Noetic Journal, 1(2): 168-173

The Dynamics of Learning to Automaticity

Richard Flor, University of Wyoming
Kevin Dooley, Arizona State University

Abstract

Schemas are an outcome of learning. Schemas facilitate recognition and retrieval in the brain, and are thought to be an important part of understanding the dynamics of the learning process. As the brain develops more efficient schemas through chunking or other mechanisms, certain tasks may become automatic, in that they require little or no attentional energy. This state of automaticity represents the outcome of a prolonged learning process. In this paper we wish to explore the dynamics of learning to automaticity. Subjects were presented a hand-tapping exercise and response time was recorded. A control group received an ad-hoc distractor (secondary) task, while the experimental group received distractor tasks much more frequently, so as to be forced into learning to some degree of automaticity. Performance on the learned task appeared to converge to a single point attractor over time. Evidence of chaotic behavior was found in some of the subjects' early performance data, indicating that perhaps the brain uses a chaotic search mechanism in searching for an optimal schematic map. Finely cusp catastrophe models fit the learning data better than any other alternative, perhaps indicating that schematic optimization follows a model of punctuated equilibrium.

1. Introduction

The human brain can perform amazing things, considering the limited data capacity of its immediate attention. If one assumes that short-term memory, or attention, is seven bits large, and the brain can process at 18 Hz, this yields and interface capable of processing 126 bps (Miller, 1956). A typical task, such as listening to another person speak, can take up to 40 bps of memory. The brain must therefore demonstrate efficiencies in recognition, retrieval, and action in order to allow the person to effectively deal with the external environment in a timely manner. The process of gaining these efficiencies is learning. It is this ability to gain control over incoming stimulus, noted over 100 years ago by William James, that is perhaps most characteristic of our cognitive functioning:

"If an act became no easier after being done several times, if the careful direction of consciousness were necessary to its accomplishment on each occasion, it is evident that the whole activity of a lifetime might be confined to one or two deeds--that no progress could take place in development" (James, 1890, p.37).

Schemata are an outcome of learning. A schema is an internal map which facilitates recognition through tags and smaller building-block schemata, and which facilitates action by linking triggers and actions (Holland, 1995). Physiologically schema are distributed electro-chemical networks embedded in connections between neurons (Gallistel, 1993). As learning occurs there may be more than one schema that evolves; in fact it is possible that schema compete during the initial stages of learning, and a single dominant schema eventually takes over.

For example, Logan (1990) in his instance-based theory of learning and development of automaticity, characterizes the process as a progression from a search algorithm schema to a direct retrieval schema. This change occurs as the number of instances (e.g. neurological firings) for a relatively fixed stimuli increase; that is as time on task increases. Logan's theory also explains why experts seem to forget all of the components steps needed to accomplish complex tasks; the schema developed and used for later performance has gained a magnitude of efficiency due to a reduction in steps.

Subsequently the brain's map representing the schema may be simplified. From an information-processing standpoint, the brain is able to chunk several distinct but interrelated ideas into a single idea, and thus reduce access and processing time; "chunking converts goal based problem solving into productions… Chunks are active processes, not declarative statements … and chunking is a form of permanent goal based caching" (Newell, 1990, p. 186). Chunking may lead to significant gains in task performance. The actual mechanics of chunking our point of debate however.

Mackay (1982) argued that practice under consistent conditions simply leads to an increase in the firing between nodes in an existing neural network; modification in the network is not necessary to explain gains in speed and accuracy. This is the worn-path view, where connections are strengthened, or made easier to follow, due to repetition. While this certainly may be one explanation for gains in processing speed, it seems likely that there exist limits in reaction times due to limitations in the neurologic and electric-chemical reaction speeds.

We may be nearing a time where efficiency gains through this mechanism may be measurable. Once the limits of this process are understood, it may be possible to factor this time out, with residual gains needing explanation by some other mechanism.

