Fluid Intelligence is Key to Successful Cryptic Crossword Solving

Kathryn J. Friedlander and Philip A. Fine
School of Psychology, University of Buckingham, UK

Correspondence: Kathryn Friedlander, kathryn.friedlander@buckingham.ac.uk

Journal of Expertise
2020. Vol. 3(2)

ISSN 2573-2773

Abstract

British-style cryptic crossword solving is an under-researched domain of expertise, relatively unburdened by confounds found in other expertise research areas, such as early starting age, practice regimes, and high extrinsic rewards. Solving cryptic crosswords is an exercise in code-cracking detection work, requiring the segregation and interpretation of multiple clue components, and the deduction and application of their controlling rules. Following the Grounded Expert Components Approach (GECA, Friedlander & Fine, 2016) an earlier survey demonstrated that solvers were typically educated to at least degree level, often in mathematics and science-related disciplines. This study therefore hypothesized that as a group they would show higher-than-average fluid intelligence compared to a general population, with experts showing higher levels than ordinary solvers. Twenty-eight crossword solvers (18 objectively defined experts, and 10 non-experts) solved a bespoke cryptic crossword and completed the Alice Heim tests of fluid intelligence (AH5), a timed high-grade test, measuring verbal and numerical (Part I) and diagrammatic (Part 2) reasoning abilities. In the 45m allowed, 17 experts and 2 non-experts correctly finished the crossword (times ranging between 11m and 40m). Both solver groups scored highly on the AH5 (both overall and for Part I) compared to manual test norms, suggesting that cryptic crossword solving has a high cognitive entry threshold. The experts scored higher than the non-experts, both overall (p = .032) and on Part I (p = .002). The overall and Part I AH5 scores correlated negatively (r_s = -.48; -.72 respectively) with extrapolated finishing times: faster finishing time being associated with higher AH5 scores. The experts and non-experts were matched in age, education, crossword solving experience, and weekly hours spent solving, leading to the suggestion that fluid intelligence differences between the groups may play an important role in cryptic crossword solving expertise. Although small in scale, the study thus adds to the growing body of literature which challenges the “deliberate practice only” framework of high expertise in a performance domain. Suggestions for future explorations in this domain are made.

Keywords

Fluid intelligence, cryptic crosswords, expert performance, Grounded Expert Components Approach (GECA), deliberate practice, problem-solving, talent

Introduction

Background: Expertise Research

Examples of technical performance experts are commonplace in everyday life—from professionals such as surgeons and lawyers, to academics and researchers in specialized fields, and thence more broadly into performance areas such as music and board games. Expertise is commonly defined as the possession of domain-specific skill-sets, knowledge, or performance levels which are demonstrably and reproducibly superior to those of most others involved in that particular domain (Ericsson & Towne, 2010; Gobet, 2015, Ch.1). This definition suggests that there is a spectrum of performance levels within a professional field, with experts lying at the far end of this. Nevertheless, it is also common to find a very small proportion of "super-experts" within a performance domain who stand out prominently, even from their expert peers. These elite performers include world-class musicians and dancers, together with individuals in "mind-game" fields such as Magnus Carlsen (chess, Gobet & Ereku, 2014; Howard, 2011), Mark Goodliffe (cryptic crosswords, Connor, 2014), Nigel Richards (Scrabble, Fatsis, 2011; Hambrick, 2015) and Kevin Ashman (UK quizzing, Waley-Cohen, 2019).

Why only some people become experts in a particular domain has intrigued psychologists for many years, and the debate relating to the importance of innate ability versus experience and environment has been at the forefront of this research. For those more concerned with understanding the general development of expert skills within a domain, the primary focus has been on "deliberate practice"—the conscious, structured, unenjoyable, and private rehearsal of domain-relevant tasks, leading to the enhancement of skills (Ericsson et al., 1993; Ericsson & Towne, 2010; Howe et al, 1998). Conversely, the "multifactorial" approach follows an individual difference line, suggesting that excellence in a particular field is driven by a helpful constellation of innate cognitive abilities, together with other environmental, motivational, and practice-related considerations (Hambrick et al., 2016; Ullén et al., 2015). The main aims of expertise research thus involve the following: first, uncovering the mechanisms by which certain individuals develop enhanced levels of performance, knowledge or skills compared to others active in that domain (Ericsson & Towne, 2010; Hambrick et al., 2016); second, exploring how the characteristics of experts differentiate them from non-experts (Friedlander & Fine, 2016; Ullén et al., 2015); and last, studying the cases of truly exceptional performers in a domain (Chi, 2006), to establish whether the "global qualities of their thinking" (Minsky & Papert, 1974, p. 59) might differ from their peers. In other words, how does expertise generally develop, why do only some people become experts, and how do we account for "super-experts"?

Expertise research uses a broad range of methodological approaches (Campitelli et al., 2015; Chi, 2011), although the choice in any particular study is largely determined by the ideological stance of the researcher (Friedlander & Fine, 2016; Hambrick et al., 2016). However, previous research has tended to apply these methodologies to a relatively restricted number of fields, primarily chess (e.g., Burgoyne, Nye, et al., 2019; DeBruin et al., 2014; Gobet & Ereku, 2014; Grabner, 2014; Howard, 2011) and music (e.g., Burgoyne, Harris, et al., 2019; Ericsson et al., 1993; Macnamara et al., 2014; McPherson & Williamon, 2015; Meinz & Hambrick, 2010; Platz et al., 2014). It is as yet unclear whether the findings of these highly practice-intensive, competitive fields, which are typically started at a very early age, will be transferable to other expertise fields without these characteristics. More recently, researchers have begun to address these issues in a wider range of alternative technical performance areas such as Scrabble (Halpern & Wai, 2007; Toma et al., 2014; Tuffiash et al., 2007), straight-definition ("US-style") crosswords (Moxley et al., 2015; Toma et al., 2014), and cryptic ("British-style") crosswords (Friedlander & Fine, 2016, 2018), together with broader professional contexts such as journalism (Wai & Perina, 2018).

This article presents an investigation of cryptic crossword expertise, specifically examining whether fluid intelligence (Gf) abilities (Cattell, 1943, 1963) underlie individual differences in levels of solving expertise, thus supporting the multifactorial account. Cryptic crosswords are popular in the UK and in countries with historically close links to Britain; unlike their "straight-definition" American counterparts, they comprise a set of quasi-algebraic, coded instructions which must be executed precisely in order to achieve the correct answer to the clue (see further Friedlander & Fine, 2016, 2018, and discussion below). We argue that the cognitive demands of solving cryptic crosswords involve the types of processing typically labelled as Gf, such that cryptic crossword solvers as a population would be expected to have higher levels than the general public, creating an "entry hurdle" for participation; and that Gf would increase in line with solving expertise. In this, we also draw on corroborative evidence from previous survey data (Friedlander & Fine, 2016) which demonstrates that cryptic crossword solvers are typically academically able individuals who pursue complex career paths in areas with high demands for problem-solving skills.

Addressing the Pitfalls of Expertise Research: Casting the Net Sufficiently Wide

A number of methodological issues have impeded progress in unravelling the antecedents of high expertise. One key limitation of many studies is the lack of in-depth understanding of the target population, leading to preconceived assumptions about the likely drivers of expertise. Furthermore, there is a danger that the selection of test paradigms may be driven more by unconscious biases related to the researchers' ideological stance on the talent/no-talent question, than by a grounded understanding of the demands of the domain itself (Friedlander & Fine, 2016).

One pertinent example of this may be found in the research domain of Scrabble (Tuffiash et al., 2007). On *prima facie* grounds, it is clear that Scrabble experts, who dedicate many hours to learning lists of Scrabble alphagrams [the alphabetically ordered letters of words], would have better orthographic word knowledge than novices, although not necessarily a better understanding of meaning or pronunciation. On this basis, Tuffiash and colleagues posited that Scrabble expertise could be fully accounted for by specialized, practice-related skills related to the pattern-recognition of potential words among a set of scrambled letters. Using Ericsson's Expert-Performance Approach (EPA, Ericsson & Smith, 1991; Ericsson & Ward, 2007), Tuffiash tested elite and average Scrabble players, together with much younger non-players, on both a Scrabble task intended to be representative of the domain (de Groot's "best-next-move" paradigm, 1946/1965), and a number of standardized verbal ability tests. Unsurprisingly, the Scrabble players outperformed the novices on Scrabble move selection and verbal tasks; and expert Scrabble players were better than less-expert players.

However, evidence from elsewhere—and particularly from interviews with Scrabble players themselves—indicates that top-flight Scrabble is much more a strategic mathematical game than a verbal one. It is, of course, a given that all world-level Scrabble players have memorized the official list of available alphagrams up to eight letters (Katz-Brown, 2006); however the role of strategy then becomes key:

> Even then, the game requires the foresight of chess and the inferential strategy of poker. I must both maximize my score on the current turn and keep strong letters on my rack to increase the probability that I can maximize my score on future turns. I further aim to squelch opponents' opportunities by guessing, based on their previous plays, which tiles they are most likely to be holding. By tracking tiles as they are played, I can also deduce exactly which tiles my opponent has in the endgame and plan my final plays accordingly. In other words, competitive Scrabble is a math game, and the level of strategy involved is one reason I keep playing (Katz-Brown - no. 36 in the world in 2014, 2006).

This claim is supported by other Scrabble experts: "It is really a game of maths - you are just taking on extra work by trying to learn all the definitions" (Paul Gallen - no. 5 in the world in 2018, Webb, 2012); and "People think Scrabble is just about words but it's the numbers that win the game, so a sound mathematical brain is an advantage" (Mikki Nicholson - no. 14 in the world in 2011, Fallon, 2010).

It is highly likely that this type of strategic/mathematical thinking in Scrabble relates far more to fluid intelligence (Gf), defined as the ability to use deliberate thought to generate solutions to novel problems, than to crystallized intelligence, defined as the ability to use previously acquired declarative knowledge and procedural skills (Cattell, 1943, 1963; McGrew, 2009). However, Gf was not explicitly explored by Tuffiash in any of his psychometric testing, because of preconceived beliefs about the nature of Scrabble expertise. Nor did the “best-next-move” paradigm (de Groot, 1946/1965) allow for the development of the type of strategic play outlined above by Katz-Brown, with the Verbal Protocol Analysis capturing only meager and functional data from the isolated challenges set, such as the strings of candidate solution words (Friedlander & Fine, 2016; Tuffiash et al., 2007).

In terms of expertise research generally, innate aptitudes are agreed to contribute strongly to Gf abilities (such as Working Memory (WM) and Executive Functions (EF)). Certainly, they are much less amenable to training than crystallized intelligence (Hambrick & Hoffman, 2016), although the contribution of the environment will still be important (Nisbett et al., 2012). While it is true that targeted EF and WM training can bring about improvements to the EF/WM task specifically being trained (Nisbett et al., 2012), there is currently little evidence of transfer to distant, or even closely, related tasks (Simons et al., 2016). Nor is there evidence that any such EF/WM training forms part of the deliberate practice regime identified in Scrabble (Tuffiash et al., 2007), with the focus being on the learning of alphagrams, thus increasing crystallized knowledge. Had the researchers tested Gf in age-matched expert and average samples, we might have expected them to find higher levels in the more expert players, implying a role for factors other than deliberate practice in expertise development, in line with the “multifactorial” view (Hambrick et al., 2016; Ullén et al., 2015). It is thus possible that confirmation bias, and a strong ideological belief in the “no-talent” approach unhelpfully constrained this research.

The Grounded Expertise Components Approach and Cryptic Crosswords

As with Scrabble (Tuffiash et al., 2007), it would have been plausible to assume that cryptic crossword expertise is also primarily concerned with the differing levels of solvers’ verbal abilities, and thus to have followed the classic EPA route, by selecting a representative task and psychometric tests based on a purely theoretical standpoint and *a priori* assumptions. It is certainly true that verbal abilities are relevant for US-style “straight-definition” crosswords, which may essentially be viewed as semantically cued retrieval tasks (Friedlander & Fine, 2016; Nickerson, 1977, 2011; Toma et al., 2014) requiring specialist crystalized vocabulary “crosswordese” (Hambrick et al., 1999; Romano, 2006).

Indeed, even for British-style cryptics, it has previously been hypothesized that cryptic crossword experts “would have particularly rich lexical networks” (Underwood et al., 1988, p. 302), although this was not actually the eventual finding of their study.

Nevertheless, Friedlander & Fine (2016) were reluctant to impose their preconceived ideas upon the direction of the present research program in this way. This reluctance was based on the conviction that objective research can be conducted on a niche population only if care is taken to characterize it carefully over a number of dimensions, leading to a grounded understanding of the motivational drivers, skill-sets, and immersion necessary for high performance in the domain (Friedlander & Fine, 2016).

Following these principles, Friedlander and Fine (2016) launched a survey to explore the broad characteristics of a wide range of experienced cryptic crossword solvers, with the aim of comparing empirically the profiles of ordinary solvers and high-end experts. During this process, they developed the Grounded Expertise Components Approach (GECA) as a modification of, and improvement to, the Expert-Performance Approach (EPA, Ericsson & Smith, 1991; Ericsson & Ward, 2007).

According to the EPA, participants are invited to the lab to conduct a “domain-representative” task, often involving one-shot challenges (Friedlander & Fine, 2016) such as the de Groot “best-next-move” paradigm (de Groot, 1946/1965), in order to explore the mechanisms of high-expert performance. These isolated challenges primarily test the ability to come up with rapid, automatic, memorized play laid down by extensive practice routines (such as chess opening gambits, Scrabble alphagrams, and other “chunked” sequences of moves): that is, problem solving which is typical of “System 1 thinking” (Evans & Stanovich, 2013). This could potentially have led to a systematic underestimation in the literature of the importance of creative, strategic, and integrated game-play (Friedlander, 2019; Friedlander & Fine, 2016). Finally, the EPA trial may be accompanied by subsidiary tests of subskills thought to be relevant on *prima facie* grounds; and is only then followed up by a questionnaire primarily intended to capture data relating to starting age, experience and levels of deliberate practice (Ericsson & Ward, 2007; Tuffiash et al., 2007).

