Routine Procedural Recipes for Rapid Learning in Choice-Reaction Tasks

Steven C. Lacey, Adam Krawitz, Jonathon J. Kopecky, David E. Kieras, & David E. Meyer
University of Michigan, Department of Psychology, Ann Arbor, Michigan, USA

Sponsored by U.S. Office of Naval Research
Grant N00014-03-1-0009

Abstract

Exactly how do people become skilled at basic choice-reaction tasks? To help answer this question, we had participants learn to perform such tasks under controlled study and test conditions with various typical mappings between visual stimuli and manual responses. In several of these cases, participants mastered individual S-R pairs very rapidly, achieving essentially asymptotic response speed and accuracy within one or two trial blocks after ten or fewer trials per pair. Frequently, their individual RT learning curves for particular S-R pairs deviated significantly from the standard power law of practice; instead, they were low and flat or had ultra-steep exponential descents to an asymptotic floor level. Our results suggest that skill acquisition in choice-reaction tasks relies on either of two strategic processes: (1) pre-established "routine procedural recipes" that efficiently interpret new declarative knowledge based on verbal task instructions; or (2) gradual creation and application of compiled procedural knowledge based on task-specific "hard-coded" production rules.

Goals of Research

1. Characterize perceptual-motor and cognitive skill acquisition in basic choice-reaction tasks
   - Describe roles of declarative and procedural knowledge precisely
   - Isolate acquisition of particular production rules for response selection
2. Understand individual differences in skill-acquisition ability
3. Formulate principles for optimal training of skilled performance
4. Generalize to complex real-world task conditions

Historical Background

Studies of skill acquisition in many tasks over past 100 years

- Sending and receiving Morse Code (Bryan & Harter, 1899)
- Tracing mirror patterns (Snoddy, 1926)
- Performing choice-reactions for spatially incompatible S-R mappings (Fitts & Seeger, 1953)
- Making cigars (Crossman, 1959)
- Searching for printed target letters (Neisser et al., 1963)
- Operating computer mice and step keys (Card et al., 1978)
- Reasoning about geometry proofs (Neves & Anderson, 1981)
- Playing the game of stair (Newell & Rosenbloom, 1981)
- Performing choice-reactions for neutral letter-finger S-R mappings (Pashler & Baylis, 1991)

Major findings and conclusions

- Skill acquisition occurs in multiple phases with intermediate plateaus
- Principal locus of improvement is in stimulus-response translation
- Power law of practice prevails pervasively for mean RT

Power Law of Practice for Mean RT

$t_n = t_\infty + (t_1 - t_\infty) \times n^{-p}$

- n is the trial block number
- t_1 is the initial mean RT on trial block 1
- t_∞ is the asymptotic mean RT as n increases
- p is an exponent that influences the rate of learning
- Absolute learning rate depends on p and $(t_1 - t_\infty)$
- Relative (normalized) learning rate decreases as n increases

Typical Parameter Values & Goodness of Fit

- t_∞ > 200ms
- 0.1 < p < 0.5
- R^2 > 0.9

Theories of Skill Acquisition

Several alternative theories, but each assumes that skill acquisition always involves a gradual transition between an initial inefficient mode of performance and a final efficient mode of performance based on specific task knowledge in long-term memory

- **ACT (Anderson et al., 1982, 1990, 1998, 2004)**

  Learning depends on compilation of "hard-coded" production rules that are optimized through processes such as rule composition and rule strengthening

  Interpretative Processing (Declarative Knowledge) → Compiled Processing (Procedural Knowledge)

- **SOAR (Newell et al., 1981, 1987, 1990)**

  Learning depends on impasse-driven creation of new production rules, which allow problem solutions to be recalled rather than deduced on subsequent occasions

  General Problem Solving (Slow & Controlled) → Routine Behavior (fast & automatic)

- **Multiple Copy Model (Logan, 1988, 1992, 1995)**

  Learning depends on cloning multiple copies of S-R associations

  Sequential Reasoning Algorithms → Automatic S-R Associations

Each theory predicts relatively shallow learning curves that conform to the power law of practice for mean RT

