League-Stepping Habits as an Escape Route (with Plateaus, Dips, and Leaps) from Stable Suboptimal Performance

Journal:	*Perspectives on Psychological Science*
Manuscript ID:	PPS-15-006.R2
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Gray, Wayne; Rensselaer Polytechnic Institute, Cognitive Science;
Keywords:	Action / Performance, Methodology: Behavioral, Application: Education, Cognition, Individual Differences, Learning (associative)
User Defined Keywords:	Deliberate Practice, stable suboptimal performance, plateaus and asymptotes, skill acquisition

League-Stepping Habits as an Escape Route (with Plateaus, Dips, and Leaps) from Stable Suboptimal Performance

Wayne D. Gray & John K. Lindstedt
Cognitive Science Department
Rensselaer Polytechnic Institute

draft of 2015/05/26 at 17:58

Author Note

This work was supported, in part, by grant N000141010019 to Wayne Gray from the Office of Naval Research, Dr. Ray Perez, Project Officer. It also benefited from support by IBM's Cognitive Science Institute. We thank Anders Ericsson, Zach Hambrick, Niels Taatgen, and Christian Fiebach for their many comments that helped clarify our thoughts and enabled us to better distinguish the focus of our work from related topics. Thanks as well to the members of the CogWorks Laboratory (Ryan Hope, Catherine Sibert, and Matthew-Donald Sangster) who read and discussed many of the papers cited in this work. (Wayne Gray may be contacted at email:grayw@rpi.edu)

Abstract

120 years ago the emergent field of experimental psychology became embroiled in debates as to whether plateaus in performance are real (or not) and if so whether they were due to periods in which league-stepping habits (originally defined by Bryan & Harter, 1897, 1899, as a *hierarchy of habits* that enabled experts to step leagues while novices were "bustling over furlongs or inches") were being acquired (or not). 20 years ago both the human-computer interaction (HCI) and cognitive science (CogSci) communities were seized with concerns over performance plateaus (i.e., extended periods of *stable suboptimal performance*) from experts. For HCI this was viewed as a systems problem and referred to as the *Paradox of the Active User*. CogSci diagnosed this as a training problem and embraced *Deliberate Practice*. After an introduction, we review this history and clarify the distinction between performance plateaus and asymptotes, suggesting that the former may be remediated by adopting new strategies and acquiring new methods whereas the latter cannot. We next sketch a broad view, a proto-theory of components of change, including performance dips, which mark attempts to break out of plateaus. We, also, review and relate our discussion of plateaus and dips to the topics of *Deliberate Practice* and *Brain Training*.

League-Stepping Habits as an Escape Route (with Plateaus, Dips, and Leaps) from Stable Suboptimal Performance

INTRODUCTION

With practice, performance whether it is with telegraphy, typing, software programs, arithmetic, programming, mnemonics, or video games generally improves. Near the dawn of experimental psychology, Bryan and Harter (1897, 1899) claimed that experts were not simply faster than novices, but had developed a *hierarchy of habits* than enabled them to *step leagues*[1] while novices were *bustling over furlongs*[2] *or inches.* They also were the first to suggest a 10-yr rule of expertise (e.g., Hayes, 1985) by observing that, "Our evidence is that it requires ten years to make a thoroughly seasoned press despatcher." And they were the first to notice plateaus of performance during training. Whether such plateaus were inevitable, due to the use of inferior methods of training, or simply artifactual was a source of some controversy (e.g., Keller, 1958; Taylor, 1943; Thorndike, 1913). However, plateaus in performance of experts were quickly established (see, e.g., Thorndike, 1913, especially, pp 178-185, *The Limit of Improvement*). Indeed, in summarizing the state of the art back then or now, I can do little better than quote Thorndike:

> It seems to me therefore that mental training in schools, in industry and in morals is characterized, over and over and over again, by *spurious limits* – by levels or plateaus of efficiency which could be surpassed. The person who remains on such a level may have more important things to do than to rise above it; the rise, in and of itself, may not be worth the time required; the person's nature may be such that he truly cannot improve further, because he cannot care enough about the improvement or cannot understand the methods necessary. But sheer absolute restraint – because the mechanism for the function itself is working as well as it possibly can work – is rare (Thorndike, 1913, p. 181).

[1] A "league" is an old English measurement corresponding to about three miles (4.83 km).
[2] A "furlong" is an old English measurement corresponding to about 220 yards (201.2 meters).

Our next section, (a) sketches the first 90 years (1897 – 1987) of scientific research on expert performance, (b) introduces the modern origins of our impatience with *stable suboptimal performance* (Fu & Gray, 2004) – the view that *expert is not good enough* and, (c) discusses two types of performance *asymptotes* and one type of *plateau*. We argue that the roots of stable suboptimal performance vary depending on whether performance has asymptoted or plateaued.

The following section, (a) introduces and elaborates three activities that we view as necessary to escape performance plateaus, it then (b) highlights the need to establish a performance baseline to establish the existence of a plateau, (c) provides two extended discussions of activities that produce league-stepping increments in performance, and (d) introduces the performance *dip* as a marker for periods of rapid change.

We next discuss three contemporary research programs from which our work draws inspiration; namely, those concerned with (a) *Brain Training* – whether by paradigm (e.g., von Bastian & Oberauer, 2014; Dunning & Holmes, 2014; Redick et al., 2012; Shipstead, Redick, & Engle, 2012) or by game (e.g., Green & Bavelier, 2003, 2012; Unsworth et al., 2015), (b) *Deliberate Practice* (e.g., Ericsson, Krampe, & Tesch-Römer, 1993; Ericsson, 2004; Hambrick et al., 2014; Macnamara, Hambrick, & Oswald, 2014), and (c) contemporary efforts in *Computational Cognitive Modeling* that seek to capture the creation and transfer of low-level mechanisms responsible for learning and transfer between tasks (e.g., Salvucci, 2013; Taatgen, 2013).

The penultimate section, *Plateaus, Dips, and Leaps*, summarizes the main lines of influence on our work and our last section provides a short summary of the paper and summarization of our conclusions.

HISTORIC AND RECENT BACKGROUND

THE PHANTOM (?) PLATEAU

Bryan and Harter's 19th century studies (1897, 1899) of telegraphic operators provide important reading for 21st century researchers. They collected data with millisecond accuracy from professional operators and trainees, gathered reports from telegraphic schools, attempted to generalize their results to training in a variety of disciplines, and provided a basic theory of skill acquisition that survived until deep into the behaviorist revolution.

Here we focus on their conclusion that in learning to become experts, their student subjects passed through several performance plateaus; that is, periods in which practice continued but performance increments halted. They hypothesized that telegraphic expertise consisted of a hierarchy of habits. The lowest level habit was the mapping of individual letters to telegraphic sendings and receivings. For example, for International Morse Code[3] three brief taps, or dots, signal an "s" whereas three longer taps, or dashes, signal an "o". Combinations of dots and dashes suffice to represent all of the letters of the alphabet and all of the numbers from 0 to 9.[4] The story told by Bryan and Harter was that after people became skilled at translating dots and dashes into letters and numbers, a performance plateau would occur while common words (e.g., "the") began to be chunked (in modern terms) and the number of acquired word chunks began to increase. Presumably the second plateau followed after all of the most common words were chunked and vanished as operators began to acquire common phrases.

By 1913, a view of performance plateaus by people who performed the same tasks daily was so well accepted that even the ever skeptical Thorndike accepted large parts of the story; for example,

> I venture to prophesy that the thousand bookkeepers in, say, the grocery stores of New York who have each had a thousand hours of practice at addition, are still, on the average, adding less than two-thirds as rapidly as they could, and making twice as many errors as they would at their limit.
>
> . . .

[3]Note that in Bryan and Harter's day the dominate Morse language was "American Morse", not "International Morse". However, subsequent researchers have not viewed the difference between American versus International Morse Code as affecting Bryan and Harter's conclusions.

[4]See Gleick (2011) for the history of the 100+ years of work that culminated in the development of the Morse code.

> It appears likely that the majority of teachers make no gain in efficiency after their third year of service, but I am confident that the majority of such teachers could teach very much better than they do.
>
> . . .
>
> Even in a game where excellence is zealously sought, the assertion that "I stay at just the same level, no matter how much I practice" probably does not often mean that the individual in question has really reached the physiological limit set for him in that function. (Thorndike, 1913, p. 179)

Although training plateaus that Bryan and Harter reported were very influential, and were reported by researchers in skills as diverse as typewriting (Book, 1908) and juggling (Swift, 1903), there was some concern at the time as to whether plateaus occurred in learning telegraphy. For example, in his influential *Psychological Bulletin* review, Taylor (1943) states that, "Although the curves of Bryan and Harter have been repeatedly presented as representing the usual course of learning in telegraphy, it is very doubtful that they are typical (p. 464)". The telegraphy findings of Bryan and Harter came under intense criticism from Keller (1958), who was a noted behaviorist writing at the height of the Behaviorist Era and at the beginning of the Information Processing Era of Psychology. As a Behaviorist, Keller's theoretical arsenal would not include *hierarchies* but did include concepts that would predict a slow but steady increase in performance with practice. Keller was also a brilliant experimentalist and an insightful critic. His critique of Bryan's studies and his interpretation of subsequent work challenged Bryan and Harter's interpretation of some of their data, especially the data supporting plateaus during learning.

Keller's review strikes us as too harsh and his conclusions too broad. In his paper he cites a footnote from Thorndike (1913) in such a way as to cast doubt on all of Bryan and Harter's conclusions. However, it is clear that Thorndike would not agree with this interpretation as half of his Chapter 6 consists of excerpts from Bryan and Harter (1899) and each of the following chapters returns to Bryan and Harter's work as a device to organize subsequent research findings in learning and performance. In any event, despite whatever merits Keller's critiques have, as Thorndike (1913) and subsequent researchers make clear, in the first half of the 20th century the evidence for plateaus in skilled performance, both among students and among skilled professionals, was abundantly clear and well documented.

As history shows, subsequent progress in cognitive science legitimized hierarchies by providing an *architecture of complexity* (Simon, 1962) that was quickly applied as a basic construct to areas as diverse as reading (LaBerge & Samuels, 1974), perceptual learning (Kellman & Garrigan, 2009), and motor performance (e.g., Memmert, Raab, & Bauer, 2006). In the 1980's, Newell (Newell & Rosenbloom, 1981) and J. R. Anderson (1982, 1987) both offered powerful mechanistic accounts of the formation of hierarchies via chunking. Since those times, many researchers have found evidence for both skill hierarchies (Logan & Crump, 2011) and plateaus in skill acquisition (Carroll & Rosson, 1987; Ericsson et al., 1993; Huang, Zimmermann, Nagapan, Harrison, & Phillips, 2013; Robertson & Glines, 1985).

MERE EXPERTISE IS NOT GOOD ENOUGH

90 years after Bryan and Harter reported finding plateaus (1897), Carroll and Rosson (1987) coined the term, *Paradox of the Active User*, to refer to the "suboptimal use of office productivity software" (e.g., spreadsheets, word processors, or more specialized systems for accounting, engineering, etc) by people who use the systems daily across the course of weeks, months, and years. Independent of these observations, a few years later, based on his studies of human expertise, Ericsson (1993) concluded that, "the belief that a sufficient amount of experience or practice leads to maximal performance appears incorrect", (Ericsson et al., 1993, p. 366). After years of lurking in the background, the plateau had returned to front and center.

