COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2016) 1–33

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12412

Plateaus, Dips, and Leaps: Where to Look for Inventions and Discoveries During Skilled Performance

Wayne D. Gray, John K. Lindstedt

Cognitive Science Department, Rensselaer Polytechnic Institute

Received 1 October 2015; received in revised form 31 May 2016; accepted 10 June 2016

Abstract

The framework of *plateaus, dips, and leaps* shines light on periods when individuals may be inventing new methods of skilled performance. We begin with a review of the role *performance plateaus* have played in (a) experimental psychology, (b) human–computer interaction, and (c) cognitive science. We then reanalyze two classic studies of individual performance to show plateaus and dips which resulted in performance leaps. For a third study, we show how the statistical methods of *Changepoint Analysis* plus a few simple heuristics may direct our focus to periods of performance change for individuals. For the researcher, dips become the marker of exploration where performance suffers as new methods are invented and tested. Leaps mark the implementation of a successful new method and an incremental jump above the path plotted by smooth and steady log–log performance increments. The methods developed during these dips and leaps are the key to surpassing one's teachers and acquiring extreme expertise.

Keywords: Expertise; Performance; Plateaus; Dips; Leaps; Breakout; Digit span; Changepoint detection; Extreme expertise; Space Fortress

1. Introduction

With practice, performance generally improves, whether it is with telegraphy, typing, programming, driving, arithmetic, laparoscopic surgery, mnemonics, air traffic control, experimental psychology paradigms, or video games. Near the dawn of experimental psychology, Bryan and Harter (1899) claimed that experts were not simply faster than novices, but had developed a *hierarchy of habits* that enabled them to *step leagues*[1] while novices were *bustling over furlongs*[2] *or inches.* And they were the first to notice that learning did not always steadily advance, but that it sometimes plateaued.

Correspondence should be sent to Wayne D. Gray, Cognitive Science Department, Rensselaer Polytechnic Institute, Troy, NY 12180. E-mail: grayw@rpi.edu

Plateaus in performance were quickly established, and by 1913 Thorndike could say:

> 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)

Plateaus of performance are the better known of three markers of skill acquisition. The modern habit of averaging data across subjects has blinded us to the other two; namely, *dips and leaps*. Dips are periods where performance suffers as new methods are being invented, tested, rejected, or accepted. Leaps are short periods where the success of a new method brings behavior out of the dip to surpass prior levels of achievement. The behavioral markers of plateaus, dips, and leaps signal periods in which the microdynamics of performance change differ—understanding what happens in these three periods will be the key to moving our science of learning from its current plateau to bring new light to understanding the development of skill, expertise, and extreme expertise.

1.1. Definitions and distinctions

Throughout this paper we make two distinctions that are worth introducing now. First, following Thorndike, we distinguish "spurious limits," which could be surpassed, from "real limits" due to "the mechanism for the function itself." The former are named *plateaus* and the latter *asymptotes*. In more modern terms, a plateau exists when there is a better method for implementing the current goal. An asymptote exists when the method(s) for achieving the goal are optimal for the given task in the given task environment.

Second, whereas much of experimental psychology studies short, repetitive tasks designed to isolate a single function of cognition (e.g., the nBack task, the AX-CPT task, visual short-term memory tasks, etc), our tasks entail complex behaviors which unfold sequentially over time. These tasks can be viewed as having a hierarchic structure (Simon, 1962) which imposes an interesting "why" and "how" relationship between superordinate and subordinate levels. For a node at level n of our hierarchy, if asked "how it does what it does," we can point to its subordinate nodes at level $n + 1$ as the *method* it invokes. If asked "why" we are doing it, we can point to its superordinate at level $n - 1$ as the *goal* we are trying to achieve. Hence, for most examples it should suffice to discuss only two levels of analysis; namely, "the why" (the local goal) and "the how" (its method).

1.2. Preview

In Section 2, we (a) sketch the first 90 years (1897–1987) of scientific research on expert performance, (b) introduce the modern origins of our impatience with *stable suboptimal performance* (Fu & Gray, 2004)—the view that *expert is not good enough*, and (c) elaborate our distinction between performance asymptotes and plateaus. We will argue that the roots of stable performance vary depending on whether performance has asymptoted or plateaued.

Section 3 introduces and elaborates the microdynamics of performance change that we view as necessary to escape plateaus.

Section 4 provides extended examples of three tasks in which the *plateaus, dips, and leaps* (PDL) are emphasized. Data for the first two were collected and initially analyzed with the traditional tools of studies of individual performance; namely self-reports and protocol analyses. The third emphasizes modern logging practices, the statistical technique of changepoint analysis, and a set of heuristics for separating periods in which new methods are being developed from periods of general fatigue or boredom with the task. After presenting these examples, we discuss what various patterns of PDLs might or might not tell us about acquisition of extreme expertise. We end the section by comparing the individual approach inherent to PDL to the more traditional approach of averaging data across all subjects.

In *Further Applications* (Section 5), we relate the PDL approach to work in Brain Training and Deliberate Practice. We then (in Section 6) provide a reprise of the major points of the paper and, in our final section (Section 7), provide a short summary and our conclusions.

1.3. Summary

PDLs are behavioral events, not hypothetical constructs. During these periods: (a) performance may be stable and/or suboptimal (plateaus) but new methods are sometimes being invented, (b) means for implementing these new methods may be developed (dips), and (c) the success of these invented methods may result in advances in performance (leaps), whereas their failures may result in a fall-back to old methods and continued plateaus.

Plateaus, dips, and leaps are objective behavioral events which are all but ignored by traditional discussions of the acquisition of skilled performance. Understanding the changes that occur during these periods is the key to understanding how experts exceed their teachers to acquire extreme expertise.

2. Historic and recent background

2.1. Plateaus in training and plateaus in performance

Observing that their student telegraph operators passed through several plateaus during training—periods in which practice continued but improvements halted—Bryan and

Harter (1897, 1899) 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 (1899) 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.

Although the plateaus that Bryan and Harter reported were very influential and plateaus were soon reported by researchers in skills as diverse as typewriting (Book, 1908) and juggling (Swift, 1903), in the decades that followed a controversy arose as to whether plateaus during *the training* of telegraph operators could be eliminated with different conditions of practice (Taylor, 1943).

In the behaviorist era, 60 years after the original study, Keller (1958) authored a blistering attack on Bryan and Harter's (1899) learning plateaus and the hierarchical goal theory which Bryan claimed that they supported. As a behaviorist, Keller's theoretical arsenal would not have included *hierarchies*, but it did include concepts that would predict a slow but steady increase in performance with practice. Today, almost 60 years after Keller's paper, it strikes us as a missed opportunity that rather than attacking the first theory of learning to emphasize hierarchical structures, that Keller did not turn, instead, to investigating which conditions of practice promoted plateaus and which did not.

To the best of our knowledge, no one in the generations of researchers between the days of Bryan and Harter (1897, 1899) and ours ever doubted the reports of performance plateaus. Indeed, in a passage that foreshadowed Carroll and Rosson's *discovery* of the *Paradox of the Active User* (Carroll & Rosson, 1987), Thorndike wrote:

> I venture to prophesy that the thousand bookkeepers in, say, the grocery stores of New York who have each had a 1,000 h 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.
>
> . . .
>
> 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)

Despite whatever merits the critiques of Bryan and Harter's theory of hierarchical structures might 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 documented.

2.2. MERE *expertise is not good enough*

Ninety 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 concluded that "the belief that a sufficient amount of experience or practice leads to maximal performance appears incorrect" (Ericsson, Krampe, & Tesch-Römer, 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, Jong, Olson, Biolsi, Rueter, & Mutter, 1993).[5]

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 methods. 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 versus 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 h 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 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 periods of stable suboptimal performance are real.[6]

2.3. Plateaus versus asymptotes

We divide periods of stable performance into two categories: (a) plateaus are periods of stable suboptimal performance, whereas (b) asymptotes are situations in which performance is stable and as good as humanly possible. 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 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.[7]

In contrast, record breakthroughs in the sister sport of high jumping were primarily driven by changes in method, the *Fosbury Flop* (contrast Figs. 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 and not human anatomy or capabilities, that limited performance. The Fosbury Flop enabled league-stepping increments in human performance that escaped a plateau.