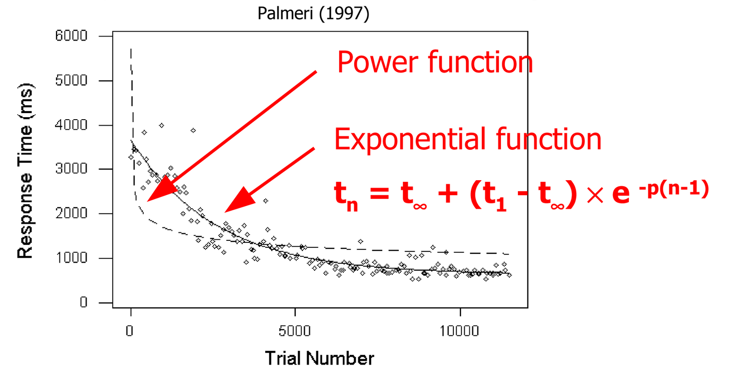
Limitations in Past Empirical Studies of Skill Acquisition

- Weak control over initial presentation of verbal task instructions and acquisition of declarative knowledge
- Few detailed analyses of practice effects at levels of individual S-R pairs and participants
- Artifacts caused by averaging RT data across various cases involving different families of learning curves

Power Law Repealed in Favor of Exponential Functions

- Heathcote et al. (2000) fit power and exponential functions to over 7,000 learning series from 475 subjects in 24 experiments
- Exponential functions fit **unaveraged** data better than power functions did
- Exponential functions have a relative learning rate that is constant (i.e., does not decrease) as n increases, unlike for power functions

Participant 2's Performance of Counting Task



New Theoretical Hypotheses & Predictions

Skill acquisition involves Routine Procedural Recipes (RPRs; cf. Kieras & Bovair, 1986, 1991)

- RPRs are sets of general production rules (i.e., if condition, then action) for performing generic categories of tasks
- Rule conditions are matched against declarative knowledge in working memory or long-term memory
- Rule conditions and actions contain variables rather than specific hard-coded exemplars
- RPRs evolve from extensive experience with generically similar task situations (e.g., human-computer interaction)
- RPRs enable extremely rapid skill acquisition with essentially flat or ultra steep RT learning curves
- Compiling and strengthening specific hard-coded production rules may be an optional strategic process; shallow learning curves that conform to the power law of practice for mean RT may NOT be obligatory

Methods

Choice-Reaction Tasks

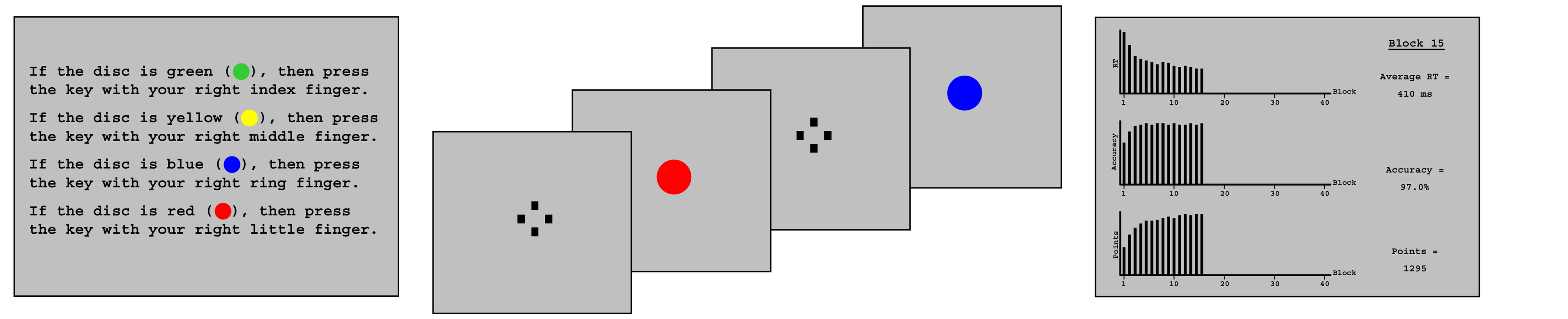
Five visual-manual stimulus-response mappings

S-R Mapping	Right Hand Response Fingers: Index	Middle	Ring	Little
Colored discs	green	yellow	blue	red
Geometric shapes	■	✚	▲	★
Letter	Q	J	X	B
Compatible spatial Locations	0---	-0--	--0-	---0
Incompatible spatial Locations	-0--	---0	0---	--0-

