Journal of Evolutionary Psychology, 2011, 249-269
DOI: 10.1556/JEP.9.2011.3.3

PREHISTORIC STONE TOOLS, CHESS EXPERTISE, AND COGNITIVE EVOLUTION: AN EXPERIMENT ABOUT RECOGNIZING FEATURES IN FLINT DEBITAGE

YVAN I. RUSSELL[1,2,*]

[1] British Academy Centenary Project, Institute of Cognitive and Evolutionary Anthropology, University of Oxford
[2] Centre for the Study of Expertise, Department of Psychology, Brunel University

Abstract. Prehistoric stone tool knapping was an expert skill. This experiment was a cognitively demanding test of modern day knappers using an adaptation of the classic Chase and Simon paradigm from chess expertise (where chess experts/novices are briefly exposed to chess positions and later asked to recall the patterns). Here, pieces of flint debitage were used instead of chess pieces, and it was a recognition test rather than recall. Expertise was measured by social status and questionnaire. Three participant groups were tested (archaeology professionals, students, and non-experts) in 15 trials each, each comprising four tasks: (1) sorting task, (2) exposure, (3) sorting task, and (4) recognition task. The sorting task (interference) required participants to sort flint debitage by size into different buckets. In the exposure task, the experimenter showed the participant three types of rock (flakes, miscellaneous rocks, cores) seriatim for 2 seconds each. The recognition task required that the participant attempt to identify previously seen rocks on three tables (flakes, miscellaneous, and cores tables). Experts performed significantly better than students and non-experts. Post-session interviews revealed a diversity of strategies, suggesting that increased expertise enhanced perception. This result parallels chess expert studies, and template theory in chess might apply to knappers.

Keywords: cognitive evolution, expertise, knapping, chess, archaeology

INTRODUCTION

Expert knowledge is a wonderful tool. It allows a person to observe stimuli and derive information that non-experts cannot. An expert radiologist, for example, can look at an x-ray of a human lung and perceive it far differently than a non-expert (MANNING et al. 2006). When asked to look for pulmonary nodules (potential cancerous growths), non-experts spend a significantly longer time looking at the x-ray, yet find fewer genuine growths than experts can *(ibid.).* People are not born with expertise. As demonstrated through decades of psychological research, expertise is based on the development of a highly structured long term memory (LTM) – some-

---

[*] This research was conducted for an MA dissertation in Cognitive Evolution, Department of Archaeology, University of Reading, UK. Address for correspondence:
E-mail: yvanrussell@gmail.com

thing which may take thousands of hours to acquire (ERICSSON et al. 1993, 2006; ERICSSON and CHARNESS 1997; GOBET 1998; GOBET et al. 2011). Yet, an expert's memory is more than a passive memory bank (STERNBERG 1997). It extends directly into the perceptual world, whereupon the expert can apply superior pattern recognition skills to relevant stimuli (e.g. GOBET and SIMON 1996a) in order to make optimal decisions (see examples in ERICSSON et al. 2006). Human expertise occurs in vastly heterogeneous domains (*ibid.;* GOBET et al. 2011), some of which are more cognitively-based (e.g. chess, mathematics, medicine, science, computer science, writing, law, history, etc.) and some of which are perceptual-motor, involving bodily movement (sports, music, dance, aviation, etc.; cf. SHALIN et al. 1997). As dissimilar to each other as the above categories of expertise may seem, the underlying cognitive organization is generally the same, even if the details are different (GOBET et al. 2011). The human mind is adaptable, allowing expertise to develop about any domain pertinent in the environment, given the opportunity to learn and practice extensively (*ibid.;* ERICSSON and CHARNESS 1997; ERICSSON et al. 2006).

Human beings appear unique in the ability to develop expertise. Although animals might, arguably, possess rudimentary forms of expertise (HELTON 2005), only human beings benefit from the ratchet effect (TOMASELLO 1999): where cultural accumulation allows successive generations to improve upon the skills developed in prior generations. However – expert or not – animals engage in 'niche construction' (LALAND 2007): they behave in ways that alter their ecological niches, and therefore alter the evolutionary selective pressures upon themselves. One way to alter the environment is through tool use, something that a variety of animals do (SEED and BYRNE 2010); but only the human lineage can be defined by its propensity to develop advanced technologies ancestrally (GAMBLE 2007) as a deliberate means of enhancing their own survivability (LALAND 2007). Human evolution can be traced through evidence of increasingly sophisticated technologies and institutions (GAMBLE 2007), and this provided the ideal cultural niche for expertise to become very important. Societies that enabled the intellectual division of labour were those that enabled expertise: they allowed a person to devote considerable mental energies to a single domain. How can we study the evolution of human expertise? In the current experiment (RUSSELL 1999), the goal was to investigate the expertise of extinct hominids through a novel application of expertise research to the archeological record, specifically to that of stone tool technology.

Stone tools manufactured during the lower Paleaolithic (2.5mya–100kya) – particularly those of the Acheulean industry – are clear examples of tools that required their makers to possess a highly sophisticated expert skill in order to create them (SCHICK and TOTH 1993; GOWLETT 1996; WYNN 1989, 2002; HARLACKER 2006; WINTON 2005; COOLIDGE and WYNN 2004, 2009; WYNN and COOLIDGE 2010). The act of knapping involves the systematic reduction of a large core by striking it with a hammer (usually a harder stone, but sometimes softer implements like antlers) and removing flakes (for a practical guide, see LORD 1993). The flakes themselves are useable as tools for butchery (SCHICK and TOTH 1993). Knapping is not a simple task. Non-expert knappers typically make various mistakes that may render the core unusable (see examples in LORD 1993; WINTON 2005). Experts need to be aware of how the stone will "react" to their hammer blows (cf. BRIL et al. 2010). Based on a learned knowledge of the likely internal composition of the stone, the "maker must be able to maintain a correct edge angle on the core, strike it with the required force at the right point, and judge where to strike" (GOWLETT 1996, p. 196) in order to produce a viable tool of the desired size and shape (for useful diagrams, see LORD 1993, Figs. 3–4).

Cognitive archaeology – the attempt to infer the cognitive abilities of extinct hominids by examining the artifacts that they created (for recent overviews, see DE BEAUNE et al. 2009; COOLIDGE and WYNN 2005, 2007, 2009; WYNN and COOLIDGE 2004, 2010) – has been enhanced by the practice of "experimental archaeology", which is the academic practice of reconstructing the usage of artifacts and by manufacturing replicas (SCHICK and TOTH 1993). This approach has been applied to stone tools, whereupon modern day knappers have taught themselves to manufacture prehistoric tools using the presumed original methodology (see SCHICK and TOTH 1993, LORD 1993). Experimental studies have shown that knapping is unquestionably an expert skill. Novice toolmakers – those who lack the necessary hours of practice compared to experts – consistently fail to produce stone tools with the regular forms consistent with the larger tradition (WINTON 2005; HARLACKER 2006). For example – in reference to handaxes – WINTON (2005) observed that novices approached the knapping task in a more disorganized fashion (not thinking of an overall plan), and were unable to create the desired extent of thinness and sharpness (the extent useful for actually using the tool for butchery) because they had not mastered their knowledge of how to strike the core in the appropriate angles and force given the original shape and composition of the core (cf. BRIL et al. 2010). Consequently, the novices produced tools that were too short, thick, irregular, and rough surfaced. Similarly, HARLACKER (2006) found that advanced knappers were able to use their know-how to extract useable material from a core far more efficiently than novices could (producing more and larger flakes from cores, making better use of the constraints of a core that is continually reducing in size).