Carroll and Rosson, who at that time worked for the IBM Watson Research Center, shared the HCI's community's concern that the expected productivity gains of the computer revolution were not occurring (e.g., Cockburn, Gutwin, Scarr, & Malacria, 2014; Doane, Pellegrino, & Klatzky, 1990; Nilsen et al., 1993).[5]

[5] Although this concern seems odd today, in the late '80s through the mid-90's the gains in productivity from

Ericsson came to his conclusions from years of study of exceptional performers. Perhaps it was very striking to him that a major construct for IQ testing, the digit span, was normed at 7 ± 2 for the entire population when, with the right type of practice, he and Chase had conclusively demonstrated that normal people could obtain a digit span of > 80 (Chase & Ericsson, 1982; Ericsson & Chase, 1982; Ericsson, Chase, & Faloon, 1980) and as high as 104 (Richman, Staszewski, & Simon, 1995). Likewise, documentation was readily available from his and others' research of performers who seemed to possess superhuman powers in narrow domains of memory expertise (e.g., Ericsson & Chase, 1982; Ericsson & Kintsch, 1995).

An interesting thing about the plateaus was that suboptimal performance was so stable. Fu and Gray (2004) reported three cases in which they diagnosed performance plateaus as due to the selection of suboptimal strategies. The first was an architect who fit Carroll and Rosson's criteria of a professional daily user of productivity software and may have met Ericsson's criteria of an expert in his field. Many of the small, basic procedures he *preferred* to use in an architectural CAD/CAM system were an order of magnitude slower than the procedures *recommended* to him during training and in the manuals. In some cases this would have been a difference of 5 vs 50 s. Although 45 s wasted does not seem like a disaster, Fu and Gray estimated that if summed across the number of times daily in which the architect used the inefficient procedures, the total would add up to a daily waste of 30 min, or 6% of an 8 hr day.

Fu and Gray's other two examples were collected from student users. In one, Education Graduate students enrolled in a semester long course were learning to use a software package to produce classroom instruction. In the other, undergraduates participated in a 60 min experiment. Neither of these sets of students would fit Ericsson's criteria for expertise nor Carroll's criteria for an "active user." However, both examples were as dramatic as that of the expert architect's and both fit the pattern of inefficiencies discovered in the architect's behavior. For all three cases, it could be demonstrated that the users knew the more optimal procedures

computers lagged the expectation of gains. For an extensive discussion of this failure, from a mid-90's perspective, see, Landauer (1995).

and would use them in certain circumstances. Perhaps as important, these latter cases support Ericsson's observation that, "most people and professionals reach a stable performance asymptote within a limited time period" (Ericsson, 2004). Practice does not make perfect. Long plateaus of stable suboptimal performance are real.[6]

PLATEAUS VERSUS ASYMPTOTES

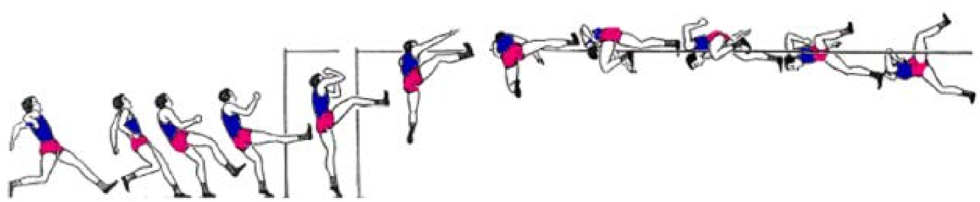

***Figure 1*.** The *Scissors and Straddle* technique for high jumping.

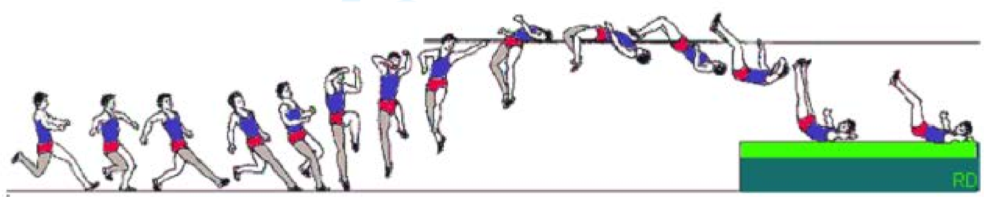

***Figure 2*.** The *Fosbury Flop* technique for high jumping.

We can divide cases of stable suboptimal performance into two categories: Asymptotes and Plateaus. To illustrate this distinction, we turn to two related sports; pole vaulting and high jumping. We can conceptualize records in pole vaunting as being limited to an asymptote defined by the technology of the pole. Hence, the history of dramatic record breakthroughs in pole vaulting is the history of technological innovations as the composition of the pole changed from ash (wood), to bamboo (a type of grass), to fiberglass/carbon. Following each of these changes in technology, pole vaulters initially used their old methods to break new records enabled by the newer material. After relatively short periods of familiarity, pole vaulters invented new methods that resulted in new rounds of record breaking as those methods were

[6]Note that Fu and Gray (2004) provide a diagnosis of the reasons for the stability of suboptimal performance; however, that discussion is beyond the scope of the current paper.

adopted and adapted by athletes (click here for YouTube video). Hence, we could say that prior to each wave of new technology, performance in pole vaulting had asymptoted as it was close to being as good as possible given the current technology.

In contrast, record breakthroughs in the sister sport of high jumping were primarily driven by changes in method, the *Fosbury Flop* (contrast Figures 1 and 2). In this case, the technique could have been deployed earlier. Therefore, records in high jumping had plateaued as it was the methods used for high jumping, not the technology, that limited performance. The Fosbury Flop enabled league-stepping increments in human performance that escaped a plateau.

Asymptotes Due to Design. An asymptote due to design is illustrated by Crossman's (1959) study of cigar rollers in Cuba. Figure 3 shows a continued increase in performance over a two year period (estimated as 3 million cigars) and then a flattening of the curve. Newell and Rosenbloom (1981, p. 7) attribute this flattening to a "known lower bound for the performance time"; namely, the "cycle time of the machine." Although we cannot prove it from these data, we would not be surprised to learn that the efficiency of eye movements, arm, hand, and figure movements continued to improve after performance had asymptoted; that is, any potential gains from log-log speedups in perception, cognition, or motor processes would have been held hostage by the cycle time of the primitive machines used by the cigar rollers.

A higher tech example of a plateau due to design is provided by CPM-GOMS[7] models of two workstations used for Telephone Company Toll & Assistance Operators (Gray, John, & Atwood, 1993). The duration of the field trial was set based on expectations that performance times on the new workstation would decrease across a four-month period. Unexpectedly, performance on the new workstation stabilized within two months with times per call slower than for the old workstation, and slow enough to increase annual operating costs by an estimated $6.2 million (in 2014 dollars).

The lack of constant small improvements over the 4 month period was puzzling. The

[7]Where 'CPM' stands for both "critical-path method" and "cognitive, perceptual, motor", whereas GOMS stands for "Goals, Operators, Methods, and Selection Rules" (see John & Kieras, 1996, for a general introduction to the GOMS family of models).

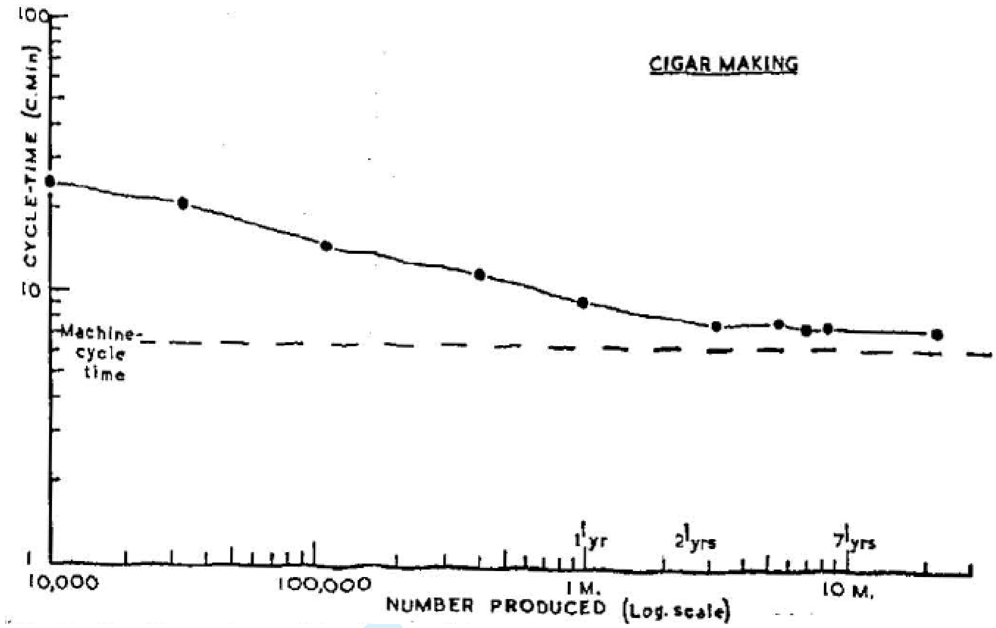

Figure 3. Crossman's (1959) cigar roller, showing a flattening of performance after 2yr and 3 million cigars due to limits in the cycle time of the machine used for cigar rolling.

designers believed that call time was driven by the number of keys per call the Operator had to press. Therefore, their design for the new workstation eliminated keypresses across all call types for what was expected to be a decrease in average call times of 4.1 s. This time savings was expected to translate into an annual savings in operating costs of $24 million (in 2014 dollars).

Our models showed that compared to the old workstation, the new workstation decreased the number of things the Operator could do in parallel. For example, the models of the old workstation showed that Operators often exchanged information with the customer while keying or while waiting for external databases (such as the credit card verfication database) to return information. Hence, although there were fewer keypresses per call, the critical path grew longer because the designers had inadvertently moved (a) more keypresses and (b) more waiting time onto the critical path.[8]

[8]Note that the complete story of the design inefficiencies of the new workstation is more complex than conveyed here, see Gray, et al. (1993) for the full story.

Asymptotes Due to Measurement Method. A clear example of false plateaus induced by measurement methods is taken from the game Space Fortess (Donchin, 1989). Most modern studies of Space Fortress use a version of the game crafted by Gopher (Gopher, Weil, & Bareket, 1994)[9]. This version kept the original metric for scoring (POINTS) and added three new ones (SPEED, CONTROL, and VELOCITY). These new scores had been introduced by Fredericksen & White (1989) as a means of focusing attention on different task components during training. However, when used as measures of performance, they have two problems. First, they are not independent of each other. Second, two of these measures asymptote even as skilled performance continues to increase.

