Journal of Expertise
2021. Vol. 4(2)
© 2021. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.
ISSN 2573-2773

Contributions of K. Anders Ericsson to Experimental Cognitive Psychology

Neil Charness

Department of Psychology, Florida State University, USA

Correspondence: Neil Charness, charness@psy.fsu.edu

Abstract

In this article I describe some of the contributions that K. Anders Ericsson has made to the development of experimental cognitive psychology in the context of understanding expert performance. I focus on three facets: use of verbal protocols, design of experimental techniques, particularly those that contrast domain-specific and domain-general capabilities, and long-term working memory theory. I also outline some of the challenges that remain for those approaches. In reviewing those contributions, I also allude to how Ericsson's work influenced my own research.

Keywords

Expertise, skill, task decomposition, Ericsson, verbal protocols, long-term working memory

Assessing Contributions

One way to assess someone's contributions to a field is to use sociometric measures such as citations. When researchers cite a given publication, it suggests that reading the cited work in some way influenced how they approached and interpreted their own work. The online tool Google Scholar shows that Ericsson's most cited work, with over 14,000 citations, is the book with Herbert Simon: *Protocol Analysis: Verbal Reports as Data*, published initially in 1984 and revised in 1993 (Ericsson & Simon, 1984, 1993). It was preceded by a *Psychological Review* article, his third most highly cited work (6800+ citations), that they published on this topic in 1980 (Ericsson & Simon, 1980). It may seem strange to focus on a technique, using think aloud concurrent or retrospective verbal protocols to trace out cognitive processes, in a review of experimental cognitive psychology. However, I argue that resurrecting a technique first developed in the context of introspection psychology has facilitated theory development from within a rich trove of experiments on expertise.

The second most cited work is a *Psychological Review* paper on expertise in musicians, with over 11,000 citations (Ericsson et al., 1993). Although that paper is likely mostly cited for the estimates of cumulative deliberate practice hours garnered from the violinists and pianists that they interviewed, there is also an experimental approach embedded within the second study that has offered a template for contrasting domain-related versus non-domain related performance. That is, although they were probably not the first researchers to introduce this paradigm of contrasting within-domain versus domain-general performance, as one could point to the very influential Chase and Simon (1973) study of chess players recalling structured and random positions, Ericsson et al. (1993) played an important role both in extending and popularizing it for the study of expert performance.

Although I will not spend as much time discussing it here as I do the other contributions, the fourth most cited work was Ericsson and Kintsch's (1995) *Psychological Review* article on long-term working memory (4600+ citations). It provided an important counterweight to then prevailing views about working memory (e.g., Baddeley, 1986), what they called short-term working memory in their paper. That publication broadened our perspective on the roles of memory structures, particularly the role of retrieval structures and memory cues, in higher order cognitive processes. Ericsson and Kintsch provided a detailed account of the important role of knowledge (in the form of chunks and retrieval structures stored in long-term memory) and retrieval cues stored in a limited capacity short-term working memory, and how such structures coupled with association processes enabled experts to escape from usual/normal limits on working memory. That article also extended Kintsch's construction-integration model of reading to explain how skilled readers could attain quick access to long-term working memory representations of prior processed text to enable them to comprehend long prose passages.

In summary, among Anders Ericsson's many contributions to experimental cognitive psychology, three stand out and provide important tools to the research community. First, the techniques for soliciting reliable and valid verbalizations during problem solving tasks, used for both concurrent and retrospective verbal protocol generation, have provided rich data sets for theory development. Second, systematic exploration of within domain and domain general correlates of expertise have also broadened our understanding of the factors supporting skilled performance. Third, the long-term working memory theory introduced an important construct, long-term working memory retrieval structures, to explain many performance feats by experts.

Verbalizations as Data

Ericsson and Simon (1980) highlighted the need for rigor in using verbalizations to develop and test theories in information processing psychology. Think aloud instructions, used by early investigators for building theories of human problem-solving (e.g., Duncker, 1945), were given strong impetus by Newell and Simon's (1972) *Human Problem Solving* tome. That volume outlined a new theory of human problem solving based on an information processing framework. In many of the example domains that Newell and Simon explored (logic problems, chess problems, cryptarithmetic problems), they asked participants to "think aloud" while problem solving and used their verbalizations to constrain the models that were generated.

However, Nisbett and Wilson (1977) soon published a very influential critique of relying on people to *explain* why they made decisions. They demonstrated through experimental manipulations (e.g., placing an object on the left versus the right) that a factor influencing a participant's preference decision process was not within conscious awareness and when asked why they preferred one of two objects, peoples' verbal reports did not mention this source of influence. Hence, Nisbett and Wilson provided strong evidence that participants were often unaware of the experimentally manipulated factors that could be shown to have influenced their choices. Use of verbalizations for theory-building in psychology had been questioned much earlier, particularly in the behaviorist era of psychology that followed the initial introspective psychology period (e.g., Lashley, 1923). Hence, the proper use of "think aloud" instructions and interpretation of the resulting verbalizations was ripe for re-examination.