2.3.1. Plateaus due to choosing the wrong goals or methods

As mentioned in Section 1, we view the plateau versus asymptote dichotomy as a useful distinction, not as a rigorous one. However, the distinction is often clear in hindsight when it can be shown that groups of individuals performing at different skill levels are doing different things. Like the Scissors and Straddle versus the Fosbury Flop high jumpers (see Fig. 2), *visually guided* typists[8] 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 touchtyping initially results in a performance *dip*, not a *leap*. However, once steady practice of the new method begins, the new touch typist catches up with and then leaps ahead of the old hunt-and-peck typist.

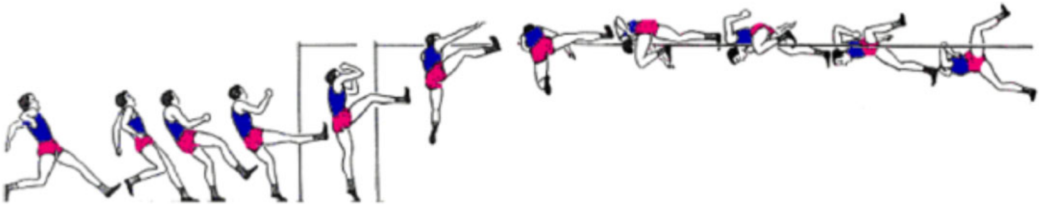

Fig. 1. The *Scissors and Straddle* technique for high jumping. (Figure used by permission of Carlos Lopez and downloaded from https://carloselopez.com/2011/12/28/change-management-and-high-jump)

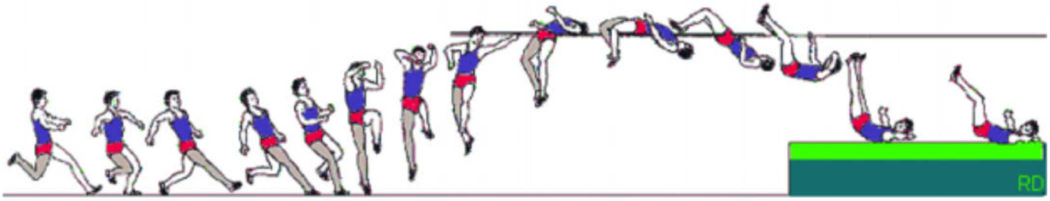

Fig. 2. The *Fosbury Flop* technique for high jumping. (Figure used by permission of Carlos Lopez and downloaded from https://carloselopez.com/2011/12/28/change-management-and-high-jump)

2.3.2. 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, Chase and Ericsson (1982) knew that both SF and DD[9] started at the normal digit span plateau of 7 ± 2, but with several hours of practice these students were able to exceed that number. 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 either (a) by the methods used to perform the task and the general unwillingness of people to put in the time and effort required to discover and practice new methods, or (b) due to our general reluctance to temporarily lose productivity from the inevitable *dips* which occur as new methods are acquired. Another way of stating this is that *a plateau exists when there is a better method for implementing the current goal. An asymptote exists when the methods for achieving the goal are optimal for the given task in the given task environment.*

3. Escaping performance plateaus by acquiring *League-Stepping Methods*

We see the acquisition of League-Stepping Methods as largely compatible with past and current work on skill acquisition. After a short overview of the three parts of our framework, we review past work on modeling skill acquisition (Section 3.2). We accept the basic framework established by Fitts and Posner (1967) and elaborated by Newell and

Rosenbloom (1981) and Anderson (1982, 1987). We are especially friendly toward the amendments to this view offered by *revisionist power law researchers* such as Rickard (1997, 1999), Delaney, Reder, Staszewski, and Ritter (1998), Haider and Frensch (2002), and Donner and Hardy (2015).

3.1. Overview

Acquiring League-Stepping Methods, which move performance off the plateau, requires three types of cognitive activities, each with its own behavioral marker:

1. Method Invention, Discovery, or Instruction: For a new method to replace an old one, someone must first wonder if there is a "better way." The period before such an invention may be marked by the *gradual plateau* of being far out on the practice curve for the current method.
2. Method Development: Having an idea of how to do something differently is far from being able to do it differently and well. As a cultural example, compare the several times which Tiger Woods sacrificed short-term gains in order to master better ways of swinging his golf club (e.g., see Eden, 2015). Although it may not be the case that new methods are always accompanied by performance dips or that all dips signal new methods, it is the case that some dips are behavioral markers of method development.
3. Practice: The behavioral marker of the early trials of practice on the new method is the leap, which takes performance with the new method beyond that of the old. However, with extended practice the new method produces fewer gains per trial and, as would be expected by traditional (Anderson, 1982, 1987; Fitts & Posner, 1967; Newell & Rosenbloom, 1981) or revisionist (e.g., Donner & Hardy, 2015; Haider & Frensch, 2002) accounts of skill acquisition, increments in practice produce smaller and smaller increases in skilled performance—which eventually reaches a new, albeit higher, plateau.

For someone who has never seen a juggler, Fig. 3 may suffice to understand the major subgoals of juggling. 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 the subgoals into the cognitive, perceptual, and motor operations that constitute the methods required to execute those subgoals. Indeed, it seems fair to say that a person's understanding of verbally or figurally presented "task goals" changes as they begin to develop the method or methods to implement them.

The method development step includes a period of adjusting the new methods 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). We believe that those combinations of new goals and methods that produce leaps in performance are what Bryan and Harter referred to as *League-Stepping Habits* and which we refer to as *League-Stepping Methods*.

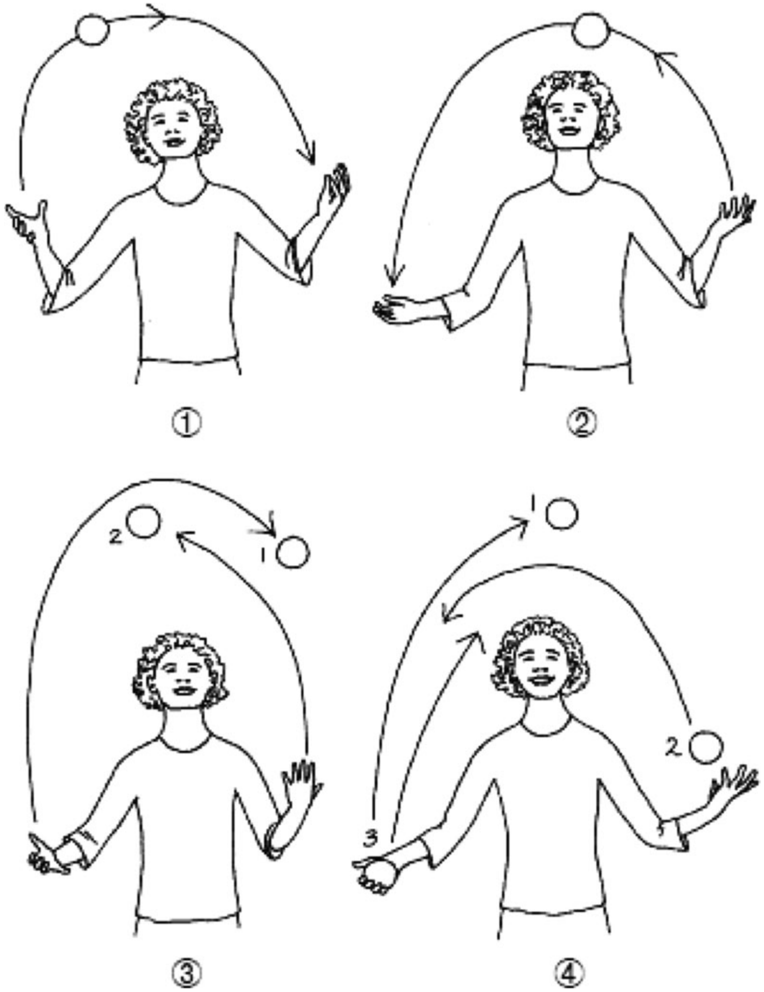

Fig. 3. How to juggle. (Figure used by permission of Andrea J. Buchanan and downloaded from http://www.daringbookforgirls.com/about-the-book/about-the-double-daring-book-for-girls/how-to-juggle/; Alexis Seabrook, illustrator).

