

# Meta-analysis of age and skill effects on recalling chess positions and selecting the best move

Jerad H. Moxley · Neil Charness

Published online: 19 March 2013  
© Psychonomic Society, Inc. 2013

**Abstract** A meta-analysis was conducted of studies that measured the effects of both age and skill in chess on the tasks of selecting the best move for chess positions (the best move task) as well as recalling chess game positions (the recall task). Despite a small sample of studies, we demonstrated that there are age and skill effects on both tasks: age being negatively associated with performance on both tasks and skill being positively associated with performance on both tasks. On the best move task, we found that skill was the dominant effect, while on the recall task, skill and age were approximately equally strong effects. We also found that skill was best measured by the best move task. In the case of the best move task, this result is consistent with the argument that it accurately replicates expert performance (Ericsson & Smith, 1991). Results for the recall task argue that this task captures effects related to skill, but also effects likely due to a general aging process. Implications for our understanding of aging in skilled domains are also discussed.

**Keywords** Aging · Chess · Best move · Recall · Expertise

In a pioneering study, Charness (1981) demonstrated that age effects in chess could be better understood by decomposing tasks and looking at differences in the method older adults use to perform the task, as well as differences in the subcomponents of skill (see Salthouse, 1984, for an example of this method in typing). Charness's (1981) main finding was that performance was not significantly related to age for a chess problem solving task where players were shown a position and asked to select the best move (henceforth in this article, the *best move task*). More significantly, the study showed that on a task where chess players were

asked to recall the configuration of game positions (henceforth in this article, the *recall task*), a task thought to demonstrate the structure of expert knowledge in chess (Chase & Simon, 1973), there were large age-related decreases in performance, controlling for skill.

While other studies have looked at differences associated with aging in chess, none that we are aware of have looked specifically at the dissociation between the recall task and the best move task with regard to aging, despite these tasks being closely related to our current theories about the role of chess knowledge. In fact, nearly all the theoretically important effects identified in one of the most important reviews of the chess literature were from the recall task (Gobet, 1998). The prominence of the recall task dates to the work of Chase and Simon (1973), who not only showed large basic differences in performance on chess recall, but also were able to explain these differences through the mechanism of better players remembering larger chunks. Recently, the best move task has become more important than it was at the time of Gobet's review for theories of skill, reflecting the observation that it predicts skill better (Charness, 1989; Pfau & Murphy, 1988) and the argument that it represents the actual skill of chess playing better (Ericsson & Smith, 1991).

This meta-analysis is motivated by multiple needs. The first is to establish accurate estimates of age effects on the two most important tasks in the chess literature: the best move task and the recall task. This should allow researchers to make better decisions on the power needed to test and account for age effects using both tasks, as well as showing to what extent these variables need to be attended to in various designs. The second reason is to motivate more reporting and research into this important topic. Despite all ages competing in chess and, at times, stark differences in performance (Charness, 1981), age is rarely controlled for in experimental studies on chess. We expect to show that age effects are present in the best move task despite inconsistent past findings, with some finding a null result

J. H. Moxley (✉) · N. Charness  
Florida State University, Tallahassee, FL, USA  
e-mail: moxley@psy.fsu.edu

(Charness, 1981), while the Amsterdam chess test (the ACT; van der Mass & Wagenmakers, 2005) showed a strong negative effect (Roring & Charness, 2007). If age has a relationship to skill and performance similar to that observed in tournament play, this would mean that controlling for age may give a better index of current skill. Additionally, we expect that the recall task will show a fairly strong relationship between age and performance. We hope to demonstrate that age is an important and helpful covariate to consider when interpreting such results. Finally, we will assume that, controlling for skill, age will follow a linear relationship with both tasks. While the relationship with age and skill is known to be a quadratic (Roring & Charness, 2007) when controlling for skill or for the skill acquisition process in chess, the relationship with age becomes a negative linear trend reflecting the improvement of younger adults to about the age of 40 and, then, a subsequent decline (Charness, Krampe, & Mayr, 1996).

## Method

In conducting the meta-analysis, we looked for studies that reported the intercorrelations of age, chess skill as measured by the chess ratings systems, and performance on either the best move task or the recall task. Keywords used were “chess and aging,” “chess and move selection,” and finally, “chess and recall.” These were used in the web of knowledge database, as well as the ProQuest dissertations and theses database. Our search found three studies, each of which had looked at both types of chess performance. These were Charness (1981), Pfau and Murphy (1988), and the ACT of van der Maas and Wagenmakers (2005). Additionally, Bilalić, McLeod, and Gobet (2009) measured all four variables, and upon being contacted, Dr. Bilalić provided the correlation table of the study for inclusion in this analysis (personal communication, February, 20, 2011). Our sample, while small, does give a very clean analysis, since all studies had each participant perform both tasks, meaning that we could directly compare the joint influence of age and skill. No doubt, more studies have measured these variables; however, this analysis will use partial correlations, meaning that we need at least three correlations: age with skill, skill with performance (on recall or best move), and age with performance. Since only a few studies have been interested in age effects, the age correlations have generally not been reported. Additionally, we are not interested in studies that looked at youth chess, since our interest was in aging.