Probably the most important contribution from Ericsson and Simon (1980; 1984) was their theory of verbalizations. It enabled them to "…specify when, where, and under what kinds of instructions informative verbal reports can be obtained…" (Ericsson & Simon, 1993; p. 9). Their theory proposed that verbalizations were similar to other cognitive processes and were well-described as sequences of heeded information flowing through cognitive structures. Verbal protocols, they argued, represented data on a par with other overt behaviors, such as sequences of key presses. Verbalizations, compared to key presses, have a considerable advantage in the richness of the data being captured, being high density sources of behavior, thereby offering considerable value for uncovering the strategies that people employ during problem solving. Hearing someone verbalize “30 x 30 = 900” as the first utterance for solving the mental arithmetic problem “what is the square of 35” suggests a different strategy for the way they solve that problem, compared to hearing “35 + 5 = 40”, the latter consistent with the use of a skilled mental calculator’s strategy (Charness & Campbell, 1988).

A second contribution beyond developing a theory of the verbalization process was methodological. Ericsson and Simon outlined the pros and cons of using concurrent and retrospective verbalizations for investigating the problem-solving process, provided advice about the timing and use of memory probes, and offered standard instructions for preparing people to think aloud, e.g., using a practice mental arithmetic task, such as ‘multiply 24 times 34’ (the Appendix in Ericsson & Simon, 1993). Those procedures have been widely adopted in applied experimental research, such as in the field of human factors for doing usability testing (Lewis, 2012).

Ericsson and Simon also brought precision to what they termed the three different levels of verbalization that experimenters had been inducing in their participants. Level 1 (direct) verbalization referred to the case when information is reproduced/spoken in the same form as “heeded” (attended). Level 2 and 3 verbalization involved recoding of information (e.g., from non-verbal representations), hence required additional processing resources to transform information before it could be spoken. For Level 2, recoding from a non-verbal code into a verbal code was thought to be the only significant transformation, perhaps slowing down the problem-solving process. In both level 1 and 2 cases heeded information is expected to remain intact. Level 3 is distinguished by the need to scan or filter information to determine if it matches the desired information to be reported. Inference or generative activities are needed, as in their example of someone being asked to report perceived traffic hazards while driving a car. The sequence of heeded/attended information would be expected to change compared to the case where people did not think aloud. As they pointed out in the prior journal article (Ericsson & Simon, 1980), asking people to report on processes that they normally do not monitor or heed, such as motives (e.g., Nisbett & Wilson’s experiments), puts the experimenter on even shakier ground for inferring internal processes and generating models of problem solving and decision making.

Their model of verbalization also enabled them to predict and examine in detail the evidence for reactivity when using think aloud instructions during problem solving activities. An early concern with think aloud instructions was the risk of changing the very process that was under investigation compared to the “usual” case of people silently engaged in the same task without verbalizing. Using their three levels of verbalization model, they concluded that when the internal representation being used was “oral”, there was some expectation that processing might be slightly slowed, but not changed significantly. When transformations were required to go from a representation that was not easily verbalized (think of a beginner describing the elements in an electronic circuit diagram), you could expect both slowing and possible change in the problem-solving search processes as limited resources are diverted to recoding. In the case of level 3 verbalizations, additional inference processes could be evoked that might change the actual sampling of information from external (environmental) and internal (working memory) sources.

It is fair to say that Ericsson and Simon (1993) helped resurrect the use of verbalization in mainstream cognitive psychology. To quote them: “In the beginning of Psychology, many influential psychologists viewed verbal reports, and more precisely introspection, as the only valid method for data collection in psychology. At a later period, during the reign of behaviorism, verbal reports were almost totally rejected as data. It is now time for verbal reports to reassume their position as a rich source of data, combinable with other data, that can be of the greatest value in providing an integrated and full account of cognitive processes and structures." (p. 373).

Unresolved Issues with Verbalization Theory and Verbal Protocol Analysis

As Ericsson and Simon were at pains to point out, their theory of verbalization needed further development, particularly in terms of testing its assumptions and further explicating the underlying mechanisms of information processing. The theory was predicated on the Newell and Simon information processing framework from the early 1970s (e.g., Newell & Simon, 1972). The last 50 years of psychological research have contributed a great deal of knowledge at the neural level of description that in some sense supersedes the symbolic processing model that undergirds the theory of verbalization. However, as Newell (1990) noted long ago, theories have different granularities and ranges in which they are useful, particularly in terms of the time scale for the underlying processes. Neural theory usually pertains to sub-second to ms time ranges (e.g., EEG data), with symbolic models particularly informative in the range of seconds to minutes for tracking problem solving processes. The basic idea that there are processing costs to transforming information from one format to another has aged reasonably successfully and I suspect that today's elaborated theories of human cognition would offer similar predictions to those of Ericsson and Simon about how the act of verbalizing would be expected to change the problem-solving process for level 1, 2, and 3 verbalizations. Building process models of cognition is still seen as an important goal (Jarecki et al., 2020).