Given the obvious level of expert skill inherent in knapping, it would be beneficial to investigate the mind of an expert knapper from the perspective of psychology. This has been done before by archaeologist Thomas Wynn. In his book, *The Evolution of Spatial Competence* (1989), he adopted Jean Piaget's framework of cognitive development as a framework for analyzing Oldowan and Acheulean stone tools. He made inferences about the "minimum necessary competence" (MNC) required to manufacture these tools (for a reevaluation of this approach, see WYNN 2002). This study continues Wynn's program of placing stone tools firmly in a psychological framework. However, instead of using Piagetian concepts, this study makes reference to the psychology of expertise. Of the various domains of expertise studies, one of the most extensively studied is chess expertise (see SIMON and GOBET 1996a; GOBET 1998). Chess masters are those who have achieved the very highest levels of skill in the game, performing with a proficiency that drastically surpasses the average. Chess experts are proficient at learning sequences of movement and even entire chess games, and they can remember them better afterwards (CHASE and SIMON 1973b). Although the game of chess may seem remote from the practice of stone tool knapping, there may be similarities in underlying cognitive structure that are worth investigating (see RUSSELL 1999; WYNN and COOLIDGE 2004, 2010; cf. GOBET et al. 2011, p. 238). According to CHASE and SIMON (1973b), "the basic ability underlying chess skill" is "the ability to perceive familiar patterns quickly" (p. 267). Like chess, stone tool knapping is also a highly visual skill, where familiar visual patterns would be seen repeatedly. These patterns are meaningful because they are a link to future actions.

The current experiment was a modification of the well-known chess paradigm of CHASE and SIMON (1973a, b; for reviews, see ERICSSON and CHARNESS 1997; GOBET et al. 2011). They studied chess players at different levels of chess expertise (e.g., "beginner", "class A player", "Master"). In their studies, the participant was asked to recall chess piece positions that he/she had recently been exposed to. For example, in an early study (CHASE and SIMON 1973b), the participant was shown a chess board with a game-on-progress on it (the positions were taken from records of previous chess games, either in middle game or end game positions). Only 5 seconds of viewing time was allowed. After this, the participant was given an empty chess board and some chess pieces, and was asked to reconstruct the chess game that he/she had just seen. The reliable result was that players with more expertise in chess always performed better on this task. But when – during the exposure phase – pieces on the board were arranged *randomly*, the expert advantage was significantly reduced. In order to explain such results, CHASE and SIMON (1973a, b) formulated the *chunking theory*, which explained expert proficiency as occurring because in the mind of the chess expert there is "a large database of chunks" (GOBET 1998, p. 118).

Chunks have become a well-established organising principle in human memory (SIMON and GOBET 2000; GOBET et al. 2001, 2011). GOBET et al. (2001) defined a chunk as "a collection of elements having strong associations with one another, but weak associations with elements within other chunks" (p. 236). This applies to virtually any type of memorising, but – in chess – we define a chunk as a configuration of chess pieces (theoretically, a chunk can be a single piece, but chunks are more useful when they are bigger). As CHASE and SIMON (1973b) describe it, the chess expert's "contents of thought are mainly these perceptual structures that skilled chess players retrieve … from long term memory" (p. 268). In playing the game, the expert sees a configuration of pieces, holds the familiar structures (chunks) in the "mind's eye" and then instantaneously begins to search through a "problem space" (possible future moves) to find a more efficient solution (CHASE and SIMON 1973b). However, the weakness of the Chase and Simon's chunking theory is that it relied too much on short-term memory (STM) as an explanatory step (see GOBET 1998). This was uncovered by psychologists (e.g. CHARNESS 1976) who interposed an 'interference task' (an unrelated cognitive task) to interfere with STM during the test. They found that the memory performance of chess experts was not substantially hindered by the interference task. This suggested that STM played a less crucial role than previously thought. Another problem with the chunking theory (see GOBET 1998) was that it lacked an organizational structure in long-term memory (LTM): chunking theory posits a large database of 'chunks' that are individually retrieved; but there is no proposal concerning how these chunks are organized.

Because of these problems, GOBET and SIMON (1996a) offered a refined version of the chunking theory called the *template theory* (also see GOBET 1998; GOBET et al. 2011). In some ways, template theory is the same as chunking theory. Like the chunking theory, template theory proposed that a pattern of chess pieces is recognised by largely by-passing STM and instead using a 'discrimination net' to permit rapid access into long-term memory (LTM). But template theory goes a step further by positing that observed chunks can be subsumed into larger mental templates. A chunk might only be one variation of a larger template. Each template consists of a basic configuration at its core. The precise position of pieces may vary. The changeable components of the templates are called *slots*, which reflect what the expert is actually seeing on the chessboard. The template is the core pattern in the chess expert's mind that can be linked to what they are seeing. The previously learned template is what enables an expert to 'see' recurrent patterns on the chess board (amidst huge variation in individual pieces and patterns) – which provides huge advantage over the novice. Part of what enables the superior performance is that the burden on short term memory (STM) is considerably reduced, because the relevant patterns can be rapidly matched to pre-existing retrieval structures (templates) in LTM. In a similar fashion to chess masters, expert stone tool knappers should possess this ability to immediately 'see' visual patterns in the flint that are recurrent during the knapping process (cf. WYNN and COOLIDGE 2004). Instead of configurations of chess pieces, in flint there are familiar patterns of cleavage, breakage, colouration, size, and shape. The specifics of the individual rock should be equivalent to mental 'chunks' but the chunks can be recognized immediately as being a variant of a typical pattern (a 'template') with a set of potential outcomes resulting from particular actions (analogous to forward thinking in chess).

The Chase and Simon paradigm has been found to be widely applicable (SIMON and GOBET 2000), and stone tool technology should be no exception (cf. GOBET et al. 2001). In the current experiment, the general Chase and Simon chess paradigm was copied step-by-step but with a number of innovations relevant to stone tool knapping. The most substantial difference is that – for practical reasons – the test is a recognition test rather than a recall test (more about this issue in the Discussion). In the experiment reported below, the hypothesis was that expert flint-knappers would outperform novices in a difficult recognition task where all participants would be briefly exposed to particular pieces of flint and then asked to choose which pieces of flint they had seen earlier from tables consisting of large numbers of displayed pieces.