3.2. The microdynamics of performance change: Three stages and their behavioral markers

We see plateaus, dips, and leaps as behavioral markers of (a) the discovery of new task goals, (b) the working out of new methods, and (c) the leap of rapid progress gained from the initial use of a better method. This view is largely compatible with the three-tier framework of Fitts and Posner (1967) in both traditional (e.g., Anderson, 1987; Newell & Rosenbloom, 1981) and revisionist (e.g., Delaney et al., 1998; Donner & Hardy, 2015; Haider & Frensch, 2002; Rickard, 1997, 1999) accounts. However, none of these works have a story about the invention or discovery of new goals and methods. While we do not yet claim to have such a story, we do claim to have a place to start looking; namely, in the dips and leaps.

As our notional plot in Fig. 4 suggests, to surpass the normal performance that Thorndike (1913) described and which Carroll and Rosson (1987) and Ericsson et al. (1993) bemoaned, we propose that it may often be necessary to discover a series of more and more advanced methods with each method being mastered or discarded based on the person's ability to invent (or otherwise acquire) and master better methods. In the next three subsections, we discuss our three behavioral markers in terms of method invention, method development, and practice.

3.2.1. Method invention at performance plateaus

Method invention 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 touchtyping, 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, 2015a) and that "the system of fingering is so simple that anybody could formulate it" (Wikipedia, 2015a).

Contrary to Mr. McGurrin, there are many systems for fingering, some of which require that each key be always typed by the same finger while others assign some keys to multiple fingers (Wikipedia, 2016). Likewise, the layout of the keys may vary with QWERTY and Dvorak being the best known. Laying out the details of a new method of typing requires much thought and hard work.

3.2.2. 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 goals and subgoals for performing a task was the only requisite for increasing expertise, then we would have a lot more extreme experts. Learning a new method to perform a skilled task requires a lot of hard work and, presumably, exceptional motivation. For example, Frank McGurrin was apparently driven to invent touchtyping by his boss, who told him that a certain young female

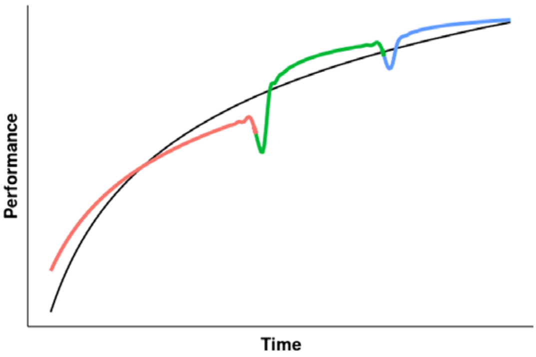

Fig. 4. Notional plot of a succession of three performance curves separated by dips and leaps.

secretary could type as fast as her boss could dictate. Rather than deflating Mr. McGurrin as his boss intended, this tale inspired him: “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” (Wikipedia, 2015a).

From Yechiam’s data (2003) (discussed earlier), we concluded that learning touchtyping 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 suffers a dip 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 are a source of discouragement for the learner, for the researcher the opposite is 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 following the adoption and practice of new and improved ones. Hence, for the researcher dips become the marker of transitional phases of performance and this paper may be viewed as a call to study the *microdynamics of performance change for individuals*; that is, a focus on individual change and differences, in addition to (not in lieu of) the more typical focus on group differences.[10]

3.2.3. Practice: The power of chunking and hierarchies

To the dismay of Keller (1958) and the behaviorists, two founding contributions of modern cognitive theory were the importance of hierarchies (Simon, 1962) and the establishment of the power law of practice (Anderson, 1982, 1987; Fitts & Posner, 1967; Newell & Rosenbloom, 1981). Together these can explain the initial performance increases with practice and its later declining returns (Card, Moran, & Newell, 1983; Newell & Rosenbloom, 1981).

Having a hierarchy of methods enables us to practice and change one part of a skill without necessarily changing all. For example, for a touchtyper, changing the layout of the letters from QWERTY to Dvorak keyboards changes the mapping of keys to letters, but it does not change the mappings of fingers to keys. From the home row, I use the same finger movement to type the first key in the row above the home row whether it be the “q” key in QWERTY or the “/” key in DVORAK. Hence, although the methods for typing words might change, the methods for pressing keys remain constant.

The power law of practice and the hierarchy of methods work together in powerful ways. For example, Delaney et al. (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 subgoals. Complementary, Rickard (1997) has shown that improvements in overall task performance may reflect improvements in two or more alternative methods. More recently, some have argued that the log-log or Power Law of learning may be apparent when performance is grouped over many subjects but may not characterize individual subjects or individual subskills (Donner & Hardy, 2015; Haider & Frensch, 2002). Likewise, it is believed that various alternative methods may co-exist (e.g., see Siegler & Stern, 1998) with the more efficient method never completely dominant.

There are some cases where the slow mechanisms of log-log learning seem capable of producing new methods through the additional mechanisms of knowledge compilation (a type of hierarchical chunking; Anderson, 1982, 1987) and memory retrieval. A case in point is the Alpha-Arithmetic task (Lovett, Reder & Lebiere, 1999; Zbrodoff, 1995) in which subjects are given tasks such a H + 3 = K or C + 2 = E and told to verify whether the equations were correct or not. Initially the count is serial with people counting up, for example, H, I, J, K; 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, practice provides a path to developing 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 the abrupt changes of the Fosbury Flop nor the radical transition from hunt-and-peck to touchtyping (see Haider & Frensch, 2002, for an extended discussion of this point).

3.3. Strategy change, subtask learning, and conclusions

We are also excited by discussions of strategy change such as Billman and Shaman (1990) and the newer focus on subtask learning such as Anglim and Wynton (2015) and Tenison and Anderson (2016). We see this work, along with the interesting theoretical developments and mathematical analyses offerred by Delaney et al. (1998), Rickard (1997, 1999), and Haider and Frensch (2002), and the recent elegant and innovative Big Data analyses of Donner and Hardy (2015), as consistent with our view that the acquisition of extreme expertise progresses through a series of plateaus, dips, and leaps resulting from the invention and implementation of new methods.

Although these recent reports advance our understanding of the computational and mathematical mechanisms of change, in all cases, the tasks being studied have been simple ones taken from the library of traditional experimental psychology. The measures discussed, while elegantly analyzed, are simple performance measures, usually response time or number correct with only one measure from each data set. The methods being modeled were well known to researchers in advance; none were discoveries. Although Haider and Frensch (2002) and Donner and Hardy (2015) highlight dips, neither discusses the role played by dips in method invention, testing, or revision.

4. Examples of plateaus, dips, and leaps in three tasks

4.1. Digit span and league-stepping methods

If this view of incrementing skilled performance 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.

4.1.1. SF: Stage 1. SF's initial set of goals lasted across the first 4 h of practice

4.1.1.1. Method invention: SF began by trying "to hold everything in a rehearsal buffer" but quickly adopted the approach 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" (Chase & Ericsson, 1982, p. 9).

Chase and Ericsson refer to the initial division of the task as "common strategies" and these 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).

4.1.2. SF: Stage 2. Meaningful units

4.1.2.1. Method invention: SF's fifth day was different as it was then when "he demonstrated the first rudimentary use of a retrieval structure"; namely, chunking three successive digits as a group while trying to hold the last 4–6 digits in his rehearsal buffer (Ericsson et al., 1980) Unfortunately, as each dot in Fig. 5 averages across 5 h of training (1 h per each day of training), any dip and subsequent leap from hour 4 to hour 5 is not visually apparent.

4.1.2.2. Method development and practice: SF practiced and improved his implementation of the chunking method to include the grouping of two successive groups of three 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 Fig. 5) by which time he was recalling 18 digits (remember that each dot on the figure represents the average score for 5 days of 1 h of practice per day).