Of the studies, only one (Charness, 1981) primarily looked at age, while the others reported it. Bilalić et al. (2009) tested whether or not players who primarily play a French defense or a Sicilian defense perform better on their

normal opening. This means that all tests were of positions from the middle of chess games from one of the two defenses. This study is the only one including an experimental manipulation. The set also varied in time constraints. On the best move task, the ACT gave 30 s per problem, Pfau and Murphy (1988) gave 5 min per problem, and both Charness (1981) and Bilalić et al. gave 10 min. There were also differences with recall time, since while both Bilalić et al. and Pfau and Murphy used 5 s, the ACT used 10 s. Charness (1981) used an unexpected recall test after a rapid position evaluation. With only four studies, it is unlikely that we could identify heterogeneity of effects. The means and standard deviations for the studies were as follows (decimals reported to two digits or as far as reported in the original study): Charness (1981) measured skill using Canadian ratings and had a mean of 1,569 ( $SD = 185$ ), with an age mean of 38.7 ( $SD = 15.05$ ); Pfau and Murphy used USCF ratings and had a mean of 1,644.7 ( $SD = 338.1$ ), with an age mean of 34.7 ( $SD = 14.9$ ); van der Maas and Wagenmakers (2005) used national ratings (mainly the Netherlands), and the subsample had a mean of 1,874.11 ( $SD = 293.25$ ) and an age mean of 30.68 ( $SD = 14.83$ ); and Bilalić et al. used FIDE ratings with a mean of 2,308 ( $SD = 160.63$ ) and an age mean of 31 ( $SD = 11$ ). For more information about the chess rating system, see Elo (1978).

Another issue arising was that the ACT gave two best move tasks and it appeared that a few participants had dropped out after the first test and had not done either the recall task or the best move task for form B; therefore, we used only results from people who had completed both forms and combined the sum scores from the two forms in our analysis to maintain complete group parity. Second, the ACT asked participants to enter the moves on a computer within 30 s, and it was apparent that failure to do this was correlated fairly significantly with age,  $r(226) = .468$ ,  $p < .001$ . A cardinal feature of adult aging is that the rate of information processing for many tasks is slowed (Jastrzembski & Charness, 2007; Salthouse, 1996), and this has been shown even for skilled chess players doing chess-related processing (Jastrzembski, Charness, & Vasyukova, 2006). This slowing seemed partly responsible for the fact that the ACT is a fairly strong outlier in the best move data. To be conservative, we will report the results with the ACT unadjusted for failure to enter a move (termed *ACT raw*) and with that factor statistically controlled (termed *ACT adjusted*). This is similar to leaving out items at the end of the test that the test taker does not reach when trying to estimate ability, which is commonly done in item response theory analysis (de Ayala, 2009). The rescore metric should not necessarily be viewed as superior, since, no doubt, it sacrifices some meaningful variance. To analyze the effect sizes, we used the DerSimonian–Laird random-effects method

(DerSimonian & Laird, 1986) as instituted in the metacore package (Laliberté, 2009) of the R-statistics program (R Development Core Team, 2011).

## Results

Table 1 shows the partial correlations from all four studies with the four effects of interest. The ACT appears to be somewhat different on the best move task with regard to age. The adjusted value appears somewhat more in line with the other studies. Table 2 shows the full results of the meta-analysis. It shows that regardless of which ACT estimate we use, there is little difference in the estimated meta-analytic effect size, likely due to the robustness of the random-effects method; however, it can be seen that the confidence interval does change noticeably. All four effects are significant, with age being a negative effect independent of ratings for both the best move task and recall, while skill is positively correlated with both (see Table 2 for all correlation estimates, confidence intervals, and *p*-values).

