Acta Psychologica 151 (2014) 16–23

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

Learning curves in highly skilled chess players: A test of the generality of the power law of practice

Robert W. Howard *

University of New South Wales, Australia

ARTICLE INFO

Article history:
Received 31 January 2014
Received in revised form 13 May 2014
Accepted 19 May 2014
Available online xxxx

PsycINFO classification:
2300
2340
2343

Keywords:
Learning curve
Power law of practice
Expertise
Skill development
Chess

ABSTRACT

The power law of practice holds that a power function best interrelates skill performance and amount of practice. However, the law's validity and generality are moot. Some researchers argue that it is an artifact of averaging individual exponential curves while others question whether the law generalizes to complex skills and to performance measures other than response time. The present study tested the power law's generality to development over many years of a very complex cognitive skill, chess playing, with 387 skilled participants, most of whom were grandmasters. A power or logarithmic function best fit grouped data but individuals showed much variability. An exponential function usually was the worst fit to individual data. Groups differing in chess talent were compared and a power function best fit the group curve for the more talented players while a quadratic function best fit that for the less talented. After extreme amounts of practice, a logarithmic function best fit grouped data but a quadratic function best fit most individual curves. Individual variability is great and the power law or an exponential law are not the best descriptions of individual chess skill development.

1. Introduction

Extensive practice can greatly improve skilled performance and a key concern in several fields is the form of the resulting learning curve. What function best interrelates performance and extent of practice? One view is that a power function fits best, whereby performance improves the most early in learning, begins to level off, and approaches asymptote. This view is summarized as the mathematical "power law of practice" (Logan, 1988; Newell & Rosenbloom, 1981).

The power law apparently describes well the development of many skills (Lacroix & Cousineau, 2006). Indeed, Ritter and Schooler (2001, P. 8602) stated "The power law of practice is ubiquitous. From short perceptual tasks to team-based longer term tasks of building ships, the breadth and length of human behavior, the rate that people improve with practice appears to follow a similar pattern".

But there are dissenters. First, Heathcote, Brown, and Mewhort (2000) argued that the power law is an artifact of data averaging, that exponential functions usually best fit individual learning curves, and that averaging many exponential curves together tends to yield a power function (see also Anderson, 2001; Murre & Chessa, 2011).

Second, the law's generality is moot. Supporting studies mostly use fairly simple skills, one or a few training sessions, and time to perform a task as the only performance measure. Data points may involve just one response to a stimulus or an average over only a few trials. With simple skills, the need to deal with novelty and complexity may diminish with practice as performance becomes automated. Principles derived from studies of simple skills may not generalize well to complex skills (Wulf & Shea, 2002). Does the power law apply to very complex cognitive skills developed over years rather than over just a session or two? Relevant longitudinal data are scarce (Ohlsson, 1992). However, Ohlsson (1992) tested the power law's generality to time to write books. A power function well fitted Isaac Asimov's total output of around 500 books over time. But this interesting finding involved just one exceptional author. Would similar results apply to output of other prolific authors? Would a power function best fit development of other very complex cognitive skills?

The issue of the power law's generality is important. Psychological laws are few and many researchers assume that the power law is correct and general. Indeed, the power law was the ninth most cited psychological law in 1990–1999 (Teigen, 2002). For instance, Ericsson, Krampe, and Tesch-Romer's (1993) Expert Performance Theory holds that "One premise of our theoretical framework is that performance increases monotonically according to the power law" (P. 384). Ritter and Schooler (2001) argue that every skill learning theory must explain the apparently ubiquitous power law, although each may do so in

* School of Education, University of New South Wales, Sydney 2052, Australia. Tel.: +61 2 9385 1954; fax: +61 2 9385 1946.
E-mail address: rwh77@outlook.com.

http://dx.doi.org/10.1016/j.actpsy.2014.05.013
0001-6918/

different ways. For instance, ACT-R theory does so by assuming that rules and memory traces are strengthened because the cognitive system is adapted to the statistical structure of the environment (Anderson & Schooler, 1991). Ohlsson's (1996) model holds that learning events are triggered by errors, that the rate of change is a function of the number of errors detected, and that the learner makes fewer errors with increasing expertise.

Another reason for interest is application. Learning curves are a major concern in training and in ergonomics (Anzanello & Fogliatto, 2011). Managers and trainers wish to know the rates at which performance may improve and what individual differences may affect skill development.

Chess playing is a complex cognitive skill which develops over decades. Play involves many component skills and some component skills may become largely automated. But the domain's great complexity means that much information-processing usually is needed during play. At top levels, players must cope with novelty and exhaustively analyze complicated positions. Organized chess also has extensive longitudinal population-level data involving an objective performance measure, a numerical performance rating (Elo, 1978) quite different from the task time measure that power law studies normally use.