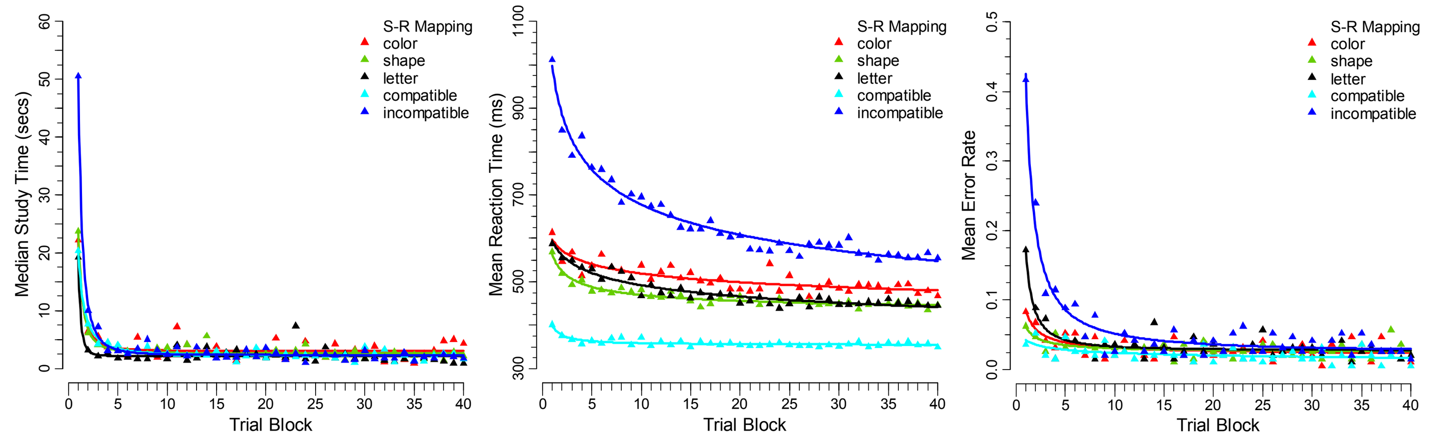
Procedural Details

- 1 Session per participant
- 40 Trial blocks per session
- 24 Trials per block
- 960 Trials per participant
- 8 Participants per S-R mapping
- Extensive feedback within and between trial blocks
  - Trial by trial auditory accuracy feedback
  - Re-display of instructions after each error
  - Graphical feedback at end of each trial block

Trial Blocks Divided Into Study, Test, and Feedback Phases



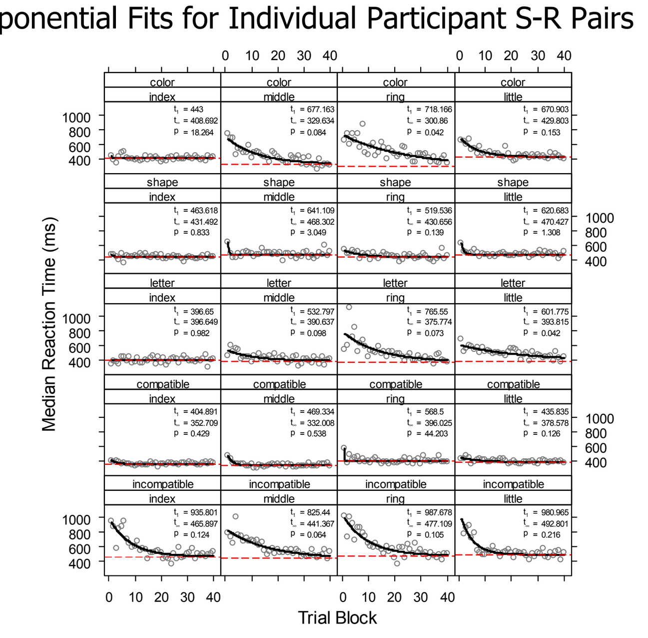
Average Learning Curves & Fitted Power Functions