An interesting question is the relative size of the effects. Inspection of the confidence intervals confirms that skill appears to be more related to the best move task, confirming the assumptions of past researchers (Charness, 1989; Ericsson & Smith, 1991) that it should best represent the skill of chess. To interpret whether age is more related to one test, we conducted a dependent samples *t*-test of the raw *z*-scores to get a test of the mean difference. Using the ACT rescore, this difference is significant,  $t(3) = 3.25$ ,  $p = .048$ , while it shows marginal significance for the raw score,  $t(3) = 2.67$ ,  $p = .075$ . Additionally, we calculated the Fisher *r* to *z* from the confidence intervals of both correlations as opposed to the sample size (Hopkins, 2007), because the test based on the sample size would yield a biased estimate in this case. This showed that the confidence interval for the difference was 95 % CI  $[-.05, -.34]$  for the ACT rescore and 95 % CI  $[-.02, -.36]$  for the ACT raw score, both of which suggest a statistically significant greater correlation with age in the recall task than in the best move task (see Cummings, 2012,

**Table 2** Results of meta-analysis

Effect	<i>r</i>	95 % CI	<i>p</i>
Age and move selection (ACT raw)	−.279	−.120 to −.420	<.001
Age and move selection (ACT rescore)	−.277	−.175 to −.373	<.001
Skill and move selection (ACT raw)	.762	.642 to .846	<.001
Skill and move selection (ACT rescore)	.758	.661 to .829	<.001
Age and chess recall	−.486	−.356 to −.597	<.001
Skill and chess recall	.512	.423 to .591	<.001

*Note.* All rating results control for age, and all age results control for rating ( $n = 343$ )

for a discussion of inferences from confidence intervals). Additionally, as can be seen from the confidence intervals, the best move task is more related to skill than to aging. The recall task appears to be approximately equally related to skill and age [ $t(3) = 0.032$ ,  $p = .768$ , for the absolute value of *z* scores], although in a different direction.

## Discussion

This meta-analysis makes several contributions. First, it establishes that the best move task can capture an age effect. While the initial study regarding this question found a null result (Charness, 1981), all four studies showed at least a negative trend (note that this was significant only in the ACT). When combined, they reveal a significant negative effect. This suggests that the task is indeed sensitive to age-related decline in performance that mirrors longitudinally measured decline (e.g., Roring & Charness, 2007). Table 3 compares the results of our meta-analysis with raw correlations from a study we are conducting comparing tournament performance with pretournament ratings across different ages. The sample includes randomly selected tournaments of 2,666 U.S. chess players over the age of 16 years. This was the entire population of people we could find on the

**Table 1** Partial correlations for the four studies included in the meta-analysis

Study	<i>n</i>	<i>r</i> & <i>mc</i>	<i>a</i> & <i>mc</i>	<i>r</i> & <i>cr</i>	<i>a</i> & <i>cr</i>
1. Charness (1981)	34	.587*	−.179	.473*	−.678*
2. Pfau and Murphy (1988)	57	.763*	−.131	.494*	−.478*
3. Bilalic et al. (2009)	24	.751*	−.228	.733*	−.522*
4. Van der Maas and Wagenmakers (2005)	228	.838* (.813*)	−.397* (−.328*)	.494*	−.410*

*Note.* *r* & *mc* = partial correlation for rating and move choice selection, *a* & *mc* = partial correlation for age and move choice, *r* & *cr* = partial correlation for rating and chess position recall, and *a* & *cr* = partial correlation for age and chess position recall. Alternate scoring controlling for degree of failure to respond in Van der Maas and Wagenmakers (2005) is shown in parentheses

\*Significance < .05

**Table 3** Results of meta-analysis on move selection (top row), as compared with classical tournament performance (bottom row), based on players age 16+

Study	Age	Skill
Meta-analysis of move selection	-.277	.758
Classical tournament performance	-.190	.826

2009 FIDE (the international sanctioning body of chess) tournament list that gave an age, who played in the U.S. and played in a tournament within 5 years of the FIDE list we used. These data also allow us to confirm that the assumption of linearity we made in our analysis at least holds in comparable data. It appears to hold since a quadratic effect of age is not significant,  $F(1, 2663) = 0.09$ ,  $p = .765$ , after controlling for skill. While we cannot confirm this in our meta-analysis, the ACT for which raw data are available also shows a quadratic effect of age that is not significant,  $F(1, 224) = 0.07$ ,  $p = .796$ , after controlling for skill. Table 3 also shows that the skill effect for the best move task is impressively close to that observed in actual tournaments. It is important to note that this analysis represents both older adults underperforming their rating and improving younger adults overperforming their rating. Thus, after controlling for chess rating, this negative linear effect of age predicting performance, controlling for rating, is consistent with past research (Roring & Charness, 2007) showing a quadratic relationship with age.