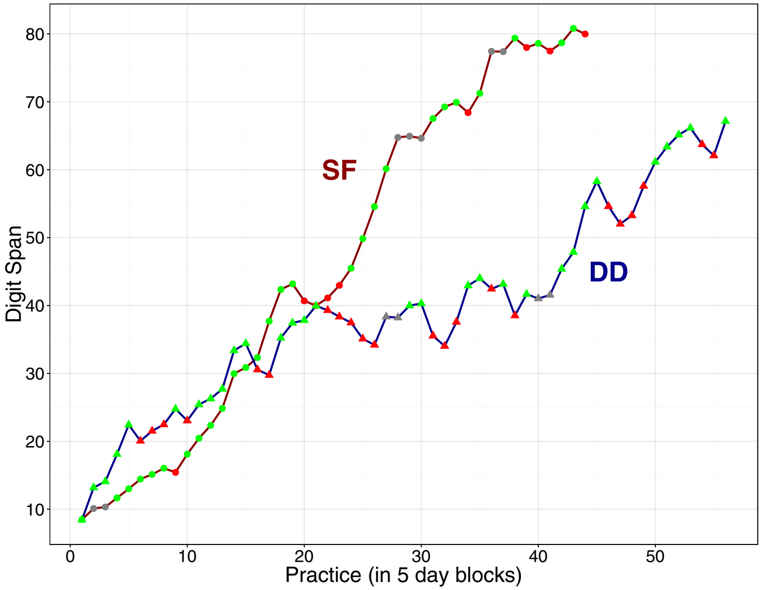

Fig. 5. Blocks of practice (each block represents 5 days of 1 h practice per day) for (a) SF (dark red line and circle points) and (b) DD (dark blue line and triangle points). The plot points are green, gray, or red and correspond to periods of rising, plateaued, or dipping performance. These three periods were calculated "by eye." A method for distinguishing between important plateaus, dips, and leaps is discussed in Section 4.3. However, the historic record from this era does not include the data we would need to apply our methods (discussed in Section 4.3) to these old data. (Data were extracted from Fig. 1 of Chase and Ericsson [1982] using the PlotDigitizer^{TM} software and should be considered as near approximations to the data in Chase and Ericsson's plots.)

4.1.3. SF: Stage 3. Supergroups

4.1.3.1. Method invention: The next advance entailed forming supergroups of two or three subgroups of three digits each. Following the output of the grouped items, he would then output the items from his rehearsal buffer.

4.1.3.2. Method development and practice: This method was perfected over the next 25 sessions (i.e., across 5 blocks in Fig. 5). The improvement seems to reflect both practice and adjustment in his methods as the number of chunks increased across these sessions. As Chase and Ericsson (1982) discuss, 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 "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 was not the case that a digit group was 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," and so on. Hence, the methods for grouping digits were being modified even as the methods for the higher level and base level grouping structures were being formulated and practiced.

4.1.4. 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 & Chase, 1982; Ericsson et al., 1980) Ericsson and Chase dwell more on the similarities than the differences between the two. However, a comparison of SF's and DD's performance in Fig. 5 shows their differences are striking. The differences that the reader will first notice are DD's faster initial ascent (blocks 0 to 19) followed by a dip from block 20–26, and a jagged period of leaps, plateaus, and dips which, by the end of the period plotted here, leaves him well below SF's level of skilled performance even though the period plotted for DD is 56 blocks and that for SF is 44 blocks. Of course, as mentioned earlier, DD did eventually exceed SF's performance and reached a digit span of 104!

Without more data, we can only speculate that DD's early ascent was aided by instruction in SF's method but beyond a certain point, the goal structure and particular methods he was trying to acquire became more and more specific to SF's prior knowledge and background and less easily adopted and adapted by DD. We will return to these speculations in Section 5.2.

4.2. *Breakout*™ *and league-stepping methods*

In *Pilgrim in the Microworld* (1983), the famed *ethnomethodologist* David Sudnow describes how he mastered the game *Breakout* (Sudnow, 1983; see Fig. 6).[11] 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.

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. We focus on his four unsuccessful attempts to develop effective methods for the goal *Where to Look*. Note that Sudnow does not present a timeline but does 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 h of practice playing Breakout.

4.2.1. *Where to look? Goal 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[12] much better and get some of the follow-ups as well. For about 20 min I sat there mesmerized, tracking the ball like my life depended on it, my entire being invested in the hypnotic pursuit

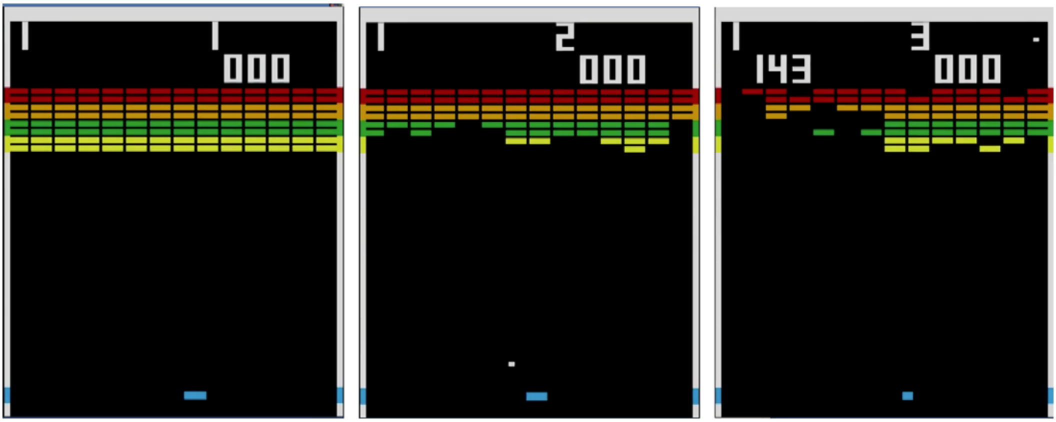

Fig. 6. Three Breakout screenshots. Blue line towards the bottom middle of each screen is the paddle which the player can move from side to side. The size of the paddle shrinks as the player's score increases (from left to right). The white dot in the middle screen is the "ball." Toward the top of the screen are eight layers of tiles (the barricade) (two layers for each of four colors). The rightmost screen shows a "breakout" through the tile barrier with the ball at the upper-right corner. The screenshots shown are from the movie at https://www.youtube.com/watch?v=hW7Sg5pXAok .

> 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)

4.2.2. *Interim*

Method 1 sufficed to solve Sudnow's problem in the early game. It also allowed him to play further than he had been able to play before. But it did not suffice to allow him to progress as far or as fast as he thought he should. Somewhat frustrated, he tells us:

> Just for kicks I covered the paddle path [the bottom of the screen, see Fig. 6] all across the bottom of the screen with an inch-thick strip of black tape. I tried playing blind, and could return only very few shots. I shortened the tape to leave a visible slot of two inches on each side of the screen, so when I was in the corners I could see the full paddle plus a bit. I swung back and forth again and again, end to end, trying to assess the gearing sensitively enough to field balls in the wide hidden area. When they came slowly, I could return about sixty percent of them, give or take a little. (pp. 46–47)

Sudnow is here describing *part-task learning* (Wickens, Hutchins, Carolan, & Cumming, 2013) to develop a method that would allow him to control the paddle without having to take his eye off the ball. Having, largely mastered this needed subskill, he goes on to develop his next goal.

4.2.3. Where to look? Goal 2

> I wondered if peripheral looking could do the job… [I] fixed my gaze right where the barricade (see Fig. 6) 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. (p. 47)

This worked for a while but…

> Then the slam. By the time my gaze could catch it, and then change over from the speed it ran to get there to the speed the ball was moving, it was all over. (p. 48)

Method failure, followed by method adjustment…

> So I'd started tracking very precisely in order to be most pointedly with it on the barricade at the instant of rebound. You don't stand still on the platform and lunge onto the train when steps come by. You make a running jump. (p. 48)

4.2.4. Plateau

After a considerable amount of time (we estimate at ≈ 50 h 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.

4.2.5. Where to look? Goal 3

Armed with a new goal, "pay attention to the paddle segments," he begins the post-Atari visit period by trying to perfect methods for ensuring that the ball strikes the paddle on exactly the "right" paddle segment. Of course, looking at the paddle negates the submethod he developed for positioning the paddle without constantly keeping his eyes on it (see Section 4.2.2).

After hours of trying to get these methods to work, he abandons this goal. 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)

4.2.6. Where to look? Goal 4

Having decided that he should not be looking at the paddle segments, he again spends many hours deciding "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 is where you had to look, and you had to look somewhere, could not 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."

4.2.7. Takeaways from Sudnow: Wrong method or wrong goal?

We count four major shifts in Sudnow's goals and methods for *where to look*. All goals emerged from his frustrations with how to track a fast-moving ball while also making sure that the paddle would be in the right place at the right time to volley the ball up to his planned target location. Each set of goals and methods were Sudnow's invention, with one—goal 3: making sure the ball hit a certain segment of the paddle—being spurred by his being told that the angle at which the ball bounced off the paddle was determined by which segment of the paddle the ball hit.