In contrast, the GECA first characterizes the population active in the domain of interest before developing testable hypotheses about expertise development in that domain, thus ensuring that these are grounded in the population data, and effectively minimizing the danger of confirmation bias. This detailed knowledge then provides the backdrop for laboratory studies, in which an integrated challenge, extended across multiple moves, is presented to the participants. Instead of using isolated tasks, this approach has the advantage of requiring participants to interact in an ecologically valid way with the full spectrum of cognitive, strategic and emotional demands of the challenge, potentially using “System 2 thinking” as well as the memorized chunks or routines of “System 1” (Friedlander, 2019; Friedlander & Fine, 2016). In common with the EPA, both experts and non-experts perform this task while being recorded, usually verbalizing their thoughts for subsequent analysis using Verbal Protocol Analysis (VPA, Ericsson & Simon, 1993; Gilhooly & Green, 1996; Green & Gilhooly, 1996). However, under the GECA, this results in much richer and more informative process-tracing data, yielding information on many facets of expert play, compared with the meager and comparatively superficial reports obtained under the EPA (Friedlander, 2019; Friedlander & Fine, 2016). Finally, psychometric sub-tests, empirically identified on the basis of the initial characterization of the population, are used to probe cognitive and strategic processes thought to contribute to the individual differences between experts and non-experts. A summary of the process is set out in Figure 1 below.

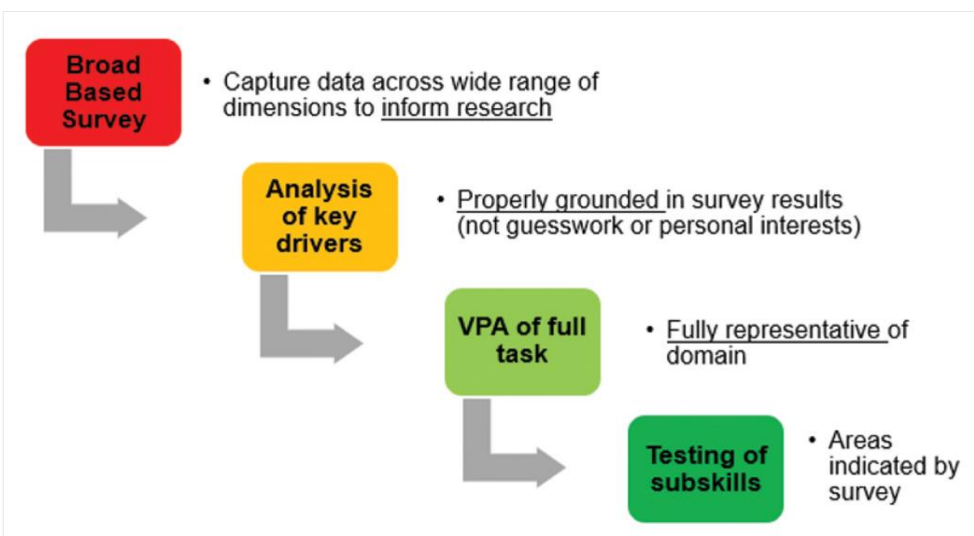

Figure 1. The stages of the Grounded Expertise Components Approach (Friedlander & Fine, 2016); “VPA” = Verbal Protocol Analysis.

Benchmarking the Levels of Expertise

In order to make a meaningful comparison between the characteristics and abilities of samples differing in expertise level, it is important that these levels can be objectively and, where possible, externally benchmarked (Friedlander & Fine, 2016). Without this, the researcher runs the risk of confounding the results due to the inaccurate assignment of participants to relevant groups. Objective benchmarking is particularly difficult in reputation-based ("r-expertise") domains such as music performance, gymnastics and diving, and business or medicine (Gobet, 2017). However, in performance-based ("p-expertise") domains such as athletics, tennis, chess, and Scrabble, an objectively accepted, quantifiable measure of expertise is typically available (Gobet, 2017). In the case of chess, research employs Elo ratings (Gobet & Charness, 2006) to assign participants to groupings; similarly, Scrabble has official tournament metrics (Tuffiash et al., 2007).

Although there is no official ranking system for cryptic crossword expertise, Friedlander & Fine (2016) developed alternative methods to categorize solvers into objectively defined expertise levels, relating to (a) the difficulty of the crossword regularly solved; (b) the speed of solving the crossword; (c) successful participation in speed-solving competitions; and (d) regular engagement in advanced cryptic crossword solving or setting (compiling) activities. For full details see Participants section, p. 111. The 805 survey respondents were thus split into three expertise categories: 179 super-expert (S) solvers, 225 high-ability (H) solvers, and 401 ordinary solvers (O). The authors knew all S solvers personally or by reputation, and their pre-eminent level of skill can be verified objectively by referring to publicly available records (Friedlander & Fine, 2016). Most solvers (729 out of 805, over 90%) had been solving cryptic crosswords for at least 10 years, regardless of expertise group, with more than half solving for over 30 years. Thus, the sample was highly experienced in the domain at all levels of expertise. This was important as it enabled a comparison of experts with equivalently experienced ordinary solvers (rather than inexperienced novices). The relative proportion of O, H, and S solvers is not representative of the general cryptic crossword population, being a product of deliberate oversampling from high-expert forum websites (Friedlander & Fine, 2016).

Cryptic Crossword Solvers Are Academically Strong and Tend Toward STEM Fields

A detailed account of many findings derived from the survey (GECA stages 1/2) has already been published (Friedlander & Fine, 2016). However, we highlight here two particularly striking results. In the first place, cryptic crossword solvers seem to be highly academically able. Over 80% of the 805 respondents, regardless of expertise group, had a university degree and 12% had PhDs. Importantly, the majority of respondents (median age 54) would have attended university at a time (1970s - 1980s) when only 10%-20% of the UK population attended (Bolton, 2012). This suggests an exceptionally high level of educational achievement for cryptic crossword solvers across the board. Survey respondents were also engaged in cognitively complex careers, as analyzed by Holland Cx ratings, with the mean and median scores of all three groupings falling close to 70, and 54% of the participants falling into the 70-79 band. Holland Cx scores range from <40 to >80: a Cx rating of 65 or higher is associated with a college degree and 4–10 years of "On-Job-Training" (Friedlander & Fine, 2016; Reardon et al., 2007).

Secondly, we also found that solvers tend to be qualified in scientific fields (Friedlander & Fine, 2016). Over half (51%) had majored in a STEM subject (science, technology, engineering, mathematics). In particular, the proportion studying mathematics at university increased markedly with cryptic crossword expertise (14% of ordinary solvers, 32% of super-experts). Overall, 56% worked in STEM, medicine, or finance, and this rose to 66% for super-experts. When STEM/finance occupations were analyzed in more detail, significantly more super-experts than ordinary solvers worked in Technology/IT (32% vs. 21%) and Banking/Accountancy (13% vs. 6%).

Conversely, only 26% had studied a “Wordsmith” subject (languages, literature, media studies, philosophy, religion) at university, and even fewer, 14%, worked in a “Wordsmith” occupation (languages, creative/media, spiritual/philosophy). Ostensibly, this seems at odds with Underwood’s prediction that rich lexical networks would be enhanced in cryptic crossword experts (Underwood et al., 1988), and suggests that there may be more important factors underlying cryptic crossword expertise than verbal abilities alone, particularly the cognitive abilities central to STEM and IT careers. Indeed, Underwood’s unsuccessful findings also led him to conclude that cryptic crossword skills are “as much bound up in the cryptic puzzle codes as they are in lexical fluency” (Underwood et al., 1988, p. 306); and intelligence has been shown to explain individual differences in both educational achievement and job complexity (Gottfredson, 1998, 2002; Rimfeld et al., 2018).

Intelligence as a Factor in Expertise Development

General intelligence is a major attribute by which individuals differ from one another. It has been defined as the “ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, [and] to overcome obstacles by taking thought” (Neisser et al., 1996, p. 77). Researchers on each side of the talent/no-talent divide have taken up strong antithetic stances on the question of whether individual differences in intelligence are related to expert performance (Grabner et al., 2007). For example, Ericsson has claimed that “there is no correlation between IQ and expert performance in fields such as chess, music, sports, and medicine” (Ericsson et al., 2007, p. 116) and that “IQ is either unrelated or weakly related to performance among experts…; factors reflecting motivation … are much better predictors of improvement” (Ericsson & Lehmann, 1996, p. 280).

Yet, psychometric “g” has been found to correlate with real-world outcomes in education and careers (Gottfredson, 1998, 2002; Rimfeld et al., 2018), and is highly predictive of the ability to earn a doctorate, publish an article, or register a patent (Lubinski et al., 2006). Furthermore, research has demonstrated that, although necessary for all domains, deliberate practice is not sufficient to produce expertise, accounting, for instance, for only 34% and 30% of the variance in expert performance in chess and music respectively (Hambrick et al., 2014; Macnamara et al., 2014). Intelligence is thus, for those supporting a “talent” approach, an attractive candidate driver of excellence in performance domains, although the relative contribution of intelligence (or any of its subordinate facets, reflecting the content-base of the challenge) is likely to vary depending upon the level of cognitive demand in any given domain (Ackerman, 2014a; Hambrick et al., 2014).

One key variable is thus the type of activity typically undertaken in the relevant domain. Intelligence has been argued to be of lesser importance in physical domains, compared to cognitive domains (Hambrick et al., 2014), and has been found, for example, to show no correlation with performance among NFL American Football players (Lyons et al., 2009). Another variable may be the persistence of task complexity and challenge (Ackerman, 2014a). Intelligence appears to confer most advantage when tasks are novel, allowing individuals to exploit learning opportunities and to pick up the rules faster during the initial stages of skill acquisition; however once learned, practice allows skills in “closed-ended” tasks, such as driving a car, to become automatized (Ackerman, 1987, 1988). Conversely, for substantially “open-ended” tasks, where the rules or conditions of the task may continue to present novel challenges (for example in post-graduate studies, in chess or in music), intelligence continues to be important (Ackerman, 2014a). Again, Cattell’s Investment theory of intelligence (Cattell, 1957, 1963) may also be relevant here: this theory suggests an influence of Gf (fluid intelligence) on Gc (crystallized intelligence), such that Gf guides the acquisition of cultural knowledge and skills through infancy into early adulthood. This in turn leads to a "Matthew effect" whereby those with higher Gf will also find it easier to acquire specialized domain knowledge (Gc) through learning (Schweizer & Koch, 2002).

Certainly, a large number of studies into chess expertise have suggested that measures of IQ correlate significantly with performance in chess (Grabner, 2014; Hambrick et al., 2014), although the evidence is somewhat mixed. Nevertheless, the results from a comprehensive meta-analysis by Burgoyne et al. (2016) demonstrate that chess skill correlates significantly and positively with four broad cognitive abilities subsumed within global IQ—Gf, Gc, Gsm (short-term memory) and Gs (processing speed) —although not with the global Full Scale IQ score itself. Each of these four components explained between 5-6% of the variance in chess skill.

In this type of "within expertise" analysis, one key point to remember is that the population being studied is already highly winnowed, producing an elite sub-population which has survived repeated rounds of competitive pre-selection, and which may therefore show "species typical traits*"* (Ackerman, 2014b, p. 3). An example of this might be basketball players, who at higher expertise levels will typically be of above-average height (Detterman et al., 1998; Howard, 2009). Where individuals are already selected for ability, the resulting correlations between achievement and ability measures will therefore be attenuated (Ackerman, 2014b; Detterman et al., 1998; Ruthsatz et al., 2008). However, as in the case of basketball players, the importance of the key trait, whether physical or cognitive, will become more apparent by contrast to the broader non-expert population than in a "within-expertise" comparison; this may suggest important entry hurdles to successful participation (Ackerman, 2014b; Detterman et al., 1998; Hambrick et al., 2014). Thus, it is important to consider key variables such as IQ and its components in normative terms, by comparison to a general population sample, not just within the context of a highly rarefied expertise sample.

Cryptic Crosswords and Fluid Intelligence

Turning to cryptic crosswords, the findings of Friedlander & Fine (2016) do fit well with what we now know about the demands of cryptic crosswords solving. Each cryptic crossword clue comprises a definition of the answer together with a set of coded instructions (the "wordplay"), which, when correctly decoded, will lead the solver to the answer (see Friedlander & Fine, 2016, 2018 for examples). Furthermore, the surface reading of the clue is often phrased in such a way as to mislead solvers by the inclusion of "red herrings" which suggest a plausible, yet unhelpful interpretation of the clue. Solving cryptic crosswords thus involves inhibiting the surface reading of the clue, which is activated highly automatically, because of a life-time's experience in parsing written text (Schulman, 1996), and then deconstructing the clue elements in order to arrive at the correct (and only) answer. The difficulty lies in recognizing the clue type and cracking the setter's code by correctly parsing the clue into definition and wordplay components.

The setter's task is therefore rather like that of a magician: to conceal the mechanisms of the deception so that they are not immediately evident (Friedlander & Fine, 2016; Kuhn, et al., 2016). Even the "definitional" element of the crossword clue might be obliquely or whimsically referenced, consciously exploiting ambiguities such as grammatical form, phrasal semantics, homophones, synonyms, and roundabout expressions (Aarons, 2015; Cleary, 1996; Friedlander & Fine, 2018). The clue type also has to be identified and interpreted, meaning that the problem space is not tightly defined, and that cryptic crosswords function as insight puzzles, requiring a representational change in problem conceptualization in order to arrive at the answer (Friedlander & Fine, 2018). All these factors mean that cryptic crosswords are typically ill-defined in solution methodology (Johnstone, 2001) and require considerable code-cracking abilities for solution. This led Friedlander and Fine (2016) to suggest that Gf might be key to solving cryptic crosswords.