PARTICIPANTS

Thirty-six volunteers were recruited from the University of Reading, the British Museum (Franks House, London), and Norfolk, U.K. (during a visit to expert knapper John Lord's house). There were 20 males and 16 females, with a mean age of 31.75 years (range 19 to 57). The participant's expertise was measured later in a questionnaire (see below), but there was a deliberate attempt to recruit professional experts by reputation (cf. HARLACKER 2006). The professional group (n = 7) included stone tool knappers, academic and non-academic archaeologists (mean age = 43.71 years). This includes re-fitters (those who re-fit the debitage to the stone to reconstruct the process). Archaeology students were also recruited (n = 19). It was not possible to presume the level of expertise in this group, because some had studied stone tools closely and some had not (mean age = 28.21). Finally, a group of self-identified non-experts (n = 10) were recruited from among the experimenter's friends and colleagues (mean age = 30.10). These were mostly students in non-archaeology fields.

MATERIALS

Figure 1 illustrates the experimental floor plan. There were five tables: sorting table, table 0 (exposure table), table 1 (flakes), table 2 (miscellaneous), and table 3 (cores). For the experimental stimuli, a large amount of broken flint had been collected in Norfolk, U.K., directly from the workshop of professional flintknapper John Lord. All of these pieces were debitage: they had been conspicuously worked upon by a stone tool knapper, but none of them had been made into a finished tool. From this collection, the debitage were classified into four types: (1) flakes, (2) cores, (3) miscellaneous stimuli, and (4) miscellaneous sorting rocks. LORD (1993) defined a *core* as "[a] carefully prepared piece of material, from which some of the removals are termed as the tools" (p. 18); and he defined a flake as "[any] piece of material detached by striking" (p. 19) (flakes are removed from cores). The pieces were mostly black or grey in colour, with numerous white and grey inclusions. Some pieces had areas of yellow cortex (the original outer casing). Photographs of the flakes, miscellaneous rocks, and the cores tables are shown in *Figures 2, 3,* and *4*, respectively.

Flakes were placed on table 1 (see *Figure 2*). These consisted of 32 rocks with a flake-like appearance. Most of these were genuine flakes (i.e. an actual product of the knapping process), but some were non-genuine (i.e. they happened to be the same shape as flakes for reasons other than knapping). Cores were places on table 3 (see *Figure 4*). These were 14 rocks that had a core-like appearance. Most of them were genuine cores from which small lithic tools have been removed. Miscellaneous stimuli rocks were placed on table 2 (see *Figure 3*). These consisted of 24 rocks that were quite dissimilar to the cores and the flakes. The shapes, sizes, and histories of these rocks were diverse. Some were sharp, angular, and fractured. Others were rounded, unbroken, and with conspicuous cortex (a remnant of the outer shell). Each piece was distinctive in some way. All of the rocks were positioned on the tables in evenly-spaced rows (four rows of flakes, three rows of miscellaneous rocks, and two rows of cores). Excluding the sorting rocks, there were 70 pieces altogether. There was conspicuous variance in size, but they were all roughly the same (about 15 cm in length and 11 cm in width). Every rock on tables 1–3 was secretly assigned an ID number that corresponded to its location of the table (this enabled the experimenter to know at all times which rocks were being used; for details of numbering system, see RUSSELL 1999, pp. 65–66).

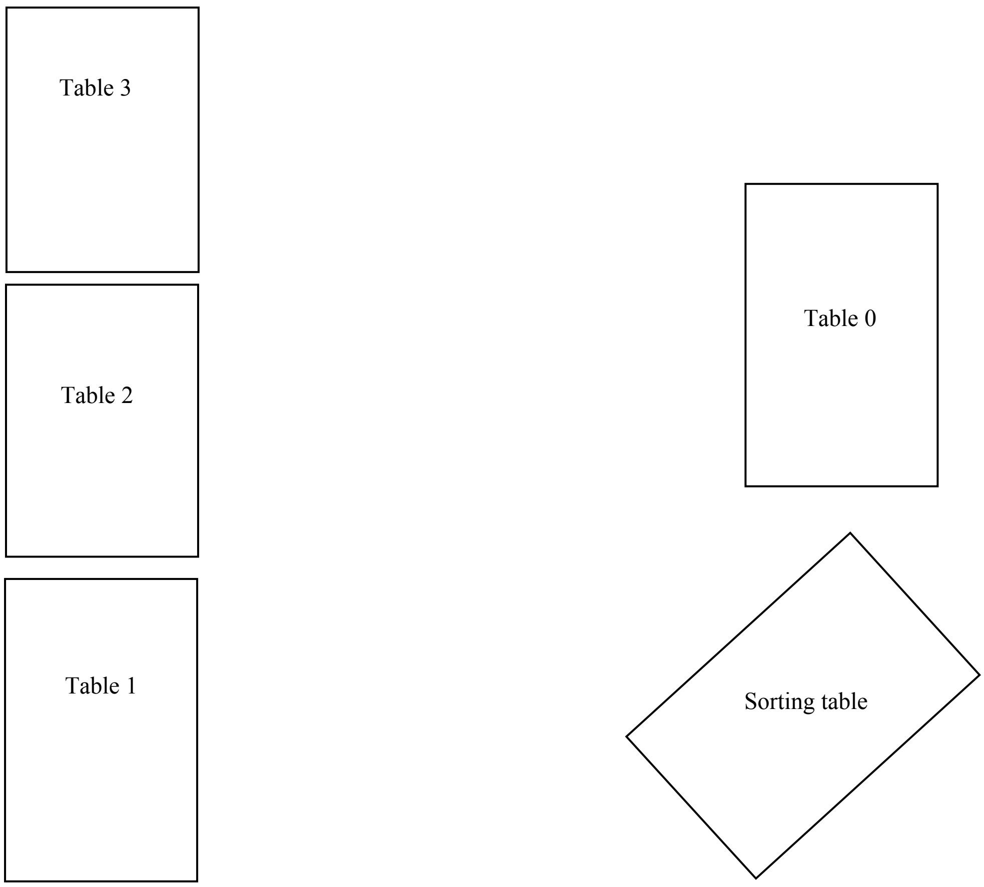

Figure 1. A diagram of the experimental floor plan

Figure 2. Flakes table

Figure 3. Miscellaneous rocks table

Figure 4. Cores table

Figure 5. Sorting table

The sorting table is shown in *Figure 5*. Miscellaneous sorting rocks were irregular pieces of broken flint of widely varying sizes. These were used in the sorting task (on the sorting table). The sorting task required three large identical yellow buckets (about 15 litres each) placed side by side on the sorting table. At the beginning of the study, all of the miscellaneous sorting rocks were in the middle bucket.