RT Learning Curves for Individual Participants & S-R Pairs

- Fit exponential and power functions to median RT data for individual participant S-R pairs across trial blocks
- Exponential functions fit better than power functions in **75%** of the cases
- Results support Heathcote et al.'s (2000) claim that the power law of practice is an artifact of averaging
- Observed forms of the RT learning curves vary significantly for different combinations of S-R mappings, response fingers, and participants
- Some RT learning curves are low and flat or very steep while others are shallow and descend more gradually
- Results suggest there are multiple optional acquisition strategies rather than a single unitary obligatory learning mechanism

Exponential Fits for Individual Participant S-R Pairs

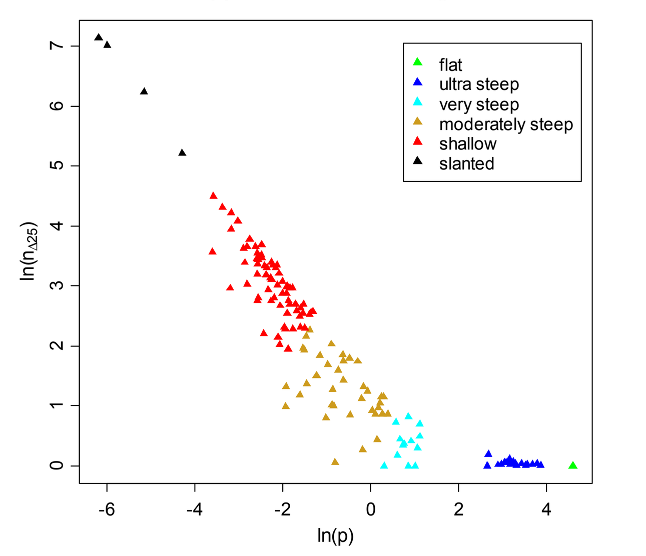


Cluster Analysis of Exponential RT Learning Curves

Statistical Procedure

- Fit exponential functions to median RTs for each combination of S-R mapping, response finger, and participant across trial blocks
- Transformed $n_{\Delta25}$ and p logarithmically
  - $n_{\Delta25}$ is the number of trial blocks needed to get within 25 ms of asymptote (t_∞)
  - p is the exponent of the function
- Submitted $\ln(n_{\Delta25})$ and ln(p) to hierarchical Partitioning Around Medoids (PAM) clustering
- Clusters identified with respect to silhouette width & proportion of variance accounted for