Though there is some debate as to the exact nature of Gf and its relationship to working memory capacity (WMC) and EF, it is generally accepted that there is a large overlap between these concepts, and that they relate to aspects of attentional control and other prefrontal cortex functions (Heitz et al., 2006; Kane et al., 2005). Broadly speaking, WM is seen to facilitate complex cognition by maintaining critical information in a highly accessible state. Thus, for those engaged in problem-solving, high WMC allows individuals to maintain the problem representation in a particularly accurate and stable form, so that solutions can be derived and tested out against the retained information (Shipstead et al., 2016). By contrast, EF refer to a set of mental abilities related to cognitive control. These include (though not exhaustively): planning; cognitive flexibility; shifting between mental sets; concept formation; inhibitory control; monitoring task performance; place-keeping ability; self-regulation; and attentional control (McCabe et al., 2010; McCloskey & Perkins, 2012; Nyongesa et al., 2019).

Cryptic crossword clues can employ a wide variety of word-play devices such as puns and double-definitions; riddles and rebus-like "word-pictures"; anagrams; charades (e.g. REIN + FOR + CEMENT = REINFORCEMENT); "sandwiched" components (e.g. EEL in RING = REELING); reversals, letter transpositions and word truncations; hidden words; and lateral thinking challenges (Biddlecombe, 2009; Friedlander & Fine, 2018). Each of these devices can be used singly, or in combination. A diverse range of cognitive abilities allied to WMC and EF is therefore likely to be involved in solving these puzzles.

For example, in order to crack the punning, double-definition, and rebus-like elements, or to interpret a more whimsically referenced definitional synonym, solvers would need to activate a wide retrieval search of semantic memory, inhibiting fixation upon incorrect, high-frequency "convergent" candidate words which might spring more readily to mind, and consciously allowing more remote "divergent" associations to be accessed (Friedlander & Fine, 2018). In this context, a review of cryptic crossword clue types and their relationship to insight puzzles (Friedlander & Fine, 2018) highlighted a number of parallels between cryptic crossword clues and (Compound) Remote Associates Puzzles (RAT(CRA) Bowden & Jung-Beeman, 2003; Mednick, 1962). These puzzles typically take the form of a triad of apparently unconnected words (e.g. *Cottage*, *Swiss*, *Cake*) which must be associated in some way with a fourth word (here *Cheese*). RAT puzzles and the closely related cryptic crossword elements identified above may be solved either through the operation of a serendipitous spreading neuronal network (Friedlander & Fine, 2018; Kenett et al., 2014; Olteţeanu & Falomir, 2015; Smith, S. M., et al., 2012) or through a more controlled generate-and-test strategy, to check out candidate solutions against each constraint for suitability (Bowden & Jung-Beeman, 2007; Friedlander & Fine, 2018; Smith, K. A., et al., 2013). Solvers may elect to switch between modes of search, depending upon the success of their approach (Bowden, Jung-Beeman, et al., 2005).

Moreover, as cryptic crosswords employ "red herring" elements and (in advanced cryptic puzzles) lateral thinking end-games, an ability to "break frame" and overcome functional fixedness is important (DeYoung et al., 2008; Friedlander & Fine, 2018). Taken as a whole, this flexibility to break through the false conceptualization of the problem, shifting to a new problem space; to inhibit unproductive avenues (Benedek et al., 2012); to accommodate "bisociation"—the perceiving of a situation in two incompatible frames of reference (Canestrari & Bianchi, 2012; Friedlander & Fine, 2018; Koestler, 1964); and to switch electively between convergent and divergent idea generation (Benedek et al., 2014; Nusbaum & Silvia, 2011) implies a highly efficient use of executive processes.

The similarity of cryptic crossword clues to algebra or computer programming has also been noted in passing (Manley, 2014); and indeed an

Australian conference paper (Simon, 2004) draws a number of close analogies between solving cryptic crossword puzzles and computer programming problems, highlighting the need for clear analytical thought and productive hypothesis testing. The algebraic/cryptographic nature of the cryptic clue means that wordplay components may be flexibly recombined or anagrammed to form new units: this particularly affects anagram, charade, sandwich, truncation, reversal, and letter-transposition clues. While many solvers use a physical jotting pad or electronic anagrammer to handle the letters, the mental ability to maintain, manipulate and integrate potentially promising combinations might be hypothesized to confer a speed advantage in solving cryptics (Friedlander & Fine, 2016). This might in turn suggest that expert solvers were using WM systems to particularly good advantage.

Finally, the nature of the crossword grid, and clue types such as hidden words, might also imply an enhanced ability to pattern-match and, most specifically, to complete word fragments provided by cross-checking letters, as for US-style crosswords (Hambrick et al., 1999; Nickerson, 1977, 2011; Thanasuan & Mueller, 2016). Efficient pattern recognition directs a more effective planned search through semantic memory, perhaps through the use of easily recognizable orthographic features (Halpern & Wai, 2007; Thanasuan & Mueller, 2016), and also involves the suppression of interference from orthographically similar, but erroneous, competitor solutions (Healey et al., 2010).

Current Study - Hypotheses

The above review has indicated that solving cryptic crosswords is likely to rely on Gf, "the ability to derive logical solutions to novel problems" (Hicks et al., 2015, p. 187). The goal of this study is therefore to compare the Gf score of super-expert (S) solvers with those of ordinary solvers (O); and, additionally, to compare overall cryptic crossword Gf scores to population norms.

Given solvers' generally high levels of educational achievement and the high proportion of those working in cognitively complex problem-solving, mathematical and intellectual professions (Friedlander & Fine, 2016), we would expect them to possess good WMC and effective EF processes, leading to higher Gf compared to the general population. Moreover, we would expect more expert cryptic crossword solvers to have even higher Gf than less expert solvers. This enables us to propose the following hypotheses:

H1. All solvers will show high Gf compared to the demographic norm.

H2. Super-expert solvers will demonstrate higher Gf than Ordinary solvers.

H3. Super-solvers will show better performance on a bespoke cryptic crossword, in terms of speed and completion success.

H4. Time taken to solve a complete bespoke cryptic crossword will correlate negatively with Gf scores, such that the higher the score on Gf, the faster an individual will be to solve the cryptic crossword.

Method

Research Design

Building on the results of the survey at GECA stage 1/2 (see above, Figure 1), this study proceeded with targeted lab-based trials exploring the mechanisms of expertise in cryptic crossword solvers (GECA 3/4). Two tasks are reported in this paper:

1. The completion of a **domain-specific representative task** (**GECA, stage 3**), while process-tracing data recordings were made. Our participants' task was to solve within 45m a complete bespoke cryptic crossword of the type and difficulty typically found in a broadsheet newspaper. We argue that this is more representative than solving single isolated clues in the absence of a grid (as for example in Deihim-Aazami, 1999; Underwood et al., 1994; Underwood et al., 1988); see further Friedlander & Fine (2016) and the comments on Tuffiash et al. (2007) and the EPA above. Solvers were asked to speak their thoughts aloud, while their actions were filmed for later transcription, and this verbal protocol analysis (VPA) data will be presented elsewhere. Participants' **solving time** and the number of **clues correctly solved** were also recorded, providing an

additional objective benchmarking criterion supporting our categorization of participants into super-expert (S) and ordinary (O) solvers; this data is reported below.

2. Prior to completing the crossword, participants also completed the AH5 (Heim, 1968) **test of fluid intelligence** (**GECA, stage 4**), together with other word-based games to be reported elsewhere.

Participants

There were 28 participants (24M, 4F), all of whom had taken part in the wide-ranging survey on crossword experience (Friedlander & Fine, 2016), and had indicated willingness to take part in further trials. Participants in the survey were obtained through adverts placed on cryptic crossword websites dedicated to the discussion of cryptic crosswords and the analysis of answers to the previous day's broadsheet puzzles. Participants were paid £20 each in defrayment of costs and time associated with travel to the University of Buckingham. The selection of trial participants within each sub-group was driven by logistical/practical considerations based on geographical proximity to the University of Buckingham, and the participants' availability. Age at the time of testing ranged from 28 to 74 years (*Mdn* = 54.5, *M* = 53.0, *SD*=10.93). Numbers of participants were constrained by the practicalities of transcribing extensive VPA material amounting to over 1hr per participant; however, polarized subgroups were deliberately invited in order to try to offset any loss of statistical power (see below).

As already discussed, criteria for assigning participants to appropriate expertise categories must be rigorous and objective. Participants were therefore categorized using the benchmarked criteria outlined in Friedlander & Fine (2016), resulting in 18 super-expert (S) solvers (15M, 3F) and 10 non-expert ordinary (O) solvers (9M, 1F).

S participants qualified by virtue of one or more of the following criteria (for more details, see Friedlander & Fine, 2016):

1. They edited or composed cryptic crosswords professionally, on at least an occasional basis, for broadsheet or specialist publications ("Pro");
2. They regularly speed-solved a broadsheet cryptic crossword in <15m; and/or had reached the final in the annual Times National Crossword Championship on at least one occasion ("Speed");
3. They had solved 42+ *Listener* (or 48+ *Magpie*) advanced cryptic crosswords correctly in 1 year and were thus named on the official roll of honour of these competitions ("Advanced"). For details of advanced cryptics, see Friedlander & Fine (2016).

The O solvers rarely completed broadsheet cryptics in under half an hour and did not tackle advanced cryptic crosswords. No High expert (H) solvers (defined as those who solve broadsheets in under 30m, but do not qualify as Super-expert) were chosen to take part in trials on this occasion. Conceptually, the two selected groups are similar to Chi's "Journeyman" (O) and "Master" (S) proficiency categories (Chi, 2006), representing a polarized sample.

Care was taken to obtain S participants who were representative of all 3 Super-expert proficiency areas to permit a more fine-grained analysis of solving style in the VPA analysis to be reported elsewhere. A number of individuals were qualified in two or more dimensions resulting in a minimum of 6 representatives in each. The resulting breakdown of super-experts by area(s) of expertise is shown diagrammatically in Figure 2.

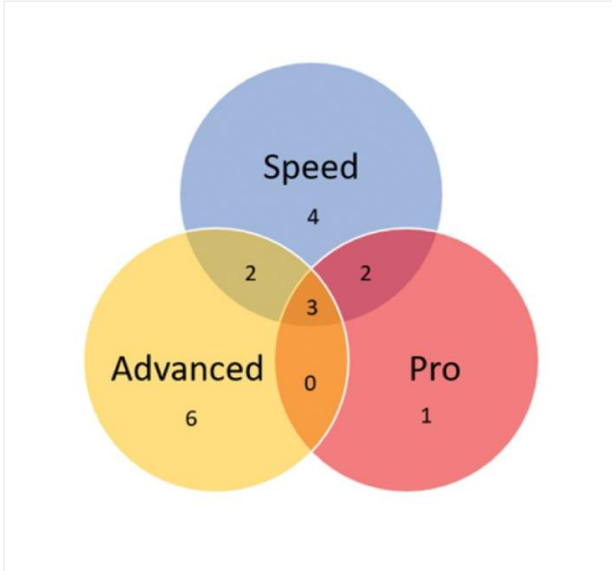

Figure 2. Numerical breakdown of Super-expert crossword participants (*n* = 18) by areas of expertise

Materials

Bespoke Cryptic Crossword

Insight puzzles are highly memorable once solved (Danek et al., 2013; Dominowski & Buyer, 2000), and for this reason it was important that solvers could not have solved the trial puzzle on an earlier occasion. Accordingly, a bespoke, professionally compiled cryptic crossword was commissioned. The crossword had to be appropriately taxing to present a reasonable challenge for expert solvers in order to preserve the richness of the VPA trace, yet simultaneously approachable by non-experts. The researchers therefore approached "Phi," a setter for the *Independent*, [London] *Times* and *Daily Telegraph* daily broadsheet newspapers, who was asked to set a typical 15 by 15 blocked cryptic crossword suitable for publication in *The Independent* [1] (which typically features a crossword of medium/hard difficulty without strong "house-style").

This crossword was piloted by both authors and by 8 independent solvers, all of whom had volunteered at survey stage to take part in later crossword research but were unable to attend in person at the Buckingham trials. Pilot solver expertise ranged from a Super-solver (Speed), who took 10m to complete the puzzle, through to a non-expert solver who took approximately 1h over two sessions and left one clue unsolved. Discussions were held with the setter to implement a few minor changes arising from pilot feedback, to ensure that the level of difficulty was appropriately pitched and that it could be reasonably completed within 45m. The crossword contained 27 clues, which is typical for this genre of puzzle.

Measurement of Fluid Intelligence (Gf)

A variety of tests are typically used for investigating Gf. Reductionist approaches employ a range of individual cognitive tasks broadly relating to WMC and attention, such as digit-span, approximate number system, block-tapping, letter set and number series tasks, together with visual short-term memory (e.g. Lane & Chang, 2018). Given the high academic achievement across the entire sample, we hypothesized for our trials that cognitively straightforward tests of WM load (e.g., simple and complex digit span tasks, or tests of visual short-term memory) would be unlikely to discriminate among groups as effectively as challenging Gf tasks, which (like cryptics) require the segregation, serialization and assembly of multiple subtask parts relating to a novel challenge, and the learning and understanding of their controlling rules (Duncan et al., 2012; Hambrick & Altmann, 2015). Tests in this category include Raven's Advanced Progressive Matrices (RAPM, Raven & Court, 1988) and the AH (Alice Heim) series of tests (Deary & Smith, 2004; Heim, 1968, 1970; Warren et al., 2004).

Gf testing using the AH4 (Heim, 1970) had already been shown not to discriminate between expertise levels of cryptic crossword solvers in the Nottingham trials (Deihim-Aazami, 1999; Underwood et al., 1994), but the results of our survey indicated that the AH4—which is designed for those who ceased education at 18—would have been wholly underpowered in that study for the assessment of such a highly academically qualified population, leading to ceiling effects acknowledged by the authors (Friedlander & Fine, 2016). A rerun of this comparison using the more appropriate AH5 test (Heim, 1968) was therefore a key priority for this research.