Cheng (1985) proposes that improvements in performance occur through more efficient restructuring, reordering, and/or reorganization of cognitive tasks. Where as Mackay's explanation lies at a neurological level, Cheng offers an example that works at a conscious, or perhaps meta-cognitive level (i.e. that which could be conscious) that addresses the procedures available to skilled and non-skilled individuals. One such case is the transition and procedure for finding the sum of five 2's: two plus two plus two plus two plus two equals 10 requires more steps then 5 x 2 =10. Newell and Rosenbloom (1981) believe that the performance program for a task becomes coded in hierarchically organized chunks of increasing size. If this were to be the case as compared to a complete restructuring into fewer steps or more efficient routing of neurological connections, one might be able to see the evidence of limitations and speed gains due to the inherent limitations at the electro-chemical level as noted above.

In order to understand how the brain is operating under varying conditions--consistent or changing--tasks can be performed in a learning environment and measures of task performance and subsequent analysis should lead to reasonable hypotheses concerning which factors account for performance gains. Mathematical models are used to quantify the relationship between the learning environment and task performance. The simplest learned model is the log-linear model (Baduri, 1992):

$$RT(t) = k * t^b \qquad [1]$$

$$Ln\,(RT(t)) = k + b\,Ln\,(t) \qquad [2]$$

where RT(t) is the task response time at time t, k and b are empirically determined constants, and t is time (trial). The log-linear model assumes rapid initial improvement followed by slow, incremental improvement. It assumes learning occurs through accumulated exposure to stimulus. If one instead assumes a more experimental discovery stage, where performance improves only gradually a first, the S-shaped learning curve is appropriate:

$$\ln\,(RT(t))=(1/M)\ \{A + B\,Ln\,(t) + C\,Ln\,(t^2) + D\,Ln\,(t^3)\} \qquad [3]$$

where M, A, B, C, and D are empirically determined. The S-shaped learning curve assumes learning occurs through contagion, were schema develop by linking to other schema.

In either case the learning process eventually converges is to a point where there is little or no schema development. Dynamically in the learning process convergence is to a point attractor; task response times remain relatively steady as long as attention due remains given. The relevant schema can be dynamically thought of as a basin of attraction (Martin, 1994). Thus our first research proposition is

P1: Performance on a learned task converges to a point attractor.

2. Learning Dynamics

Statistically, convergence is indicated by weak stationarity: mean, variance, and covariance remain constant and independent of time (Kaplan and Glass, 1995). Dynamically, convergence is in part indicated by the lack of divergence, or chaotic dynamics. While performance may converge to a point attractor, it may go through more complex dynamics as schema are chunked and formulated. It is feasible to think that during the initial stages of learning, when the brain is testing competing schema and processing the schema in a much more cumbersome manner, that both convergence and divergence would be present. The presence of divergent dynamics is an indication of chaos. The detection of chaotic dynamics has both practical as well as the practical relevance.

P2: Initial performance on a learned task exhibits chaotic dynamics.

Systems in a chaotic state can exhibit at high degree of sensitivity in one moment and in extreme degree of robustness the next. The output of a chaotic system is point by point unpredictable, but forms a recognizable pattern over time if observed properly. The discovery of chaos leads to a rejection of the random hypothesis. This in turn can change our assumptions about the underlying dynamics of the system in question. There are some who hypothesize that systems actually purposely migrate to such a chaotic state, in that it gives the system maximum flexibility and learning capability (Kauffman, 1995).

One way to characterize the temporal dynamics is to calculate the dimension of the system's attractor (its historical trajectory). If it comes out to be fractional and the system exhibits sensitivity to small changes (this can also be calculated explicitly), then the system may be considered as to be behaving in a chaotic state. Numerous different algorithms can be applied to determine with some degree of confidence whether or not the system is indeed chaotic, or more conservatively, nonlinear.

There is some reason to believe that the early stages of the learning process may be chaotic (Goertzel, 1995). In the case of an individual learner, a reason for such chaotic dynamics might be that while the brain is searching for an optimal chunking pattern by which the store instructions for the task, a chaotic search mechanism would tend to be optimal. Because of coupling between the neurological and physiological systems, a chaotic search mechanism may also lead to chaotic vibrations in the outcome of the task.