The issue of the extent of reactivity for concurrent verbalizations during problem solving is still an open issue particularly in the context of individual differences. To give one example, my lab found reactivity in the sense of changed problem solving processes only on one of several types of problem tasks and only in older adults (Fox & Charness, 2010) when comparing solution time and accuracy in silent and think aloud conditions. By asking older adults to think aloud while solving the Raven's Progressive Matrices problems, we added nearly a standard deviation to their "fluid" (cognitive) ability levels. We were able to replicate that finding in a second sample. That striking result led Fox to collaborate with another of my students who had taken Ericsson's Protocol Analysis graduate course on a review paper with Ericsson that further developed our understanding of when to expect reactivity when thinking aloud (Fox et al., 2011).

In sum, it is relatively easy to build a case for why Ericsson's most highly cited publication deserved the attention of the research community. Understanding problem solving processes is still an important goal in experimental cognitive psychology. Particularly in educational settings, diagnosing incorrect reasoning behavior, that is, troubleshooting faulty problem solving, remains an important task for instructors. Having someone verbalize while problem solving can provide direct evidence for "bugs" in understanding (Brown & Van Lehn, 1980) or in implementing procedures (e.g., fraction addition: Braithwaite & Siegler, 2020) compared to a simpler, but perhaps less efficient method: drawing inferences from carefully crafted multiple choice questions on an exam.

Although recording of verbalizations, followed by transcription, then translation into problem behavior graphs, followed by analysis of the statistics of search behavior has the advantage of providing deep insights into problem solving processes, that technique has disadvantages compared to other methodologies. It is a very resource intensive process that requires skilled personnel to carry out reliably, compared to automated recording of keystrokes with computer-assisted experiments. As someone who spent months coding and creating problem behavior graphs for 136 chess protocols (Charness, 1981), I can testify to the investment cost in using verbal protocol analysis. It is one of many methods to triangulate between data and theory. Experimental manipulations can lead to strong inference for theory building in a way that descriptive techniques such as protocol analysis cannot easily do. However, by combining the two approaches researchers can have the best (and worst) of both worlds.

Decomposing Expertise in the Laboratory

The second most-cited publication, Ericsson et al. (1993), represented a watershed for the understanding of the development of expert performance. It introduced techniques for studying what was called "deliberate practice" in violinists and pianists using a combination of retrospective recall of practice behaviors and matching the estimates for the most recent year to diaries of current practice activities (unsurprisingly finding over-estimates with retrospective recall). The premise that expertise differences are mostly due to differences in types of practice, and specifically to deliberate practice, remains a controversial topic in expertise research and other papers in this issue are no doubt discussing the significance of and evidence for that claim. In a reply to Gardner (1995), we (Ericsson & Charness, 1995) observed: "Only careful observation and study of differences in the type and number of activities associated with the longitudinal emergence of abilities and performance in normal and "very talented" children will allow us to determine the potential and possible limits of explanations based on characteristics acquired through focused and extended activity" (p. 804).

However, here I want to focus on study two in Ericsson et al. (1993) that popularized a quasi-experimental paradigm[1] for assessing the relationship between skill level and domain-general versus domain-specific abilities. The advantages and disadvantages of experimental and individual difference approaches to research have been debated for years (e.g., Cronbach, 1957). In this paper one sees a classic example of using both experimental manipulations of task types (domain-general, domain-specific) and examining correlations both between tasks and across skill levels.

At the beginning of the paper, Ericsson and colleagues discuss the Galton model of natural ability and its role in the achievement of "eminence." The theory was comprised of three facets: a genetically determined limit for an ability (an innate "capacity"), someone's extent of motivation ("zeal" for exercising the ability), and the actual amount that the ability is exercised ("power of doing a great deal of laborious work"). This hypothesized combination of stable genetic and malleable environmental influences on capacity/performance, suggests a strategy of parsing out performance components into domain-related (primarily influenced by practice history) and domain-general (primarily genetically influenced) components.