It is likely that with a reasonable number of well-designed chess problems, researchers can expect to capture skill at a level comparable to that for performance in a tournament. This is very important because the best move task allows for experimental control and manipulation and allows researchers to measure processes through eye movements, verbal protocols, or neurological data (see Schult-Mecklenbeck, Kühberger, & Ranyard, 2011, for discussions of the application of these and other process-tracing methods to decision-making research). Additionally, the best move task's significantly greater correlation than for the frequently used recall task argues for the best move task as the gold standard task for representing chess skill in the laboratory and for observing the cognitive processes associated with chess skill (Ericsson & Moxley, 2011; Ericsson & Smith, 1991).

The age effect seems large in comparison with the tournament results. This may be due to peculiarities of the individual studies; for instance, the larger effect size for the ACT could reflect a differential aging effect on speeded tasks (Salthouse, 1996), although we do not have enough studies to confirm this. It should be noted that in the best move task, the players must orient to an unfamiliar board without the benefit of prior planning. It is possible that this

is more difficult for older adults. Even within the best move task, it has been shown that players perform better when they are playing a move from an opening that they are more familiar with (Bilalić et al., 2009). It is possible that the age effect will interact with either familiarity of position or amount of prior planning from previous moves. In fact, a recent study by Vasyukova (2012) showed that solving two consecutive positions significantly changes the form of the search, particularly in stronger players. These effects may be related to memory as well. For instance, older players might rely on a familiar set of openings to maintain their comfort level with positions. These are important questions for future researchers, although we would caution that to be able to measure such a difference in correlations on the basis of the size of our estimates, one would need more power than is typically found in chess studies.

The recall task is also an effective measure of skill in chess. However, our analysis shows that age should be controlled for when it is used. Unfortunately, the small set of studies identified that controlled for age suggests that this is rarely done. According to this analysis, age is almost identically related to task performance as is skill. The effect of age is much greater in the recall task than that observed in tournament performance. It also appears to be larger than that found on the best move task, suggesting that the task picks up on the effect of skill plus something else that is also related to age. Recently, Smith, Gobet, and Lane (2007) simulated the aging effect in recall performance using CHREST, a computational model representing template theory (Gobet & Simon, 1996), a theory that argues that chess skill is based on the accumulation of complex schemata and patterns in long-term memory. Their simulation best fit the aging effect in recall by decreasing working memory, a capacity that declines with age (Salthouse, 2010). The results of this simulation, in the context of our study, would suggest that working memory capacity is strongly related to recall of chess boards, but less so to the actual selection of moves (or that working memory is not a unitary capacity and that it varies across task types). This would be consistent with the theoretical idea that search and selection of the best move are supported by the use of long-term memory representations, such as those suggested by template theory (Gobet & Simon, 1996) or long-term working memory theory (Ericsson & Kintsch, 1995). Charness (1981) also tested recognition memory and showed that it did not show a significant age-related decline. The dissociation of age effects in recall and recognition memory has been observed in nonexperts (Craig, 1994), with a suggested cause being that recall requires more cognitive resources and that the resources recall requires are fewer in older adults. Smith et al.'s (2007) simulation is consistent with this interpretation and shows how chess recall might help identify what cognitive resources are overtaxed in older adults.



Other cognitive factors might also play a critical role in explaining age effects. Mireles and Charness (2002) had success in predicting recall of chess opening sequences as a function of both age and skill, particularly for interactions of these two variables, using recurrent neural net simulations that varied the size of the knowledge base (skill) and also parameters that could be interpreted as reflecting age-related changes in speed of processing, decay rate for memory, inhibition failure, neural noise, and brain lesions. Again, our results suggest that these general age processes change recall performance more than they affect best move performance. One explanation is that these parameters affect performance more in a relatively new task (chess players rarely explicitly must recall a position) than in a vastly overlearned task such as searching for the best move.

It should be noted that chess memory involves more than memory for chess position, as in the experiments included in this meta-analysis. For instance, recent research has shown a skill difference in opening knowledge (Chassy & Gobet, 2011), which involves memorization of set sequences of moves from a stable opening. In the original work on pattern recognition in chess (Chase & Simon, 1973), a dissociation between these types of memory was noted because, whereas recall memory for randomly generated positions was greatly reduced for experts, memory for randomly generated opening sequences showed a striking expert advantage. It is possible that memory for this type of information will have a different pattern of association with age, possibly even increasing with age as many forms of crystallized knowledge do. Looking at aging effects in the openings area may allow better understanding of how highly overlearned knowledge decays, just as studying recall tasks allows testing of deficits in older adults memory in domains of skill.