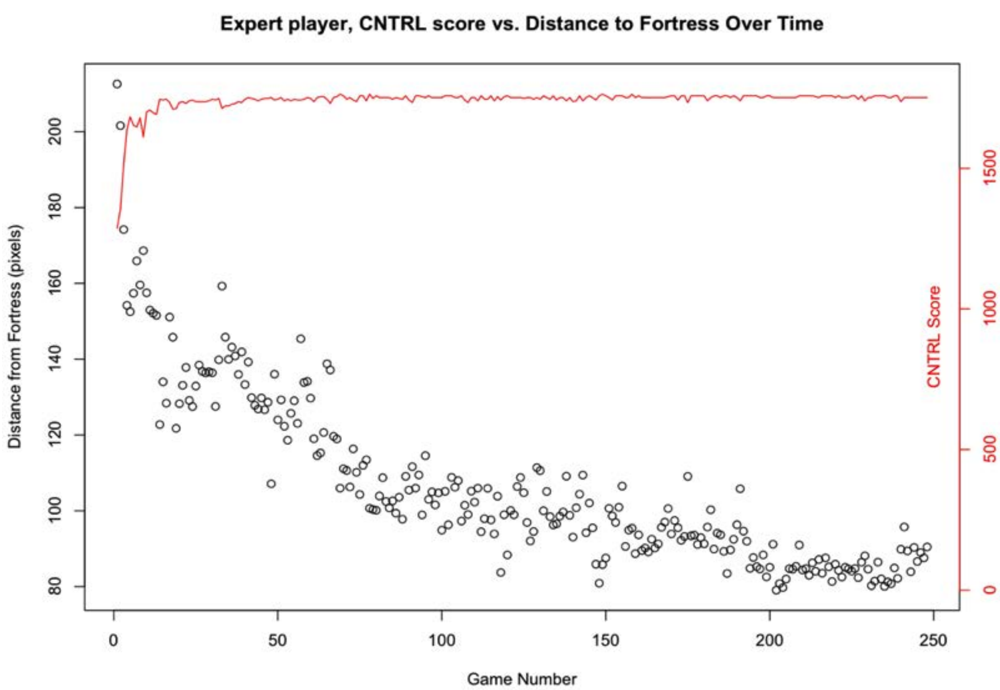

***Figure 4*.** Plot of one player's score on the CNTRL measure (shown in red at the top of the plot and on the right y-axis) versus the distance he is actually flying from the Fortress (shown in black at the bottom and on the left y-axis). Although the player's distance follows the log-log law of improvement, his score (shown in red) rises rapidly across his first 10 games and then asymptotes from about game 30 to game 248. (Based on unpublished data collected by, Destefano, 2010).

[9]Several criticisms of this version are provided by Donchin (1995).

As one example, Figure 4 shows the relationship between the CONTROL score and the distance of the player's ship from the centrally located Space Fortress (which gives the game its name). The control measure was originally intended to help teach novice players to adopt the strategy of flying close to the Fortress (Frederiksen & White, 1989). Although it may have served that purpose for novices, as Figure 4 shows, experts continue to fly closer and closer long after they have asymptoted on the CNTRL measure. Unfortunately, many game studies attempt to relate improvements within the game to external measures of cognitive abilities (see e.g., Boot et al., 2010). It is clearly the case that such efforts are doomed to false negatives as the measure asymptotes long before the skill plateaus.

Plateaus Due to Strategy: Performance. In distinguishing the causes of stable suboptimal performance, the difference between a plateau and an asymptote may be hard to discern, especially in laboratory studies which only collect a few hundred performance trials [unlike Crossman (1959), not all researchers are fortunate enough to have data that continue over 3 million trials]. However, the distinction is clear in hindsight when it can be shown that groups of individuals performing at different skill levels are following different strategies. Like the Scissors and Straddle versus the Fosbury Flop high jumpers (see Figure 2), *visually guided* typists[10] perform far below the level of touch typists. Indeed, Yechiam, Erev, Yehene, and Gopher (2003) tell us that "Following a long period of touch-typing training, typists reach an average speed of 60 to 70 wpm, whereas the average speed of very experienced visually guided typists is much lower (approximately 30–40 wpm)." Interestingly and importantly, the shift to touch typing initially results in a *performance dip*, not a *performance leap*. It is only through motivation and continued practice that the new touch typist becomes as fast and, eventually, faster than the old hunt and peck typist.

Limits to Human Performance: True Human Asymptotes. For completeness, it is necessary to mention true asymptotes on human performance; that is, performance limits not due to artifact design, system design, measurement error, or strategy but which reflect

[10] Also known as *hunt and peck*, *search and peck*, or *Eagle Finger* typists.

optimization of all system components, limited only by the bounds of human physical and mental machinery. Such limits are undoubtedly close in many sports where, barring the discovery of strategies such as the Fosbury Flop or the introduction of negative friction swimwear, it seems likely that future world records will reflect millisecond increments in performance times, which would have more to do with enlarging the pool of competitors and resources expended per competitor than it would with any dramatic breakthrough in training or technique. Presumably, as the number of trials for a performance increment increases logarithmically, improvements would be more than balanced by barriers imposed by age or the human lifespan.

Discussion.

The difference between a plateau and asymptote is made clear by the existence of extreme experts with a known history of transcending the plateau. Hence, Ericsson and Chase knew that both SF and DD[11] started at the normal digit span plateau of 7 ± 2. As Figure 5 shows, SF went through a second plateau at around 18 digits, a third plateau around a digit span of 43, and then slowly ascended to a span of around 82 before leaving the project. Note as well that what we are calling plateaus are often composed of jagged periods of dips and leaps. These dips and leaps are present for both subjects but are especially prominent for DD. These changes were accompanied by SF's reports of the invention of new strategies for accomplishing the digit span task. SF's maximum digit span was not an asymptote as in a later publication Richman et al. (1995) report that DD eventually achieved a digit span of 104! That may or may not have been an asymptote rather than a plateau, but we will never know as at that point DD left the project. We return to Chase and Ericsson (1982) and to SF and DD, below. However, knowing that even one person was able to transcend normal, strongly implies that 7 ± 2 is not an asymptote imposed by the hardware of the human brain. Rather, it is a performance plateau imposed by the strategies used and the general unwillingness of people to put in the time and effort required to discover or otherwise acquire and practice new strategies while also paying the

[11]The two students who become their digit span experts.

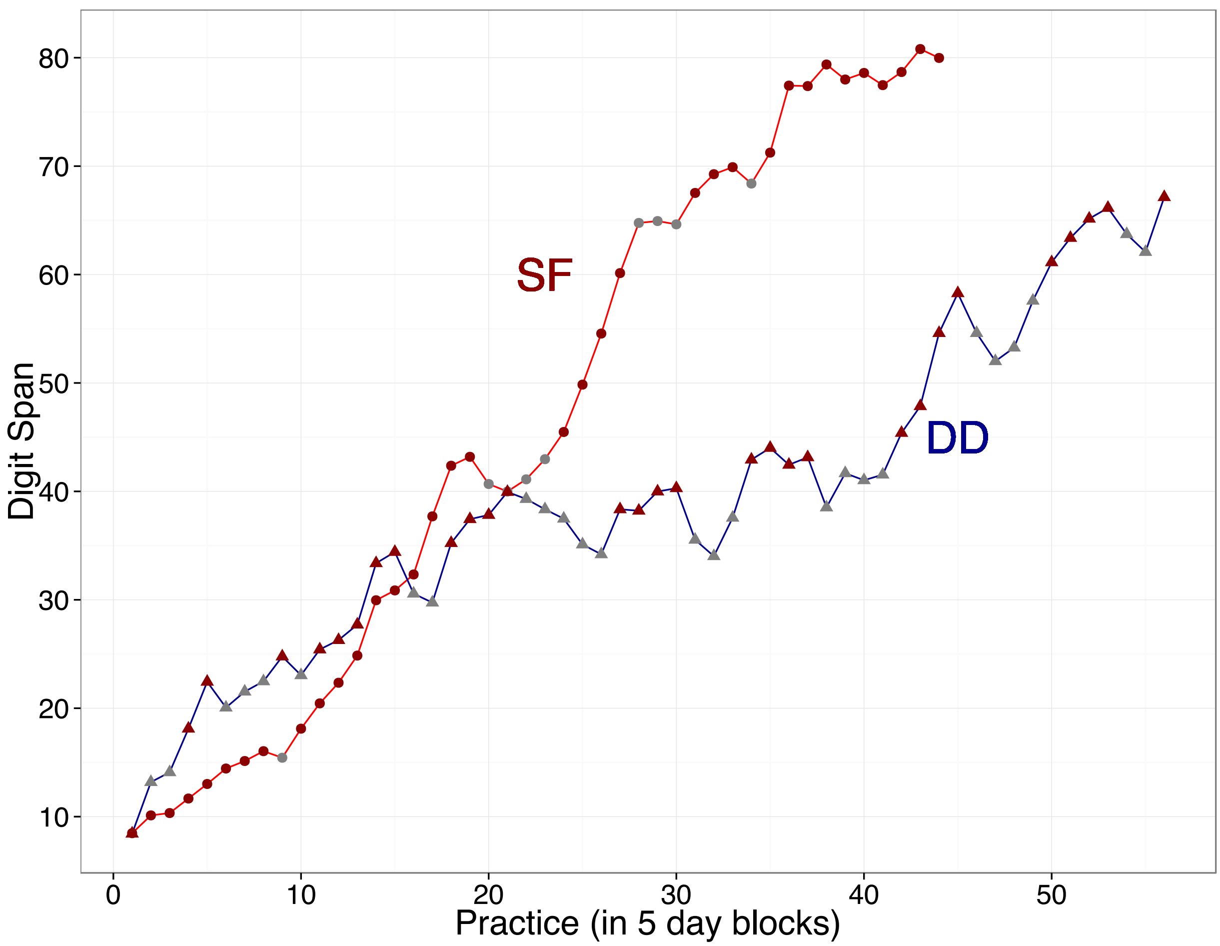

***Figure 5*.** Blocks of practice (each block represents 5 days of 1 hr practice per day) for (a) SF and (b) DD. The plot points are red or gray. Red points indicate periods during which performance improved. Gray points show performance dips (see text in the Summary & Conclusions section for discussion). (Data were extracted from Figure 1 of Chase and Ericsson (1982) using the PlotDigitizer™ software and should be considered as near approximations to the data in Chase and Ericsson's plots.)

additional costs of loss of productivity due to dips in performance as small variations on new methods are tested, rejected, acquired, and automated.

ESCAPING PERFORMANCE PLATEAUS BY ACQUIRING *LEAGUE STEPPING HABITS*

We propose that acquiring League Stepping Habits, which move performance from the plateau of stable suboptimal performance, requires three types of cognitive activities:

1. Strategy Acquisition: The discovery of, or instruction in, a *better* strategy.
2. Method Development: The invention, discovery, or instruction of methods required to implement the better strategy and to adapt it to variations in the structure of the task environment, as needed.
3. Drill & Skill: The extended practice that produces automation and log-log increments in skilled performance.

Being told "how to juggle" is the acquisition step (perhaps as provided by Figure 6) while incremental improvements in juggling is the Drill & Skill step. However, the hard work in juggling and the point at which most would-be jugglers quit, is the translation of verbal or pictorial knowledge of a strategy into the cognitive, perceptual, and motor methods required to execute that strategy. Indeed, it seems fair to say that a person's understanding of a verbally or figurally presented strategy changes as they attempt to develop a method to implement it and begin the drill & skill phase of practice. Included in this step would be rules for adjusting the method to differences in the task environment. For juggling this might include different objects (e.g., balls, rings, or clubs), differences in the weight or size of these objects, as well as differences in the number of objects (i.e., 3, 4, or more)! Hence, to a large degree, it is the implementation and tuning of methods that is the key to acquiring League Stepping Habits.