AH5 test (Heim, 1968)

The AH5 is a test of fluid intelligence intended to be used to distinguish between a selected population of highly intelligent people, such as university students and research workers. Heim characterizes the demands of the test as follows:

> "In devising the test items, the aim has been to raise the level of difficulty by increasing the complexity and closeness of the reasoning involved whilst losing nothing of its cogency. ...As in the intelligence tests devised for the less highly selected groups, the stress is largely on deductive reasoning. Other qualities required for success in AH5 include accurate observation, meticulous attention to instructions and ability to appreciate shades of meaning. Increased difficulty [...] has been achieved by requiring the subject often to "hold in his head" two or more opposing ideas ... to apprehend "second order" notions....and, mentally, to reverse a given order of things" (Heim, 1968, pp. 1-2).

Warren et al. also compare the AH5 to the RAPM, stating, "The Alice Heim 5 test (AH5) similarly requires identification and application of simultaneous patterns to complete verbal, numerical, and geometric sequences" (Warren et al., 2004, p. 1447).

The AH5 consists of two parts, each taking 20m, and administered one after another. Part 1 contains verbal and numerical items; Part 2 contains diagrammatic non-verbal items. Each part consists of 36 items, split into 9 items for each of 4 types. The AH5 uses a timed "spiral omnibus" design (Deary & Smith, 2004) such that the types are alternated in order, as the difficulty progressively increases. Prior to commencement of each test part, participants are given 8 practice items, 2 for each type, and there is no time limit for these practice items.

Part 1 (verbal and numerical) item types are as follows:

1. Directions, involving meticulous attention to complex instructions, potentially including sequencing pieces of information and having a high working memory load;
2. Verbal analogies, requiring the discernment and then application of a specific relationship between two words;
3. Numerical series, where the candidate has to determine one (or more) specific numbers missing from a given series, but with traps for the unwary, requiring careful attention to instructions;
4. Similar relationships, in which candidates are provided with a pair of words which they must relate in the same way (either synonyms or antonyms) to one of 5 potential matches.

Part 2 (diagrammatic non-verbal) item types are as follows:

1. Analogies - as above but with figures, normally involving some combination of reflection, rotation, diminution or enlargement;
2. Series, where the candidate has to determine the rule linking the given diagrams to decide which of a number of given items comes later in a sequence;

3. Directions, which include 2 different spatial tasks, though both again requiring careful attention to instructions, such as the interpretation of reflected items or the mental assembly of partially indicated shapes;
4. Features in common, conceptually similar to Similar relationships above, where candidates are required to determine which of 5 given diagrams either do, or do not, contain the feature in common in a pair of probe items.

The AH5 has good test-retest reliability (Heim, 1968) over a period of weeks (Cane & Heim, 1950) and a year (Watts, 1954). Many of the questions are multiple choice, with one out of a variable number of possibilities being the correct answer. For a small number of the questions, there are no suggested answers, and the participant has to propose the solution themselves.

Procedure

The data for this article was collected as part of a larger study investigating cryptic crossword solvers. Ethical approval was obtained from the relevant institutional committee for all parts of the study. Participants were tested individually, in dedicated lab facilities at the University of Buckingham, at a mutually convenient time. After giving informed consent, the participants completed the two parts of the AH5, following the standard guidance given in the manual, starting with Part 1. Twenty minutes was allowed for each part, plus time allowed for practice questions, and participants were free to tackle the questions in any order. Answers were handwritten, and any rough work was allowed on the answer paper.

Once the AH5 was completed, and a rest-break offered, video and audio recording commenced, with the express consent of the participant. The researcher withdrew from the room at this point, but viewed the proceedings from the control room, through a one-way mirror. Instructions were relayed via the lab sound system. The participants were asked to speak all thoughts aloud, following standard VPA procedure (Ericsson & Simon, 1993; Gilhooly & Green, 1996; Green & Gilhooly, 1996). The participants carried out two brief speak-aloud word games lasting approximately 20m in total (which will be reported elsewhere). They then had a maximum of 45m to complete the bespoke cryptic crossword, solving as much of the crossword as they could, as quickly and accurately as possible, in the time allowed. The VPA analysis of the cryptic crossword solving processes will be reported elsewhere (see also Friedlander & Fine, 2016), but the time taken to complete, and number of clues correctly answered, are important for this article.

Participants were then debriefed in a concluding extended conversation covering aspects of the crossword just solved, and the participants' general thoughts on expertise in crossword solving, which was video recorded, but will not form part of this discussion. The entire procedure took approximately 2h 30m for each participant.

Results

Given this is an expertise study involving only 28 participants, where the sample size was constrained (as is common in expertise studies) by the need to acquire a highly expert population and to transcribe extensive recorded material, we have followed the approach of "retiring statistical significance" (Amrhein et al., 2019; Campitelli, 2019). Although we include p values, effect sizes and confidence intervals, we do not therefore ascribe the term "significant" to the analyses.

Characterization of Two Groups of Interest

The O (n = 10) and S (n = 18) groups were compared on a number of demographic criteria to ensure that any differences on the AH5 test were not due to confounding variables. An independent samples t-test showed that the age of O (M = 53.3, SD = 12.24) and S (M = 52.8, SD = 10.50) solvers were equivalent (t (26) = .106, p = .92, Cohen's d = .04, 95% CI [-.76, .84]). Similarly, the solving experience of the two groups did not differ, either in terms of years' solving (O M = 32.9yrs, SD = 13.25; S M = 39.8yrs, SD = 11.41; t (26) = 1.444, p = .16, Cohen's *d* = .56, 95% CI [-.26, 1.37]) or hours solving per week (O *M* = 7.6h, *SD* = 3.47; S *M* = 7.5h, *SD* = 3.32; *t* (26) = .096, *p* = .924, Cohen's *d* = .04, 95% CI [-.76, .84]). Overall, therefore, the two groups were matched on age and solving experience, and had on average been solving cryptic crosswords for over 3 decades. Gender breakdown of participants reflected typical male preponderance in the solving population (Friedlander & Fine, 2016). Age, solving experience and hours spent solving per week were consistent with findings of the broader population from which this sample was taken (survey participants, all groups: Age *M* = 52.1*;* Yrs solving *M* = 31.4yrs; Hours spent solving per week *M* = 7.27h, Friedlander & Fine, 2016).

The participants as a whole were highly academically qualified, with 23 out of the 28 participants (82%) having at least an Honors Degree, and 12 with Masters or Doctoral qualifications (S 9/18 (50%); O 3/10 (30%)). Nineteen (68%) had studied STEM subjects (S 13/18 (72%); O 6/10 (60%)); an equivalent number in each group worked in STEM areas or finance (S 13/18 (72%); O 6/10 (60%)). Only 4 participants had studied Wordsmith subjects (such as Literature and Languages), and only 2 worked in Wordsmith-related areas. Thus, a greater proportion of experts than non-experts both studied and subsequently worked in STEM or finance-related areas. Job complexity (Cx) was broadly equivalent across the two groups, with O participants following slightly more complex careers (S *M* = 68.9, *SD* = 6.17; O *M* = 72.1, *SD* = 4.65; (*t* (25) = 1.398, *p* = .174, Cohen's *d* = .58, 95% CI [-0.25, 1.41]). One S participant's occupation ("Cryptic crossword compiler") could not be assigned a Holland Cx rating. Overall these academic and workplace statistics were consistent with the larger survey population from which this sample was selected (Friedlander & Fine, 2016).

Performance on the AH5 Test

All 28 participants took the AH5. Out of a maximum of 72, a mean of 44.0 (*SD* = 9.42) items were correctly completed in the time limit, ranging from 27 to 65 for the individual solvers. For Part 1, participants correctly completed a mean of 22.4 (*SD* = 5.61) items out of 36, ranging from 13 to 34; and for Part 2, a mean of 21.6 (*SD* = 5.06) items, ranging from 12 to 34. Details of mean scores by expertise groups, together with comparison populations from the AH5 manual, are given in Table 1.

Table 1. AH5 mean scores (*SD* in brackets)

	n	AH5 Part 1	AH5 Part 2	AH5 overall
Crossword sample				
Ordinary	10	18.3 (3.77)	20.7 (3.92)	39.0 (6.86)
Super-Expert	18	24.7 (5.20)	22.2 (5.64)	46.8 (9.63)
Total crossword sample	28	22.4 (5.61)	21.6 (5.06)	44.0 (9.42)
Comparison with other high ability norms*				
Oxford Science Scholarship students	360	21	23.9	44.9 (8.44)
Oxford Architecture students	402	18	23.5	41.5 (5.95)
Oxford Zoology students	139	17.5	21	38.5 (6.74)
Cambridge Arts students	118	18.5	18.2	36.7 (7.17)

* Comparison totals are taken from the AH5 manual (Heim, 1968, table 3, p. 10). *SD* is only available for the overall *M*.

There was little difference between the overall performance of crossword solvers on the two AH5 parts (*t* (27) = .787, *p* = .44, Cohen's *d* = .13, 95% CI [-.41, .66]), and performance on the two parts was strongly correlated (Pearson's *r* (28) = .56, 95% CI [.23, .77], *p* = .002). Thus, participants tended to be of a fairly consistent standard across the whole AH5.

Comparison of Overall Crossword Group with Normed Samples

In terms of the overall group mean (44.0, *SD* = 9.42), cryptic crossword solvers compared very favorably with Oxford and Cambridge students in the Heim manual (1968), falling just short of the highest listed score (that of Oxford Science Scholarship students, *n* = 360, *M* = 44.9, *SD* = 8.44) - see Table 1 above. This is the highest normed sample mean recorded in the AH5 manual, exceeding other means recorded in the manual for high-grade engineering students (*n* = 1,375, *M* = 40.6, *SD* = 7.58), and well exceeding other groups such as medical students (*n* = 866, *M* = 37.5, *SD* = 7.53) and PG arts teacher-trainees (*n* = 559, *M* = 34.6, *SD* = 7.54). Super-solvers exceeded this score (*M* = 46.8, *SD* = 9.63*)*, thus becoming the highest scoring available sample.

Heim notes that science and arts disciplines perform differently on Part 1 and Part 2 of the AH5, with arts students typically gaining a higher mean on Part 1; and science/architecture/design performing better on Part 2, which is spatially driven (Heim, 1968). The crossword sample as a whole show roughly equivalent scores to scientific populations on Part 2, as might be expected from their typical degree subject and occupational background, yet (in common with the Science Scholars) also perform well on Part 1 scores, with Super-experts scoring outstandingly on this part.

One important point to note is the age difference between these comparison groupings (presumably a young undergraduate sample aged around 18-22yrs) and the crossword sample (mean age 53yrs). Given that fluid intelligence is known to peak from 20yrs and then to decline with age (Deary, 2014; Rabbitt, 1993) on a relatively stable trajectory from baseline (Staff et al., 2018), this implies that the crossword sample in earlier life might have performed at an exceptionally able level. Heim includes AH5 statistics for mature students (university not specified) age 19-32yrs (*n* = 104, *M* = 27.8, *SD* = 7.67) and 33-45yrs (*n* = 109, *M* = 24.9, *SD* = 7.22), and features the frequency distribution curve of this combined group (*n* = 213), together with data for the Oxford Science Scholarship students (*n* = 360), within the manual (Heim, 1968, section XI, p. 19, second unnumbered figure). This figure is replicated below in Figure 3, with the addition of equivalent crossword solver data. From this, it is evident that the crossword population as a whole is performing at a highly superior level, even given their relative age disadvantage. The "double-spiked" profile of the crossword solvers' frequency distribution curve is discussed in the next section.

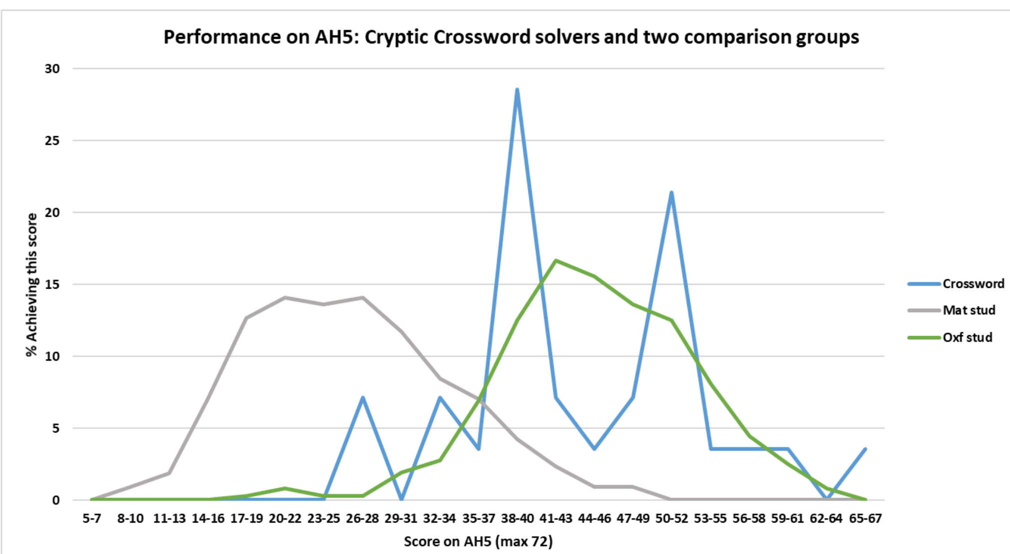

Figure 3: Frequency distribution curves showing performance of mature students, Oxford Science Scholars and crossword solvers (combined groups, *n*=28) on the AH5.

Between-Groups Comparison of AH5 Scores for Cryptic Crossword Solvers

As anticipated in Hypothesis 2, overall AH5 performance was better for S (M = 46.8, SD = 9.63) than O solvers (M = 39.0, SD = 6.86): t (26) = 2.264, p = .032, Cohen's d = 0.94, 95% CI [.09, 1.78]. Although S solvers scored more highly than O on both parts of the test, this difference was clearly driven by performance on Part 1 of the AH5 (t (26) = 3.394, p = .002, Cohen's d = 1.40, 95% CI [0.51, 2.30]; S M = 24.7, SD = 5.20; O M = 18.3, SD = 3.77), as the groups hardly differed for Part 2 (t (26) = .728, p = .473, Cohen's d = .30, 95% CI [-.50, 1.11]).