His goal changes appear to be coterminous with his attempts to develop methods for those goals and he seems to attribute his initial dips in performance to his failure to correctly implement the new method. However, what triggered his goal shifts?

Goal 1 solved a problem that he had in the early game. However, after mastering that part of the game he found himself stuck a little farther on. We score this as a failure of goal 1 to scale up to meet the greater demands of a more advanced game state.

The failure of goal 1 led him to experiment by taping over most of the paddle area—meaning that he could not see the paddle's location except at the extreme left and right sides of the screen (about one paddle's width on each side). In cognitive science terms, this experiment gave him the idea of disassociating visual attention from the point of fixation (Nakayama & MacKeben, 1989) so that the eye tracked the fastest moving object, the ball, with covert attention being used to monitor the paddle position. If this is what he did, it strikes us as an effortful method to implement for any period of time.[13] This new goal allowed him to make progress; however, another impasse (perhaps due to the effort involved in tracking two objects at once?) led him to seek out the Atari programmers who told him about the paddle areas.

Prior to his Atari visit, he must have had some control on the direction the ball went after it hit the paddle; that is, it seems likely that he had some implicit knowledge of the paddle zones. However, having explicit knowledge led him to develop his third "where to look" goal of focusing on the paddle to make sure the ball hit the "right" segment. He spent hours trying to develop a method that would make this goal work but, in the end, decides that the goal was fatally flawed.

His fourth “looking goal” required him to fixate 1/2 inch above the paddle. However, his simple statement of the goal omits several complexities. Keeping the fixation 1/2 inch above one moving target (the paddle) requires overriding the exogenous influence to look at the more rapidly moving ball. Hence, although he had found “a looking method,” we are not surprised that “it didn’t work.”

4.3. Changepoint analysis: Partially automating anomaly detection

The work we reviewed in the above two subsections is inspiring but depressing. Inspiring as it seems to validate our focus on PDLs as behavioral markers for changepoints in task performance. Depressing as the workflow seems more like hand-to-hand combat with nature to discover one or two secrets than it does our more typical workflow of designing a model, based on theory, and determining how well that model predicts the data and supports the theory. So the bad news is that wringing patterns out of data collected from individuals can be hard work. The good news is that it does not have to be as hard as it was in the days of SF and DD (Chase and Ericsson’s subjects) or as it was for us in wringing patterns out of Sudnow’s narrative account of his self-study.

4.3.1. A four-point approach to changepoint analyses

The tools we advocate are (a) a thorough and thoughtfully designed logging system that logs all system events, user events, and screen objects, (b) statistical software for changepoint analyses, and (c) multiple measures of performance and multiple features based on the log files of task performance. After these tools are in place, the least automated phase (and the analyst’s hard work) begins; namely, (d) looking for periods across task performance where some features and measures plateaued, others dipped, others leaped, and trying to interpret what is found.

The changepoint statistic is run on each measure and feature independently of the others (e.g., see the three analyses of three measures and features of Space Fortress data shown in Fig. 7). By our usage, if the changepoint statistics show that all features and measures decrease in the same time span, then the player probably fell asleep or was otherwise inattentive to the task. On the other hand, if some measures increase while others decrease, then something interesting may be going on and the analyst will want to study the varying features and measures to determine what that might be.

To be clear, the changepoint statistic is computed one measure or one feature at a time and it looks at trends over time in the performance of that one feature. The different features and measures may correlate with each other. For example, in the game Space Fortress, Total Score is generally correlated with the number of mines destroyed. However, Total Score might stay high even while the number of mines destroyed dips, if the mine dip came during a time in which the number of fortresses destroyed leaped. (A clear example of this is provided by Destefano and Gray [2016, Fig. 7] and is discussed next.)

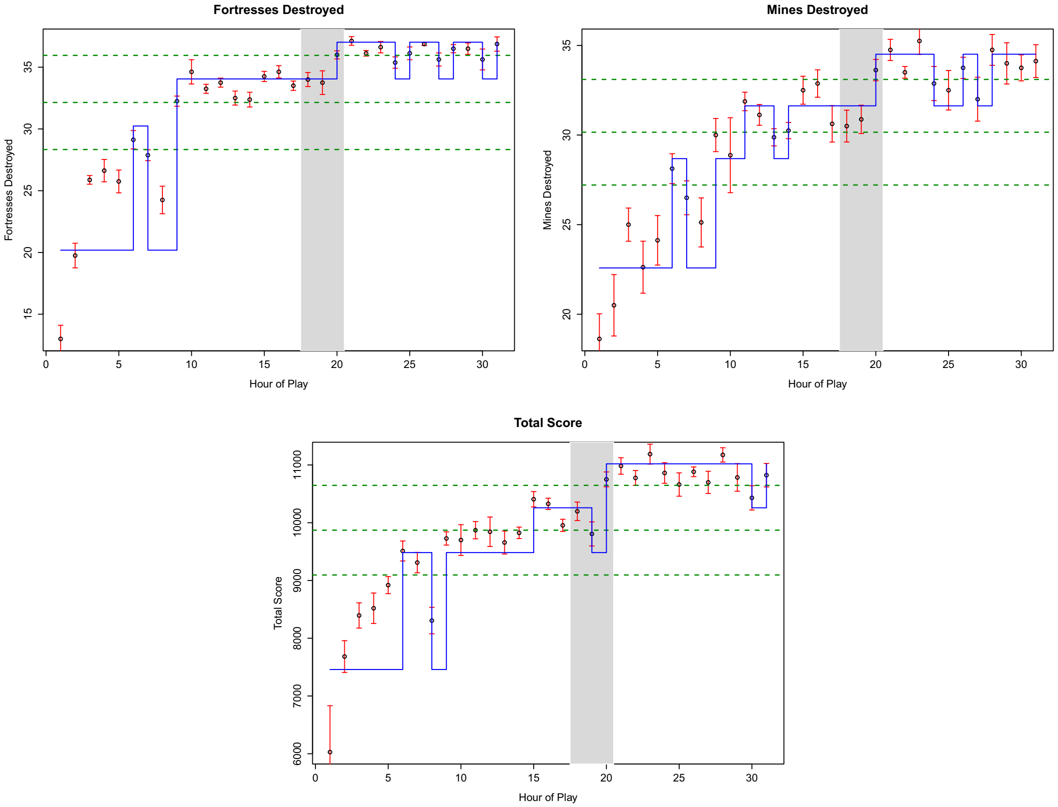

Fig. 7. A SAX changepoint analysis showing two features (Fortresses Destroyed and Mines Destroyed) and one measure (Total Score). All data are from Player 3534. The gray bars denote hours 19 and 20 of gameplay, circles with red bars are score and standard error (each hour is an average of eight games). Blue lines represent patterns of dips and leaps as found by the SAX analysis. (Figures are from Destefano and Gray, 2016.)

4.3.2 PDLs in Space Fortress

Our example is taken from Destefano and Gray (2016) and is based on data collected from one subject, playing the game Space Fortress (Destefano, 2010; Donchin, 1995) over 31 h long sessions. Five measures and 12 features were logged. This provided the data set. Several statistical methods exist for changepoint analyis; here we used the Symbolic Aggregate ApproXimation (SAX) method (Keogh, 2007; Lin, Keogh, Lonardi, & Chiu, 2003).

Although we have said nothing about the game of Space Fortress, the reader can easily spot some strange “goings on” with for player 3534 in the gray-bar area of Fig. 7, around hours 18–20 of game play. His Total Score, which he should be trying to maximize as that is the top goal of the game, clearly dips and leaps during this period (see the gray area), whereas the number of Fortresses and Mines he destroys each exit from a long plateau (more or less constant across the last 10 h for Fortresses Destroyed and for the last

7 h for Mines Destroyed) to a performance leap. As discussed in some detail by Destefano and Gray (2016), this player discovers an obscure unwritten rule of Space Fortress, one which even the programmer/experimenter was unaware. The story told by Destefano and Gray is that during this time period, the player discovered the obscure rule and established a new flight pattern which enabled him to take advantage of the rule. Clearly, without the detailed data and without the SAX changepoint analysis we would have never looked at hours 18–20 for this player, we would have never known about this rule, and we would never have appreciated that the dip in score signaled the period of exploration for a better flight pattern that, once it was established, enabled the score to leap.