Chaotic dynamics have been found in numerous studies of neuron level activity. Freeman (1994) states, concerning the utility of chaos: "certain advantages may accrue by virtue of the breath of spectrum or immunity to entrapment of the dynamics and limit cycle tractors. The maintenance of set points near separatrixes is between different chaotic basins of attraction may, because of sensitivity to initial conditions, allow amplification of small fluctuations into large microscopic patterns. Perhaps the most compelling advantage may lie in the capacity of chaotic systems to create as well as to destroy information. This property may constitute the key feature by which memories are constructed from the raw materials of sensory input that survives the residues of synaptic changes left behind after traumatic experiences or sleep" (p.303). Kowalik and Elbert (1994) use measures of "chaoticness" to determine phase transitions in brain activities.

At a psycho-motor task level, Cooney and Troyer (1994) performed a simple symbol-memory test and found response time behaved in a chaotic manner, with low dimensionality. This lies in stark contrast to traditional assumptions about the learning process. Clayton and Frey (1994) however performed a replicate of the study and did not find any evidence of chaos; their analysis indicated the subsequent dynamics were random. Subsequent tests showed that dynamics post-mastery could perhaps be described by colored noise (Clayton and Frey, 1997).

At the macro-level of learning a number of different studies concerning research and development processes have found chaotic and divergent behavior to be present in the early stages of development while more orderly behavior (convergent) was present at the later stages of development (Cheng and Ven de Ven, 1997). Guastello (1995) has also found chaotic dynamics in a number of different learning environments.

As performance on the learned task converges, a state of mastery is achieved. Further exposure to the learning task results in chunking rather than improvement in performance per se. Less and less conscious effort is required, and processing and action become automatic. This state, known as automaticity, is an important state of learned behavior, where a task can be performed with little are no conscious attention (LaBerge and Samuels, 1974). A skill such as driving is thought to be automatic to a competent adult, as one can simultaneously perform other tasks that require significant conscious attention. In work situations, tasks learned to automaticity can be performed rapidly with few errors. Conversely, unlearning for an automatic state can be quite difficult. There has been very little empirical work done on learning to automaticity.

The link between limitations in human attention and automatic processing are strong (Norman and Bobrow, 1975). Resource dependent tasks are those that are affected by the amount of potential energy, or cognitive resources, one applies to its performance. Resource insensitive tasks are unaffected by the amount of intentional effort one applies to the task. Developing automaticity may be seen as a progressive movement from initial task performance being resource limited to skilled performance that is resource insensitive (Norman and Bobrow, 1975).

While other explanations are feasible, chunking is a reasonable model of why tasks can be made automatic. Chunking may tend to follow a model of punctuated equilibrium, where long periods of stasis are interrupted by short periods of rapid change (Gould, 1980). This mimics the evolutionary dynamics seen in genetic systems (Holland, 1995), and also observed in the artifacts of human societal development (Mazur, 1989). Catastrophe models model the phenomenon of discontinuous and abrupt change. The cusp model, for example, is described by two stable modes of behavior. Change between the two states occurs as a function of the two control parameters, asymmetry (A) and bifurcation (B). When B is low, change is smooth and a function of A. When B is high, change still occurs as A changes, but in this case is more discontinuous and hence, the Cusp. For example, obe of Guastello's experiments (1995) shows that when an operator is working with a light load, that the risk of having an accident is a smooth function of the amount of environmental hazards present; when load is high, however, risk is discontinuous and better described by two attractors (low-risk and high-risk) separated by a cusp.