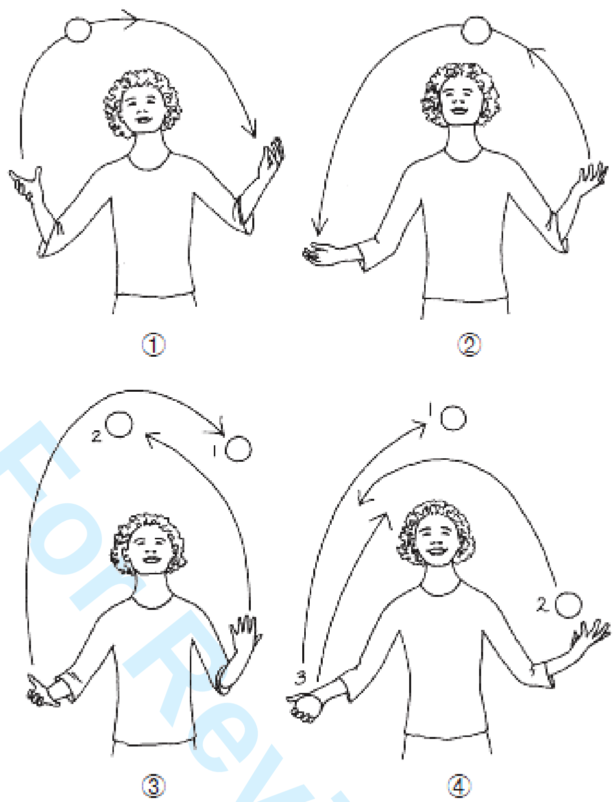

Figure 6. http://www.daringbookforgirls.com/about-the-book/about-the-double-daring-book-for-girls/how-to-juggle/

The Three Activities of Deliberate Practice

Strategy Acquisition. Strategy acquisition can be an extended effort requiring much search, trial, and error or it can be something that seems obvious or something which we are told. For example, Frank Edward McGurrin, who is often credited with inventing touch typing, said, "I do not take any great credit for having thought of operating without looking at the keyboard for it is simply a matter of common sense" (Wikipedia, 2014). I will not argue with Mr. McGurrin here but history is quite clear that the inventor of the typewriter did not have touch typing in mind when he invented it. Hence, acquiring strategies by being told is easy (as was the case for Mr. McGurrin), coming up with an original strategy is generally more difficult. In either case, however, the proof of the strategy lies in finding a method that can implement it.

Method Development and *Dips* in Performance. If sweat were the only requisite for increasing physical fitness then saunas would be a lot more popular than they are. Likewise, if knowing the best strategies was the only requisite for increasing expertise then we would have a lot more extreme experts. In Mr. McGurrin's case, after telling us that acquiring the strategy of touch typing was easy, he also tells us that, "and the system of fingering is so simple that anybody could formulate it" (Wikipedia, 2014).

Apparently Mr. McGurrin was driven to invent touch typing by his boss, who told him that a certain young female secretary could type as fast as her boss could dictate. Rather than deflating Mr. McGurrin as his boss intended, this tale inspired him as, "boy like, I made up my mind that whatever a girl could do I could do, so I set to work to learn to operate without looking at the keyboard."

From Yechiam's data (2003) (discussed earlier), we concluded that learning touch typing after having mastered hunt and peck typing (as would have been the case for Mr. McGurrin) requires acquiring and mastering difficult methods. Indeed, these methods are so difficult that performance initially dips until the methods are worked out and begin to become automated. These dips can be discouraging and appeals to "future time savings" often do not suffice, with the new ways often abandoned for the old. However, from Mr. McGurrin's story we might conclude that mastering a difficult method requires finding the right motivation.

Although dips may be a source of discouragement for the learner, for the researcher the opposite may be true. We propose that dips are markers of periods of exploration, method formulation, trial and error, and end with either a drop back to tried and true methods or a leap forward to the formulation of new and improved ones. Hence, for the researcher dips become the marker of transitional phases of performance and a call to focus microgenetic methods (Siegler, 1991; Agre & Shrager, 1990) of data collection and analyses on these periods of transition.

Drill & Skill: Log-Log Learning. Two complementary contributions of modern cognitive theory are the establishment of the power law of practice (Newell & Rosenbloom, 1981; J. R. Anderson, 1982, 1987) and the importance of *methods* as the key to task performance (Newell & Simon, 1972; Newell, 1973). The former entails the hierarchical chunking of a task's various cognitive, perceptual, and action subcomponents. The two work together in powerful ways. For example, Delaney, Reder, Staszewski, and Ritter (1998) have shown that log-log improvements in complex task performance are more properly considered as log-log increments in methods for each of the various substrategies. Complementary, Rickard (1997) has shown that improvements in simple tasks may reflect the independent log-log performance increments from the acquisition of methods for two (or more) alternative strategies. In these cases, various alternative strategies may co-exist (e.g., see Siegler & Stern, 1998) with the more efficient strategy never completely dominant.

There are some cases where the slow mechanisms of log-log learning seem capable of producing new strategies through the additional mechanisms of knowledge compilation (a type of hierarchical chunking, J. R. Anderson, 1982, 1987) and memory retrieval. A case in point is the Alpha-Arithmetic task (Zbrodoff, 1995; Lovett, Reder, & Lebiere, 1999) in which subjects are given tasks such a H + 3 = ? or C + 2 = ? and told to find the answer by counting up from the first letter so that H + 3 = K and C + 2 = E. Initially, the count is serial; however, with practice, the direct lookup method for addition (i.e., given two small numbers such as 3 + 2 retrieve their sum) comes to be used for Alpha-Arithmetic. A few additional assumptions (not very extreme but beyond the scope of the current discussion) suffice to switch the model (and presumably the human) from the counting method to direct retrieval. Hence, with commonly accepted assumptions, Drill & Skill does provide a path to learning some types of new methods. However, although these mechanisms might account for incremental improvements leading to world records in, say, pole vaulting, they could never generate a Fosbury Flop nor could they transition from hunt-and-peck to touch typing!

Practical Constraints on Defining a Performance Baseline for the Acquisition of League Stepping Habits

After a strategy has been implemented by an efficient method, the Drill & Skill phase yields diminishing returns and it may be difficult or impossible to discriminate a plateau from a very flat part of the log-log learning curve. Fortunately, such a discrimination is not required by our theory.

For example, for the Telephone Operator workstation (Gray et al., 1993) discussed earlier, we had estimated that the average operator handled 800 calls per day. Hence, after 12 days on the new workstation the operators would be at log 4 with approximately 10,000 calls, after 125 days at log 5 with 100,000 calls, and after 5.2 years (assuming a work year of 240 days) at log 6. When we figured in the normal turnover in personnel and miscellaneous changes in staffing and procedures, it struck us as very unlikely that we would have noticed any performance increment on the new workstation beyond the first four months of the trial (i.e., approximately log 5).

The study of League Stepping Habits requires a definition of the practical plateau; that is, a stable performance base from which to measure the impact on performance of acquiring the candidate League Stepping Habit. For digit span experts, this might be the population estimate of 7 ± 2. For surgeons, this might be at the beginning of their residency (after graduation from medical school and the completion of their internship). For architects, this might be graduation with a Bachelor of Fine Arts. For teachers, this might be after their 3rd year of service (as per Thorndike's comment). For musicians, this might be competing and placing in a regional competition. For Chess players this might be receiving an Elo rating of 1200 (or 2000 or 2300 or . . .).

Given the difficulties inherent to studies of longterm effects, estimates of plateaus must be somewhat opportunistic. However, the notion of League Stepping Habits gains in power and meaning when the base from which it is measured can be shown to be stable across successive days, weeks, or months.

Digit Span and League Stepping Habits

If this view of Deliberate Practice has a *poster child* or *children* it would be SF and, possibly, DD. SF was introduced to the world by Ericsson et al. (1980). DD arrived a few years later in Chase and Ericsson (1982) and starred in Richman et al. (1995). As Chase and Ericsson tell us,

> When we first started this experiment, we simply wanted to run a subject for a couple of weeks to see if it was possible to increase the memory span with practice and, if so, whether we could use the subject's retrospective reports to figure out how it happened (Chase & Ericsson, 1982, p. 8).

The rest, as they say, is history . . . or science.

Chase and Ericsson read digits to SF at the rate of 1 per sec, followed immediately by ordered recall. If all digits were recalled correctly, the length of the next run of digits was increased by 1. If all were not correct, the next run was decreased by 1.

SF: Stage 1. SF's initial strategies lasted across the first four days.

Strategy Acquisition. SF began by trying "to hold everything in a rehearsal buffer" but quickly adopted the strategies of separating "one or two groups of three digits each in the beginning of the list, concentrate on these sets first and then set them 'aside' somewhere, and then hold the last part of the list in the rehearsal buffer; at recall, retrieve and recall the initial sets while simultaneously concentrating on the rehearsal buffer, and then recall the rehearsal"[12] (Chase & Ericsson, 1982, p. 9).

Method Development and Drill & Skill. Chase and Ericsson refer to the initial two strategies as "common strategies" and are presumably ones with which SF had prior familiarity (though not necessarily in the context of the digit span task). Therefore the incremental improvements across days 1-4 reflect a combination of adapting these methods to the digit span task and log-log increments due to practice. However, by Day 4, "SF reported that he had reached his limit and no further improvements were possible" (pp. 9-10).

[12] At hour per day of practice, this point was reached at $^4/_5{}^{ths}$ of the first 5-day block shown in Figure 5.

SF: Stage 2. Meaningful units.

Strategy Acquisition. SF's 5thday was different as it was then when "he demonstrated the first rudimentary use of a retrieval structure"; namely, chunking 3 successive digits as a group while trying to hold the last 4-6 digits in his rehearsal buffer (Ericsson et al., 1980).

Method Development and Drill & Skill. SF practiced and improved his implementation of the chunking method to include the grouping of two successive groups of 3 running times, while continuing to rehearse the last few digits. However, a plateau was reached around block 8 or 9 (see SF's curve in Figure 5) by which time he was recalling 18 digits (remember that each dot on the Figure represents the average score for 5 days of one hr practice per day).

SF: Stage 3. Supergroups.

Strategy Acquisition. The next advance entailed forming supergroups of two or three subgroups of 3 digits each. Following the output of the grouped items, he would then output the items from his rehearsal buffer.

Method Development and Drill & Skill. This method was perfected over the next 25 sessions (i.e., across 5 blocks in Figure 5). The improvement seems to reflect both Drill & Skill and adjustment in his methods as the number of chunks increased across these sessions. As Chase and Ericsson discuss (1982), initially SF (who is described as a "very good long-distance runner") used running times as the basis for his chunks. Later he added "years", and later still added "ages" for "for digit groups that could not be coded as running times".

Chase and Ericsson take pains to point out that the semantic memories used were very rich. It is not simply the case that a digit group is encoded as a good running time for a "mile" but as a good "high school" time, "near world-record" time, "training time for a marathon", etc. Hence, the methods for grouping digits were being enriched even as the methods for the higher level and base level grouping structures were being formulated and practiced.

DD. At some point during SF's data collection, Ericsson and Chase took on a new subject, DD. Like SF, DD was a runner. However, this time rather than rely on discovery learning, they taught SF's methods to DD. In their published reports (e.g. Chase & Ericsson, 1982; Ericsson et al., 1980; Ericsson & Chase, 1982) Ericsson and Chase dwell more on the similarities than the differences between the two. However, as a comparison of SF's and DD's performance in Figure 5 shows, their differences are striking. We will have more to say about these differences in a later section.