4.3.3. Changepoint analyses

Destefano and Gray (2016) used SAX for their analyses, but many methods exist including several boutique methods created by cognitive researchers (e.g., Donner & Hardy, 2015; Gallistel, Fairhurst, & Balsam, 2004). A well-written survey of several methods used by financial analysts is provided by Kipnis (2015a). Other techniques are discussed by Killick and Eckley (2014) and Kipnis (2015b).[14]

A takeaway from this section is that our four-point approach to changepoint analyses does not guarantee that periods of method invention and implementation will be found or, if found, that the methods used will be interpretable by the researcher who is not also an expert in the domain being studied. The only thing we are sure about is that shining our light in the places identified by the four-point approach is better than searching under the lightpole.

4.4. Dips and Leaps mark the spots where research must focus

Plateaus, dips, and leaps are behavioral markers of activities crucial to developing expertise. However, although they tell us where to look for periods of change, they do not signal the nature of that change or even whether the change is due to efforts to develop and practice new methods or, rather, to a decline in motivation or an increase in fatigue.

We postulate that dips and leaps could signal one of several phenomena:

- Dip + Leap: Might signal the adoption of better goals and sound methods for implementing those goals. For these cases, the dips might signal an initial decline in performance as variations of new methods are tried on variations of the task; for example, different sequences of numbers for our digit span experts or different trade-offs between the use of overt and covert attention to the ball and paddle in Breakout. The leap would signal the use of a method that implements the goal.
- Dip + Return to Plateau: This is an ambiguous case as it might signal fatigue or loss of motivation followed by rest or the return of motivation. On the other hand, it might signal a period of exploration that does not result in a new and more successful method, the abandonment of that method, and a return to an earlier one.

- No Dip + Leap: The cost of implementing a new method may not be great enough to create a dip, especially if the new method is significantly less costly than the old. Such might be the case if the architect reported in the Fu and Gray study (see Section 2.2) were reminded of the much shorter (5 s) menu-based method at the point in time when he would otherwise have selected the longer (55 s) preferred method. This great time saving would show a performance leap without a performance dip. Of course, in Reinforcement Learning terms, the difficulty in selecting a known new goal would be in overcoming the much greater strength of the older goal at the point in time when the two are competing to be selected.

4.5. Marking the spot versus missing the point

There is, of course, a trade-off between averaging over data to extract a performance mean and looking at individual performance; that trade-off is illustrated in Fig. 8.

Each colored line, with colored points, plots the score for one of nine Space Fortress (Destefano, 2010; Donchin, 1995) players across 31 h of play (with each dot averaging the scores of eight games played in each hour). The black line with white dots plots the mean score across all nine players. Although it has some jiggle, the black line is well behaved and provides the illusion of a smooth and negatively accelerating increase in game performance. In contrast, this nice story is not supported by any of the individual lines. Instead, when we look closely, as we did in Section 4.3, we see players struggling with plateaus, dips, and leaps that, taken together, show progress, inventions, discoveries, and much hardwork but which by no means convey the impression of inevitable and smooth performance increases with practice.

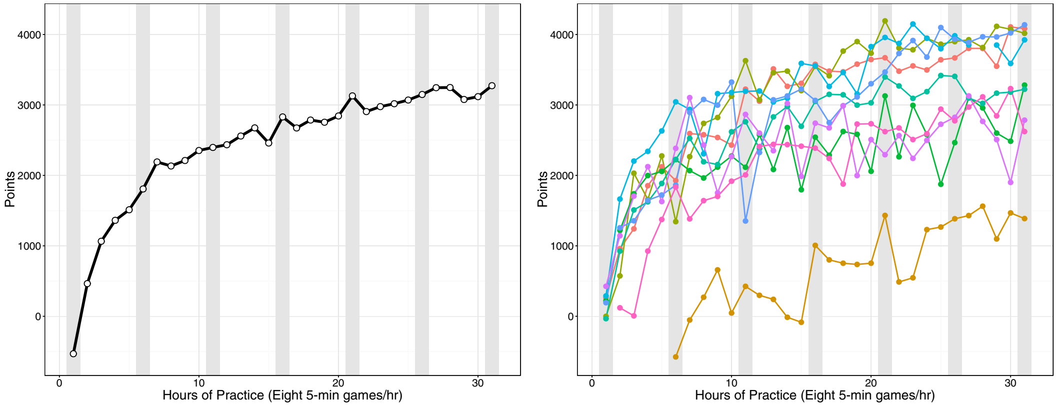

Fig. 8. Contrast between growth in performance averaged across all players (left plot—showing the black line with white dots) and variability in the growth of individual performance (right plot—colored lines with colored dots). Data come from nine people who played Space Fortress (Donchin, 1995) for 31 h (eight games per each hour, no more than 1 h per day). Note that to better highlight the plateaus, dips, and leaps in individual data, we truncated the range of the y-axis to -600. (From Destefano & Gray, 2016.)

5. Further applications: Relating plateaus, dips, and leaps to Brain Training and Deliberate Practice

In this section, we briefly mention and relate our work to two current strands of research; namely, (a) Brain Training (by experimental paradigms and by video games) and (b) Deliberate Practice.

Our intent is not to provide an exhaustive review of these topics but to use our discussion to highlight contributions that a focus on PDLs might make to these efforts.

5.1. Brain Training

We define *Brain Training* as the view that population norms in the performance of a variety of basic tasks, such as those used in experimental psychology or to measure intelligence, reflect 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. Various tasks including traditional research paradigms such as[15] nBack, AX-CPT, and so on, commercial products, and commercial video games have been cited as leading to general improvements in human performance.

5.1.1. Research focus of the Brain Training literature

Much research on Brain Training analyzes performance of the tasks being trained at a very high level. These studies take as their main focus transfer on various psychometric pre- and post-tests. Hence, *if people do training task A, then post-tests on the psychometric tasks are better than the pre-tests*. If people play first-person shooters a lot (or for at least 15 h), their scores on basic tests of memory and performance change (e.g., Anderson & Bavelier, 2011). If people do a certain type of nBack for 15 h, their scores on an IQ test might improve (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008) but more likely not (Redick et al., 2013).

5.1.2. Brain Training and plateaus, dips, and leaps

Although to our intellectual 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 famous dichotomy of Descartes), 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 vault example, these limits would reflect human limits 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 when pole vaulters switched from wood poles to bamboo or fiberglass ones.

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, then any higher level skills that require these low-level goals would be enhanced.

For the human mind, the distinction between a plateau or asymptote can be subtle as often changes in both capacity and efficiency are involved with one component being more dominant than the other. So acquiring the Fosbury Flop requires using the human body in a new way—this is a strategy shift—a change in efficiency in the use of pre-existing components. However, any strengthening of the muscles required to execute the Fosbury Flop or any increased flexibility in those muscles represents a change in capacity. For pole vaunting, the change in the composition of a pole is the major capacity shift (adopting the view that the unit of analysis is "human + equipment" as per Hutchins [1995]) and changes in movement (perhaps an extra "twist" which takes advantage of the new material's increased flexibility over the old while going over the top) would represent a change in efficiency.

5.1.3. Conclusions on Brain Training

Our reading of this voluminous literature[16] 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 is clear that very little, if any, of this work focuses on the performance of individuals, as advocated here. Another way of saying this is that the majority of research in this area focuses on outcomes, not processes. Until we understand the processes involved and whether and how they impact human asymptotes or human plateaus, the promise of brain training will not be fulfilled.

5.2. 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[17] and widely discussed by numerous researchers across many fields.[18] It is generally viewed as an activity which increments expertise in a well-defined domain.

To some degree, PDL may be viewed as a complement to DP. PDL provides a framework within which we can explore the acquisition of new goals and methods. DP provides a framework for structuring practice. Hence, it is easy to believe that the products of PDL might be applied to develop DP methods for new domains. However, although the example of the digit-span experts SF and DD (Chase & Ericsson, 1982) seems to support this possible application of our work, it also raises questions which need to be addressed before this view can be accepted.