A pair of gloves was offered to the participant as optional protective wear. Table 0 was used for the exposure activity. Three upside-down containers were used to cover the pieces of debitage during the exposure activity. These containers varied slightly in size, but were approximately 22 cm in length, 20 cm in width, and 16 cm in height.

The experimenter scored correct or incorrect responses on a data sheet (shown in RUSSELL 1999, p. 139), which was partitioned into three sections: flake, miscellaneous rocks, and cores. On the left column of each section were the ID numbers for the rocks to be used. The middle column was a space to record which rock the participant had pointed to. The rightmost column was used to mark the answer as correct or incorrect. There were 15 rows of flakes, miscellaneous, and cores, corresponding to 15 trials in the study. Also, a new questionnaire was devised to determine the extent of expertise in each participant. There were 33 questions, but some of these questions were later found to be redundant or not useable (see discussion in RUSSELL 1999; for full text of questionnaire, see *ibid.,* pp. 118–129). Hence, only 11 questions are reported here. These are shown in *Table 1* (columns on left).

PROCEDURE

Every participant was individually tested using the same floor plan (*Figure 1*). The participant was told that the study was about "recognizing patterns in rocks". At the beginning of the session, the experimenter identified and allowed the participant to briefly look at tables 1, 2, and 3. The entire session lasted about 40 minutes and consisted of 15 trials. Every trial had four tasks: (1) sorting task, (2) exposure, (3) sorting task, and (4) recognition task. After this sequence was repeated 15 times, the participant was interviewed, given the questionnaire, and then debriefed.

The sorting task was an "interference" task, meant to interfere with the participant's memory during the recognition task. This was modeled after interference tasks in chess expertise, where "[many] types of interpolated activity have been used: counting backwards by 3s and 7s, copying symbols, classifying symbols, and so forth" (CHARNESS 1976, p. 644). The sorting task – created specifically for this study – was conducted at the sorting table (see *Figure 1*). The participant encountered three large yellow buckets *(Figure 5).* The middle bucket was filled with flint debitage of various shapes and sizes. These were acquired from the same source as the stimulus rocks (see Materials), and therefore looked generally the same as the stimulus rocks (although they were more variable in size). The other two buckets were empty. Participants were asked to relocate the rocks according to size: large and small. The large rocks were to be placed in the left bucket; the small rocks in the right bucket. The experimenter offered no criteria for deciding how to classify size (it was entirely the decision of the participant). The experimenter briefly demonstrated the sorting task (using unambiguously sized rocks) and then asked the participant to begin sorting.

After about 10 seconds, the sorting task was interrupted, and the participant was led to table 0. While the participant had been concentrating on the sorting task (with back turned to experimenter), the experimenter had placed one rock under each upside-down container. In each trial, container 1 covered a flake (taken from table 1), container 2 covered a miscellaneous rock (from table 2), and container 3 covered a core (from table 3). The experimenter explained that rocks 1, 2, and 3 were rocks removed from tables 1, 2, and 3, respectively. Then, he lifted up the three containers in sequence, exposing each rock for two seconds. During the exposure, he verbally identified the rocks as either rocks 1, 2, or 3, allowing the participant to know which table it was from. At this time, the participant was not allowed to look at tables 1, 2, or 3.

After the exposure, the participant was asked to continue the sorting task. During the sorting task, the experimenter returned the rocks on table 0 to their original positions on tables 1, 2, and 3. Then, after the participant had been sorting for 10 seconds, the experimenter asked the participant to walk to table 1 (flakes, shown in *Figure 2*). Here, he asked the participant to point to the first rock that he/she had just seen on table 0. There was no specific time limit, but the participant was asked to make the decision quickly and to guess the rock if they were not sure. After pointing to a rock, the experimenter did not indicate if the choice was correct. Then, the participant was led to table 2 (miscellaneous rocks, shown in *Figure 3*) and was asked to point to the second rock from table 0. Then, the participant was led to table 3 (cores, shown in *Figure 4*) and was asked to point to the third rock from table 0. After all three rocks had been pointed out, the participant was led back to the sorting task to begin the cycle anew (sorting, exposure, sorting, recognition).

After fifteen cycles had been completed, the experimenter sat down with the participant and asked them to describe their mental experience in the task. This was an open-ended question, where the experimenter took notes. Then, the participant was asked to fill out the questionnaire. Finally, the participant was debriefed, during which the experimenter explained the purpose of the study, and asked the participant how effective the sorting task had been as a distraction from the recognition task.

RESULTS

Independent variables were the questionnaire items and participant status. Summary questionnaire results are shown in *Table 1* (left columns). The dependent variable was the number of recognitions for flakes, miscellaneous rocks, and cores. Because there were 15 trials, the possible scores were from 0–15. The overall mean score of correct *flake* recognitions (first table on floor plan) was 9.69 (SD = 2.42). The overall mean score of correct *miscellaneous rocks* recognitions (second table) was 10.94 (SD = 2.46). The overall mean score of correct *core* recognitions (third table) was 11.08 (SD = 2.05). The overall mean score of all correct recognitions (maximum score: 45) was 31.78 (SD = 5.58).

For the yes/no questions (see *Table 1*), independent samples *t*-tests were conducted to compare recognition scores against the questionnaire items. *Table 1* (right columns) shows the results for recognition of flakes, miscellaneous rocks, cores, and total. For question 3 (past experience in creating stone tools), there was a significantly higher recognition score for flakes, $t(34) = 2.042$, $p = .049$, but not for miscellaneous rocks $t(34) = –0.141$, $p = .889$, or cores, $t(34) = –389$, $p = .700$. For question 6 (taught someone to manufacture stone tools), there were no significant differences for flakes, $t(34) = 0.889$, $p = .380$, for miscellaneous rocks, t(34) = 0.969, $p = .340$, or cores, $t(34) = –0.108$, $p = .915$. For question 7 (illustrated stone tools), there was a significantly higher score for flakes, $t(34) = 2.442$, $p = .020$, and cores, $t(34) = 2.610$, $p = .013$, but not for miscellaneous rocks, $t(34) = 1.564$, $p = .127$. The overall score (all three tables) was also significant, $t(34) = 2.824$, $p = .008$. For question 8 (re-fitted stone tools), there were no significant differences for flakes, $t(34) = 1.297$, $p = .203$, for miscellaneous rocks, $t(34) = 0.240$, $p = .812$, or cores, $t(34) = –0.181$, $p = .857$. For question 9 (writing about stone tools), there were no significant differences for flakes, $t(34) = 0.892$, $p = .378$, for miscellaneous rocks, $t(34) = 1.417$, $p = .166$, or cores, $t(34) = 0.041$, $p = .968$. For question 10 (*published* writing about stone tools), there were no significant differences for flakes, $t(34) = 1.879$, $p = .069$, for miscellaneous rocks, $t(34) = 1.745$, $p = .090$, or cores, $t(34) = .102$, $p = .893$. However, for the overall score (all three tables), there was a significant difference $t(34) = 2.326$, $p = .026$. There were no significant sex differences for any scores.