In conclusion, this analysis demonstrates that aging affects performance on the best move task. This effect is very similar to that shown in tournaments and suggests that laboratory research in chess can accurately measure both skill and age. We believe that research in chess should focus mainly on understanding the processes that lead to accurate move selection using the best move task, particularly using experimental manipulations, as some researchers are already doing. The recall task may be an extremely effective task for studying age and skill differences, since it appears more sensitive to them than do other chess tasks. This age by task effect appears to be robust, so attempts to use it to disentangle the processes underlying skilled performance across the lifespan may bear fruit and allow us to understand better how older adults maintain one aspect of performance to a much greater degree than others and how memory changes with age in general.

## References

- Bilalić, M., McLeod, P., & Gobet, F. (2009). Specialization effects and its influence on memory and problem solving in expert chess players. *Cognitive Science*, 33, 1117–1143.
- Charness, N. (1981). Aging and skilled problem solving. *Journal of Experimental Psychology: General*, 110, 21–38.
- Charness, N. (1989). Expertise in chess and bridge. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 183–208). Hillsdale, NJ: Erlbaum.
- Charness, N., Krampe, R., & Mayr, U. (1996). The role of practice and coaching in entrepreneurial skill domains: An international comparison of life-span chess skill acquisition. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (pp. 51–80). Mahwah, NJ: Erlbaum.
- Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.
- Chassy, P., & Gobet, F. (2011). Measuring chess experts' single-use sequence knowledge: An archival study of departure from 'theoretical openings'. *PLoS One*, 6, e26692. doi:10.1371/journal.pone.0026692
- Craik, F. I. M. (1994). Memory changes in normal aging. *Current Directions in Psychological Science*, 3, 155–158.
- Cummings, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177–188.
- Elo, A. E. (1978). *The rating of chessplayers, past and present*. New York: Arco Publishing.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211–245.
- Ericsson, K. A., & Moxley, J. H. (2011). Thinking aloud protocols: Concurrent verbalizations of thinking during performance on tasks involving decision making. In M. Schulte-Mecklenbeck, A. Kueberger, & R. Ranyard (Eds.), *Handbook of process tracing methods for decision research* (pp. 89–114). New York: Psychology Press.
- Ericsson, K. A., & Smith, J. (1991). Prospects and limits in the empirical study of expertise: An introduction. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 1–38). Cambridge: Cambridge University Press.
- Gobet, F. (1998). Expert memory: A comparison of four theories. *Cognition*, 66, 115–152.
- Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, 31, 1–40.
- Hopkins, W. G. (2007). A spreadsheet for deriving a confidence interval, mechanistic inferences and clinical inferences from a p value. *Sportscience*, 11, 16–20.
- Jastrzembski, T. S., & Charness, N. (2007). The Model Human Processor and the older adult: Parameter estimation and validation within a mobile phone task. *Journal of Experimental Psychology: Applied*, 13, 224–248.
- Jastrzembski, T., Charness, N., & Vasyukova, C. (2006). Expertise and age effects on knowledge activation in chess. *Psychology and Aging*, 21, 401–405.
- Laliberté, E. (2009). metacor: Meta-analysis of correlation coefficients. R package version 1.0-1. <http://CRAN.R-project.org/package=metacor>
- Mireles, D. E., & Charness, N. (2002). Computational explorations of the influence of structured knowledge on age-related cognitive decline. *Psychology and Aging*, 17, 245–259.

- Pfau, H. D., & Murphy, M. D. (1988). Role of verbal knowledge in chess skill. *American Journal of Psychology*, 101, 73–86.
- R Development Core Team. (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>
- Roring, R. W., & Charness, N. (2007). A multilevel model analysis of expertise in chess across the life span. *Psychology and Aging*, 22, 291–299.
- Salthouse, T. A. (1984). Effects of age and skill in typing. *Journal of Experimental Psychology: General*, 113, 345–371.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103, 403–428.
- Salthouse, T. A. (2010). *Major issues in cognitive aging*. New York, NY: Oxford University Press.
- Schult-Mecklenbeck, M., Kühberger, A., & Ranyard, R. (Eds.). (2011). *A handbook of process tracing methods for decision research*. New York: Taylor & Francis.
- Smith, R. L., Gobet, F., & Lane, P. C. (2007). An investigation into the effect of ageing on expert memory with CHREST. In *Proceedings of the seventh UK workshop on computational intelligence*.
- van der Maas, H. L. J., & Wagenmakers, E. J. (2005). A psychometric analysis of chess expertise. *The American Journal of Psychology*, 118, 29–60.
- Vasyukova, E. E. (2012). The nature of chess expertise: Knowledge or search. *Psychology in Russia: State of the Art*, 5, 511–528.