The cusp catastrophe model can be defined by (Guastello, 1995):

$$\text{Ln}(RT(t)) = B0 + \text{Ln}(RT(t-1)) + B1 * Q1 + B2 * Q2 * \text{Ln}\, R(RT(t-1)) + B3 * \text{Ln}(RT^2(t-1)) + b4 * \text{Ln}(RT^3(t-1)) + B5 * Q3 \qquad [4]$$

where be B0 through B4 are empirically determined. Q1 is the asymmetry parameter (=t), Q2 is the bifurcation parameter (equals 0 when there is no secondary distracted task, and equals 1 when there is, and Q3 equals 1 when there is an error and 0 there is not. Thus we expect the following:

P3: A catastrophe model can be used to model the discontinuance dynamics of learning to automaticity.

3. Experimental Design

Fifty-five college students participated in a learning experiment that involved learning a hand-tapping pattern. The subjects were shown 10 numbers on one card corresponding to their left hand and 10 numbers on another. Subjects had to tap the correct sequence on a keyboard. The keyboard was set out so as to collect data on response times and errors. All subjects performed the task to mastery in a first session. It was noted that a significant difference existed in tapping style: simultaneous tappers using both hands at once were significantly quicker than sequential tappers.

An experimental group performed a second session under a dual task condition, and went to automaticity. The secondary task involved recalling a set of instructions from memory until the primary task was finished, and then successfully carry out the said task. A secondary task was involved in about half of all the trials. The control group performed a second session under the same initial conditions as the first session. Data was collected on the number of errors, and response time. Mastery was defined as the point, after an initial practice, where a subject could perform the primary task correctly three times in a row without cue cards. Automaticity was defined at several points, the first being that point, after mastery, where two consecutive trials with distractors were performed at or below mastery speed (Flor, 1994).

4. Results

For brevity sake, only the results from 7 experimental subjects are presented: trial A (learning to automaticity) for a sequential and simultaneous tapper, and trial B (control group, with a few occurrences of a distracted task) with sequential tappers. Figure 1 shows the response times for one of the control group members; trials where errors occurred are omitted.

One first questions whether the log-trial model or the S-shaped model is a better describer of the decay dynamics. Additionally three other models are tested. First two linear terms are added to the log-trial model, in order to explicitly model the presence of an error event and/or a distractor task. Secondly, the effect of the distractor task was changed from a linear impact (that only affected RT(t)) to an exponential impact (effect = exp(-b * time since distractor test)). Lastly, a cusp catastrophe model was fit.

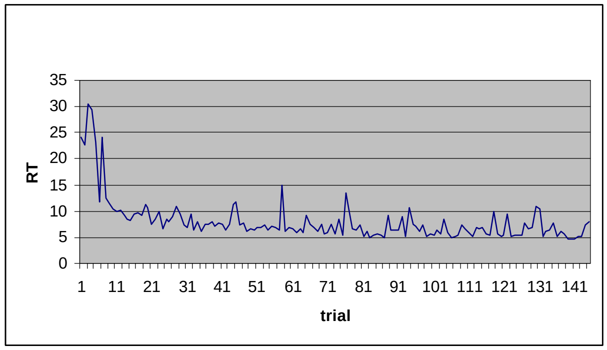

Figure 1 Response Times for Typical Subject

The log-trial model works as well (r^2 between 0.28 and 0.71) as the S-shaped model, and has fewer parameters, so by parsimony it seems to be superior. The addition of linear terms which capture the error or distractor event improve the model fit for the most part (r^2 between 0.47 and 0.83), so that will be the model form of choice. The exponential terms did not improve model fit. Finally, a cusp model was not a good fit. This does not put proposition three (that learning can be represented by a catastrophe model) at risk, because only a few trials had distractor tasks present in the control group.

In order to test the first proposition--that performance in a learned task converges to a point attractor, we can first investigate the form of the model:

$$Ln (RT(t)) = k + b * Ln (t) + c * error + d * distractor \quad [5]$$

Where B, C, and D are empirically determined and error and distractor are indicator (0,1) variables indicating the presence or absence of an error or distractor event. Taking the derivative:

$$d(Ln (RT(t))/dt = b/t \quad [6]$$

Thus as t increases, the change in RT decreases to zero, indicating a point attractor. One can also calculate the Lyapunov exponent of the time series data. If the exponent is positive, it indicates divergence of nearby trajectories; if negative it indicates convergence of nearby trajectories. An algorithm to calculate the exponent, especially tailored to small data sets, was used (Rosenstein et al., 1993). Results indicated negatives exponents for each of the test subjects in the control group, indicating support for proposition 1.