As an example, for piano performance, one might see genetics as setting the limit for tapping speed (finger-movement dexterity), advantaging some people over others in their asymptotic ability to trill one or more keys rapidly. A difference in basic tapping speed limits would be expected to show up in piano key pressing, in typing, and perhaps in reaction time to strike a single button repeatedly. Genetics might set the limit for such dexterity in the same way that it might set the limit for reach for keys across a long piano keyboard as a function of hand anatomy. Ericsson would no doubt challenge my characterization, arguing that most people, at least early in the life span, have enormous adaptive capability that can override the influence of genetically controlled presets. He would probably point to the changes seen in the brains of string players in the sensory homunculus with much more tissue becoming devoted to the fingering hand but not the bowing one, or the changes in specific muscle fiber tracts from fast twitch to slow twitch fibers, depending differentially on which muscle groups have been stressed by physical exercise (e.g., back for rowers vs. legs for runners). Anders enjoyed arguing with anyone who would engage with him about fundamental processes limiting adaptation effects.

In Study 2 the authors assessed two groups of pianists, experts and amateurs, on experimental tasks that had been developed in dissertation work by Krampe. The tasks comprised a complex movement coordination task using a piano keyboard to carry out nine keystrokes single handedly (left or right) and bimanually as mirror image movements or different movements in opposite hands. Fingers were assigned to the numbers 1-9 and the number sequence appeared on a computer screen. In the second session (a week or so later), a technically easy musical piece (Bach's Prelude No. 1 in C-major) was performed three times in succession following a 15-minute familiarization and practice period. This task was followed by a standardized test of perceptual-motor speed, the digit-symbol substitution task, then a two-choice reaction time task, and finally by finger tapping as quickly as possible for 15 s using right, left, and alternating forefingers on a keyboard. These tasks were expected to tap psychomotor performance that was either related to music performance (e.g., playing a Bach prelude), or unrelated to music performance (digit-symbol substitution, choice reaction time). The remaining tasks were thought to be skill related because they tapped some of the components that underly playing music skillfully (e.g., complex coordination skills between and within hands for pianists).

The sample used was somewhat small by today's standards (12 experts, 12 amateurs), though not small for a study of experts, so lacked power to find skill effects and interactions on these tasks unless they were large in magnitude. Research on expertise is inherently difficult because of the scarcity of experts. For instance, in an attempted replication of study one by Macnamara and Maitra (2019), the authors report difficulties in recruiting an equivalent size sample of experts (it took years). Even more unfortunately for the sake of comparing the studies, they were unable to match violinist skill in the same way (prizes won in competitions). Nonetheless, the paradigm of trying to find domain-related and domain-unrelated tasks to trace out underlying processes, and potentially, genetic versus environmental influences on performance, is a powerful one. Use of near and far transfer tasks is quite common today in studies of individual differences, such as for skill differences, and particularly in intervention studies to assess the impact of what was learned during training.

Noteworthy too was bringing a representative task from the domain into the lab, performance of a Bach Prelude, and the attempt to assess quality of the interpretation, paying careful attention to reliability of measurement across judges, as well as assessing the consistency of the dynamic changes across repetitions. Use of this type of task (with obvious face validity) provided a standard for many other studies of expert performance. Looking for consistency across trial repetitions highlighted Ericsson and colleagues' definition of expert performance: *consistently* superior performance for representative tasks from the domain. One of the truly challenging aspects of studying expert performance is finding ways to bring representative tasks into the laboratory, somewhat easier to do with domains like music performance or chess playing, than with athletic performance, particularly for team sports.

Nonetheless, piano soloists do not always play solo pieces, so tasks such as coordinating performance with other musicians (e.g., in an orchestra, or a chamber music group) represent an untapped dimension of music performance in this highly cited study. Similarly, letting one (simple) piece stand for the full repertoire that performers typically master in classical music has risks too, though even here two of the amateurs were dropped from consideration because of their difficulty performing it fluently enough for three repetitions. This attrition underscores the challenge of meeting Ericsson's criterion of finding representative tasks that individuals across the skill range can perform, and explains why researchers usually opt for novices, not beginners, as a control group in skill comparison studies.

Unresolved Issues with Representative Tasks and Decomposition of Skills

Defining representative tasks within the diverse set of domains that occupy expertise researchers is still more an art than a science, though having experts rate prospective tasks for representativeness may offer a path forward. (See the range of domains discussed in the *Handbook of Expertise and Expert*

Performance, *2nd Edition*, by Ericsson et al., 2018.) Nonetheless, there is some inherent circularity in using experts to rate tasks as representative if the domain itself suffers from confusion about defining who is an expert (e.g., difficulties in determining if there is investment expertise, or how to measure coaching expertise).

Even in domains where expertise has a validated scale with interval properties (e.g., Elo ratings in chess), deciding on representativeness of a task often requires expert judgments about domain situations. For instance, the choose-a-move task from an unfamiliar chess position has many facets. The task of choosing the best move on its face is a highly representative task for playing chess, given that choosing the best move is iterated from starting position to final position in real games. Still, which part of the chess game should positions be taken from? If you choose positions from the opening stage of a game, players may not be equally familiar across different skill levels. Worse yet, as Bilalić et al. (2009) have shown, move quality can vary drastically for openings that are within or outside of even a strong player's opening repertoire. Player knowledge can be very specialized even at the same level of skill.