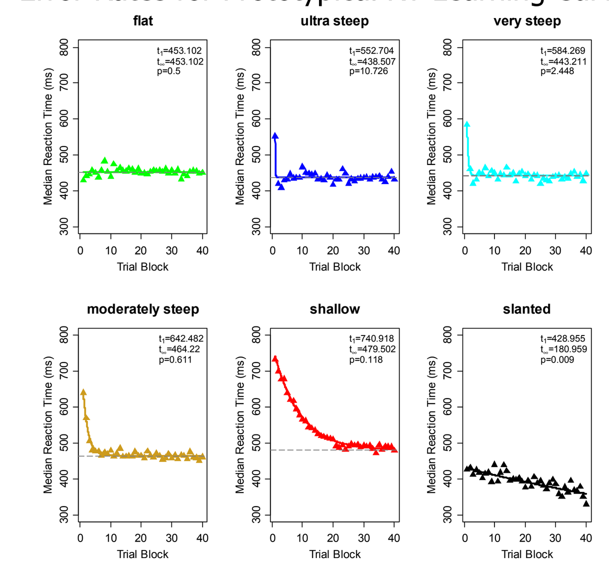
Prototypical RT Learning Curves



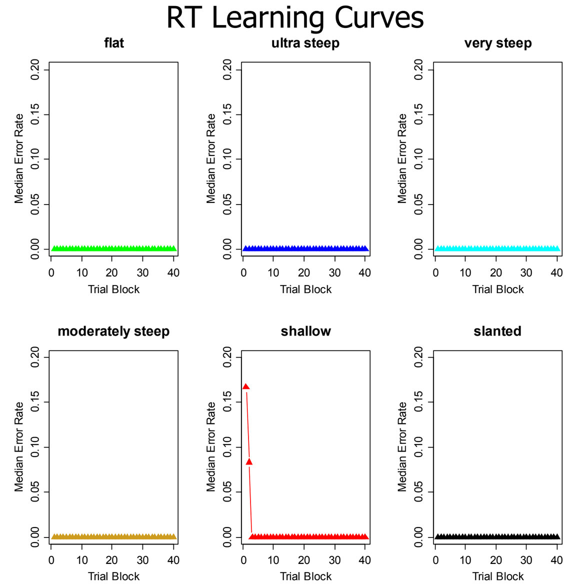
Types of RT Learning Curves

- PAM clustering revealed 6 distinct types of RT learning curves
  - **Flat**, **ultra steep**, and **very steep** curves with <u>virtually instantaneous learning</u>
  - **Moderately steep**, **shallow**, and **slanted** curves with <u>gradual learning</u>
- Some types of RT learning curves (e.g., flat, ultra steep, & slanted) deviate dramatically from the power law of practice
- Different types of RT learning curves do **NOT** stem from systematic speed-accuracy tradeoffs

Error Rates for Prototypical RT Learning Curves



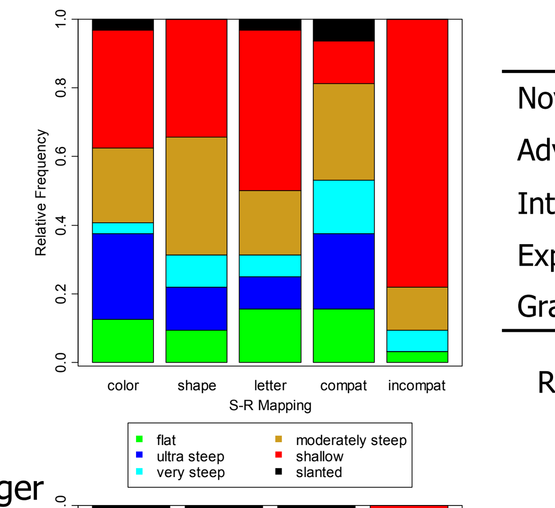
Error Rates for Prototypical RT Learning Curves



Relationships Among Curve Types & Task Factors

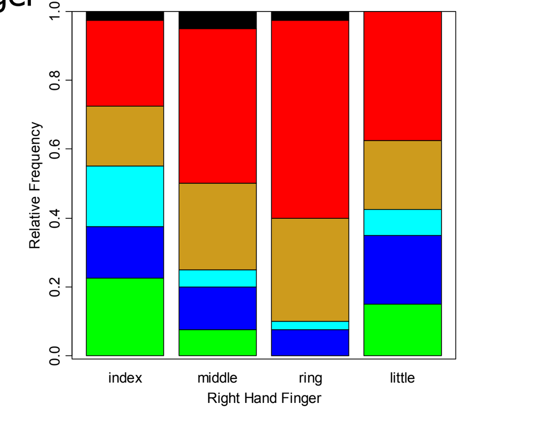
Relative Frequency of Curve Types by S-R Mapping

- Flat, ultra steep, and very steep RT curves (virtually instantaneous learning) occurred frequently for all S-R mappings except the incompatible one
- Shallow RT curves (gradual learning) occurred disproportionately often for the incompatible mapping
- The relationship between curve type and S-R mapping is reliable; $\chi^2(20,N=160)=41.3985$, p<0.01
- The choice of learning strategies is constrained, in part, by the difficulty of the S-R mapping



Relative Frequency of Curve Types by Response Finger

- Flat, ultra steep, and very steep RT curves (virtually instantaneous learning) occurred disproportionately often for the index and little response fingers
- Shallow RT curves (gradual learning) occurred most frequently for ring finger responses
- The relationship between curve type and response finger is reliable; $\chi^2(15,N=160)=27.7368$, p<0.05
- The choice of learning strategies is constrained, in part, by the response finger