Breakout™ and League Stepping Habits

In *Pilgrim in the Microworld* (1983), the famed *ethnomethodologist*, David Sudnow describes how he mastered the game *Breakout* (Sudnow, 1983, see Figure 7).[13] In style, the book is a first-person narrative of an obsession with mastering this video game, recounted in chronological order with occasional digressions.

For those too young or too old to have played Breakout, the game was developed by Atari as a console game (played in bars or arcades) around 1976. Sudnow's Breakout seems to be the first version released for the home Atari computer. It differs from a later home version that was called SuperBreakout.[14]

As a first person account, by a trained and expert observer, Sudnow provides a motherlode of detail that might be mined again and again for different purposes. For our purposes, we focus on his development of five successive methods of *Where to Look* which, as his skill progressed, were refined and reused in the service of different strategies. Note that Sudnow does not present a timeline to us but mentions the passage of time and hours practiced several times in his book. Our best guestimate is that the book covers a period of 100-125 hours of practice playing Breakout.

Where to Look? Method 1.

> *Method 1.* In a half hour of just "concentrating" I'd refined the instruction. I discovered if I told myself to "glue my eye to the ball" I could start fielding first slams much better and get some of the follow-ups as well. For about twenty

[13] Thanks to Stuart Reeves for pointing me towards this book and Sudnow's other work.

[14] Those wishing to familiarize themselves with Breakout will find a fairly good copy of the 1976 arcade game at: <YouTube of the 1976 Atari Breakout> (start at 7:20 for the color overlay version).

> minutes I sat there mesmerized, tracking the ball like my life depended on it, my entire being invested in the hypnotic pursuit of that pea-sized light. Kneading my eyeballs into the guts of its movement like following a guy in a fast crowd where a momentary diversion would lose him, I soon got to hold on to a four- or five-round volley of fast ones. (p. 45) . . .

Where to Look? Method 2.

> I wondered if peripheral looking could do the job. . . . [I] fixed my gaze right where the barricade [see Figure 7 leftmost] touches the edge of the screen on the right, stared intently there without moving my eyes, and served a shot. I returned it. In fact I could play through a long volley gluing my eye away from the ball. Peripheral vision sufficed. (pp. 46-47)

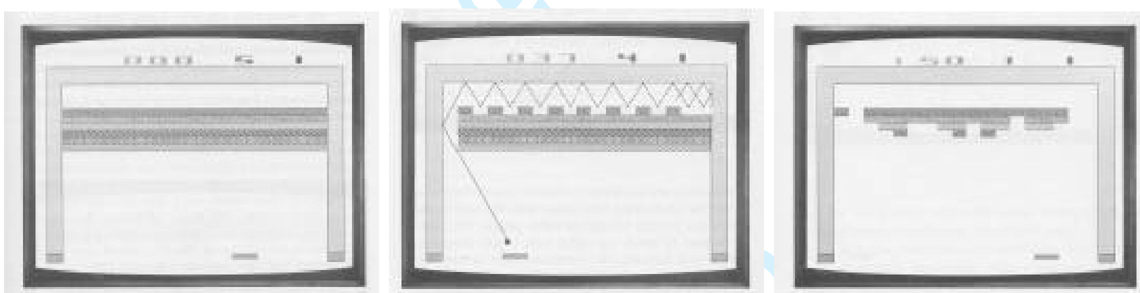

Figure 7**.** (a) Initial board, (b) After a column has been cleared, this shows the path for the ball to "breakout" above the bricks and remove bricks from the top, (c) Board showing remaining "bricks" sometime later in the game. (All figures from Sudnow, 1983)

Plateau. After a considerable amount of time (we estimate at $\approx$ 50 hrs of play), Sudnow becomes stuck in a performance plateau. In frustration with his lack of progress, he contacts Atari and makes an appointment to talk with some of the Breakout programmers. As he tells us, "I'd been playing the computer for several weeks, couldn't clear the screen, wasn't getting anywhere" (p. 103).

One thing he learned from the programmers "was that the paddle was divided into five discrete portions, each of which angled the ball a certain way" (p. 92). Given how much attention Sudnow gave to the paddle early on, and given how much practice he had with it since that time, this news came as a revelation.

Where to Look? Method 3. During this post-Atari visit period, a large part of his time becomes focused on making sure that the ball strikes the paddle at exactly the "right" paddle segment for that play. After hours of trying to get this method to work, he abandons it. In his words,

> Knowledge about the paddle's programmed subdivisions and angles no more truly aids the task at hand than a knowledge of physics could help you line up a certain point on a bat with the ball in order to hit to the field (p. 122).

Where to Look? Method 4. Having decided that he should not be looking at the paddle segments, he, again, spends many hours trying to decide "where to look" complaining that, "The targets aren't easy to fix on, the bricks aren't marked." This time, he concludes that the right spot is,

> [I]n the range of about a half inch above the paddle as the ball came down. Not at the paddle itself, not jumping up and down to the barricade, and certainly not from the floor, but most intently just before the point of contact. That's where you had to look, and you had to look somewhere, couldn't look nowhere. This time I was certain the focus took place right there, just above the paddle. I found a looking method. (p. 124)

Unfortunately for Sudnow, in the very next paragraph he tells us, "But it didn't work."

Takeaways from Sudnow. We count four major shifts in his methods for *where to look*. These methods were in service of four major strategy shifts during his 100+ hours of play. It seems fair to say that all of his shifts were motivated by performance plateaus. It is also clear that many of his shifts initially resulted in performance dips. Sudnow accepted these dips as the cost of implementing a new method. However, after some point in time, after some internal criterion was reached for "time to success" or "time spent without success", he would abandon the unsuccessful method or, if successful, move his practice and search for methods onto the next phase of the game.

Dips: X Marks the Spot Where Research Must Focus

Return to Figure 5 for a moment and look again on the acquisition curves documenting SF's and DD's rise towards extreme expertise in digit span. Last time we pointed out that SF's rising digit span (shown by the small circles) went through several plateaus and we highlighted these by changing their color from red to gray. DD's plateaus (the gray triangles) are more striking than SF's as their periods tend to be longer.

Perhaps not as remarkable as the rises and plateaus are the *dips*: periods during which performance declined. Remembering that each triangle or circle summarizes 5 days of practice, even a small dip represents a period where performance became worse *despite 5 more hours of practice* rather than better.

We ran into dips earlier in our discussions of the decline in performance that skilled hunt & peck typists face when learning to touch type. Figure 5 documents this phenomena in a very different domain. We postulate that dips could represent one of several phenomenon:

- Assuming the adoption of better strategies and sound methods for implementing those strategies, the dips might represent an initial decline in performance as new methods become automatized through Drill & Skill and performance rises to and then beyond its old bounds.

- Much of our discussion of SF, DD, and Sudnow emphasized that method development was hard and somewhat uncertain work. These dips could represent periods during which methods were being developed, adjusted, rejected until a variation was found that would implement the new strategy.

- Strategy development is hard work and we might expect self-directed strategy development to result in numerous dead ends when a method cannot be easily found that would implement the strategy.

The first explanation would account for dips when provably better strategies and methods are taught such as touchtyping (Yechiam et al., 2003) or better interface methods are added to a favorite computer program (Cockburn et al., 2014; Scarr, Cockburn, Gutwin, & Quinn, 2011). The second and third explanations suggest that dips are the cauldron in which strategies are being discovered or created and candidate methods for implementing those strategies are being engineered, tested, rejected, or accepted. In other words, *dips* may be marking the periods during which some of the most interesting, important, and least understood phenomenon in cognitive science occur.

Plateaus of Stable- and Dips of Unstable-Suboptimal Performance

Although Digit Span entails a mental skill and Breakout a perceptual-motor one, both show the development of skilled performance and plateaus. In the digit span case the task began with performance at the *normal* human plateau of 7 ± 2. Presumably because he was the sole subject whose retrospective reports were solicited by the experimenters on a daily basis, SF was motivated to discover, implement, and rehearse new methods for task performance even though the costs of these explorations and trials were dips in his digit span. Unlike SF, DD started his task knowing that performance could be better than 7 ± 2 and being provided with a roadmap of methods that, if acquired and automated, could boost performance. DD seems to have struggled more in adopting SF methods than SF did in creating them. However, eventually he persevered through the dips to exceed his instructor (as documented in Richman et al., 1995).

Sudnow knew that Breakout could be played much better by others than he could play it out of the box. As he initially elected to develop his own strategies and methods, his self-reports provide a valuable record of the exploration of the space of low-level methods for optimizing motor control and visual control as well as higher-level methods for imposing a sequential structure on the task space. Some plateaus, such as a reliance on Hunt & Peck typing, allow their task to be completed; albeit, slowly and inefficiently. Other plateau's such as Sudnow's plateaus in Breakout, end the task well before it is completed (i.e., won). Perhaps because of the nature of his plateaus, Sudnow showed a remarkable tolerance for dips while he invented a series of strategies and methods for playing through to the end of the game.

TRENDS IN BRAIN TRAINING, DELIBERATE PRACTICE, AND COGNITIVE MODELING

Our revival of performance plateaus is an attempt to focus new eyes on an old problem. Following 100 years after Bryan and Harter (Bryan & Harter, 1897, 1899) and Thorndike (Thorndike, 1913) and 25 years after Carroll and Rosson (1987) and Ericsson et al. (1993) our effort is a reboot that benefits from years of additional research, a swirl of contemporary theory and research, as well as a popularization (and commercialization) of cognitive, social, and behavioral research. Below we briefly mention and relate our work to three current strands of research; namely, (a) Brain Training (by experimental paradigms and by video games), (b) Deliberate Practice, and (c) contemporary efforts in Computational Cognitive Modeling. Our intent is not to provide an exhaustive review of these topics but to use our discussion to highlight contributions that a focus on plateaus, dips, and leaps might make to these efforts.

Brain Training

We define *Brain Training* as the collection of research (and commercial products) that see population norms in the performance of a variety of basic tasks, such as those used in experimental psychology or to measure intelligence, as reflecting surmountable limits on human performance that can be overcome by extended training. The focus of Brain Training is on low level cognitive functions that increase very general purpose functionality such as decreases in response time or increases in response accuracy in basic choice tasks, decreases in retrieval time or increases in retrieval accuracy in basic memory tasks, and/or increases in the ability to attend to a wider field of view or to better focus on just one particular location or event in attention tasks.

Although to our ancestors there was some doubt as to whether change in mental states or abilities reflected physical changes in the brain or changes in some other state of being (e.g., as in Descarte's famous dichotomy), there is no doubt in our time that all changes in performance, memory, attention, and whatever, result in or follow from some change in the human brain. However, there are subtleties in how we interpret these changes that, for our current concerns, run along a dimension defined at one end by enhancements in *brain capacity* and at the other by enhancements in *brain efficiency*.

We view limits due to brain capacity as reflecting asymptotic performance. Much like our pole vaunt example, these limits would reflect inherent human limitations in cognitive control, perception, memory, or so on. The metaphor here would be that increased capacity (increased speed of processing, interconnections among cells in one part of the brain, increased connections between different parts of the brain) would provide the potential for a general increase in performance, much like the boost obtained in pole vaunting in switching from a wood pole to a bamboo or fiberglass one.