Thus, the tradition has been to present unfamiliar middle game positions to players for solving the problem of choosing the best move. Even after fixing the type of position, what time controls should be allowed for choosing a move? In most of 2020, during the Covid-19 pandemic, in-person classical chess tournaments featuring standard tournament chess all but ceased. In standard tournament play players have about 2 hr each to make the first 40 moves in a game. Some chess tournaments moved online and changed to fast time controls such as rapid: 10-60 minutes/game; blitz: < 10 minutes/game, and bullet: <3 minutes/game, all variants of chess. In other words, time controls shifted from averaging about 180 s per move for normal tournament chess to about 10 s per move for blitz. Although skill levels at these four different time controls are likely significantly correlated, they are not perfectly so. For instance, I analyzed rating data for different formats from a sample of the 37 top-rated players in the world (taken from https://2700chess.com on Dec. 22, 2020). Performance at the two slowest time controls, classical and rapid ratings, are significantly more highly correlated ($r = .67$) than classical and blitz ratings ($r = .39$), $z = 2.199$, $p = .014$ (with the two faster controls rapid and blitz ratings correlated, $r = .56$). Hence, having data about speeded decision making seems necessary to fully capture expertise in choosing a move in chess given that in high-level play today ties in slower classical tournaments are often decided by rapid chess or blitz chess mini matches.

The idea of decomposition of complex cognitive tasks into elementary components helped initiate the cognitive revolution (e.g., Chase, 1978; Sternberg, 1975) and eventually led to models that could make quantitative predictions for the time to complete routine cognitive tasks (Card et al., 1983). At the same time, attempts were made to blend the individual difference correlation approach to task decomposition (Sternberg, 1977). The Ericsson et al. (1993) paper represents an attempt to use inter-task correlations to uncover skill-related and skill-unrelated components for complex tasks. This approach raises interpretative concerns. For instance, tapping speed, though faster in expert than amateur pianists, did not predict group membership once accumulated deliberate practice levels were accounted for in a multiple regression. However, the same multiple regression procedure showed that complex movement coordination remained a significant predictor after accounting for accumulated practice. This result led the authors to conclude that tapping speed, a potential genetically determined capability, was probably more strongly environmentally influenced (by deliberate practice), hence was not a critical factor in piano performance expertise. Similarly, the non-significant correlations between expertise and both digit-symbol substitution and choice reaction time suggested that domain-general cognitive and motor capabilities (candidate abilities for being genetically determined) played little role in expertise.

Such correlates approaches suffer from several weaknesses. First, because expertise is rare and samples tend to be small, siding with the null hypothesis, the failure to find a significant correlation, is risky for drawing inferences about the role of basic abilities. Further, as Ackerman (1992) showed for skill acquisition in an air-traffic control task over a short, but intense training period (22 hours of practice rather than thousands of hours of practice for both amateur and expert pianists), different cognitive abilities predicted performance early (e.g., general ability) versus late (psychomotor ability) in training. That is, skill difference studies run the risk that the predictor set that they examine may not be valid for skill levels other than those sampled, here, young amateur pianists with significant musical experience and young piano soloist prospects.

Such difficulties are not insurmountable. If the researcher is willing to examine relationships across a broader range of skill, at the risk of capturing non-asymptotic performance from those still developing their skill, then larger sample sizes that are adequately powered to detect small effect sizes become possible (e.g., Burgoyne et al., 2019; Charness et al., 2005). It should be noted that skill usually varies over an enormous range when it comes to expert performance. In chess, Elo scale ratings vary between about 1000 for a beginning tournament player and about 2800 for a world champion, with a standard deviation of about 200 points, representing a range of about nine standard deviations. Hence, it is not surprising that early studies with just three chess players, a novice, an expert, and a master, were able to identify important cognitive mechanisms underlying chess skill, such as the role of chunks (Chase & Simon, 1973). Similarly, sampling a wider variety of tasks to correlate with expertise, particularly ones that might be expected to be stable (genetically determined limits?) versus ones that can be expected to be reactive to training, can help identify ability-level facets of expertise.

Long-term Working Memory

The publication on long-term working memory by Ericsson and Kintsch (1986) is notable for extending Chase and Ericsson's (1982) skilled memory theory by synthesizing many difficult to explain phenomena about memory performance from both the expertise and the reading comprehension literatures. I focus briefly on the ideas that they presented in the context of explaining expert performance. One important goal for any theory of expert performance is to account for how experts seemingly circumvent the temporary storage limitations of working memory capacity (e.g., Miller's [1956] famous 7 plus or minus 2 chunks).

