects at all performance levels are moderately confident that they can determine the caliber of players who produced the initial chess positions. However, subjects at all levels, especially below the level of Expert (2000), are poor at estimating ratings from a static position. Seeing moves in addition to the initial position is essential to accurate recognition of playing level. Estimation error decreases after seeing the first pair of half-moves, and decreases again after the second pair of moves is revealed. No further "adjustments" follow the third pair of moves. Stronger players make superior estimates in response to the initial position as well as the subsequent moves. This suggests that players may be characterized by both the moves they contemplate and the positions they create.

What is the process by which subjects estimate relative playing strength? I have suggested that they judge a position relative to themselves. They might ask, "What is the likelihood that I would get into a position like this, or make moves like these?" If a position that looks like the player's own is seen, or moves that are fully understood or have been predicted are observed, then the position might be identified as representing the player's own strength. Conversely, judges might identify a position or moves not resembling their own as coming from a very different performance level. There is evidence for both of these schemas. I propose that the use of one's self as a standard of performance be called a *self-reference heuristic.*

Most likely, the higher rated players use both the positive and negative self-reference schemas for the initial position and each of the revealed moves. For correlations of error and rating difference, the 5 highest rated subjects show strong positive correlations (.70 to .48) across positions and moves, suggesting the consistent use of the self-reference heuristic. The middle-level players show a strong positive correlation for the initial position (.76), but not for subsequent moves (.26 to .01), suggesting that they use themselves as a point of reference for familiar positions but not moves. The lowest category shows no evidence of ever using the self-reference heuristic (.24 to .05). These correlations are only suggestive, however, because there are too few subjects in each of the three categories for a meaningful statistical inference.

Analysis in terms of a self-reference heuristic would account for the fact that the higher rated players are better able to relate to moves arising from all performance levels. Although all judges are capable of evaluating others of their own level, the expert can relate to a broader range of levels. They are the only ones who, at one time or another, have been rated lower as well as higher.

Most disciplines have the concept of "mastery," and the notion that expertise can be determined and emulated. In general, an isolated and static end product seldom indicates the level of expertise of the process, regardless of the discipline being judged. The present experiment indicates that an essential variable in evaluating expertise is the observation of performance in process.

Notes

The author wishes to thank Tamir Druz and Alex Galkovich for their work in collecting data.

Correspondence concerning this article should be addressed to Robert I. Reynolds, P.O. Box 754, Santa Ynez, CA 93460. Received for publication May 31, 1990; revision received May 6, 1991.

References

Charness, N. (1981). Visual short-term memory and aging in chess players. *Journal of Gerontology, 36,* 615–619.

Chase, W. G., & Simon, H. A. (1973a). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–272). New York: Academic Press.

Chase, W. G., & Simon, H. A . (1973b). Perception in chess. *Cognitive Psychology, 4,* 55–81.

Davis, H. L., Hoch, S. J., & Ragsdale, E. K. E. (1986). An anchoring and adjustment model of spousal predictions. *Journal of Consumer Research, 13,* 25–37.

Elo, A. (1978). *The rating of chessplayers, past and present.* New York: Arco.

Friedlander, M. L., & Phillips, S. D. (1984). Preventing anchoring errors in clinical judgment. *Journal of Consulting and Clinical Psychology, 52,* 366–371.

Friedlander, M. L., & Stockman, S. J. (1983). Anchoring and publicity effects in clinical judgment. *Journal of Clinical Psychology, 39,* 637–643.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Holding, D. H. (1989). Evaluation factors in human tree search. *American Journal of Psychology, 102,* 103–108.

Holding, D. H., & Reynolds, R. I. (1982). Recall or evaluation of chess positions as determinants of chess skill. *Memory & Cognition, 10,* 237–242.

Neisser, U. (1976). *Cognition and reality.* New York: Freeman.

Reynolds, R. I. (1982). Search heuristics of chess players of different calibers. *American Journal of Psychology, 95,* 383–392.

Reynolds, R. I. (1991). The application of a search heuristic by skilled problem solvers. *Bulletin of the Psychonomic Society, 29,* 55–56.

Tversky, A., & Kahneman, D. (1982). Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 3–22). New York: Cambridge University Press.