When viewed in the context of the frequency distribution curves seen in Figure 3, the difference in combined Part 1 and Part 2 scores goes some way towards explaining the "double-spiked" profile of the overall mean scores: the O and S groups should be viewed as distinct populations (see Figure 4).

Differences Between Subtest Scores for Cryptic Crossword Groups

Performance on the AH5 subtests was also analysed, to explore which of the subsidiary tasks were particularly associated with expert performance. Results are set out in Table 2 below.

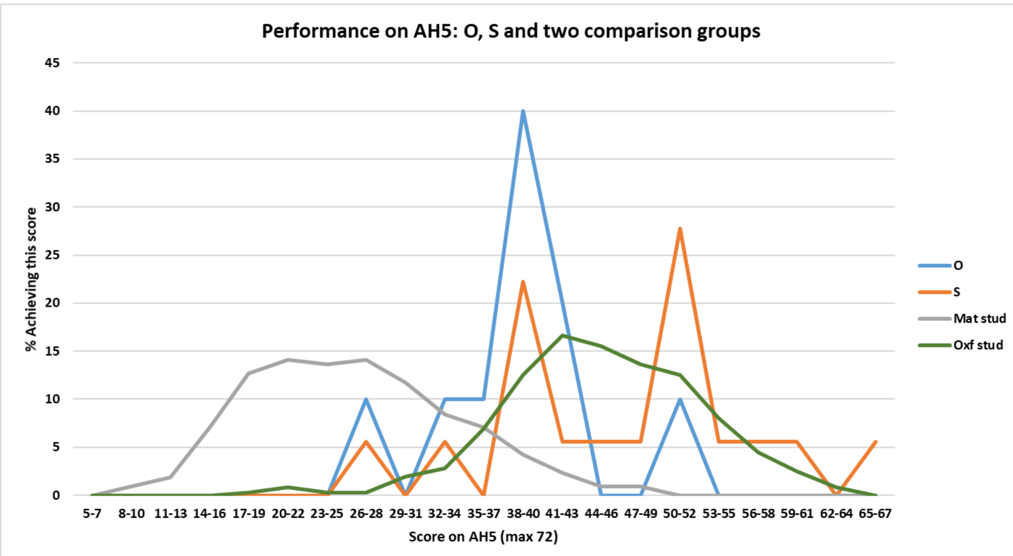

Figure 4. Frequency distribution curves showing performance of mature students, Oxford Science Scholars and crossword solvers (O n=10 and S n=18) on the AH5.

Table 2. AH5 subtest mean scores by crossword expertise group (SDs in brackets)

	Ordinary	Super-experts	All Solvers
Part 1 - verbal / numerical			
Similar Relationships	6.2 (2.25)	7.5 (1.58)	7.0 (1.92)
Directions*	3.7 (1.16)	5.6 (1.62)	4.9 (1.71)
Verbal Analogies*	5.7 (0.82)	6.9 (1.02)	6.5 (1.11)
Numerical Series*	2.7 (1.49)	4.7 (2.27)	4.0 (2.23)
Part 2 - non-verbal diagrammatic			
Analogies	6.1 (1.52)	6.6 (1.42)	6.4 (1.45)
Series	6.0 (1.25)	5.4 (1.92)	5.6 (1.70)
Directions*	4.1 (2.18)	5.7(1.84)	5.1 (2.09)
Features in Common	4.5 (1.65)	4.4 (2.48)	4.5 (2.19)

* p < .05

As a whole, crossword solvers performed best on tasks of analogical reasoning (whether verbal or non-verbal) and on "similar relationships" between a verbal pair and a target word with 5 potential matches. As can also be seen from Table 2, Super-experts performed better on 6 of the 8 subtests than Ordinary solvers, including all those in Part 1, and two of the subtests in Part 2, with Ordinary solvers scoring slightly higher on the Series and Features in Common subtests in Part 2.

Data was not normally distributed for all subtest/expertise combinations (as assessed by Shapiro-Wilk's test), and as a result all results shown are bootstrapped [BCa CI 95%]. Independent *t*-tests showed that Super-experts performed considerably better on five of the subtests, four of which were in Part 1. These subtests were as follows:

Part 1 Directions (t (26) = 3.19, p = .005, Cohen's d = 1.32, *MDiff* = 1.86, 95% CI [.88, 2.76])

Verbal Analogies (t (26) = 3.15, p = .007, Cohen's d = 1.28, *MDiff* = 1.19, 95% CI [.29, 2.04])

Numerical Series (t (26) = 2.52, p = .018, Cohen's d = 1.05, *MDiff* = 2.02, 95% CI [.66, 3.35])

Similar Relationships (t (26) = 1.79, p = .085, Cohen's d = 0.67, *MDiff* = 1.3, 95% CI [-.56, 3.02])

Part 2 Directions (t (26) = 2.09, p = .046, Cohen's d = 0.80, *MDiff* = 1.62, 95% CI [-.02, 3.29])

All effect sizes were medium or large. Group differences were small on the other three Part 2 subtests.

Completion and Solving Times for the Commissioned Cryptic Crossword

As shown in Figure 5, 19 of the 28 solvers finished the crossword in the 45m allowed: 17 experts (of 18) and 2 non-experts (of 10). However, this includes 1 expert who finished in under 28m, but post-trial inspection of the grid revealed one error. A chi-square analysis demonstrated a strong association between expertise group and completion (χ^2 (1) = 16.33, p < .001, Cramer's V = .76, a large effect size). Standardized residuals indicated that O solvers were very much more likely to fail to solve the crossword (z = 2.7), thus validating the initial assignment of cryptic crossword solvers to expertise groups using Friedlander & Fine's (2016) criteria.

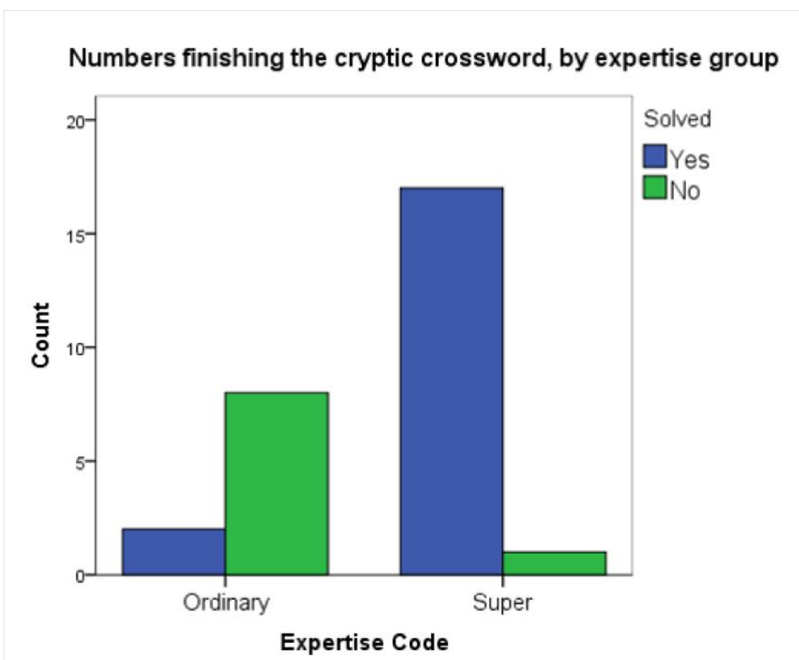

Figure 5. Numbers of cryptic crossword solvers by expertise group finishing the crossword within 45m

Those who finished the crossword took between 647s (10m47s) and 2430s (40m30s). The 10 solvers who did not complete correctly had between 1 and 13 clues left blank or incorrect. In total, out of a possible 756 clues (28 participants, solving 27 clues each), 74 (9.8%) were either omitted or incorrectly solved.

Although perfectly possible to investigate correlations between finishers' solving times and their AH5 scores, this has the disadvantage of ignoring 9 solvers in the analysis (primarily O solvers). Therefore, extrapolated solving times were calculated for all non-finishers as follows. Solvers were assumed to solve clues at a consistent speed, and the number of clues correctly solved in the 45m was noted. This allowed a mean "solving time per clue" to be calculated and, for the non-finishers, added for each unsolved or incorrect clue to the 45m. Additionally, an extrapolated time for the expert solver who finished incorrectly was calculated by assuming they would have taken an average ("per clue") additional time to solve one extra clue, had this error been pointed out at the time.

For all participants, extrapolated solving times now ranged from 647s (10m47s) to 5207s (86m47s). The mean was 2250s (37m30s) with a *SD* of 1374s (22m54s). Extrapolated solving times correlated very strongly with number of clues correctly solved (r_s = -.821, p < .001), confirming the validity of the extrapolation method. Details of extrapolated mean solving times and number of correctly solved clues are given below by expertise group in Table 3.

Table 3. Extrapolated solving times and numbers of correctly solved clues, by expertise group (*SD*s in brackets)

	n	*M*	Min	Max
Solving Time (s)				
Ordinary	10	3762 (985)	2264	5207
Super-Expert	18	1410 (625)	647	3038
Total crossword sample	28	2250 (1374)	647	5207
Clues correctly solved (n)				
Ordinary	10	20.0 (4.83)	14	27
Super-Expert	18	26.8 (0.73)	24	27
Total crossword sample	28	24.4 (4.37)	14	27

A comparison of S and O extrapolated solving times and clues correctly solved was conducted. Data was not normally distributed, and so bootstrapping [BCa CI 95%] was applied. As anticipated in Hypothesis 3, Super-solvers were considerably faster to complete the crossword than Ordinary solvers (t (26) = 7.75, p = .001, Cohen's d = 2.85, *MDiff* = 2352, 95% CI [1620, 3055]) and completed more clues correctly (t (9.23) = 4.41, p = .018, Cohen's d = 1.97, *MDiff* = 6.78, 95% CI [3.46, 9.74]).

Correlation of Crossword Solving Speed and Scores on the AH5

Spearman's correlations were conducted between extrapolated solving times and AH5 performance. Spearman's non-parametric were chosen over Pearson's parametric correlations as the method of calculating extrapolated solving times was fairly arbitrary in terms of absolute times, but rational in terms of relative times, such that the fewer clues a solver completed, the longer their time. Negative correlations imply that a shorter solving time is associated with a higher Gf score.

As anticipated in Hypothesis 4, extrapolated solving time correlated negatively with both overall AH5 performance (r_s = -.48, p = .011) to a moderate effect size, and with Part 1 AH5 performance (r_s = -.72, p < .001), to a strong effect size: see below, Figure 6. However, they did not correlate with Part 2 AH5 performance (r_s = -.05, p = .814).

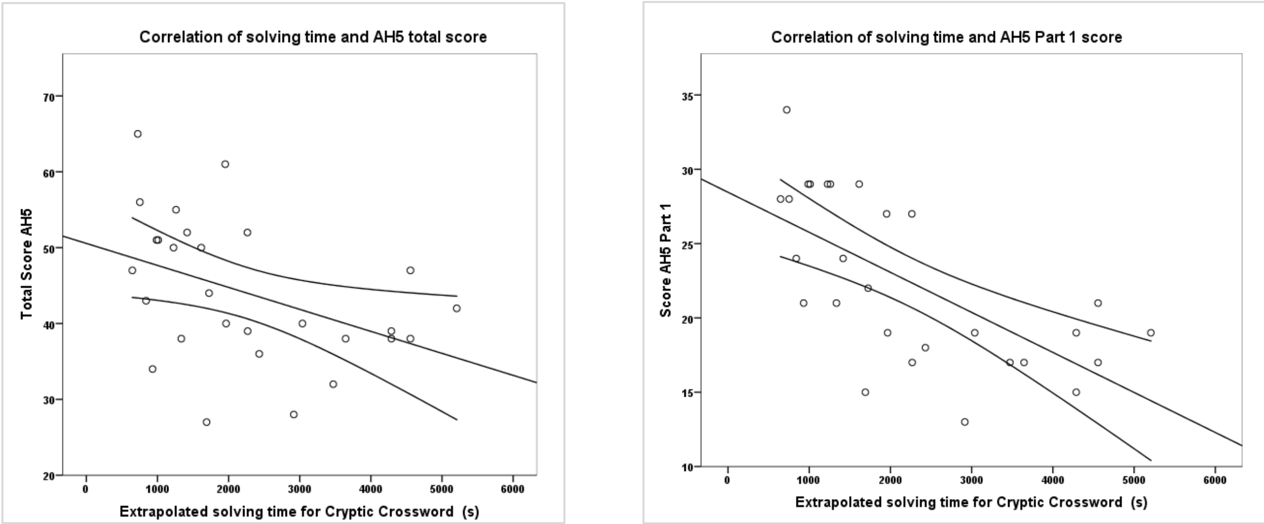

Figure 6a and b. Correlations of solving time with overall AH5 score (a) and AH5 Part 1 (b), including lines indicating CI (95%) of the mean

The 8 individual AH5 subtest scores were also investigated in the same way. Extrapolated solving time correlated strongly and negatively with all 4 Part 1 subtests (similar relationships: r_s = -.52, p = .005; directions: r_s = -.65, p < .001; verbal analogies: r_s = -.66, p < .001; numerical series: r_s = -.53, p = .004, all to a strong or moderate effect size). However correlations with Part 2 subtests, were much weaker with directions being the strongest (r_s = -.29, p = .13, a small-medium effect size).

Spearman's correlations were also investigated between the number of correctly solved clues and AH5 scores. The pattern was the same as that for extrapolated completion times, inasmuch as the number of correct solutions was positively correlated with AH5 scores overall (r_s = .49, p = .008) and for Part 1 (r_s = .65, p < .001), but this was not the case for Part 2 (r_s = .14, p = .492).

Discussion

This study followed the "Grounded Expertise Components Approach" (Friedlander & Fine, 2016), employing the results of a detailed and wide-ranging survey to determine key aspects of follow-up trials in the lab. These elements involved a challenge which was truly representative of domain skill (the completion of a full, professionally-compiled, cryptic crossword) and a battery of tasks including completion of a test of fluid intelligence designed to discriminate amongst high-ability populations (the AH5, Heim, 1968). This novel approach enabled the formulation of hypotheses empirically grounded in the survey results, which were upheld by the subsequent lab-based trials.