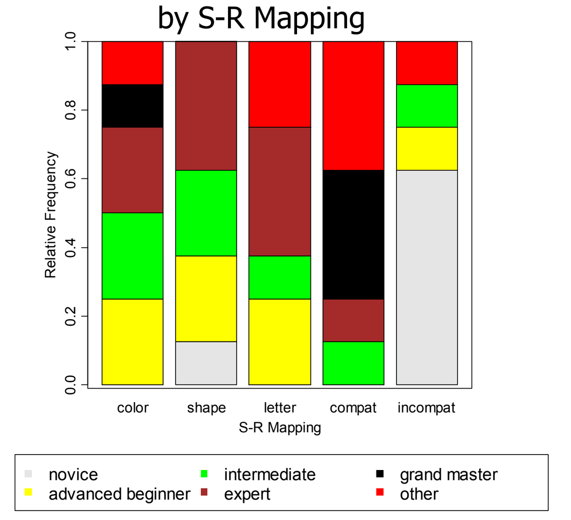


Typology of Participants

- Participants were assigned to categories based on which types of learning curves occurred for the index, middle, ring, and little response fingers
- 82.5% of the participants (n=40) were categorized as either **Novices**, **Advanced Beginners**, **Intermediates**, **Experts**, or **Grand Masters**
- There was a hierarchical relationship between these participant categories and the response fingers for which learning was virtually instantaneous ($n_{\Delta25}$<2) as shown in the table below

Participant Category	RT Learning Curves with $n_{\Delta25}$<2
Novice	none
Advanced Beginner	index
Intermediate	index, little
Expert	index, middle, little
Grand Master	index, middle, ring, little

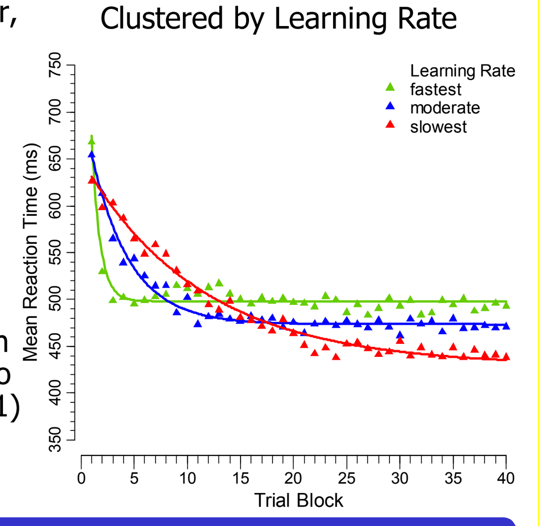
Relative Frequency of Participant Categories by S-R Mapping



Asymptotic RT: No Pain, No Gain

- For each combination of S-R mapping and response finger, we identified 3 clusters of curves based on $n_{\Delta25}$, corresponding to:
  1. Fastest learners
  2. Moderate learners
  3. Slowest learners
- In 19 of the 20 S-R mapping by response finger combinations the curves with the slowest learning rate also had the shortest t_∞ (p<0.001)
- Slow and steady wins the race

Average RT Learning Curves When Clustered by Learning Rate



Summary of Results

- Declarative knowledge was acquired very rapidly for most S-R mappings except the incompatible one, as evidenced by graphs of mean study times over trial blocks
- Mean RTs and error rates were affected reliably by S-R mappings
- Power functions fit average RT learning curves well for each S-R mapping
- Nevertheless, for individual participants and S-R pairs, 75% of the cases were fit better by exponential functions rather than power functions
- Individual RT learning curves fell into distinct clusters, including flat, ultra steep, very steep, moderately steep, shallow, and slanted ones
- Overall, 33% of 160 cases involved virtually complete learning after fewer than 10 practice trials, and in 20% of the cases, learning was essentially instantaneous (i.e., $n_{\Delta25} \approx 0$)
- Systematic relationships occurred between S-R mappings, response fingers, participants, and relative frequencies of the curve types
- When adjusted for main effects of S-R mapping and response finger, relatively slow gradual learning yielded shorter asymptotic RTs than did essentially instantaneous learning

Conclusions

Two Modes of Skill Acquisition and Performance

1. Interpretative processing through Routine Procedural Recipes
   - Optional strategy for skill acquisition
   - Enables ultra-fast learning under some conditions
   - Facilitates positive transfer across similar generic task contexts
   - Limited by declarative WM capacity
   - Culminates in relatively long asymptotic RTs
   - Source of individual differences caused by variable WM capacity
2. "Compiled" processing of procedural knowledge based on sets of task specific hard-coded production rules
   - Optional strategy for skill acquisition
   - Used instead of interpretative processing through RPRs
   - Hinders positive transfer across similar generic task contexts
   - Unconstrained by declarative WM capacity
   - Involves slower learning; requires extensive practice
   - Culminates in relatively short asymptotic RT