Limits due to brain efficiency reflect plateaus. Such plateaus would reflect limits in efficiency in the performance of certain low level cognitive tasks. If more efficient methods for performing those low level tasks could be found and practiced, than any higher level skills that require these low level components would be enhanced. In our example, the Fosbury Flop would represent an increase in efficiency.

Experimental Psychology Tasks for Brain Training. Probably the poster child for the *changes in general capacity* approach was the Jaeggi, Buschkuehl, Jonides, and Perrig (2008) study which was interpreted as showing a half standard deviation increase in IQ following 15 hours of training on a specific experimental psychology task; namely, the spatial nBack. It has since been shown that this otherwise very careful group of researchers fell prey to a number of subtle experimental design problems and that when those are corrected the effect goes away (Redick et al., 2012). More recent meta-analyses (e.g., von Bastian & Oberauer, 2014; Dunning & Holmes, 2014; Shipstead et al., 2012) have cast doubt on the more facil of many similar claims. However, the attempts to show (in the scientific literature) or claim (in the public arena) increases in mental capacity due to simple *brain exercises* continues unabated and, as advertisements in various media suggests, seem to have become a commercial success.

Games as Brain Training. The Green and Bavelier (2003) paper was a seismic event in the experimental psychology community and the world press as it showed behavioral differences on a variety of low-level, experimental psychology tasks between experienced players of *First Person Shooter* games and non-players. Perhaps more impressive was the data which showed that non-gamers, after playing such games for about 15-hrs, showed notable changes in their performance on these same sets of measures that moved them in the direction of those with hundreds and thousands of hours on these games. It is important to stress that the claims made by many of the proponents of *Games as Brain Training* are nuanced and focused on specific effects such as perceptual learning (Green, Li, & Bavelier, 2010) or specific components of cognitive control (Strobach, Frensch, & Schubert, 2012). Of course, these (along with other) claims have not gone unchallenged (e.g., Boot, Kramer, Simons, Fabiani, & Gratton, 2008; Unsworth et al., 2015) and, lately, much of the discussion has turned to methodological issues (e.g., Boot, Simons, Stothart, & Stutts, 2013; Green, Strobach, & Schubert, 2014; Latham, Patston, & Tippett, 2013).

Conclusions on Brain Training. Much research on Brain Training analyzes performance at a very high level; *if people do task A then scores on task B get better*. If people play first person shooters a lot (or for at least 15 hrs), their scores on basic tests of memory and performance change (e.g., A. F. Anderson & Bavelier, 2011). If people do a certain type of nBack[15] for 15 hrs their scores on an IQ test might improve (Jaeggi et al., 2008) but more likely not (Redick et al., 2012).

Given our focus on understanding Strategies, Methods, and Drill & Skill for remediating performance plateaus, the reader will not be surprised that what strikes us as most remarkable about the Brain Training literature is the general lack of attention to details at the cognitive,

[15]The nBack is a memory task involving the continuous updating and reporting of a list of items.

perceptual, and motor level of analysis – this is Newell's level of immediate behavior [e.g., see Figure 6 of Newell and Card (1985) and Chapter 5 of Newell (1990)] where milliseconds do matter (Gray & Boehm-Davis, 2000). Our reading of many of the above papers is that it does not seem as if subjects in most (any?) of these studies are getting instruction on what to do or, perhaps rather, "how to do it". Likewise, it does not seem that careful microgenetic analyses (Agre & Shrager, 1990; Siegler, 1991) of what people actually do in these tasks, before or after training, are conducted; that is, no focus on "methods" and little focus on "strategy".

Deliberate Practice

> Definition: "To assure effective learning, subjects ideally should be given explicit instructions about the best method and be supervised by a teacher to allow individualized diagnosis of errors, informative feedback, and remedial part training. The instructor has to organize the sequence of appropriate training tasks and monitor improvement to decide when transitions to more complex and challenging tasks are appropriate." (Ericsson et al., 1993, p. 367)

Deliberate Practice (DP) is a theoretically important concept that has been embraced by a number of communities of practice[16] and widely discussed by numerous researchers across many fields[17]. It is generally viewed as an activity which increments a well defined expertise.

Ericsson has suggested (Personal Communication, 2015.03.22) that *Teacher-Directed DP* might be contrasted with *Self-Directed DP*. This definition might apply to (a) what people in

[16] As a very informal survey, on 2015.05.10 we did 9 Google searches using the term "deliberate practice <>"; where <> was "teacher", "golf", "skiing", "frisbee", "marathon", "poker", "mathematics", "second language learning", and "russian literature." Of these terms, all except "russian literature" produced thousands to millions of hits that when spot checked (from those returned on the first page of results) showed a preponderance of pages talking about the use of DP in that field.

[17] Macnamara et al. (2014, p. 1608) report finding 4,200 citations for Ericsson et al. (1993) in an April 2014 Google Scholar search. Our search of the more conservative Web of Science on May 10, 2015 found 1,744 citations. Clearly both of these are very impressive numbers.

Brain Training tasks do when left to formulate their own strategies and methods for task performance, (b) to our discussion of SF, DD, and Sudnow (above), (c) to what we must do when confronted with advances in human-computer interactions where new technologies are being used to implement new functionality (Cockburn et al., 2014), or (d) to areas where extreme experts create new methods rather than simply following those created by others. An example of the later might be Dehane's example of mathematicians:

> Very little evidence exists that great mathematicians and calculating prodigies have been endowed with an exceptional neurobiological structure. Like the rest of us, experts in arithmetic have to struggle with long calculations and abstruse mathematical concepts. If they succeed, it is only because they devote a considerable time to this topic and eventually invent well-tuned algorithms and clever shortcuts that any of us could learn if we tried, and that are carefully devised to take advantage of our brain's assets and get round its limits. (Dehaene, 2011, p. xxi)

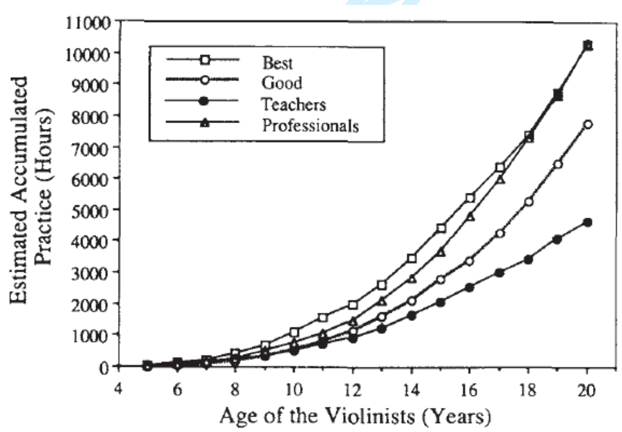

***Figure 8*.** Accumulated amount of practice alone (on the basis of estimates of weekly practice) as a function of age for the middle-aged violinists (△), the best violinists (□), the good violinists (○), and the music teachers (●).(Figure 9 from Ericsson, Krampe, & Tesch-Römer, 1993)

Deliberate Practice as Domain-Specific Hard Work. In the original document (Ericsson et al., 1993) and in others since, DP has been defined as requiring the repeated modification of, “one’s strategies, processes, and representations over [an] extensive period of training” (Tuffiash, Roring, & Ericsson, 2007). However, the relationship between the periods where such strategies, processes, and representations are acquired and the periods of routine practice of these methods is seldom mentioned. Indeed, Figure 8 (which is a copy of Figure 9 from Ericsson et al., 1993, p. 379) converts retrospective reports of hours of practice per week into an estimate of accumulated practice by years of age for (a) individual music students (divided into "best" and "good" students), (b) music teachers, and (c) professional musicians) from age 5 to 20. The caption refers to this plot as the “[a]ccumulated amount of practice alone.” However, the important data is not and could not be reported by a retrospective study; namely, *how much of that practice alone was DP and how much was other activity!* This observation is not new to us but is an often overlooked but acknowledged limit of the original work:

> It is important to note that our study shows only that the amount and distribution of practice is related to the level of performance of adult musicians. In fact, many additional factors consistent with the skill-acquisition framework could attenuate the differences among our three groups. (Ericsson et al., 1993, p. 380)

Controversies in Deliberate Practice. Ericsson tends to take a hardline and argue that what separates experts from others in their field is the amount of DP – period. This strong position opens the work up to two types of criticism; (a) one is based on the relationship between retrospective accounts of hours of practice and the amount of DP represented by those hours, and (b) the other comes from those who see a role for talent (e.g., Ackerman, 2014), intelligence (Grabner, 2014), nurture (Plomin, Shakeshaft, McMillan, & Trzaskowski, 2014), and so on, in the ascension to expert performance that DP does not accommodate.

Proponents of both types of critiques have taken Macnamara’s recent meta-analysis (Macnamara et al., 2014) of DP and performance across multiple domains as support for the view that DP may be necessary but not sufficient to explain expertise. Macnamara et al. found that estimates of hours of DP correlated 0.51 for game expertise, 0.46 for music, 0.42 for sports, 0.21 for education, and 0.05 for professions. These estimates and others have led Macnamara and colleagues to counter Ericsson's strong claims with strong claims of their own; namely;

> [T]he claim that individual differences in performance are largely accounted for by individual differences in amount of Deliberate Practice is not supported by the available empirical evidence. (Macnamara et al., 2014, p. 1616)

For comparison, a recent meta-analysis of DP, which focused solely on musical achievement, found the hours of DP correlated 0.61 with achievement (Platz, Kopiez, Lehmann, & Wolf, 2014). This analysis harvested 13 studies that satisfied the dual criteria of "reported durations of task-specific accumulated practice as predictor variables and objectively assessed musical achievement as the target variable." Their carefully nuanced conclusion is:

> In summary, it is incorrect to interpret our findings ($r_c = 0.61$) as evidence that DP explains 36% of the variance in attained music performance. Instead, it is correct to state that the currently trackable correlation between an approximation of deliberate practice with indicators such as solitary study or task-relevent training experiences is related to measurements of music performance with $r_c = 0.61$.

Challenges for Deliberate Practice. Apparently, whether the DP glass is half-full or half-empty depends on where you sit. 100 years after Galton's emphasis on heredity, it is clear that many researchers and much of the popular press believe that domain expertise is due to the hard work required by a type of practice labeled as DP. However, it is also clear that the construct of DP is both (a) hard to measure and (b) underspecified.

DP is hard to measure as it is extremely difficult to obtain reliable measures for any activity that extends across years and decades (as explicitly recognized in our quote from Platz et al., 2014). Even in the case of Chase and Ericsson's digit span experts, it seems likely that there were some periods spanning days or perhaps weeks when most of their hours could be labeled as DP and other periods when they could not.

To our minds, underspecification is DP's biggest weakness. Without a strong cognitive model of strategy discovery, method development, and practice it is not clear to us what the relationship should be between numbers of hours of practice and increments in expertise. Clearly, "more is usually better", but increments in some skills must come faster than others. For example, anecdotally, expert pianists have told us of the years they spent in a form of (what they referred to as) DP to get small hands to have the reach to play pieces by Chopin and others which, decades and more ago, were assumed to require large ones. Reportedly, this was a long period in which hours and days of practice were traded for millimeter increments in reach. Just as assuredly, there must have been other lessons more easily acquired that resulted in notable performance increments across the span of hours, not months. Simply summarized, studies of Deliberate Practice cannot rest content with hour counts but must focus on hour contents.