For continuous variables (see *Table 1)*, Pearson correlations were used. For recognition test scores for flakes, miscellaneous rocks, and cores, there were no significant correlations for question 1 (self-rated knowledge of stone tools), question 2 (self-rated knowledge of stone tool manufacture), or question 4 (self-rated ability to manufacture stone tools). Age was another continuous variable. Here, there were no significant correlations for any recognition task scores – but surprisingly there was a significant positive correlation between age and question 1 ($r = .461$, $p = .005$), question 2 ($r = .360$, $p = .031$), and question 4 ($r = .734$, $p = .010$). The final continuous variable – question 5 (estimated lifetime number of attempts to manufacture stone tools) – was problematic because the most expert stone tool knapper (John Lord) said "too many to count". For an alternative analysis, this variable was split into three groups: zero attempts ($n = 23$), 1–29 attempts ($n = 6$), and 30 or more attempts ($n = 7$). No significant trends were found in this new analysis.

In addition to using the questionnaire, the recognition scores were also analysed according to the three groups in our initial selection criteria: (1) professional archaeologists and knappers, (2) archaeology students, and (3) non-experts. *Table 2* lists the recognition test scores for these groups. The archaeology students were the most heterogenous in their experience with stone tools (only 42% had ever attempted knapping). Self-reported knowledge of stone tool knapping (question 2) for each group is shown in *Table 2*. For the comparison between professionals and students, there was a significant difference for overall score (all three tables), $t(34) =$

Table 1. Comparison of questionnaire items to recognition task scores

Number	Question	Type of answer	Mean score (SD)	Yes	No	Mean recognition score – Flakes	Mean recognition score – Misc. rocks	Mean recognition score – Cores	Mean recognition score – All (out of 45)
1	Self-rated knowledge of stone tools	Scale 0 – 10	4.19 (1.94)						
2	Self-rated knowledge of stone tool manufacture	Scale 0 – 10	3.64 (2.45)						
3	Past experience in creating stone tools	Yes/No		17	19	Yes: 10.53 (2.53) No: 8.95 (2.12)	Yes: 10.88 (2.57) No: 11.00 (2.43)	Yes: 10.94 (1.75) No: 11.21 (2.32)	Yes: 32.47 (5.52) No: 31.16 (5.70)
4	Self-rated ability to manufacture stone tools*	Scale 0 – 10	1.06 (1.10)						
5	Lifetime number of attempts to make stone tools*	Estimated number	31.08 (30.46)†						
6	Taught someone else to manufacture stone tools	Yes/No		6	30	Yes: 10.50 (3.33) No: 9.53 (2.24)	Yes: 11.83 (2.40) No: 10.77 (2.47)	Yes: 11.00 (1.67) No: 11.10 (2.14)	Yes: 33.50 (6.22) No: 31.43 (5.49)
7	Illustrated stone tools	Yes/No		12	24	Yes: 11.00 (2.63) No: 9.04 (2.07)	Yes: 11.83 (2.08) No: 10.50 (2.55)	Yes: 12.25 (1.54) No: 10.50 (2.04)	Yes: 35.17 (4.59) No: 30.08 (5.32)
8	Attempted to re-fit stone tools	Yes/No		13	23	Yes: 10.38 (3.15) No: 9.30 (1.87)	Yes: 11.08 (2.33) No: 10.87 (2.58)	Yes: 11.00 (2.04) No: 11.13 (2.10)	Yes: 32.62 (6.64) No: 31.30 (4.98)
9	Written about stone tools	Yes/No		21	15	Yes: 10.00 (2.90) No: 9.27 (1.53)	Yes: 11.43 (2.15) No: 10.27 (2.76)	Yes: 11.10 (2.14) No: 11.07 (1.98)	Yes: 32.62 (5.89) No: 30.60 (5.05)
10	Published writing about stone tools	Yes/No		6	30	Yes: 11.33 (3.14) No: 9.37 (2.17)	Yes: 12.50 (1.38) No: 10.63 (2.53)	Yes: 12.33 (1.51) No: 10.83 (2.07)	Yes: 36.33 (4.23) No: 30.87 (5.41)
11	Age	Number	31.75 (10.31)						

* participants who had answered no to question 3 were automatically assigned the value 0.
† missing top value because the most accomplished stone knapper said “too many to count”.

2.296, $p = .031$, but not individually for flakes, $t(34) = 1.833$, $p = .079$, miscellaneous rocks, $t(34) = 1.759$, $p = .091$, or cores, $t(34) = 1.466$, $p = .156$). For the comparison between professionals and non-experts, there was a significant difference for overall score (all three tables), $t(34) = 3.262$, $p = .005$, and also there was a significance difference for flakes, $t(34) = 2.200$, $p = .044$, miscellaneous rocks, $t(34) = 2.358$, $p = .032$, and cores, $t(34) = 2.167$, $p = .047$. For the comparison between students and non-experts, there were no significant differences.

Table 2. Comparison of participant status to recognition task scores

Status	Self reported knowledge of stone tool knapping (scale 0–10)	Mean recognition score – Flakes (Max = 15)	Mean recognition score – Misc. rocks (Max = 15)	Mean recognition score – Cores (Max = 15)	Mean recognition score – All (max = 45)
Professional academics and knappers	6.14 (2.48)	11.43 (2.88)	12.57 (1.27)	12.29 (1.38)	36.57 (3.91)
Students	3.42 (1.92)	9.26 (2.60)	10.95 (2.30)	10.95 (2.25)	31.16 (5.73)
Non-experts	2.30 (2.21)	9.30 (0.95)	9.80 (2.90)	10.50 (1.84)	29.60 (4.60)

Finally, the post-experimental interview provided a wealth of descriptive self-reported information about what the participants paid attention to during the experiment (for individual reports, see RUSSELL 1999, pp. 142–152). There was considerable diversity in the terminology that participants used and how much detail they offered. Many participants described the precise features in the stone to which they attended. These could be partitioned into either *global features* (overall shape and size) or *small details* (markings on the stone). Regarding global features, 52.7% paid attention to overall shape. Only 19.4% paid attention to size (rocks within tables were generally the same size). Regarding the small details, 80.6% of individuals paid attention to the small details. Small details could be further partitioned into either two-dimensional markings (e.g., “spots”, “speckles”, “aureolae”, “chalk marks”, etc.) or three-dimensional topographic features (e.g., “scars”, “indentations”, “raised area”, “flat area”, “ripples”, “pointy bits”, “bulb of percussion”, etc.). 50.0% of all individuals paid attention to two-dimensional markings, whereas only 36.1% noticed the topographic features. 8.3% ignored small markings altogether and paid attention exclusively to global features. Conversely, 27.8% ignored global features altogether and paid exclusive attention to small details. 58.3% paid simultaneous attention to global features and *two-dimensional* small details, but only 22.2% paid simultaneous attention to global features and *topographic* details. Some specific small details included inclusions (foreign objects embedded in the stone, e.g. fossils), cortex, and colour. 25.0% attended to inclusions, 36.1% to cortex, and 44.4% to colour. Some participants mentioned an overall strategy for the task. For example, 27.8% tried to attach names to individual stones and features therein. Non-experts used vague words to describe the features (e.g. "shaped like a horn"), whereas professionals and students used academic words. Another strategy was a chaining strategy (linking together the features of the three rocks), used by 8.3% of participants. Other participants had no strategy at all, simply hoping that the features would "jump out at them". Finally, we asked them to report the effectiveness of the "sorting task" as interference. Here, 16.7% reported that it was very distracting, 66.7% reported mild distraction, and 16.7% reported no distraction at all. Four individuals reported that they stopped paying attention to the distraction task part way through the procedure in order to focus on the recognition task.