Cryptic Crossword Solvers Do Show Elevated Gf Compared to Demographic Norms (H1)

Our findings in this study supported the first hypothesis - that cryptic crossword solvers from both expertise groups would show elevated Gf compared to the general population. This premise had been grounded in the survey results, which found that cryptic crossword solvers were generally academically able adults pursuing cognitively complex professions (Friedlander & Fine, 2016). In the lab trials, overall scores on the AH5 for the cryptic crossword solvers compared very favourably with Oxford and Cambridge student norms listed in the Heim manual, falling just short of the highest listed norm in the manual (that of Oxford Science Scholarship students). This was all the more remarkable, given the difference in average age between the student population (assumed to be 18-22yrs old) and the crossword sample (mean age 53yrs), given that fluid intelligence is argued to decline with age (Deary, 2014; Rabbitt, 1993) on a relatively stable trajectory from baseline (Staff et al., 2018). As a sub-group, cryptic crossword Super-experts exceeded even the mean scores of the Oxford Science Students, thus becoming the highest scoring sample we are currently aware of. However, Ordinary solver scores were also elevated: they remained comparable to Oxford student groupings such as Zoology students. Graphically presented frequency distribution data demonstrated that they appeared to be a distinct population to Super-experts, but still performed at a highly superior level, well above Heim's listed population of mature university students (aged 19-45yrs). This appears to confirm that there is a fluid intelligence threshold for entry into the domain, even at "Ordinary solver" level.

Particular Cognitive Strengths of Cryptic Crossword Solvers

In our survey, cryptic crossword solvers of all levels were predominantly qualified in STEM subjects and continued to work in STEM and financial areas post-university. This trend towards STEM increased with expertise, and Super-experts were significantly more likely to have studied Maths and to have worked in the areas of IT or Banking/Accountancy than the other groups (Friedlander & Fine, 2016). In line with these survey findings, the crossword sample showed roughly equivalent scores to Heim's scientific populations on Part 2, typically thought to favour scientific and spatial thinkers. Additionally, Super-experts did score higher on Part 2 than Ordinary solvers. Nevertheless, crossword solvers as a whole also scored strongly on Part 1 of the AH5, which is concerned with verbal and numerical data, typically favouring arts participants (Heim, 1968), with Super-experts scoring outstandingly on this part, and considerably better than Ordinary solvers. Again, this finds some support in our survey, given that—outside their scientific careers—participants frequently engaged with word-based and cultural hobbies coded as "A" activities (Arts based) under the RIASEC Holland coding system (Holmberg et al., 1997). It is also reasonably safe to assume that the process of successful crossword completion will at least partly involve the possession of richer semantic networks (Friedlander & Fine, 2016).

Across both parts of the AH5, crossword solvers at both expertise levels scored highest on tasks of analogical reasoning ("Analogies" whether verbal or diagrammatic) and on "Similar relationships" between a verbal pair which they had to relate in the same way to 5 potential matches. These tests all require an individual to identify a common relational system between two given instances, and then to generate further inferences driven by these commonalities (Gentner & Smith, 2012). The cognitive processes involved can be characterized by reasoning approaches such as mapping, inference, abstraction and evaluation (De Acedo Lizarraga et al., 2011), facilitating hypothesis formation, the consideration of alternatives, and the understanding of new problems as something familiar (de Fátima Morais, 2009). Analogical thinking is thus seen by some as a core component of scientific creativity and high fluid intelligence (De Acedo Lizarraga et al., 2011; Gentner et al., 2001; Gentner & Smith, 2012), associated with greater interconnectivity of remote associations within the brain (Geake, 2008; Green et al., 2012).

Why might cryptic crossword puzzlers have a particular affinity for this type of reasoning? The discussion above highlighted a number of parallels between cryptic crossword clues and (Compound) Remote Associates Puzzles (RAT/CRA) (Bowden & Jung-Beeman, 2003; Mednick, 1962). In general terms, RAT puzzles employ similar associative processes to the "definition" in cryptic crosswords, and to "double-definition" and punning clues. Impasse in these crossword elements, as for RAT puzzles themselves, may arise from a fixation on more readily available incorrect words, which block access to the more remotely associated words needed for the solution (Friedlander & Fine, 2018; Gupta et al., 2012). This is equally the case for the more complex AH5 "analogies" and "similar relationships" questions, which employ a high level of deliberate distractors and intrusive elements, requiring suppression and the avoidance of fixation, together with increasingly tangential associations with the correct target word.

Within-Expertise Comparison: Super-Experts Have Higher Gf Than Ordinary Solvers (H2)

Our study cannot definitively prove that this keen ability to think associatively and analogically is an innate aptitude of cryptic crossword solvers, rather than a skill honed by decades of engagement with cryptic crossword puzzles. However, the between-group comparison of solvers lends considerable support to the "aptitude" argument. Our second hypothesis posited that Super-expert solvers would demonstrate higher Gf than Ordinary solvers, and this was demonstrated for the AH5 as a whole, and for Part 1 scores in particular (the groups did not differ statistically on Part 2 scores overall). Given that the groups were fully matched on key demographic criteria such as age, years solving, and hours spent solving each week, and indeed had both been solving for over 3 decades, practice effects are highly unlikely to account for performance differences between the solver groups, suggesting that there is indeed an innate component which leads to the development of crossword expertise. In fact, in common with studies in other fields (e.g. Gobet & Campitelli, 2007; Hambrick et al., 2014; Staff et al., 2019), Ordinary-level performers had typically engaged with the domain for over 13,000h by the time of the trial (M = 7.6h/w x52 x 32.9y), well exceeding the "10,000 hour rule" (Gladwell, 2008; Hambrick et al., 2016), but had not progressed to higher expertise levels, as the "deliberate practice" account would predict.

Differences Leading to The Super-Solver Superiority on AH5

Although both solver groups performed particularly highly on subtests employing the ability to think analogically and associatively ("Analogies," whether verbal or diagrammatic, "Similar relationships"), notable differences between the solver groups only appeared in five primary areas: "Directions" (verbal and non-verbal), "Verbal analogies," "Similar Relationships," and "Numerical series." Super-experts outperformed Ordinary solvers on all five of these areas, with effect sizes ranging from medium to large. Skill sets involved in "Analogies" and "Similar Relationships," and their relationship to cryptic crossword solving have already been discussed above. In terms of "Directions," the AH5 test employs multiple strategies to distract the solver with deliberately complex challenges, requiring attention to detail, the retention of values and instructions during sequencing and organization, and the resistance of intrusion from similar, but subtly different previous items. This loads very heavily, therefore, on Working Memory and on Executive Functions such as focus, maintenance, inhibition, disengagement, place-keeping, evaluation and sequencing/ organization. Similar skills are tested in the "Numerical series" tests, with some tests presenting deliberate traps, requiring focus and attention to evade them successfully.

As noted above, cryptic crossword clues present an infinitely varied range of quasi-algebraic coded instructions, distracting the solver through deliberate red herrings, which must be inhibited if progress is to be made. Clue types can be used singly or combined in multi-part instructions, but must always be deduced and segregated through the analytical deconstruction of the clue itself, in order to deduce the governing rules (Friedlander & Fine, 2016, 2018). This requires the non-literal interpretation of individual clue components, overriding the natural reading and "deep structure" of the text, which is tacitly invoked through a life-time's experience of reading (Aarons, 2015). Instructions must then be mentally maintained, and executed precisely, in order to arrive at the correct answer to the clue (Friedlander & Fine, 2016). A diverse range of cognitive abilities, allied to the Working Memory and Executive Function skills involved in the AH5 subtests, is therefore likely to be involved in solving these puzzles. This may in turn explain why Super-expert performance is associated with superior outcomes on these subtasks of the AH5. Again, the equivalence between the groups, in terms of solving experience and other key demographic factors, makes it unlikely that cryptic crossword solving in itself had produced this group difference on the AH5; and indeed brain-training literature in general has not supported such transfer effects for WM/EF (Simons et al., 2016).

Super-Experts Perform Better on the Domain Representative Task, Which Is Correlated with Gf Scores (H3/4)

As expected (Hypothesis 3), Super-solvers were demonstrably better than Ordinary solvers at solving the bespoke crossword, in terms of crossword completion during the time limit, extrapolated speed of solving and number of clues completed. This validates the initial assignment of cryptic crossword solvers to expertise groups using Friedlander & Fine's (2016) criteria. Hypothesis 4 proposed that the time taken to solve a complete bespoke cryptic crossword would correlate negatively with Gf scores, such that the higher the score on Gf, the faster an individual would be to solve the cryptic crossword. This was again supported in our trials: times correlated negatively with the AH5 overall and AH5 Part 1, though not with Part 2 scores. A similar pattern was observed for the AH5 subtests, with all those in Part 1 showing correlations with solving speed, but only "Directions" in Part 2 showing a correlation, with small-medium effect size. This implies that those differences on the AH5 which distinguished between Super-experts and Ordinary solvers are also associated with the efficient solving of cryptic crosswords, and that Gf, particularly when associated with verbal/numerical rather than spatially oriented challenges, is highly relevant to the domain.

Limitations

The study was based on a small sample of 28 participants, since numbers were constrained by the practicalities of transcribing extensive VPA material arising from the video-recorded tasks, and by the difficulties of recruiting a high-expert population. For this reason, results can only be interpreted as indicative; and indeed, we have "retired statistical significance" in line with best practice in small expertise studies (Campitelli, 2019).

In order to mitigate against the small sample size, we also deliberately invited two polarized subgroups - Super-experts and Ordinary solvers - to take part. Participants were drawn from the original survey population, which had responded to open invitations on a wide variety of web-based platforms covering the entire range of crossword difficulty. The survey population thus represented the full spectrum of crossword solving expertise, and their high academic achievements were not a product of "snowballing" within academic circles or personal contacts. Care was taken to make sure that the sample in this study was as representative as possible of the general survey population from which participants were drawn, and results indicated that the sample matched the survey population in a number of key demographic and experience-related factors. Additionally, Super-experts were drawn from the full range of expertise proficiency areas—not just "speed solving"—with at least 6 representatives in each dimension of expertise. Finally, invitations were extended to participants on a non-systematic basis, within the broad expertise groupings, largely based on their geographic proximity to the University premises and availability during the trial period.

The split of participants into expertise groupings was based on previously established criteria (Friedlander & Fine, 2016) which are a pragmatic blend of "reputation-based" and "performance-based" metrics (Gobet, 2017). Although Super-experts were all known to the researchers either personally or by reputation, Ordinary solvers were assigned to this category purely on the basis of their self-assessed responses to the original survey. Nevertheless, we have no reason to believe that Ordinary solvers had cause to engage in false modesty during the survey process; and indeed the results of the "domain-representative task" - solving a professionally compiled cryptic crossword of medium difficulty - emphatically endorsed the assignment of participants to their expertise groups, with Super-experts being considerably more likely to complete the puzzle within 45 minutes.

Because of the large number of non-completions within the Ordinary solver sample, extrapolated completion times were used to explore the correlation between the AH5 scores and the time taken to tackle the crossword. This somewhat arbitrary method of calculating absolute finishing times was mitigated against by using non-parametric statistical analysis, which would have used relative times in an ordinal fashion to calculate the correlation.

It is possible that the “speak-aloud” process, and the knowledge that the session was being video-recorded, may have impacted adversely on the absolute crossword-solving times of participants. However, there is no reason to suggest that Ordinary solvers would have been particularly disadvantaged by this process (indeed, in many ways the Super-experts had more to lose in these timed trials, and might therefore have been more inhibited by the presence of a camera). We therefore believe that our comparison of relative (rather than absolute) solving speeds remains valid. Indeed, research by Gilhooly (2007) indicated that the “think aloud” protocol does not cause verbal overshadowing affecting the fluency or novelty of idea production in a divergent thinking task, which may suggest that interference would be minimal (so also Ericsson & Simon, 1993). Additionally, we deliberately arranged for participants to engage in two video-recorded “speak-aloud” word-games as a warm-up process (lasting 20m in total) in order to familiarize themselves with the setting and procedure.

Conclusion and Future Directions

Research into expert performance has traditionally focused upon a limited number of domains, often exploring a restricted set of factors based on *a priori* assumptions about the skill sets required for excellence in the field. Cryptic crosswords bring fresh perspectives to the debate: the domain is typically unburdened by intensive practice regimes, has a comparatively late starting age for engagement, and is driven more by intrinsic motivators than by the lure of monetary reward or international prestige (Friedlander & Fine, 2016). In this small-scale study, we have demonstrated that fluid intelligence appears to be fundamentally important both to ordinary-level engagement in the domain, and to the development of high expertise, thus adding to the growing body of literature which challenges the “deliberate practice” framework of high expertise (Hambrick et al., 2016). Given the small sample size, a crucial next step will be the replication of these results in follow-up studies, to confirm the importance of the relationship between Gf and success in cryptic crossword solving.

Other future directions of research will include the analysis of the VPA trace recorded during the solving of the bespoke cryptic crossword, to explore whether different solver expertise groups go about solving in distinctive ways. We also intend to explore a number of sub-skills strongly indicated by the results of this research program, such as the importance of remote associations to the cryptic crossword solving process; the triggering of “insight moments” and their relationship to expertise; and the need for resistance to red herrings and intrusion implied by the clue format. Finally, we intend to explore the construct of “deliberate practice” and the extent of its relevance to the cryptic crossword solving community. From this we hope to present a multi-faceted understanding of the drivers of excellence in this novel and relatively unexplored domain, which may in turn refine our understanding of expertise in other less familiar domains, pursued out of the limelight of intense competition.

Endnote

1. The crossword was eventually published in the *Independent* on 25 November 2011 as #7835. It has been blogged by Phi subsequently (Henderson, nd) with the full pdf and solution to the puzzle.