Returning again to Fig. 5, we see that, after having been taught SF's methods, DD's rate of progress over the first five blocks (25 days) of practice was faster than SF's had been. However, from block 6 to block 15 this boost vanished so that by block 15 the two were equal in their digit spans despite DD's presumed advantage of having been taught SF's methods. At block 21, their learning curves cross with DD entering a performance plateau that lasts until around block 42. Although there are many reasons why a college student's attention might wander from the focus of becoming a digit span master, this result may provide a cautionary tale. It might be the case that *someone else's expert methods* will only take us so far as the process of discovering and constructing these methods may depend on using the skills and knowledge we already possess in new and different ways; that is, the discovery process for human skill acquisition may be, in part, one of discovering how to apply old skills and knowledge in new ways. If this was the case for DD, then it might be that his early boost from using SF's methods led him down a dead end as he was unable to extend those methods in the same way that SF extended them. By this story, DD would not have surpassed SF (as he eventually did) by bettering SF's methods, but by forking those methods early on to develop new methods based on his own prior experiences and skills.

5.3. *Summary of related areas*

In this section, we have reviewed work in two areas we see as closely aligned with our interests in League-Stepping Methods. The area of Brain Training focuses on outcomes, not processes. A focus on individual performers, via PDL, may prove practically as well as theoretically beneficial. The relationship between DP and PDL seems like a straightforward case of PDL developing domain-specific insights that can be used by DP to boost expertise. However, it also raises some deep questions regarding how much the discovery of new methods relies on knowledge that is specific to the individual discoverer versus how generally other people can take these discoveries and use them as their own.

6. Plateaus, dips, and leaps—The essential elements of league-stepping methods

> Something old, something new, something borrowed, something blue. (The four "somethings" from old English folklore, Wikipedia, 2015b)

We have talked about "somethings old." Specifically, we began our essay with 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 embraced a methodology that requires a detailed examination of individual data. Although always a minority report, this methodology has an old tradition within the young disciple of Cognitive Science (e.g., Anzai & Simon, 1979; Newell & Simon, 1972; Siegler, 1987; VanLehn, 1991). We quickly added in something less old, but still old, the concern with stable suboptimal performance that was 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 toward human factors engineering (e.g., Cockburn et al., 2014; Gray, John, & Atwood, 1993; Scarr, Cockburn, Gutwin, & Quinn, 2011). To put this as clearly as we can, *a plateau exists when there is a better goal (or better method for implementing an existing goal). An asymptote exists when the goal and its method(s) is optimal for the given task in the given task environment.*

Something else "new" was the recognition that dips are not simply a concern for their demotivating effect on learners (as per our discussion of Cockburn et al., 2014; Yechiam et al., 2003) but when viewed as part of the PDL triad they are the behavioral signal of periods of experimentation, discovery, trial and error, and successive approximations to developing League-Stepping Methods.

Likewise, as improbable as it may seem, despite 120 years of Experimental Psychology research on learning and transfer, looking at dips and leaps is something new.

Our "something borrowed" would be the power law of practice (Anderson, 1982, 1987; Fitts & Posner, 1967; Newell & Rosenbloom, 1981) and the importance of *a hierarchy of methods* as the key to task performance (Card et al., 1983; Newell, 1973; Newell & Simon, 1972; Simon, 1962). Indeed, two of our behavioral stages, the leap and the plateau, fit somewhat comfortably within revisionist (i.e., Donner & Hardy, 2015; Haider & Frensch, 2002) interpretations of practice. However, we reserve judgment on this point until we have a better understanding of the processes of method development which, at least sometimes, are marked by the dip.

"Something blue" pushes the metaphor a bit but might apply to bringing a focus on the PDL Triad to research in Brain Training and Deliberate Practice.

7. Summary and conclusions

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 goals, 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 (nee, habits) became fashionable following Bryan and Harter's (1897, 1899) work, enjoyed a revival in the 1980s and 1990s, and is overdue for another. As the first 120 years of experimental research on skill acquisition has shown, even when we tell people the goal and teach them the methods, they will often satisfice with stable suboptimal performance (Fu & Gray, 2004). Leaps which surpass suboptimal performance require motivation, effort, new ways of dividing or aggregating task performance, some combination of invention, instantiation, and refinement of methods, and practice.

Beyond the cognitive science challenge of understanding the process of goal 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 touchtyping, SF's going way beyond 7 ± 2, and Sudnow's dogged pursuit of expertise at Breakout, the link between 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 our review and framework are as follows:

- The difference between a plateau and an asymptote may be hard to discern—especially when there is no higher reference point. However, when trying to decide why performance has flat-lined, these categories lead us to ask two types of questions—is the limit due (1) to a failure of imagination/motivation/cost-benefit considerations (potential gains not worth the effort) or (2) to limits of technology/the design of interactive systems/the nature of the brain? Exploring these distinctions may lead us to a training (plateau) or a design (asymptote) solution.
- Rather than stable plateaus, SF's and DD's plots (as well as the individual data shown in Fig. 8) show dips that might be attributable to an ongoing process of abandoning old goals and methods to work out the kinks in new ones. Sudnow's self-reports suggest that most of his experiments with methods resulted in, at least initially, poorer performance.
- Not every hour of practice is equal. Some focus on inventing new goals, others on implementing different methods, others on practicing methods until the return on this effort seems minimal, while others slip away in daydreams. Hence, hour counts of practice will always be an imperfect predictor of hour contents.
- Although plateaus have been known for over 100 years, dips and leaps have not. Dips and leaps are behavioral markers of possible periods of exploration and change. Identifying such markers requires a focus on (a) microdynamics, (b) a paradigm shift to focus on individual, not group, performance, and (c) acquisition of new analytic and statistic tools to speed the process of identifying and classifying periods of change (Destefano & Gray, 2016; Gray & Destefano, 2016).

- Long-term studies of skill acquisition are required if we are to pursue the microdynamics of goal discovery and method invention. Fortunately, inside the laboratory, whatever the relative merits of the brain capacity versus brain efficiency debates, the surge of interest in long-term Brain Training studies and in long-term video game studies seems ideal for capturing detailed longitudinal data (Gray, in preparation). Outside the laboratory, Big Data, based on harvesting and analyzing commercial brain training data (as shown by Donner & Hardy, 2015) and online gaming records (as shown by Huang, Zimmermann, Nagapan, Harrison, & Phillips, 2013), has already yielded clear evidence of plateaus in individual performance.

In conclusion, the markers provided by plateaus, dips, and leaps provide a behavioral framework within which to build theories of the acquisition of expertise. For the researcher, dips become the marker of transitional phases of performance and this paper may be viewed as a call to study the *microdynamics of performance change for individuals*. The hierarchy of methods developed during these dips and leaps are the key to acquiring League-Stepping Methods.

One hundred years ago, researchers interested in goal discovery, invention, and implementation dropped their keys and have been looking for them ever since under the lamp of averaging-across subjects. It is now time to stop looking for them there and to bring our lamps over to where we dropped them, over here, in one of these plateaus, dips, or leaps.

Acknowledgments

This work was supported, in part, by grant N000141310252 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 Rick Cooper, Anders Ericsson, Christian Fiebach, Zach Hambrick, Ray Perez, Stephen Reed, Jim Spohrer, Niels Taatgen, and several anonymous reviewers 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) with whom we read and discussed many of the papers cited in this work.

Notes

1. A "league" is an old English measurement corresponding to about 3 miles (4.83 km).
2. A "furlong" is an old English measurement corresponding to about 220 yards (201.2 m).

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.
5. Although this concern seems odd today, in the late '80s through the mid-90's the gains in productivity from computers lagged the expectation of gains. For an extensive discussion of this failure, from a mid-90s perspective, see Landauer (1995).
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.
7. To those of us who pay attention to the sport once every 4 (or more) years, pole vaunting may seem like a simple activity with few methods or subgoals. In constrast to this view, McGinnis (2007) provides a task analysis that divides pole vaulting into one category of personal characteristics of "The Vaulter," one category of how to handle the equipment, and five sequential subgoals during the pole vault itself. Each of these catergories or subgoals is in turn divided into 1 to 7 sub-subgoals with a short descriptive method provided for each.
8. Also known as *hunt and peck*, *search and peck*, or *Eagle Finger* typists.
9. The two students who became their digit span experts.
10. This distinction is often referred to as idiographic versus nomothetic. By those names, it has a long and contentious history in psychology. For historic as well as recent sources, see Barlow and Nock (2009), Lamiell (1998), Münsterberg (1899), Roberts and Newton (2001), Underwood (1975), and Vogel and Awh (2008).
11. Thanks to Stuart Reeves for pointing us toward this book and Sudnow's other work.
12. There is no human opponent in Breakout. However, the game is capable of returning a ball either slowly or fast. What Sudnow calls a "slam" would be a very fast return.
13. It would be interesting to verify the human ability to make such a dissociation between two moving objects over a prolonged time interval.
14. An introductory survey of changepoint methods for psychological researchers was provided by Gray and Destefano (2016). Slides for that talk will be available at http://www.rpi.edu/~grayw/pubs/.
15. nBack and AX-CPT are memory tasks often used to study control of cognitive processing.
16. An earlier version of this section contained a long list of citations; however, while still being incomplete, that list was longer than this section. Hence, we leave it to the interested reader to dig into this literature.
17. As a very informal survey, on 2016.05.30 we did nine 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.