Emerging Model-Based Approaches to *What Transfers?*

In fairness to the Brain Training and Deliberate Practice researchers, those both pro and con, running and analyzing a properly controlled brain training study or a long term study of DP is a major undertaking and it may be too much to ask any one set of researchers to account for all aspects of the behavior of the tasks they study. Presumably, such an accounting would come from the cognitive modeling community.

Indeed, 33 years after Marr's famous book (1982), a recent collection of papers (edited by Peebles & Cooper, 2015) take on this task by revisiting Marr's three levels of explanation which are typically translated into three approaches to cognitive modeling. We refer the reader to several papers in that collection which are most relevant to bridging the gap between a cognition that is grounded in neurons but optimized to achieve certain functions in the world; namely, French and Thomas (2015), Love (2015), and Cooper and Peebles (2015).

Modeling the transfer of low level constructs such as interactive routines (Gray & Boehm-Davis, 2000) or declarative knowledge, to higher-level performance is a difficult task. A vast amount of such work can be summarized as yielding many individual successes but no general approaches that would generate the diversity of all, most, or even many of the experimental paradigms and action games used by the Brain Training community or expertise in tasks such as piano playing covered by the DP community. However, progress is being made.

Mileposts towards this destination are provided by Salvucci (2013) on modeling transfer of interactive routines and by Taatgen (2014) on modeling transfer of declarative knowledge. Rather than starting with a blank slate, as so many modelers do, both Salvucci and Taatgen embed their work as modifications to the ACT-R architecture of cognition (J. R. Anderson, 2007). This strategy enables them to reuse the basic control structure of ACT-R and many of ACT-R's existing modules. Hence, rather than focusing on modeling one or two interesting phenomena, both Salvucci and Taatgen focus on changes at the architectural level that, when a new model is created and run, produce procedural (Salvucci) or declarative (Taatgen) knowledge, some of which can be harvested and reused by (i.e., transferred to) other models. In the context of this essay, we cast this work as holding *capacity* of cognition constant and exploring architectural changes that would allow humans to increase *efficiency*.

Summary of Trends

In this section, we have reviewed trends in three areas that seem interrelated to plateaus and dips during the development of League Stepping Habits. Although our 3-tier approach of Strategies, Methods, and Drill & Skill seems relevant to each, as far as we can determine, none of these areas have considered our signature concern; namely, the empirical phenomena of plateaus, dips, and leaps in the acquisition of skilled performance.

PLATEAUS, DIPS, AND LEAPS – THE ESSENTIAL ELEMENTS OF LEAGUE-STEPPING HABITS

Something Old, Something New, Something Borrowed, Something Blue. (The four "somethings" from old English folklore, Wikipedia, 2015)

We began our essay with "something old", a review of performance plateaus and the role they played in the early development of the psychology of skill acquisition (Bryan & Harter, 1897, 1899; Thorndike, 1913). We quickly added in something less old, but still old, the concern with stable suboptimal performance that is encapsulated by Carroll and Rosson's (1987) phrase, "the paradox of the active user."

This led us to "something new", the distinction between performance plateaus and performance asymptotes – an essential distinction for those who believe that performance could or should be better than it is. One route, plateaus, leads towards human training (e.g., Ericsson et al., 1993), whereas the other route, asymptotes, leads towards human factors engineering (e.g., Cockburn et al., 2014).

Something else "new" was the recognition that dips in performance were not simply a concern due to their possibility of demotivating learners (as per our discussion of Cockburn et al., 2014; Yechiam et al., 2003) but should be viewed as periods of experimentation, discovery, trial & error, and successive approximations to developing league-stepping habits. The capture and study of these dips may well be the most important and most overlooked task of the past 120 years of Experimental Psychology.

Our "something borrowed" would be the three-tier approach of Fitts and Posner (1967) to skill acquisition. Of course, this is a venerable workhorse that has been borrowed before (J. R. Anderson, 1982, 1987; Newell & Rosenbloom, 1981) and we benefited greatly from those explorations and explications.

"Something blue" pushes the metaphor a bit but would apply to bringing a focus on plateaus, dips, and leaps to research in Brain Training and Deliberate Practice. It also might apply to our hopes that the maturing field of computational modeling is about to turn its attention to a general account of skill development, integration, and transfer. However, as this is a core problem in cognitive science, the chances are that although a new age of experimentation with modeling formalisms has arrived (Salvucci, 2013; Taatgen, 2014), success is not yet within sight.

SUMMARY AND CONCLUSIONS

League Stepping Habits are relevant to issues of Brain Capacity, Brain Efficiency, Deliberate Practice, and most any other way of phrasing the processes of human learning. From a practical sense, it is vital to distinct between performance plateaus versus asymptotes. Asymptotes cannot be remediated by training but they might be remediated by design (Cockburn et al., 2014). In an active and motivated learner, dips in performance may signal a period of exploration, development, and testing of new ways of dividing the task, new strategies, and/or new methods.

> Early investigators of extended skill acquisition . . . revealed subjects' active search for methods to improve performance and found that changes in methods could often be related to clear improvements. [More recent studies] have also shown that subjects actively try out different methods and refine methods in response to errors and violated expectations. (Ericsson et al., 1993)

The focus on methods became fashionable following Bryan and Harter's work (1897, 1899), enjoyed a revival in the 1980's and 1990's, and is overdue for another. As the first 120 years of experimental research on skill acquisition has shown, even when we tell people the strategy and teach them the methods, they will often satisfice with stable suboptimal performance (Fu & Gray, 2004). Surpassing suboptimal performance requires motivation, effort, new ways of dividing or aggregating task performance, exploring new strategies, some combination of invention, instantiation, and refinement of methods, and much Drill & Skill.

Beyond the cognitive science challenge of understanding the process of strategy and method discovery, refinement, and practice, we also need to focus on identifying plateaus and motivating individuals to go beyond plateaued methods. Like Mr. McGurrin's invention of touch typing, SF's going way beyond 7 ± 2, and Sudnow's dogged pursuit of expertise at BreakOut, the idea that a task could be done better and the motivation to do it better are vitally important and little understood.

The conclusions we draw from this review are as follows:

- The difference between a plateau and an asymptote may be hard to discern – especially when there is no higher reference point.

- Bryan and Harter may be right that plateaus mark periods in which new league-stepping strategies and methods are being developed. However, rather than stable plateaus, SF's and DD's plots show performance dips that might be attributable to abandoning learned strategies and methods to work out the kinks in new ones. Sudnow's self-reports, though qualitative rather than quantitative, also lead us to conclude that most of his experiments with methods resulted in, at least initially, poorer performance.

- Not every hour of practice is equal. Some focus on mastering new strategies, others on implementing different methods, others on practicing new methods until they are automated. Hence, hours of practice per se will always be an imperfect measure of performance change – whether due to Brain Training or Deliberate Practice.

- Performance dips are our current candidate for markers of periods of exploration and change. These will be difficult to identify and study as they will require collection and analysis of data at a microlevel typically not pursued in experimental psychology, at least not over long periods of time (though see Gray, 2000; Gray & Boehm-Davis, 2000; Lindstedt & Gray, 2015; Schoelles & Gray, 2001).

- The discussion of genetic and environmental contributions to skilled performance is important and will not soon be resolved (as per the recent special issue of *Intelligence*, see Detterman, 2014).

- Performance plateaus within individuals or within groups of individuals, such as is the case for many Brain Training tasks, provide cases where the phenomenon of plateaus, dips, and leaps in skilled performance can be studied with some precision in the laboratory. The use of computer games as experimental paradigms in these studies may facilitate the long term study and microgenetic analysis of such data in the university laboratory. Big Data, based on the harvesting and analysis of online gaming records, has yielded clear evidence of plateaus (Huang et al., 2013, p. 702) and may facilitate this endeavor (Gray, in preparation).

References

Ackerman, P. L. (2014). Nonsense, common sense, and science of expert performance: talent and individual differences. *Intelligence*, 1–12. doi:http://dx.doi.org/10.1016/j.intell.2013.04.009

Agre, P. E. & Shrager, J. (1990). Routine evolution as the microgenetic basis of skill acquisition. In *12th annual conference of the cognitive science society* (pp. 694–701). Cognitive Science Society.

Anderson, A. F. & Bavelier, D. (2011). Action game play as a tool to enhance perception, attention, and cognition. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 307–329). Information Age Publishing.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369–406. doi:10.1037//0033-295X.89.4.369

Anderson, J. R. (1987). Skill acquisition - compilation of weak-method problem solutions. *Psychological Review*, *94*(2), 192–210. doi:10.1037//0033-295X.94.2.192

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford series on cognitive models and architectures. New York: Oxford University Press.

Book, W. F. (1908). The psychology of skill: with special reference to its acquisition in typewriting. *Bulletin No. 53, Psychological Series No. 1*.

Boot, W. R., Basak, C., Erickson, K. I., Neider, M., Simons, D. J., Fabiani, M., . . . Kramer, A. F. (2010). Transfer of skill engendered by complex task training under conditions of variable priority. *Acta Psychologica*, *135*(3), 349–357.

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica*, *129*(3), 387–398. doi:{10.1016/j.actpsy.2008.09.005}

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The Pervasive Problem With Placebos in Psychology: Why Active Control Groups Are Not Sufficient to Rule Out Placebo

Effects. *Perspectives on Psychological Science*, *8*(4), 445–454. doi:{10.1177/1745691613491271}

Bryan, W. L. & Harter, N. (1897). Studies in the physiology and psychology of the telegraphic language. *Psychological Review*, *4*(1), 27–53. doi:10.1037/h0073806

Bryan, W. L. & Harter, N. (1899). Studies on the telegraphic language: the acquisition of a hierarchy of habits. *Psychological Review*, *6*(4), 345–375. doi:10.1037/h0073117

Carroll, J. M. & Rosson, M. B. (1987). Paradox of the active user. In J. M. Carroll (Ed.), *Interfacing thought: cognitive aspects of human-computer interaction*. Cambridge, MA: MIT Press.

Chase, W. G. & Ericsson, K. A. (1982). Skill and working memory. *The Psychology of Learning and Motivation*, *16*, 1–58. doi:10.1016/S0079-7421(08)60546-0

Cockburn, A., Gutwin, C., Scarr, J., & Malacria, S. (2014). Supporting novice to expert transitions in user interfaces. *ACM Computing Survey*, *47*(2), 1–36. doi:10.1145/2659796

Cooper, R. P. & Peebles, D. (2015). Beyond single-level accounts: the role of cognitive architectures in cognitive scientific explanation. *Topics in Cognitive Science*, *7*(2), 243–258. doi:10.1111/tops.12132

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. *Ergonomics*, *2*(2), 153–166. doi:10.1080/00140135908930419

Dehaene, S. (2011). *The number sense: how the mind creates mathematics* (2nd). Oxford University Press.

Delaney, P. F., Reder, L., Staszewski, J., & Ritter, F. (1998). The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science*, *9*(1), 1–7. doi:{10.1111/1467-9280.00001}

Destefano, M. (2010). *The mechanics of multitasking: the choreography of perception, action, and cognition over 7.05 orders of magnitude* (Doctoral dissertation).

Detterman, D. K. (Ed.). (2014). Acquiring expertise: ability, practice, and other influences [Special issue]. *Intelligence*, *45*. doi:http://dx.doi.org/10.1016/j.intell.2014.02.004

Doane, S. M., Pellegrino, J. W., & Klatzky, R. L. (1990). Expertise in a computer operating system: conceptualization and performance. *Human-Computer Interaction*, *5*(2-3), 267.