Acknowledgements

This research was made possible through the generosity of the Dennison Research Fund at the University of Buckingham, which enabled us to defray participants' costs of travel to the University lab and to commission the cryptic crossword. The authors are also indebted to a great number of people in the cryptic crossword community. We wish to thank in particular the owners and administrators of the websites who allowed us to advertise for participants (*Times for the Times, 15 Squared, Big Dave's Crossword Blog,* and *Derek Harrison's Crossword Centre*); John Green for advertising the survey with the annual *Listener* statistics; Peter Biddlecombe for recruiting volunteers at the *Times National Crossword Championship*; and the team at the *Magpie* crossword magazine. A final thank you is due to all the participants who have taken part in our research, and particularly to those who gave up their time to attend these lab trials. Many have now contributed to more than one phase in our crossword studies, and we are extremely grateful for their continued faith in the value of our research and their enthusiasm to take part.

Authors' Declarations

The authors declare that there are no personal or financial conflicts of interest regarding the research in this article.

The authors declare that they conducted the research reported in this article in accordance with the [Ethical Principles](#) of the Journal of Expertise.

The authors declare that they are not able to make the dataset publicly available but are able to provide it upon request. The cryptic crossword featured in this research, including solution, is available at http://phionline.net.nz/my-other-puzzles/independent-newspaper/independent-7835/ (abbreviated as "Henderson, nd," in Endnote).

References

Aarons, D. L. (2015). Following Orders: Playing Fast and Loose with Language and Letters. *Australian Journal of Linguistics, 35*(4), 351-380. https://doi.org/10.1080/07268602.2015.1068459

Ackerman, P. L. (1987). Individual Differences in Skill Learning: An Integration of Psychometric and Information Processing Perspectives. *Psychological Bulletin*, *102*(1), 3-27. https://doi.org/10.1037/0033-2909.102.1.3

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of experimental psychology: General, 117*(3), 288.

Ackerman, P. L. (2014a). Facts are stubborn things. *Intelligence*, *45*, 104-106. https://doi.org/10.1016/j.intell.2014.01.002

Ackerman, P. L. (2014b). Nonsense, common sense, and science of expert performance: Talent and individual differences. *Intelligence*(45), 6-17. https://doi.org/10.1016/j.intell.2013.04.009

Amrhein, V., Greenland, S., & McShane, B. (2019). Retire statistical significance. *Nature, 567*, 305-307.

Benedek, M., Franz, F., Heene, M., & Neubauer, A. C. (2012). Differential effects of cognitive inhibition and intelligence on creativity. *Personality and Individual Differences, 53*(4), 480-485. https://doi.org/10.1016/j.paid.2012.04.014

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., & Neubauer, A. C. (2014). Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence, 46*, 73-83. https://doi.org/10.1016/j.intell.2014.05.007

Biddlecombe, P. (2009). *Yet Another Guide to Cryptic Crosswords (YAGCC) - Clue Types*. http://www.biddlecombe.demon.co.uk/yagcc/YAGCC2.html

Bolton, P. (2012). *Education: historical statistics* (SN/SG/4252). House of Commons Library, London. http://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN04252

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers, 35*(4), 634-639. https://doi.org/10.3758/BF03195543

Bowden, E. M., & Jung-Beeman, M. (2007). Methods for investigating the neural components of insight. *Methods, 42*(1), 87-99. https://doi.org/10.1016/j.ymeth.2006.11.007

Bowden, E. M., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight *Trends in Cognitive Sciences, 9*(7), 322-328 https://doi.org/10.1016/j.tics.2005.05.012

Burgoyne, A. P., Harris, L. J., & Hambrick, D. Z. (2019). Predicting piano skill acquisition in beginners: The role of general intelligence, music aptitude, and mindset. *Intelligence, 76*, 101383. https://doi.org/10.1016/j.intell.2019.101383

Burgoyne, A. P., Nye, C. D., Macnamara, B. N., Charness, N., & Hambrick, D. Z. (2019). The Impact of Domain-Specific Experience on Chess Skill: Reanalysis of a Key Study. *The American Journal of Psychology, 132*(1), 27-38. https://doi.org/10.5406/amerjpsyc.132.1.0027

Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., & Hambrick, D. Z. (2016). The relationship between cognitive ability and chess skill: A comprehensive meta-analysis. *Intelligence, 59*, 72-83. https://doi.org/10.1016/j.intell.2016.08.002

Campitelli, G. (2019). Retiring Statistical Significance from Psychology and Expertise Research. *Journal of Expertise, 2*(4).

Campitelli, G., Connors, M. H., Bilalić, M., & Hambrick, D. Z. (2015). Psychological perspectives on expertise. *Frontiers in Psychology, 6*, 258. https://doi.org/10.3389/fpsyg.2015.00258

Cane, V., & Heim, A. W. (1950). The effects of repeated retesting: III. Further experiments and general conclusions. *Quarterly Journal of Experimental Psychology, 2*(4), 182-197.

Canestrari, C., & Bianchi, I. (2012). Perception of Contrariety in Jokes. *Discourse Processes, 49*(7), 539-564. https://doi.org/10.1080/0163853X.2012.710524

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*(3), 153.

Cattell, R. B. (1957). *Personality and motivation structure and measurement*. World Book.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1-21.

Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics, Ch2. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. Feltovich (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 21-30). Cambridge University Press.

Chi, M. T. H. (2011). Theoretical perspectives, methodological approaches, and trends in the study of expertise. In Y. D. Li & G. Kaiser (Eds.), *Expertise in mathematics instruction* (pp. 17-39). Springer.

Cleary, J. (1996). Misleading Contexts: The Construction of Ambiguity in the Cryptic Crossword Clue. *Edinburgh Working Papers in Applied Linguistics, 7*, 14-29.

Connor, A. (2014, 20 October). Crossword blog: Watching a champion solver at work. *The Guardian*. http://www.theguardian.com/crosswords/crossword-blog/2014/oct/20

Danek, A. H., Fraps, T., von Müller, A., Grothe, B., & Öllinger, M. (2013). Aha! experiences leave a mark: facilitated recall of insight solutions. *Psychological Research, 77*(5), 659-669. https://doi.org/10.1007/s00426-012-0454-8

De Acedo Lizarraga, M. L. S., De Acedo Baquedano, M. T. S., & Closas, A. H. (2011). Development and validation of a questionnaire to assess analogy and creativity in open problem-solving. *Studia Psychologica, 53*(3), 235.

De Bruin, A. B., Kok, E. M., Leppink, J., & Camp, G. (2014). Practice, intelligence, and enjoyment in novice chess players: A prospective study at the earliest stage of a chess career. *Intelligence, 45*, 18-25. https://doi.org/10.1016/j.intell.2013.07.004

de Fátima Morais, M., Almeida, L., Azevedo, M., & Nęcka, E. (2009). Cognitive dimensions of creativity: What makes the difference between creative and non-creative university students? *Polish Psychological Bulletin, 40*(2), 55-61.

de Groot, A. (1946/1965). *Thought and choice in chess*. Mouton Publishers. https://doi.org/10.2478/s10059-009-0022-2

Deary, I. J. (2014). The stability of intelligence from childhood to old age. *Current Directions in Psychological Science, 23*(4), 239-245. https://doi.org/10.1177/0963721414536905

Deary, I. J., & Smith, P. (2004). Intelligence Research and Assessment in the United Kingdom. In R. J. Sternberg (Ed.), *International handbook of intelligence* (pp. 1-48). Cambridge University Press.

Deihim-Aazami, C. (1999). *Cognitive expertise in solving crossword puzzles* [PhD, University of Nottingham].

Detterman, D. K., Gabriel, L. T., & Ruthsatz, J. M. (1998). Absurd environmentalism (a response to Howe, Davidson and Sloboda, 1998). *Behavioral and Brain Sciences, 21*(03), 411-412. https://doi.org/10.1017/S0140525X98271238

DeYoung, C. G., Flanders, J. L., & Peterson, J. B. (2008). Cognitive abilities involved in insight problem solving: an individual differences model. *Creativity Research Journal, 20*(3), 278-290. https://doi.org/10.1080/10400410802278719

Dominowski, R. L., & Buyer, L. S. (2000). Retention of problem solutions: the re-solution effect. *American Journal of Psychology, 113*(2), 249-274. https://doi.org/10.2307/1423730

Duncan, J., Schramm, M., Thompson, R., & Dumontheil, I. (2012). Task rules, working memory, and fluid intelligence. *Psychonomic Bulletin & Review, 19*(5), 864-870. https://doi.org/10.3758/s13423-012-0225-y

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and Exceptional Performance: Evidence of Maximal Adaptation to Task Constraints. *Annual Review of Psychology, 47*, 273-305.

Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007). The making of an expert. *Harv Bus Rev, 85*(7-8), 114-121.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis* (Revised ed.). The MIT Press.

Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits*. Cambridge University Press.

Ericsson, K. A., & Towne, T. J. (2010). Expertise. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(3), 404-416. https://doi.org/10.1002/wcs.47

Ericsson, K. A., & Ward, P. (2007). Capturing the naturally occurring superior performance of experts in the laboratory: Toward a science of expert and exceptional performance. *Current Directions in Psychological Science, 16*(6), 346-350. https://doi.org/10.1111/j.1467-8721.2007.00533.x

Evans, J. S., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science, 8*(3), 223-241. https://doi.org/10.1177/1745691612460685

Fallon, A. (2010, 31st Oct). Transsexual Mikki Nicholson, who only began playing Scrabble five years ago, beats opponent Mark Nyman to take British national crown. *The Guardian*. https://www.theguardian.com/lifeandstyle/2010/oct/31/scrabble-national-champion-crown-mikki

Fatsis, S. (2011, September). An Outtake from Word Freak: The Enigmatic Nigel Richards. *The Last Word - The Independent Tournament SCRABBLE® Newsletter*. http://www.thelastwordnewsletter.com/Last_Word/Archives_files/TLW%20September%202011.pdf

Friedlander, K. J. (2019, Jan 8th). Solving the Puzzle of Expertise Research. *CREATE Ψ* https://createpsy.com/2019/01/08/solving-the-puzzle-of-expertise-research/

Friedlander, K. J., & Fine, P. A. (2016). The Grounded Expertise Components Approach in the novel area of cryptic crossword solving. *Frontiers in Psychology, 7*, 1-21. https://doi.org/10.3389/fpsyg.2016.00567

Friedlander, K. J., & Fine, P. A. (2018). “The Penny Drops”: Investigating insight through the medium of cryptic crosswords. *Frontiers in Psychology*, 1-22. https://doi.org/10.3389/fpsyg.2018.00904

Geake, J. G. (2008). The neurobiology of giftedness. [Paper presentation]. 10th Asia Pacific Conference on Giftedness, Singapore.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (2001). *The analogical mind: Perspectives from cognitive science*. MIT press.

Gentner, D., & Smith, L. H. (2012). Analogical reasoning. In V. Ramachandran (Ed.), *Encyclopedia of human behavior* (pp. 130-136). Elsevier.

Gilhooly, K. J., Fioratou, E., Anthony, S. H., & Wynn, V. (2007). Divergent thinking: strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology, 98*(4), 611-625. https://doi.org/10.1348/096317907X173421

Gilhooly, K. J., & Green, C. (1996). Protocol analysis: theoretical background. In J. T. E. Richardson (Ed.), *Handbook of qualitative research methods for psychology and the social sciences.* (pp. 43-54). BPS Books.

Gladwell, M. (2008). *Outliers: The Story of Success*. Little, Brown and Company.

Gobet, F. (2015). *Understanding Expertise: A Multi-disciplinary Approach*. Palgrave Macmillan.

Gobet, F. (2017). Three views on expertise: Philosophical implications for rationality, knowledge, intuition and education. *Journal of Philosophy of Education, 51*(3), 605-619. https://doi.org/10.1111/1467-9752.12253

Gobet, F., & Campitelli, G. (2007). The role of domain-specific practice, handedness, and starting age in chess. *Developmental Psychology, 43*(1), 159-172. https://doi.org/10.1037/0012-1649.43.1.159

Gobet, F., & Charness, N. (2006). Expertise in Chess. In K. A. Ericsson, N. Charness, P. Feltovich, & R. R. Hoffman (Eds.), *Cambridge handbook on expertise and expert performance* (pp. 523-538). Cambridge University Press.

Gobet, F., & Ereku, M. H. (2014). Checkmate to deliberate practice: The case of Magnus Carlsen. *Frontiers in Psychology, 5*, 10-12. https://doi.org/10.3389/fpsyg.2014.00878

Gottfredson, L. S. (1998). The General Intelligence Factor. *Scientific American Presents, 9*(4), 24-29. http://www.udel.edu/educ/gottfredson/reprints/1998generalintelligencefactor.pdf

Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. *Human Performance, 15*(1-2), 25-46.

Grabner, R. H. (2014). The role of intelligence for performance in the prototypical expertise domain of chess. *Intelligence, 45*, 26-33. https://doi.org/10.1016/j.intell.2013.07.023

Grabner, R. H., Stern, E., & Neubauer, A. C. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica, 124*(3), 398-420. https://doi.org/10.1016/j.actpsy.2006.07.008

Green, A. E., Kraemer, D. J., Fugelsang, J. A., Gray, J. R., & Dunbar, K. N. (2012). Neural correlates of creativity in analogical reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(2), 264. https://doi.org/10.1037/a0025764

Green, C., & Gilhooly, K. J. (1996). The practical use of protocol analysis: promises and pitfalls. In J. T. E. Richardson (Ed.), *Handbook of qualitative research methods for psychology and the social sciences.* BPS Books.