The above information was compared against their actual performance in the recognition task. There were no significant differences in performance for any of the above strategies – with one exception: individuals who paid attention to size performed better on table 2 (miscellaneous rocks) than those who did not, $t(21.167) = 2.229$, $p = .037$ (equal variances not assumed). We also compared the self reported strategies against their questionnaire results (*Table 1*). For the strategy of comparing both *overall* and *topographical* features combined, there was a significant difference between individuals who had taught others to create stone tools (question 6, *Table 1*) and those who had not, $t(29.000) = 3.247$, $p = .003$. This was also true among those who had written about stone tools (question 9, *Table 1*) compared to those who had not, $t(20.000) = –3.508$, $p = .002$. The teachers also paid more attention to inclusions, $t(31.521) = 1.127$, $p = .019$. There was also a significant difference between illustrators and non-illustrators (question 7, *Table 1*) in paying attention to overall shape, $t(23.408) = 2.550$, $p = .018$. There was also a significant difference between re-fitters and non-re-fitters (question 8, *Table 1*) in paying attention to cortex, $t(34) = –2.530$, $p = .016$. There were no sex differences in strategies.

DISCUSSION

Experts performed better than non-experts in the recognition task, as defined by their career status. Specifically, professionals outperformed students and non-experts. These results mirrored those of the chess studies. Because no age and sex differences were found, performance can be safely attributed to expertise alone (however, the sample size was perhaps not large enough to find effects for sex). We also found a few differences linked to the participants' replies in the questionnaire (see *Table 1*). Here, we found superior performance among those who had illustrated stone tools and those who had published writing about stone tools. The overall procedure in this experiment was modeled after the CHASE and SIMON (1973a, b) paradigm, but substantially altered in some respects. The most important difference was that our study had a recognition test as the dependent variable – in contrast to Chase and Simon who asked participants to reconstruct previously viewed chess positions on the chess board (i.e. putting the pieces in place). Most studies of chess expertise involve recall rather than recognition. There are a few exceptions to this rule (e.g. GOLDIN 1979, who found that experts are superior at recognition too; cf. ERICSSON et al. 2006, p. 528). GOBET and SIMON (1996b) confirmed that pattern recognition is the crux of a chess master's superior game play (and in fact, allows successful play with minimal forward planning). This supports the idea that recognition, as used here, is also a valid measure (recall and recognition can be just two different ways of remembering the same information). Obviously, a recall experiment would have been more congruent with the Chase and Simon paradigm (and be a tougher test than a recognition task). Also, a benefit of using recall instead of recognition is that novices would have a more difficult time devising deliberate strategies (i.e. a way to achieve high scores without having real expertise). Yet, it would have been impossible to conduct a recall test in the current study (because it would have required participants to actually manufacture new versions of the stone tools). The recognition test in this experiment was innovative, not following the usual form of recognition testing in psychology (see GOBET et al. 2011) due to the unusual apparatus involved.

This experiment was cognitively demanding, and the participant self reports revealed a diversity of coping strategies. Although the open ended nature of the reports was problematic for analysis, it allowed a wealth of useful detail. No strategy was found to be a 'magic bullet' in isolation. The only exception was that those who focused on size performed better on table 2 (miscellaneous rocks). This is probably because the rocks on table 2 varied in size more than the other tables. Successful recognition apparently involved a suite of strategies working together, rather than a single strategy. This is implied in the result that teachers and re-fitters paid attention simultaneously to both global features (shape, size) and *topographical* details on the stone (in contrast to two-dimensional marks). There were also indications of how a person's past experience caused them to differentially pay attention to different features of the stone. Hence, illustrators paid close attention to overall shape, re-fitters paid close attention to cortex, and knapping teachers paid close attention to topography and inclusions. People were guided by their past experiences when attending to features in the rocks. This is illustrated by one professional who proved utterly unable to recognize rocks from table 1 (flakes, shown in *Figure 2*). He explained the cause of this inability: it occurred because flakes were completely unimportant to him in his career (in his words, "the lowest of the low").

What made the procedure cognitively demanding was the need to correctly identify forty-five pieces of flint within tight time frames. The initial exposure time to the rock had been brief (2 seconds), allowing very little time for explicit cogitation. It is this brevity that allowed us to separate experts from non-experts. Part of what enables the superior performance of an expert is that the burden on short term memory (STM) is considerably reduced, because the relevant patterns can be rapidly matched to pre-existing retrieval structures (templates) in LTM (GOBET and SIMON 1996a). This is the same rationale as in chess experiments, where exposures to chess stimuli are brief (e.g. CHASE and SIMON 1973a, b used 5 second exposures). Interestingly, some of the highest scoring archaeologists and knappers claimed to have no explicit strategy at all when making their recognition choices. Apparently, they were relying on their considerable pattern recognition skills to do the work for them. Other high scorers *did* use explicit mental strategies. The highest scorer in the group (91% correct) described her strategy as based on attending to many features at once: global features (size, shape), small details including detailed labeling of features (e.g. inclusions were either "cherty" or "smooth"), fractures distinguished by how they were formed, and the presence or absence of ripples (see RUSSELL 1999, pp. 145–146). In her case, she benefited from the ability to combine an explicit naming strategy with her implicit pattern recognition skills. Most experts, however, were able to put forth very little mental effort and still attain a high score. This is in contrast to many of the novices who put considerable effort into the task, but could not score as highly as the experts.

The demands of this experiment can be interpreted from the perspective of 'cognitive load theory' (VAN MERRIËNBOER and SWELLER 2005), studied in educational psychology. The procedure in the current experiment was deliberately designed to present a very high cognitive load on the participant. Furthermore, there was performance pressure (cf. BEILOCK 2008): the participant was under some extent of additional stress due to the experimenter's tacit expectations for the participant to perform well. Research in performance pressure *(ibid.)* has shown that stressful situations have the greatest impact upon those who need to rely on their STM (instead of LTM) in order to solve a task. The capacity of STM (or working memory, WM) is not what separates experts from non-experts: all humans have the same limitations in that respect (GOBET and SIMON 1996a; VAN MERRIËNBOER and SWELLER 2005). Increasing the cognitive load (e.g. number of elements presented simultaneously) causes problems for non-experts because it increases the burden on STM (VAN MERRIËNBOER and SWELLER 2005); whereas for the experts the burden on STM is minimal because the stimuli is rapidly matched to LTM structures *(ibid.;* GOBET and SIMON 1996a; GOBET et al. 2011).