Given the still ongoing debate about the relative contributions of nature versus nurture in expert performance, it was unclear whether experts represented exceptions to information processing limits, such as working memory capacity of seven plus or minus two chunks. Chase and Simon's work (1973) seemed to indicate that chess experts did not have exceptional memory, and that differential chunk size could explain their recall performance advantage over novices when asked to reproduce a briefly seen unfamiliar structured (as opposed to random) chess position. However, it was clear that chunking theory needed to be tweaked to account for ability to recall random material after training, such as for the case of digit span experts, or even for recall of chess positions following a period of interpolated processing that should have wiped out information thought to have been stored in short-term memory (Charness, 1976).

The Ericsson and Kintsch theory postulated that experts and skilled readers generated integrated memory representations in long-term memory, not in short-term working memory. The long-term memory retrieval structures that experts used consisted of some combination of (1) simple domain-specific hierarchical structures with retrieval cues associated with units of encoded information (their Figure 1), and (2) more complex structures consisting of knowledge-based associations that linked patterns and schemas via encoded associations to units of newly encoded information (their Figure 4).

The construct of long-term working memory retrieval structures provided a promising explanation for many amazing recall and computational feats, such as digit-span experts recalling 100 rapidly presented random digits, chess grandmasters recalling the exact sequence of moves from a just played chess game of about 80 plies (40 moves each for white & black sides), or lightning calculators quickly solving multi-digit multiplication problems mentally.

Unresolved Issues with Long-term Working Memory Theory

Views of memory structures and processes have changed significantly since the Ericsson and Kintsch (1995) publication. One of the enduring problems of information processing theories has been the potential trade-offs between invariant structures (hardware) and acquired knowledge (software) in the models. One can account for performance data by postulating a relatively simple memory structure, a single limited-capacity short-term memory, coupled with relatively complex symbolic data structures that contain encoded information from different modalities (vision, hearing). Or one could propose a more complex working memory structure composed of several different memories, such as Baddeley's working memory system comprised of sub systems: an articulatory loop and a visuo-spatial sketchpad that are specialized for their respective input modalities.

There continues to be a fractionation of memory systems from the so-called modal model of Atkinson and Shiffrin (1968) to the many different variants seen today. Studies have shown that individual differences in executive function, the ability to control attention, may underly some aspects of working memory capacity and higher order cognitive processes (Engle, 2018). Further, advances in neuroscience now permit construction of models of memory via fMRI data at the brain structure level, for instance, the Nee and Jonides (2011) model postulating different roles for prefrontal and hippocampal structures in short and long-term memory processes.

Long-term working memory theory was conceptualized in a somewhat vague way given its aim of accounting for a very broad range of memory phenomena from expert performance to reading comprehension processes, hence it has never been very easy to test. Nonetheless, the potential importance of identifying hypothesized long-term working memory retrieval structures provides a strong incentive to researchers to use experimental manipulations to trace out the mechanisms supporting skilled performance.

Future Directions

If we are to build on Ericsson and colleagues' achievements in understanding the development of expertise in the context of cognitive mechanisms that may be changing with different types of practice (Ericsson & Harwell, 2019), there are better tools available than a correlates approach. Among these methods are small scale intensive studies of a single individual or a few individuals as they progress over an extended period, for instance, digit span experts (Ericsson, Chase, & Faloon, 1980) or memorizing experts (Ericsson et al., 2004). Once a potential model for performance has been hypothesized, such as the role of encoding techniques and memory retrieval structures, the experimenter can systematically manipulate the materials to be memorized and show that the manipulation degrades performance in the expected direction.

One good example was the case of memory training for a college track athlete who used his prior knowledge of running times to help recode long sequences of random digits into meaningful chunks. When the digit sequences were manipulated by Ericsson and colleagues to include those that could not be represented as running times, digit recall dropped strikingly. Similarly, when switching a memory expert who was skilled at memorized long sequences of digits to memorizing letters, memorizing efficiency showed a precipitous decline, until he learned new techniques for recoding letters into digits.

Cause and effect manipulations are ultimately much more powerful tools for identifying within individual mechanisms that can explain skilled performance than looking at associations such as inter-individual differences in cumulative deliberate practice, birth month, handedness, numerical ability, etc. These intensive study approaches are not perfect either, as one cannot tell whether the mechanism identified is specific to a "talented" individual (such as the college runner SF) who persisted in the study or generalizes to most people. For instance, there were two dropouts from the second digit span study conducted by Ericsson and his colleagues in which the study participant DD reached a span of 100 digits (Ericsson & Staszewski, 1989). However, the major advantage of such studies is that they can track down specific mechanisms, such as encoding strategies and retrieval structures, and build testable theories about task performance. When the goal of the research is theory building, having more detailed theories than Ericsson seemed comfortable with, for instance, implemented as computer simulations (Richman et al., 1995), provides a distinct advantage. When the model can learn the task similarly to the human, it at last partly resolves the sufficiency portion of a "necessary and sufficient" causal explanation for a phenomenon.