18. Macnamara, Hambrick, and Oswald (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 30, 2016 found 2,013 citations. Clearly, both of these are very impressive numbers.

References

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, 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). Charlotte, NC: Information Age Publishing. ISBN 978-1-61735-410-6 (e-book).

Anglim, J., & Wynton, S. K. A. (2015). Hierarchical bayesian models of subtask learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 957–974.

Anzai, Y., & Simon, H. A. (1979). Theory of learning by doing. *Psychological Review*, *86*(2), 124–140. doi:10.1037//0033-295X.86.2.124

Barlow, D. H., & Nock, M. K. (2009). Why can't we be more idiographic in our research? *Perspectives on Psychological Science, 4*(1), 19–21. doi:10.1111/j.1745-6924.2009.01088.x

Billman, D., & Shaman, D. (1990). Strategy knowledge and strategy change in skilled performance: a study of the game othello. *American Journal of Psychology*, *103*(2), 145–166. doi:10.2307/1423140

Book, W. F. (1908). *The psychology of skill: with special reference to its acquisition in typewriting*. Missoula, MT: University of Montana Publications in Psychology.

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

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

Carroll, J. M., & Rosson, M. B. (1987). *Paradox of the active user*. J. M. Carroll (Ed.). 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

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. PhD thesis, Rensselaer Polytechnic Institute, Troy, NY.

Destefano, M., & Gray, W. D. (2016). Where should researchers look for strategy discoveries during the acquisition of complex task performance? The case of space fortress. In Papafragou, A., Grodner, D., Mirman, D., and Trueswell, J. C., (eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 668–673). Austin, TX. Cognitive Science Society.

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. (1995). Video games as research tools: The Space Fortress game. *Behavior Research Methods, Instruments, & Computers*, *27*(2), 217–223.

Donner, Y., & Hardy, J. L. (2015). Piecewise power laws in individual learning curves. *Psychonomic Bulletin & Review*, *22*(5), 1308–1319. doi:10.3758/s13423-015-0811-x

Eden, S. (2015). *Stroke of madness*. Available at http://espn.go.com/golf/story/_/id/8865487/tiger-woods-reinvents-golf-swingthird-career-espn-magazine. Accessed October 28, 2015.

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

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

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.

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

Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(36), 13124–13131. doi:10.1097/pnas.0404965101

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

Gray, W. D. (in preparation). Games-XP: Action games as cognitive science paradigm. *Topics in Cognitive Science*.

Gray, W. D., & Destefano, M. (2016). Searching not under the lightpole but where we dropped our keys: Using Changepoint Detection to shine the light on periods of strategy invention and change. In Paper presented at the 57th Annual Meeting of the Psychonomic Society.

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

Haider, H., & Frensch, P. (2002). Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997,1999), Delaney et al. (1998), and Palmeri (1999). *Journal of Experimental Psychology-Learning Memory and Cognition*, *28*(2), 392–406. doi:10.1037//0278-7393.28.2.392

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. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 695–704). New York: ACM. doi:10.1145/2470654.2470753

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, *19*(3), 265–288. doi:10.1207/s15516709cog1903_1

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.

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

Keogh, E. (2007). Mining shape and time series databases with symbolic representations. [accessed August 12, 2016]. Retrieved from http://www.cs.ucr.edu/~eamonn/SAX.htm

Killick, R., & Eckley, I. A. (2014). changepoint: An R Package for Changepoint Analysis. *Journal of Statistical Software*, *58*(3), 1–19.

Kipnis, I. (2015a). An introduction to change points (packages: ecp and BreakoutDetection) [accessed January 1, 2016]. Retrieved from https://quantstrattrader.wordpress.com/2015/01/21/an-introduction-to-change-points-packages-ecp-and-breakoutdetection/

Kipnis, I. (2015b). *Pelting a competing changepoint algorithm*. [accessed January 1, 2016]. Retrieved from https://quantstrattrader.wordpress.com/2015/02/09/pelting-a-competing-changepoint-algorithm/

Lamiell, J. T. (1998). “Nomothetic” and “idiographic.” *Theory & Psychology*, *8*(1), 23–38. doi:10.1177/0959354398081002

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

Lin, J., Keogh, E., Lonardi, S., & Chiu, B. (2003). A symbolic representation of time series, with implications for streaming algorithms. *Proceedings of the 8th ACM SIGMOD workshop on research issues in data mining and knowledge discovery* (pp. 2–11). DMKD ’03. San Diego, CA: ACM.

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

McGinnis, P. M. (2007). Mechanics of the pole vault: Mechanical bases of effective pole vaulting technique [accessed October 4, 2015]. Available at http://www.usatf.org/groups/coaches/library/2007/Pole%20Vault%20Training/2007NPEP-McGinnis.pdf.

Münsterberg, H. (1899). Psychology and history. *Psychological Review*, *6*(1), 1–31. Retrieved from http://search.ebscohost.com.libproxy.rpi.edu/login.aspx?direct=true&db=psyh&AN=1926-02912-001&site=ehost-live&scope=site

Nakayama, K., & MacKeben, M. (1989). Sustained and transient components of focal visual-attention. *Vision Research*, *29*(11), 1631–1647. doi:10.1016/0042-6989(89)90144-2

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., & 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). New York: ACM. doi:10.1145/169059.169126

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*(2), 359–379. 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

Rickard, T. C. (1999). A CMPL alternative account of practice effects in numerosity judgment tasks. *Journal of Experimental Psychology-Learning Memory and Cognition*, *25*(2), 532–542. doi:10.1037//0278-7393.25.2.532

Roberts, M. J., & Newton, E. J. (2001). Understanding strategy selection. *International Journal of Human-Computer Studies*, *54*(1), 137–154.

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. New York: ACM. doi:10.1145/1978942.1979348

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*(3), 250–264.

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.

Sudnow, D. (1983). *Pilgrim in the microworld*. New York: 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

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

Tenison, C., & Anderson, J. R. (2016). Modeling the distinct phases of skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(5), 749–767. doi:10.1037/xlm0000204

Thorndike, E. L. (1913). *Educational psychology. volume II: The psychology of learning*. New York: Teachers College, Columbia University.

Underwood, B. J. (1975). Individual-differences as a crucible in theory construction. *American Psychologist*, *30*(2), 128–134.

VanLehn, K. (1991). Rule acquisition events in the discovery of problem-solving strategies. *Cognitive Science*, *15*, 1–47.

Vogel, E. K., & Awh, E. (2008). How to exploit diversity for scientific gain: Using individual differences to constrain cognitive theory. *Current Directions in Psychological Science*, *17*(2), 171–176.

Wickens, C. D., Hutchins, S., Carolan, T., & Cumming, J. (2013). Effectiveness of part-task training and increasing-difficulty training strategies: A meta-analysis approach. *Human Factors*, *55*(2), 461–470. doi:10.1177/0018720812451994

Wikipedia. (2015a). Frank Edward McGurrin—Wikipedia, the free encyclopedia [accessed January 9, 2015]. Available at http://en.wikipedia.org/w/index.php?title=Frank_Edward_McGurrin&oldid=626496435

Wikipedia. (2015b). Something old—Wikipedia, the free encyclopedia [accessed April 30, 2015]. Available at http://en.wikipedia.org/w/index.php?title=Something_old&oldid=656686892

Wikipedia. (2016). Touch typing—Wikipedia, the free encyclopedia [accessed April 30, 2015]. Available at https://en.wikipedia.org/wiki/Touch_typing. Accessed February, 22, 2016.

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