In order to test proposition two, that performance on a learned task may behave chaoticly at the initial stages of learning, the data was segmented into two parts. Since mastery was achieved relatively early on after five to 15 trials, there was no way to test for chaos pre-mastery. Since the overall experiment for the control group lasted 140 trials, it was somewhat arbitrarily decided to take the first 50 observations and analyze those for the presence of chaotic dynamics. Residuals from model 5, rather than the raw data, were used. A program by Rosenstein et al. (1993) was used to find the exponent and the fractal dimension of the time series. A delay time was chosen equal to the point where the sample auto correlations diminished to zero. For subjects 34,35,38, and 44(subject 52 had too many missing data points) this corresponded to a delay lag of 6,3,6 and 1 respectively.

Analysis showed that the initial first 50 points performance data for subjects 34 and 35 were possibly chaotic with dimension 3.5; similar analysis of subject 38 and 44 did not demonstrate chaos. To further analyze whether chaos was present or not less, Guastello's (1995) method was used (see Johnson and Dooley, 1996, for details). For the initial learning data only subject 34's showed chaos, with dimension around 2.5; the others appeared random. Additionally, all the end data proved to be random.

Thus, there is further evidence to support a single attractor (convergence) at the end of learning, but there is inconclusive evidence as to whether chaotic dynamics are present in the early stages of learning. In order to test proposition three, that chunking occurs catastrophically, we fit the model shown in equation 4 and compare the goodness of fit of that model to its most direct linear alternative:

$$\text{Ln}(RT(t)) = a + b * \text{Ln}(t) + c * Q2 + d * Q3 \qquad [7]$$

where a, b, c, and d are empirically determine, Q2 is the bifurcation parameter and Q3 is the error indicator.

The model is conceptually shown in Figure 2. When bifurcation is low (no distractor is present), progress is smooth, and response times decrease with experience. When bifurcation is high (distractor task present), improvement in performance is slower, until chunking occurs, and then a drastic improvement is seen. The results from fitting equations 4 and 7 to the data for subjects 13 and 25 support proposition three, that learning is catastrophic. For subject 13, the catastrophe model had an r^2 of 0.65, compared to the linear alternative of 0.55; for subject 25, the catastrophe model had an r^2 of 0.48 while the linear alternative was 0.38.

5. Summary

Learning has always been thought of as a convergent process, so it is not surprising to find strong support for our first proposition, that performance on a learned task is driven to a point attractor. In reality however there is often little attention paid to the rate of learning; interventionists and educators should monitor the rate of said convergence to understand when additional training in practice has limited benefits.

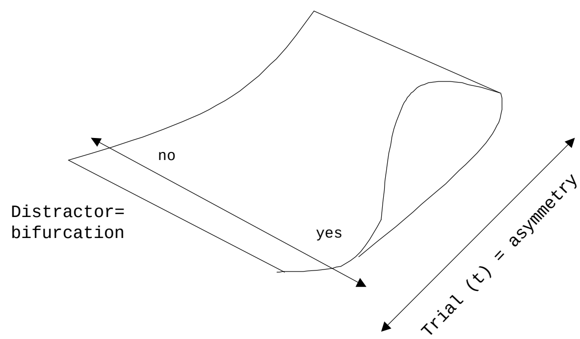

Figure 2 Cusp Catastrophe Model of Learning Curve

There was inconclusive evidence concerning our second proposition, namely that performance dynamics will tend to show chaotic behavior early in learning. Of the four subjects tested, two did not show such tendencies, one did, and one was inconclusive.