In my interactions with Anders, he clearly preferred theories that made ordinal-level predictions, arguing that quantitative models almost always were going to lead to untenable predictions. That preference may have represented a fear that the baby (theory) would too easily be thrown out with the bathwater (easily disconfirmed quantitative predictions). I vaguely remember having to push to introduce the term "outliers" as a way of quantifying expert performance levels in Ericsson and Charness (1994).

From the perspective of developing a field, a skill difference approach initially can be beneficial in much the same way that epidemiological analyses can identify promising targets for more detailed study. An experimental follow-up of cognitive correlates can lead to identifying cognitive mechanisms in much the same fashion as the identification of a geographic source for cholera outbreaks can lead to the identification of specific pathways (e.g., civic water supplies) that enable pathogens to transmit the disease. Both have their value, and Ericsson and colleagues offered salutary examples of each approach.

One problem with intensive studies, particularly if a researcher wants to sample more than a few people, is expense, requiring grant funding. Perhaps surprisingly for someone so productive, Ericsson did not hold many externally funded research grants. In conversations with him over the early years he spent at Florida State University, I had asked him why he did not apply for grant funding very often. (He did have continuous, but modest funding through his FSU Conradi Eminent Scholar account.) He noted that he had observed colleagues at Colorado spending what he felt was too much time on grant-related activity that was not research-related, so seeking grants was generally not a good use of his time. As it turns out for many of his keen observations, research studies carried out in the past decade or so suggest that he was right. Survey studies conducted by the Federal Demonstration Partnership show that federally funded researchers in the USA report spending up to 44% of their time meeting requirements of the grant rather than conducting active research: https://thefdp.org/default/assets/File/Documents/FDP%20FWS%202018%20Primary%20Report.pdf.

Nevertheless, given the resource intensive nature of experimental work for longer term skill acquisition studies, more funding would clearly be welcome for advancing a research agenda to uncover the mechanisms supporting expert performance. Despite weak evidence to this point about the influence of innate factors in skill acquisition, external funding will be critical for exploring hypothesized neural underpinnings and any associated genetic mechanisms supporting advancement through deliberate practice (e.g., development of prodigies: Marion-St-Onge et al., 2020). Funding will also be essential for supporting detailed studies of practice regimens, particularly for understanding the feedback loops between student and coach/teacher that Ericsson hypothesized were necessary for promoting efficient progress along the road to excellence. While Ericsson can no longer participate in the endeavor, the future of expertise research looks particularly promising in a world where excellence, and the ability to go beyond what others have achieved, remains an increasingly prized attribute for our society.

Endnote

The study was quasi-experimental because the variable, skill, was an observed variable, whereas the specific tests given to participants were manipulated variables.

Author Note

This work was conducted while the author was supported in part by a grant from the National Institute on Aging, under the auspices of the Center for Research and Education on Aging and Technology Enhancement (CREATE), 4 P01 AG 17211.

ORCID iD

Neil Charness
https://orcid.org/0000-0002-1002-3439

Author's Declarations

The author declares there are no personal or financial conflicts of interest regarding the research reviewed in this article.

The author declares that the research review reported in this article was conducted in accordance with the Ethical Principles of the *Journal of Expertise.*

References

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: dynamics of ability determinants. *Journal of Applied Psychology, 77*, 598-614. https://doi.org/ 10.1037/0021-9010.77.5.598

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence, *The psychology of learning and motivation: II.* Academic Press. https://doi.org/10.1016/S0079-7421 (08)60422-3

Baddeley, A. D. (1986). *Working memory*. Oxford University Press.

Bilalić, M., McLeod, P., & Gobet, F. (2009). Specialization effect and its influence on memory and problem solving in expert chess players. *Cognitive Science, 33*, 1117–1143. https://doi.org/10.1111/j.1551-6709.2009.01030.x

Braithwaite, D. W., & Siegler, R. S. (2020). Putting fractions together. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000477

Brown, J. S., & Van Lehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science*, *4*, 397-426. https://doi.org/10.1207/s15516709cog0404_3

Burgoyne, A. P., Nye, C. D., Charness, N., & Hambrick, D. Z., (2019). The Impact of Domain-Specific Experience on Chess Skill: Re-Analysis of a Key Study. *The American Journal of Psychology, 132,* 27-38. https://www.jstor.org/stable/10.5406/amerjpsyc.132.1.0027

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Lawrence Erlbaum Associates.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory, 2*(6), 641–653. https://doi.org/ 10.1037/0278-7393.2.6.641

Charness, N. (1981). Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 467-476. https://doi.org/ 10.1037//0096-3445.110.1.21

Charness, N., & Campbell, J. I. D. (1988). Acquiring skill at mental calculation in adulthood: A task decomposition. *Journal of Experimental Psychology: General, 117*, 115-129. https://doi.org/10.1037/0096-3445.117.2.115

Charness, N., Tuffiash, M., Krampe, R., Reingold, E. M., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology, 19,* 151-165. https://doi.org/10.1002/acp.1106

Chase, W. G. (1978). Elementary information processes. In W.K. Estes (Ed.) *Handbook of learning and cognitive processes,* Vol. 5 (pp.