The present study had two aims. The first was to test the generality of the power law to the development of chess skill. International chess has many participants and whether good fits to a power function in grouped data emerge from averaging individual exponential curves can be tested with a large sample, although with a much more complex skill than those that power law studies typically use. Furthermore, some but not all studies with other skills have compared fits only of power and exponential functions. Other relationships are possible and even might be likely in some circumstances, such as after extended practice. For example, a quadratic function may eventually fit better, where learners reach asymptote and then show declining performance perhaps because of decreasing motivation or other factors. Fits of power, logarithmic, quadratic and exponential curves were tested here.

The second aim was to examine effects of two variables on the form of the best fitting function; extent of innate talent and extreme amount of practice. Talent and performance level may interact in complex ways and may produce different learning curve shapes. Perhaps different functions best fit learning curves of participants with differing degrees of talent.

1.1. FIDE (the international chess federation) ratings

Elo (1978) and Howard (2006) describe FIDE ratings in detail. FIDE rates skill on a scale from 1200 to about 3000. The highest-ever FIDE rating is just below 2900. A rating may change after each game according to result and the opponent's rating and thus gauges skill at a given time. For example, if a player rated 2300 beats one rated 2400, her rating rises some points while her opponent's rating falls some points. Defeating a much higher rated player adds more points and defeating a much lower rated player adds relatively few points. Wins against higher rated players increase a rating progressively while losses to lower rated players decrease it. A rating is considered reliable after 25 initial games.

The first official FIDE rating list appeared in 1970. FIDE tallies numbers of games in a rating period only from July 1985. Rules to get on and stay on the rating list have changed. Females once needed a minimum rating of 2000. Males once needed a minimum 2200, but from 1993 FIDE dropped the minimum rating periodically to a very modest 1200 in 2012. The ratings in a given list are approximately normally distributed. The mean rating in January 2012 was 1956.66 ($SD = 239.05$, $N = 139{,}000$). Howard (2006) placed FIDE data into the database used here and which is available from the author. The rating system is not perfect but measures well skill at a given time, and expertise researchers use chess data extensively (Charness, 1992).

1.2. Defining practice in international chess

Psychologists traditionally defined practice as actually performing a task, such as batting a ball (Smyth, 1975). Howard (2009) dubbed this sense "actual practice". Studies testing the power law typically use this definition. The skill examined in the present study is playing FIDE rated games and so "practice" in this traditional sense is playing FIDE rated games. There also is "deliberate practice" which Ericsson and Ward (2007) defined as "a (typically planned) training activity aimed at reaching a level just beyond the currently attainable level of performance by engaging in full concentration, analysis after feedback, and repetitions with refinement". Various activities may constitute deliberate practice for chess players, such as playing over grandmaster games and studying opening manuals. However, playing rated games appears to be the most important factor in chess skill development (Howard, 2012, 2013). Indeed, it's a well-known principle that learning mostly occurs on the job; e.g. while actually playing rated games here. But other activities such as chess study and playing unrated games also may affect chess skill. However, such activities could not possibly be measured over the years covered in the present study and so practice was given only the traditional definition.

1.3. Forms of mathematical model for international chess performance

The power law has various mathematical forms. The simplest and most oft-used is;

$$T = a\ P^{b} \tag{1}$$

where T is the performance measure (normally being time to perform the task), P is a practice measure (number of trials or time spent practicing) and a and b are fitted parameters. Parameter a is the curve's starting point and b is the skill improvement rate. Other forms applied to time-based measures of skill have additional parameters which can be difficult to fit. A three parameter version is;

$$T = K + a\ P^{b} \tag{2}$$

where K is asymptotic performance level. Participants in studies supporting the power law often are not trained to asymptote and this parameter often represents a physical limit beyond which a participant cannot go (Heathcote et al., 2000). If a participant physically cannot perform a task faster than in three seconds, for example, adding this value to the remaining term (which involves task performance time) makes sense.

A four parameter version, the "full" power law, adds a parameter E for amount of relevant prior experience before training;

$$T = K + a(P + E)^{b} \tag{3}$$

Which power law model should be used for FIDE ratings? The simplest form (Eq. (1)) was appropriate here for several reasons. A key principle of curve-fitting is that fitted parameters must be scientifically plausible (Motulsky & Christopoulos, 2003). The asymptote parameter K makes sense with task time as the performance measure but does not make sense with FIDE ratings. First, all participants in the present study were examined up to performance asymptote anyway (although the exact asymptotic rating varied across participants) and so an asymptote parameter was redundant. As noted, many studies do not take participants to asymptote and K is added to estimate asymptotic values. Second, when task time is the performance measure, an asymptote parameter represents a physical limit and the value can be added usefully to task performance time. But the arithmetic does not work the same way with FIDE ratings. One cannot usefully add a rating value to the remaining term. The top rating on record is around 2900 and if K is constrained to be no more than 2900 the remaining term