What exactly are the chunks and templates in these stones? Important to remember is that the focus of the current study was on the features *within the stone*, rather than between stones. This means that a 'template' could be no bigger than a single stone itself. Admittedly, there are still many unknowns in this researcher's attempt to identify chess-equivalent chunks and templates within these stones. For chess, it is easier to identify a perceptual chunk: it is a specific two-dimensional arrangement of chess pieces (GOBET 1998; SIMON and GOBET 2000; GOBET et al. 2011). In contrast, the features of the stone are in three dimensions, including the unseen interior of the rock (see WYNN 1989). Moreover, the recurrent 'chunks' in pieces of knapped flint have not been catalogued by the way chess piece positions have. This is especially true because most of the stimuli in the current experiment consisted of debitage (pieces that had been discarded), rather than finished tools. This means that even experts will have frequently seen novel combinations of features in this study. Nonetheless, superior performance amongst the expert was shown. This is consistent with research in chess expertise, showing that experts have superior memory even for random combinations (GOBET et al. 2001), because the pieces often line up coincidentally into familiar patterns. In a knapping context, we could identify a 'chunk' (equivalent to a chess chunk) as a particular combination of features of the rock, incorporating 'global' aspects of shape and size but also outward features of the stone that give clues about the internal composition of the stone, and the way that the rock has already been knapped. A chunk will be equivalent to any pattern seen on a rock (e.g. inclusion + certain type of fracture + certain type of edge). In the current study, novices more often focused on individual features, ignoring other aspects of the stone surface. This is equivalent to looking at a chunk with only one 'piece' (e.g. an inclusion by itself). This is congruent with research by CHASE and SIMON (1973b), who found that beginning chess players utilized much smaller chunks than experts (often consisting of only a single piece) because "beginners don't have access to many patterns in long-term memory" *(ibid.* p. 254). According to the older chunking theory, the chess expert possesses in his mind "a large database of chunks, indexed by a discrimination net" (GOBET 1998, p. 118). The discrimination net is the mechanism that allows rapid recognition, because it functions to rapidly recognize patterns and then link them to patterns stores in long term memory. Non-experts simply have not developed this discrimination net, and therefore their exposure to the lithic patterns placed a huge burden on their working memories (cf. VAN MERRIËNBOER and SWELLER 2005).

If the pattern on the stone is a familiar pattern, then it can be matched to a 'template' in the expert's memory (GOBET and SIMON 1996a): a *core structure* with modifiable slots that can vary somewhat. In template theory, chunks are recognized but they are not stored in memory as a singular pattern. Instead, they are exemplars which are matched to the closest-fitting template where the core characteristics are the same, even if some of the details vary *(ibid.).* Template theory in chess (GOBET and SIMON 1996a) might offer a valuable perspective on defining exactly how knapping are utilizing their enhanced memory. Before continuing this discussion, a semantic distinction should be made, given that WYNN and COOLIDGE (2004) have denied that stone tool knappers would have a 'mental template' in advance of producing the stone. In template theory, the word usage is different. It is not about working in accordance to a strict mental blueprint. Instead, it is about *using* previously formed templates in order to evaluate newly encountered patterns to previously encountered patterns. With regards to knapping, template theory should be more relevant than the older chunking theory (CHASE and SIMON 1973a, b). As mentioned earlier, template theory in chess is a modification of chunking theory, incorporating the chunk recognition process into a more flexible system (GOBET and SIMON 1996a; GOBET 1998; GOBET et al. 2011). In flint knapping, the knapper will never see precisely the same two patterns twice, but will have repeatedly been exposed to core patterns with constantly varying details. Template theory fits the knapping situation better than chunking theory because it would mean that the uniqueness of new chunks is not problematic. Newly encountered patterns do not need to be treated like isolated occurrences. Instead, the expert can recognise it as a variant of something seen earlier. Moreover, this ability to recognise patterns is also memorable because it tells the expert what to do next. Like that of a chess expert, a knapping expert can use their previously learned templates to enable a search through the "problem space" (possible future moves) to figure out exactly what knapping movements to do next (cf. CHASE and SIMON 1973b), including what angles to strike at and how forcefully to strike (cf. LORD 1993; BRIL et al. 2010).

Can we refer to prehistoric stone tool knappers as experts? Would *Homo erectus* have had the same underlying mechanisms of expertise that we find in modern humans? COOLIDGE and WYNN (2005, 2007, 2009; WYNN and COOLIDGE 2004, 2010) postulate that the tool making abilities of modern humans evolved in the *Homo* lineage as the result of enhanced abilities in WM and executive function, which enabled the expert planning of tasks. Their supposition generates fascinating questions. What effects would a reduced WM (relative to humans) have on the ability to knap? Presumably, it would be an attentive bottleneck in learning – but what does it imply for the ability to create the LTM structures required for expertise? Given the crucial role of LTM structures in the development of expertise, it would be beneficial to also look beyond WM and executive function and ask about the capacity to form templates than enable superior expert performance (sensu GOBET and SIMON 1996a). The current study has found evidence that experts have been able to attain high scores by taking advantage of the automatisation of skill (STERNBERG 1997) that is characteristic of expertise, and this is likely because they were easily able to overcome the cognitive load and identify chunks and templates quickly. Modern human expertise – by definition – refers to a vast diversity of skills which are domain-specific and only attainable through extensive practice (see ERICSSON et al. 2006). If we dissect lithic expertise among modern humans into its psychological components, then we can begin to investigate the archaeological record for evidence on the extent to which prehistoric knappers possessed modern expertise when manufacturing their artifacts.

ACKNOWLEDGEMENTS

Many thanks to Prof. Steven Mithen for enabling this project, and to the staff and fellow students involved in the MA Cognitive Evolution 1998–1999 at Reading University. Also thanks to John and Val Lord, Jill Scott of Franks House, and to the esteemed professionals who participated in this study (J.L., S.J.M., J.S., R.J., N.A., A.P., etc.), as well as the archaeology students and friends who participated. Thank you to Prof. Fernand Gobet and anonymous reviewers for comments on an earlier draft.

REFERENCES

BEILOCK, S. L. (2008): Math performance in stressful situations. *Current Directions in Psychological Science*, 17, 339–343.

BRIL, B., REIN, R., NONAKA, T., WENBAN-SMITH, & DIETRICH, G. (2010): The role of expertise in tool use: Skill differences in functional action adaptations to task constraints. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 825–839.