Donchin, E. (1989). The learning-strategies project: introductory remarks. *Acta Psychologica*, *71*(1-3), 1–15.

Donchin, E. (1995). Video games as research tools: the space fortress game. *Behavior Research Methods, Instruments, & Computers*, *27*(2), 217–223.

Dunning, D. & Holmes, J. (2014). Does working memory training promote the use of strategies on untrained working memory tasks? *Memory & Cognition*, *42*(6), 854–862. doi:10.3758/s13421-014-0410-5

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, *79*(10, S), S70–S81. doi:{10.1097/00001888-200410001-00022}

Ericsson, K. A. & Chase, W. G. (1982). Exceptional memory. *American Scientist*, *70*(6), 607–615.

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*(4448), 1181–1182.

Ericsson, K. A. & Kintsch, W. (1995). Long-term working-memory. *Psychological Review*, *102*(2), 211–245. doi:{10.1037//0033-295X.102.2.211}

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. doi:{10.1037/0033-295X.100.3.363}

Fitts, P. M. & Posner, M. I. (1967). *Human performance*. Belmont, CA: Brooks Cole.

Frederiksen, J. R. & White, B. Y. (1989). An approach to training based upon principled task decomposition. *Acta Psychologica*, *71*(1-3), 89–146. doi:{10.1016/0001-6918(89)90006-1}

French, R. M. & Thomas, E. (2015). Interactive effects of explicit emergent structure: a major challenge for cognitive computational modeling. *Topics in Cognitive Science*, *7*(2), 206–216. doi:10.1111/tops.12135

Fu, W.-T. & Gray, W. D. (2004). Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. *Cognitive Science*, *28*(6), 901–935. doi:10.1207/s15516709cog2806_2

Gleick, J. (2011). *The information: a history, a theory, a flood*. Pantheon.

Gopher, D., Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors*, *36*(3), 387–405.

Grabner, R. H. (2014). The role of intelligence for performance in the prototypical expertise domain of chess. *Intelligence*, 1–8. doi:http://dx.doi.org/10.1016/j.intell.2013.07.023

Gray, W. D. (2000). The nature and processing of errors in interactive behavior. *Cognitive Science*, *24*(2), 205–248.

Gray, W. D. (Ed.). (in preparation). Action games as cognitive science paradigm.

Gray, W. D. & Boehm-Davis, D. A. (2000). Milliseconds matter: an introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied*, *6*(4), 322–335.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world performance. *Human-Computer Interaction*, *8*(3), 237–309. doi:10.1207/s15327051hci0803_3

Green, C. S. & Bavelier, D. (2012). Learning, Attentional Control, and Action Video Games. *Current Biology*, *22*(6), R197–R206. doi:{10.1016/j.cub.2012.02.012}

Green, C. S. & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537.

Green, C. S., Li, R., & Bavelier, D. (2010). Perceptual learning during action video game playing. *Topics in Cognitive Science*, *2*(2), 202–216.

Green, C. S., Strobach, T., & Schubert, T. (2014). On methodological standards in training and transfer experiments. *Psychological Research*, *78*(6), 756–772. doi:10.1007/s00426-013-0535-3

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*, 1–12. doi:http://dx.doi.org/10.1016/j.intell.2013.04.001

Hayes, J. R. (1985). Three problems in teaching general skills. In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 2). Hillsdale, NJ: Erlbaum.

Huang, J., Zimmermann, T., Nagapan, N., Harrison, C., & Phillips, B. C. (2013). Mastering the art of war: How patterns of gameplay influence skill in Halo. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 695–704). New York, NY, USA: ACM. doi:10.1145/2470654.2470753

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833.

John, B. E. & Kieras, D. E. (1996). Using goms for user interface design and evaluation: which technique? *ACM Transactions on Computer-Human Interaction*, *3*(4), 287–319.

Keller, F. S. (1958). The phantom plateau. *Journal of the Experimental Analysis of Behavior*, *1*(1), 1–13. doi:10.1901/jeab.1958.1-1

Kellman, P. J. & Garrigan, P. (2009). Perceptual learning and human expertise. *Physics of Life Reviews*, *6*(2), 53–84. doi:10.1016/j.plrev.2008.12.001

LaBerge, D. & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. doi:{10.1016/0010-0285(74)90015-2}

Landauer, T. K. (1995). *The trouble with computers: usefulness, usability, and productivity*. Cambridge, MA: MIT Press.

Latham, A. J., Patston, L. L. M., & Tippett, L. J. (2013). Just how expert are "expert" video-game players? Assessing the experience and expertise of video-game players across "action" video-game genres. *Frontiers In Psychology*, *4*, 1–3. doi:{10.3389/fpsyg.2013.00941}

Lindstedt, J. K. & Gray, W. D. (2015). MetaT: Tetris as an Experimental Paradigm for Cognitive Skills Research. *Behavior Research Methods*. doi:10.3758/s13428-014-0547-y

Logan, G. D. & Crump, M. J. C. (2011). Hierarchical control of cognitive processes: the case for skilled typewriting. *Psychology of Learning and Motivation*, *54*, 1–27. doi:10.1016/B978-0-12-385527-5.00001-2

Love, B. C. (2015). The algorithmic level is the bridge between computation and brain. *Topics in Cognitive Science*, *7*(2), 230–242. doi:10.1111/tops.12131

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: an act-r perspective. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active maintenance & executive control* (pp. 135–182). New York: Cambridge University Press.

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. doi:10.1177/0956797614535810

Marr, D. C. (1982). *Vision: a computational investigation into the human representation and processing of visual information.* New York: W. H. Freeman.

Memmert, D., Raab, A., & Bauer, P. . (2006). Laws of practice and performance plateaus concerning complex motor skills. *Journal of Human Movement Studies*, *51*(4), 239–255.

Newell, A. (1973). You can't play 20 questions with nature and win: projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). New York: Academic Press.

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

Newell, A. & Card, S. K. (1985). The prospects for psychological science in human-computer interaction. *Human-Computer Interaction*, *1*(3), 209–242.

Newell, A. & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Lawrence Erlbaum Associates.

Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Nilsen, E., Jong, H., Olson, J. S., Biolsi, K., Rueter, H., & Mutter, S. (1993). The growth of software skill: a longitudinal look at learning & performance. In *Proceedings of the interact '93 and chi '93 conference on human factors in computing systems* (pp. 149–156). Amsterdam, The Netherlands: ACM. doi:10.1145/169059.169126

Peebles, D. & Cooper, R. P. (2015). Thirty years after marr's vision: levels of analysis in cognitive science. *Topics in Cognitive Science*, *7*(2), 187–190. doi:10.1111/tops.12137

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), 1–13. doi:10.3389/fpsyg.2014.00646

Plomin, R., Shakeshaft, N. G., McMillan, A., & Trzaskowski, M. (2014). Nature, nurture, and expertise: response to ericsson. *Intelligence*, 1–3. doi:http://dx.doi.org/10.1016/j.intell.2014.01.003

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., . . . Engle, R. W. (2012). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology: General*. doi:10.1037/a0029082

Richman, H. B., Staszewski, J. J., & Simon, H. A. (1995). Simulation of expert memory using EPAM-IV. *Psychological Review*, *102*(2), 305–330. doi:{10.1037//0033-295X.102.2.305}

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology - General*, *126*(3), 288–311. doi:{10.1037/0096-3445.126.3.288}

Robertson, R. J. & Glines, L. A. (1985). The phantom plateau returns. *Perceptual and Motor Skills*, *61*(1), 55–64.

Salvucci, D. D. (2013). Integration and Reuse in Cognitive Skill Acquisition. *Cognitive Science*, *37*(5), 829–860. doi:{10.1111/cogs.12032}

Scarr, J., Cockburn, A., Gutwin, C., & Quinn, P. (2011). Dips and ceilings: understanding and supporting transitions to expertise in user interfaces. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 2741–2750). CHI '11. Vancouver, BC, Canada: ACM. doi:10.1145/1978942.1979348

Schoelles, M. J. & Gray, W. D. (2001). Argus: a suite of tools for research in complex cognition. *Behavior Research Methods, Instruments, & Computers*, *33*(2), 130–140.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is Working Memory Training Effective? *Psychological Bulletin*, *138*(4), 628–654. doi:{10.1037/a0027473}

Siegler, R. S. (1991). The microgenetic method: a direct means for studying cognitive development. *American Psychologist*, *46*(6), 606–620.

Siegler, R. S. & Stern, E. (1998). Conscious and unconscious strategy discoveries: a microgenetic analysis. *Journal of Experimental Psychology-General*, *127*(4), 377–397.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, *106*(6), 467–482.

Strobach, T., Frensch, P. A., & Schubert, T. (2012). Video game practice optimizes executive control skills in dual-task and task switching situations. *Acta Psychologica*, *140*(1), 13–24. doi:{10.1016/j.actpsy.2012.02.001}

Sudnow, D. (1983). *Pilgrim in the Microworld*. New York, NY: Warner Books.

Swift, E. J. (1903). Studies in the psychology and physiology of learning. *The American Journal of Psychology*, *14*(2), 201–251. doi:10.2307/1412713

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review*, *120*(3), 439–471. Retrieved from http://search.ebscohost.com.libproxy.rpi.edu/login.aspx?direct=true&db=pdh&AN=2013-19654-001&site=ehost-live

Taatgen, N. A. (2014). Between architecture and model: strategies for cognitive control. *Biologically Inspired Cognitive Architectures*, *8*, 132–139. doi:http://dx.doi.org/10.1016/j.bica.2014.03.010

Taylor, D. W. (1943). Learning telegraphic code. *Psychological Bulletin*, *40*(7), 461–487. doi:10.1037/h0054172

Thorndike, E. L. (1913). *Educational Psychology Vol II: The Psychology of Learning*. NYC: Teachers College, Columbia University.

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. doi:{10.1037/1076-898X.13.3.124}

Unsworth, N., Redick, T. S., McMillan, B. D., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2015). Is playing video games related to cognitive abilities? *Psychological Science*, 1–16. doi:10.1177/0956797615570367

von Bastian, C. C. & Oberauer, K. (2014). Effects and mechanisms of working memory training: a review. *Psychological Research-Psychologische Forschung*, *78*(6, SI), 803–820. doi:{10.1007/s00426-013-0524-6}

Wikipedia. (2014). *Frank Edward McGurrin — Wikipedia, The Free Encyclopedia*. [Online; accessed 9-January-2015]. Retrieved from http://en.wikipedia.org/w/index.php?title=Frank_Edward_McGurrin&oldid=626496435

Wikipedia. (2015). Something old — wikipedia, the free encyclopedia. [Online; accessed 30-April-2015]. Retrieved from %5Curl%7Bhttp://en.wikipedia.org/w/index.php?title=Something_old&oldid=656686892%7D

Yechiam, E., Erev, I., Yehene, V., & Gopher, D. (2003). Melioration and the transition from touch-typing training to everyday use. *Human Factors*, *45*(4), 671–684. doi:10.1518/hfes.45.4.671.27085

Zbrodoff, N. J. (1995). Why is $9 + 7$ harder than $2 + 3$: Strength and interference as explanations of the problem-size effect. *Memory & Cognition*, *23*(6), 689–700. doi:{10.3758/BF03200922}