19-90). Lawrence Erlbaum Associates.
Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation, Vol. 16* (pp. 2-58). Academic Press. https://doi.org/10.1016/S0079-7421(08)60546-0
Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55-81. https://doi.org/10.1016/0010-0285(73)90004-2
Cronbach L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671–684. https://doi.org/10.1037/h0043943
Duncker, K. (1945). On problem solving. *Psychological Monographs, 58* (5), ix-113. https://doi.org/10.1037/h0093599
Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science*, *13*(2), 190-193. https://doi.org/10.1177/1745691617720478
Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*, 725-747. https://doi.org/10.1037/0003-066X.49.8.725
Ericsson, K. A., & Charness, N. (1995). Abilities: Innate talent or characteristics acquired through engagement in relevant activities? *American Psychologist, 50*, 803-804. https://doi.org/10.1037/0003-066X.50.9.803
Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science, 208*, 1181-1182. https://doi.org/10.1126/science.7375930
Ericsson, K. A., Delaney, P. F., Weaver, G., & Mahadevan, R. (2004). Uncovering the structure of a memorist's superior "basic" memory capacity. *Cognitive Psychology*, *49,* 191–237. https://doi.org/10.1016/j.cogpsych.2004.02.001
Ericsson, K. A., & Harwell, K. (2019). Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: Why the original definition matters and recommendations for future research. *Frontiers in Psychology, 10*, 2396. https://doi.org/10.3389/fpsyg.2019.02396
Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211-245. https://doi.org/10.1037/0033-295X.102.2.211
Ericsson, K. A., Hoffman, R., Kozbelt, A., & Williams, M. (Ed.). (2018). *Cambridge handbook of expertise and expert performance, 2nd Edition.* Cambridge University Press.
Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251. https://doi.org/10.1037/0033-295X.87.3.215
Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Bradford Books.
Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). MIT Press.
Ericsson, K. A., & Staszewski, J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235–267). Erlbaum.
Fox, M. C., & Charness, N. (2010). How to gain eleven IQ points in ten minutes: Thinking aloud improves Raven's Matrices performance in older adults. *Aging, Neuropsychology and Cognition*, *17*, 191-204. https://doi.org/10.1080/13825580903042668.
Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137,* 316-344. https://doi.org/10.1037/a0021663
Gardner, H. (1995). Why would anyone become an expert? *American Psychologist, 50*, 802–803. https://doi-org.proxy.lib.fsu.edu/10.1037/0003-066X.50.9.802
Jarecki, J.B., Tan, J.H. & Jenny, M.A. (2020). A framework for building cognitive process models. *Psychonomic Bulletin & Review*, *27***,** 1218–1229. https://doi.org/10.3758/s13423-020-01747-2
Lashley, K. S. (1923). The behavioristic interpretation of consciousness ii. *Psychological Review, 30*, 329–353. https://doi.org/10.1037/h0067016
Lewis, J. R. (2012). Usability testing. In G. Salvendy (Ed.) *Handbook of human factors and ergonomics* (*4th Ed.*) (pp. 1267-1312). John Wiley.
Macnamara, B. N., & Maitra, M. (2019). The role of deliberate practice in expert performance:

Revisiting Ericsson, Krampe, & Tesch-Römer (1993). *Royal Society Open Science.* 190327. http://dx.doi.org/10.1098/rsos.190327

Marion-St-Onge, C., Weiss, M. W., Sharda, M., & Peretz, I. (2020). What makes musical prodigies? *Frontiers in Psychology, 11*, 566373. https://doi.org/10.3389/fpsyg.2020.566373

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97. https://doi.org/10.1037/h0043158

Nee, D. E. & Jonides, J. (2011). Dissociable contributions of prefrontal cortex and the hippocampus to short-term memory: Evidence for a 3-state model of memory. *NeuroImage, 54,* 1540-1548. https://doi.org/10.1016/j.neuroimage.2010.09.002

Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259. https://doi.org/10.1016/j.concog.2006.09.004

Richman, H. B., Staszewski, J. J., & Simon, H. A. (1995). Simulation of expert memory using EPAM IV. *Psychological Review, 102*, 305-330. https://doi.org/10.1037/0033-295x.102.2.305

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Lawrence Erlbaum Associates.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27*, 1-32. https://doi.org/10.1080/14640747508400459

Submitted: 31 January 2020
Revision submitted: 7 February 2021
Accepted: 10 February 2021