CHARNESS, N. (1976): Search in chess: Age and skill differences. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 641–653.

CHASE, W. G. & SIMON, H. A. (1973a): Perception in chess. *Cognitive Psychology*, 4, 55–81.

CHASE, W. G. & SIMON, H. A. (1973b): The mind's eye in chess. In: Chase, W. G. (ed.): *Visual Information Processing*. New York: Academic Press, pp. 215–281.

COOLIDGE, F. L. & WYNN, T. (2005): Working memory, its executive function, and the emergence of modern thinking. *Cambridge Archaeological Journal*, 15, 5–26.

COOLIDGE, F. L. & WYNN, T. (2007): Did a small but significant enhancement in working-memory capacity power the evolution of modern thinking? In: Mellars, P., Boyle, K., Bar-Yosef, O. & Stringer, C. (eds): *Rethinking the Human Revolution*. Cambridge: McDonald Institute for Archaeological Research, pp. 79–126.

COOLIDGE, F. L. & WYNN, T. (2009): *The Rise of Homo Sapiens: The Evolution of Modern Thinking*. Chichester, West Sussex/Malden, MA/Oxford: John Wiley & Sons.

DE BEAUNE, S. A., COOLIDGE, F. L. & WYNN, T. (eds) (2009): *Cognitive Archaeology and Human Evolution*. Cambridge: Cambridge University Press.

ERICSSON, K. A. & CHARNESS, N. (1997): Cognitive and developmental factors in expert performance. In: Feltovich, P. J., Ford, K. M. & Hoffman, R. R. (eds): *Expertise in Context: Human and Machine*. London/Menlo Park, CA/Cambridge, MA: AAAI Press / MIT Press, pp. 3–41.

ERICSSON, K. A., CHARNESS, N., FELTOVICH, P. J. & HOFFMAN, R. R. (eds) (2006): *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge: Cambridge University Press.

ERICSSON, K. A., KRAMPE, R. T. & TESCH-RÖMER, C. (1993): The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406.

GAMBLE, C. (2007): *Origins and Revolutions: Human Identity in Earliest Prehistory*. Cambridge: Cambridge University Press.

GOBET, F. (1998): Expert memory: a comparison of four theories. *Cognition*, 66, 115–152.

GOBET, F., LANE, P. C. R., CROKER, S., CHENG, P. C.–H., JONES, G., OLIVER, I. & PINE, J. M. (2001): Chunking mechanisms in human learning. *Trends in Cognitive Science*, 5, 236–243.

GOBET, F. & SIMON, H. A. (1996a): Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, 31, 1–40.

GOBET, F. & SIMON, H. (1996b): The role of recognition processes and look-ahead search in time-constrained expert problem solving: Evidence from grand-master-level chess. *Psychological Science*, 7, 52–55.

GOBET, F., CHASSY, P. & BILALIĆ, M. (2011): *Foundations of Cognitive Psychology*. London: McGraw-Hill.

GOLDIN, S. E. (1979): Recognition memory for chess positions: Some preliminary research. *American Journal of Psychology*, 92, 19–31.

GOWLETT, J. (1993): Mental abilities of early *Homo*: Elements of constraint and choice in rule systems. In: Mellars, P. & Gibson, K. (eds): *Modelling the Early Human Mind*. Cambridge: McDonald Institute for Archaeological Research, pp. 191–215.

HARLACKER, L. (2006): Knowledge and know-how in the Oldowan: An experimental approach. In: Apel, J. & Knutsson, K. (eds): *Skilled Production and Social Reproduction: Aspects of*

Traditional Stone-Tool Technologies, Uppsala: Societas Archaeologica Upsaliensis, pp. 219–243.

HELTON, W. S. (2005): Animal expertise, conscious or not. *Animal Cognition*, 8, 67–74.

LALAND, K. N. (2007): Niche construction, human behavioural ecology and evolutionary psychology. In: Dunbar, R. I. M. & Barrett, L. (eds): *The Oxford Handbook of Human Evolutionary Psychology*. Oxford: Oxford University Press, pp. 35–47.

LORD, J. W. (1993): *The Nature and Subsequent Uses of Flint, Volume I: The Basics of Lithic Technology*. Norfolk, UK: John Lord.

MANNING, D., ETHELL, S., DONOVAN, T. & CRAWFORD, T. (2006): How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography*, 12, 134–142.

RUSSELL, Y. I. (1999): *Lithic Expertise and a Debitage Recognition Task.* Unpublished M.A. dissertation, University of Reading.

SCHICK, K. D. & TOTH, N. (1993): *Making Silent Stones Speak*. London: Weidenfeld and Nicolson.

SEED, A. & BYRNE, R. (2010): Animal tool-use. *Current Biology*, 20, R1032–R1039.

SHALIN, V. L., GEDDES, N. D., BERTRAM, D., SZCZEPKOWSKI, M. A. & DUBOIS, D. (1997): Expertise in dynamic, physical task domains. In: Feltovich, P. J., Ford, K. M. & Hoffman, R. R. (eds): *Expertise in Context: Human and Machine*. London/Menlo Park, CA/Cambridge, MA: AAAI Press / MIT Press, pp. 195–217.

SIMON, H. A. & GOBET, F. (2000): Expertise effects in memory recall: A reply to Vicente and Wang (1998). *Psychological Review*, 107, 593–600.

STERNBERG, R. (1997): Cognitive conceptions of expertise. In: Feltovich, P. J., Ford, K. M. & Hoffman, R. R. (eds): *Expertise in Context: Human and Machine*. London/Menlo Park, CA/Cambridge, MA: AAAI Press / MIT Press, pp. 149–162.

TOMASELLO, M. (1999): *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

VAN MERRIËNBOER, J. J. G. & SWELLER, J. (2005): Cognitive load theory and complex learning: Recent developments and future directions. *Educational PsychologyReview*, 17, 147–177.

WINTON, V. (2005): An investigation of knapping-skill development in the manufacture of Palaeolithic handaxes. In: Roux, V. & Bril, B. (eds): *Stone Knapping: The Necessary Conditions for a Uniquely Hominin Behaviour*. Cambridge: McDonald Institute for Archaeological Research, pp. 109–116.

WYNN, T. (1989): *The Evolution of Spatial Competence*. Urbana, IL: University of Illinois Press.

WYNN, T. (2002): Archaeology and cognitive evolution. *Behavioral and Brain Sciences*, 25, 389–438.

WYNN, T. & COOLIDGE, F. L. (2004): The expert Neanderthal mind. *Journal of Human Evolution*, 46, 467–487.

WYNN, T. & COOLIDGE, F. L. (2010): How Levallois reduction is similar to, and not similar to, playing chess. In: Nowell, A. & Davidson, I. (eds): *Stone Tools and the Evolution of Human Cognition*. Boulder: University Press of Colorado, pp. 83–103.