Gupta, N., Jang, Y., Mednick, S. C., & Huber, D. E. (2012). The road not taken: Creative solutions require avoidance of high-frequency responses. *Psychological Science, 23*(3), 288-294. https://doi.org/10.1177/0956797611429710

Halpern, D. F., & Wai, J. (2007). The world of competitive Scrabble: Novice and expert differences in visuospatial and verbal abilities. *J Experimental Psychology: Applied, 13*(2), 79-94. https://doi.org/10.1037/1076-898X.13.2.79

Hambrick, D. Z. (2015, 21 April). Winning SCRABBLE and the Nature of Expertise. *Scientific American*. https://www.scientificamerican.com/article/winning-scrabble-and-the-nature-of-expertise/

Hambrick, D. Z., & Altmann, E. M. (2015). The role of placekeeping ability in fluid intelligence. *Psychonomic Bulletin & Review, 22*(4), 1104-1110. https://doi.org/ 10.3758/s13423-014-0764-5

Hambrick, D. Z., & Hoffman, R. R. (2016). Expertise: A Second Look. *IEEE Intelligent*

Systems, 31(4), 50-55. https://doi.org/10.1109/MIS.2016.69

Hambrick, D. Z., Macnamara, B. N., Campitelli, G., Ullén, F., & Mosing, M. A. (2016). Beyond born versus made: A new look at expertise. In B. H. Ross (ed.) *Psychology of learning and motivation* (Vol. 64, pp. 1-55). Elsevier. https://doi.org/10.1016/bs.plm.2015.09.001

Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence, 45*, 34-45. https://doi.org/10.1016/j.intell.2013.04.001

Hambrick, D. Z., Salthouse, T. A., & Meinz, E. J. (1999). Predictors of Crossword Puzzle Proficiency and Moderators of Age–Cognition Relations. *Journal of Experimental Psychology*, *128*(2), 131-164.

Healey, M. K., Campbell, K. L., Hasher, L., & Ossher, L. (2010). Direct evidence for the role of inhibition in resolving interference in memory. *Psychological Science, 21*(10), 1464-1470. https://doi.org/10.1177/095679761038212

Heim, A. W. (1968). *AH5 Group Test of High-grade Intelligence: Manual. Revised Edition*. National Foundation for Educational Research, Slough.

Heim, A. W. (1970). *AH4 Group Test of General Intelligence: Manual*. National Foundation for Educational Research, Slough.

Heitz, R. P., Redick, T. S., Hambrick, D. Z., Kane, M. J., Conway, A. R. A., & Engle, R. W. (2006). Working Memory, executive function, and general fluid intelligence are not the same. *Brain and Behavioural Sciences, 29*(2), 135. https://doi.org/10.1017/S0140525X06319036

Henderson, P. (nd). Independent 7835. *Phi Crosswords*. http://phionline.net.nz/my-other-puzzles/independent-newspaper/independent-7835/

Hicks, K. L., Harrison, T. L., & Engle, R. W. (2015). Wonderlic, working memory capacity, and fluid intelligence. *Intelligence, 50*, 186-195. https://doi.org/10.1016/j.intell.2015.03.005

Holmberg, K., Rosen, D., & Holland, J. L. (1997). *The Leisure Activities Finder, SDS Form R* (2nd ed.). Psychological Assessment Resources, Inc.

Howard, R. W. (2009). Individual differences in expertise development over decades in a complex intellectual domain. *Memory & Cognition, 37*, 194-209. https://doi.org/10.3758/MC.37.2.194

Howard, R. W. (2011). Does high-level intellectual performance depend on practice alone? Debunking the Polgar sisters case. *Cognitive Development, 26*(3), 196-202. https://doi.org/10.1016/j.cogdev.2011.04.001

Howe, M. J. A., Davidson, J. W., & Sloboda, J. A. (1998). Innate talents: Reality or myth? *Behavioral and Brain Sciences, 21*(3), 399-406; responses at 407-442. https://doi.org/10.1017/S0140525X9800123X

Johnstone, A. H. (2001). Can Problem Solving Be Taught? Proceedings from the Nyholm Symposium: Are we teaching our students the skills they need? *University Chemistry Education, 5*(2), 69-73.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 66-71. https://doi.org/10.1037/0033-2909.131.1.66

Katz-Brown, J. (2006, Nov 14th). No Loss for Words: Why I play competitive Scrabble. *MIT Technology Review*. https://www.technologyreview.com/s/406847/no-loss-for-words/

Kenett, Y. N., Anaki, D., & Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Frontiers in Human Neuroscience, 8*, 407. https://doi.org/10.3389/fnhum.2014.00407

Koestler, A. (1964). *The act of creation*. Hutchinson.

Kuhn, G., Olson, J. A., & Raz, A. (2016). Editorial: The Psychology of Magic and the Magic of Psychology. *Frontiers in Psychology, 7*(1358). https://doi.org/10.3389/fpsyg.2016.01358

Lane, D. M., & Chang, Y.-H. A. (2018). Chess knowledge predicts chess memory even after controlling for chess experience: Evidence for the role of high-level processes. *Memory & Cognition, 46*(3), 337-348. https://doi.org/10.3758/s13421-017-0768-2

Lubinski, D., Benbow, C. P., Webb, R. M., & Bleske-Rechek, A. (2006). Tracking exceptional human capital over two decades. *Psychological Science, 17*(3), 194-199. https://doi.org/10.1111/j.1467-9280.2006.01685.x

Lyons, B. D., Hoffman, B. J., & Michel, J. W. (2009). Not much more than g? An examination of the impact of intelligence on NFL performance. *Human Performance, 22*(3), 225-245. https://doi.org/10.1080/08959280902970401

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science, 25*(8), 1608-1618.

Manley, D. (2014). *Chambers Crossword Manual* (5th Revised ed.). Chambers.

McCabe, D. P., Roediger III, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology, 24*(2), 222. https://doi.org/10.1037/a0017619

McCloskey, G., & Perkins, L. A. (2012). *Essentials of executive functions assessment* (Vol. 68). John Wiley & Sons.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1-10. https://doi.org/10.1016/j.intell.2008.08.004

McPherson, G. E., & Williamon, A. (2015). Building gifts into musical talents. In G. E. McPherson (Ed.), *The child as musician: A handbook of musical development* (pp. 340-360). OUP.

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review, 69*(3), 220-232. https://doi.org/10.1037/h0048850

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological Science, 21*(7), 914-919. https://doi.org/10.1177/0956797610373933

Minsky, M., & Papert, S. (1974). Artificial intelligence: Condon lectures. *Eugene, Oregon: Oregon System of Higher Education*.

Moxley, J. H., Ericsson, K. A., Scheiner, A., & Tuffiash, M. (2015). The Effects of Experience and Disuse on Crossword Solving. *Applied Cognitive Psychology, 29*(1), 73-80. https://doi.org/10.1002/acp.3075

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*(2), 77-101. https://doi.org/10.1016/bs.plm.2015.09.001

Nickerson, R. S. (1977). Crossword puzzles and lexical memory. In S. Dornič (Ed.), *Attention and Performance VI: Proceedings of the Sixth International Symposium on Attention and Performance, Stockholm, Sweden, July 28-August 1, 1975* (pp. 699-718). Lawrence Erlbaum.

Nickerson, R. S. (2011). Five down, Absquatulated: crossword puzzle clues to how the mind works. *Psychonomic Bulletin Review, 18*(2), 217-241. https://doi.org/10.3758/s13423-011-0069-x

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: new findings and theoretical developments. *American Psychologist, 67*(2), 130. https://doi.org/10.1037/a0026699

Nusbaum, E. C., & Silvia, P. J. (2011). Are intelligence and creativity really so different? Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence, 39*(1), 36-45. https://doi.org/10.1016/j.intell.2010.11.002

Nyongesa, M. K., Ssewanyana, D., Mutua, A. M., Chongwo, E., Scerif, G., Newton, C. R.

J. C., & Abubakar, A. (2019). Assessing Executive Function in adolescence: A scoping review of existing measures and their psychometric robustness. *Frontiers in Psychology, 10*(311). https://doi.org/10.3389/fpsyg.2019.00311

Olteţeanu, A.-M., & Falomir, Z. (2015). comRAT-C: A computational compound Remote Associates Test solver based on language data and its comparison to human performance. *Pattern Recognition Letters, 67,* 81-90. https://doi.org/10.1016/j.patrec.2015.05.015

Platz, F., Kopiez, R., Lehmann, A. C., & Wolf, A. (2014). The influence of deliberate practice on musical achievement: a meta-analysis. *Frontiers in Psychology, 5,* 646. https://doi.org/10.3389/fpsyg.2014.00646

Rabbitt, P. (1993). Does it all go together when it goes? The Nineteenth Bartlett Memorial Lecture. *The Quarterly Journal of Experimental Psychology, 46*(3), 385-434.

Raven, J. C., & Court, J. H. (1988). *Raven's progressive matrices and vocabulary scales*. Oxford Psychologists Press.

Reardon, R. C., Bullock, E. E., & Meyer, K. E. (2007). A Holland perspective on the US workforce from 1960 to 2000. *The Career Development Quarterly, 55*(3), 262-274. https://doi.org/10.1002/j.2161-0045.2007.tb00082.x

Rimfeld, K., Malanchini, M., Krapohl, E., Hannigan, L. J., Dale, P. S., & Plomin, R. (2018). The stability of educational achievement across school years is largely explained by genetic factors. *NPJ science of learning, 3*(16), 1-10. https://doi.org/10.1038/s41539-018-0030-0

Romano, M. (2006). *Crossworld: One Man's Journey into America's Crossword Obsession.* Broadway Books.

Ruthsatz, J., Detterman, D., Griscom, W. S., & Cirullo, B. A. (2008). Becoming an expert in the musical domain: It takes more than just practice. *Intelligence, 36*(4), 330-338. https://doi.org/10.1016/j.intell.2007.08.003

Schulman, A. (1996). The Art of the Puzzler. In M. P. Friedman & E. C. Carterette (Eds.), *Cognitive Ecology: Handbook of Perception and Cognition* (2nd ed., Vol. 16, pp. 293-321). Academic Press.

Schweizer, K., & Koch, W. (2002). A revision of Cattell's Investment Theory: Cognitive properties influencing learning. *Learning and Individual Differences, 13*(1), 57-82.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science, 11*(6), 771-799. https://doi.org/10.1177/1745691616650647

Simon. (2004). The cryptic crossword puzzle as a useful analogue in teaching programming. In R. Lister & A. Young, ACM International Conference Series ACE '04: The Sixth Australasian Conference on Computing Education, Dunedin.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do “brain-training” programs work? *Psychological Science in the Public Interest, 17*(3), 103-186. https://doi.org/10.1177/1529100616661983

Smith, K. A., Huber, D. E., & Vul, E. (2013). Multiply-constrained semantic search in the Remote Associates Test. *Cognition, 128*(1), 64-75. https://doi.org/10.1016/j.cognition.2013.03.001

Smith, S. M., Sifonis, C. M., & Angello, G. (2012). Clue insensitivity in remote associates test problem solving. *The Journal of Problem Solving, 4*(2), 28-49. https://doi.org/10.7771/1932-6246.1124

Staff, R. T., Hogan, M. J., Williams, D. S., & Whalley, L. J. (2018). Intellectual engagement and cognitive ability in later life (the “use it or lose it” conjecture): longitudinal, prospective study. *British Medical Journal, 363*, k4925. https://doi.org/10.1136/bmj.k4925

Staff, T., Gobet, F., & Parton, A. (2019). Investigating the Period of Practice Needed to Acquire Expertise in Great Britain 2012 Track and Field Olympic Athletes. *Journal of Expertise, 2*(3).

Thanasuan, K., & Mueller, S. T. (2016). Investigating and simulating the effect of word fragments as orthographic clues in crossword solutions. In D. Reitter & F. E. Ritter (Eds.), Proceedings of the 14th International Conference on Cognitive Modeling (ICCM 2016).

Toma, M., Halpern, D. F., & Berger, D. E. (2014). Cognitive Abilities of Elite Nationally Ranked SCRABBLE and Crossword Experts. *Applied Cognitive Psychology, 28*(August), 727-737. https://doi.org/10.1002/acp.3059

Tuffiash, M., Roring, R. W., & Ericsson, K. A. (2007). Expert performance in SCRABBLE: implications for the study of the structure and acquisition of complex skills. *Journal of Experimental Psychology: Applied, 13*(3), 124-134. https://doi.org/10.1037/1076-898X.13.3.124

Ullén, F., Hambrick, D. Z., & Mosing, M. A. (2015). Rethinking Expertise: A Multifactorial Gene–Environment Interaction Model of Expert Performance. *Psychological Bulletin, 142*(4), 427-446. https://doi.org/10.1037/bul0000033

Underwood, G., Deihim, C., & Batt, V. (1994). Expert performance in solving word puzzles: from retrieval cues to crossword clues. *Applied Cognitive Psychology, 8*(6), 531-548. https://doi.org/10.1002/acp.2350080602

Underwood, G., MacKeith, J., & Everatt, J. (1988). Individual differences in reading skill and lexical memory: the case of the crossword puzzle expert. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: current research and issues* (Vol. 2: Clinical and educational implications, pp. 301-308). Wiley.

Wai, J., & Perina, K. (2018). Expertise in journalism: Factors shaping a cognitive and culturally elite profession. *Journal of Expertise, 1*(1), 57-78.

Waley-Cohen, J. (2019, 30th April). Top performers on UK broadcast quizzes. *QuizQuizQuiz*. https://quizquizquiz.com/2019/02/top-uk-quizzers/

Warren, R. E., Allen, K. V., Sommerfield, A. J., Deary, I. J., & Frier, B. M. (2004). Acute hypoglycemia impairs nonverbal intelligence: importance of avoiding ceiling effects in cognitive function testing. *Diabetes Care, 27*(6), 1447-1448. https://doi.org/10.2337/diacare.27.6.1447

Watts, K. P. (1954). The effect of a university training in mathematics on the scores in an intelligence test. *British Journal of Educational Psychology, 24*(1), 32-37.

Webb, S. (2012, 5th Nov). Countdown winner crowned UK Scrabble champion with incredible 71 point word “neonates” (but do you know what it means?) *Mail Online*. https://www.dailymail.co.uk/news/article-2228031/Countdown-winner-crowned-UK-Scrabble-champion-incredible-71-point-word.html

Accepted: 16 January 2020
Revision received: 13 March 2020
Revision accepted: 13 March 2020