Finally there is some evidence that learning to automaticity can be best described dynamically by a catastrophe model. We found that the data supported such a model over its best linear alternative. One additional theoretical advantage of the catastrophe model is its hysteresis. Just as the number of trials must significantly increase in order for chunking to occur, it is equally difficult to unlearn the tasks relative to the schema. Unlearning was not tested in this experiment.

Further research should examine additional subjects, point-wise dimension estimates, multiple levels of distractor tasks, and unlearning. Additional replicates in a different experimental environment would also enhance validity.

References

Badiru, A. (1992). Computation survey of univariate and multivariate learning curve models. IEEE Transactions on Engineering Management 39(2): 176-188.

Cheng, P. (1985). Restructuring vs. automaticity: alternative accounts of skill acquisition. Psychological Review 92(3): 414-423.

Cheng, Y. and A. H. Van de Ven, (1996), "Learning the Innovation Journey: Order out of Chaos?" Organization Science 7(6): 593-614.

Clayton, K., and Frey, B. (1994). Inter- and Intra-trial dynamics in memory and choice. Society for Chaos Theory in Psychology and Life Science. Baltimore, MD: June.

Clayton, K., and Frey, B. (1997). Studies of mental noise. Nonlinear Dynamics, Psychology, and Life Sciences 1(3): 173-180.

Cooney, J., and Troyer, R. (1994). A dynamic model of reaction time in a short-term memory test. Journal of Experimental Child Psychology.

Flor, R. (1994). An assessment of individual differences in the development of automaticity in a psychomotor learned task and the effect of original learning environment on memory. Unpublished Ph.D. dissertation, University of Minnesota.

Freeman, W. (1994). Characterizations of state transitions in spatially distributed, chaotic, nonlinear dynamical systems in cerebral cortex. Integrative Physiological and Behavioral Science 29(3): 451-482.

Gallistel, W. (1993). The Organization of Learning. Cambridge, Mass: MIT Press.

Goertzel, B. (1995). Evolutionary dynamics in minds and immune systems. In Chaos Theory in Psychology, Abraham and Gilgen (eds.), Westport, CT: Greenwood.

Gould, S. (1980). Is a new and general theory of evolution emerging? Paleobiology 6:119-130.

Gaustello, S. (1995). Chaos, Catastrophes, and Human Affairs. Mahwah, NJ: Erlbaum.

Holland, J. (1995). Hidden Order. Reading, Mass: Addison-Wesley.

James, W. (1890). The Principles of Psychology. NY: Holt.

Johnson, T., and Dooley, K. (1996). Looking for chaos in time series. Nonlinear Dynamics in Human Behavior, W. Sulis & A. Combs (eds), Singapore: World Scientific: 44-76.

Kaplan, D., and Glass, L. (1995). Understanding Nonlinear Dynamics. NY: Springer-Verlag.

Kauffman, S. (1995). At Home in the Universe. Oxford: Oxford Press.

LaBerge, D., and Samuels, J. (1974). Towards a theory of automatic processing in reading. Cognitive Psychology 6: 293-323.

Logan, G. (1990). Repetition priming and automaticity. Cognitive Psychology 22: 1-35.

Mackay.

Martin, B. (1994). The schema. Complexity: Metaphors, Models, and Reality. Cowan, Pines, and Meltzer (eds.), Reading, Mass: Addison-Wesley, p. 263-277.

Mazur, A. (1989). Periods and question marks in the punctuated equilibrium of human social behavior. In The Dynamics of Evolution. Somit and Peterson (eds.), Ithaca, NY: Cornell).

Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review 63: 81-97.

Newell, A. (1990). Unified Theories of Cognition. Cambridge, Mass: Harvard.

Newell, A., and Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In Cognitive Skills and Their Acquisition, J. Anderson (ed.), Hillsdale, NJ: Erlbaum.

Norman, D., and Bobrow, W. (1975). On data-limited and resource-limited processes. Cognitive Psychology 7:44-64.

Rosenstein, M., Collins, J., and DeLuca, C. (1993). A practical method for calculating the largest Lyapunov exponent from small data sets. Physica D 65: 117-134.