Transfer of Training and its Effect on Learning Curves

Craig P. Speelman
Edith Cowan University

Kim Kirsner
University of Western Australia

The Newell and Rosenbloom (1981) depiction of the power law of learning implies that improvements in task performance that result from practice can be described by a power function with one variable for amount of practice. We suggest that performance on all but the simplest of tasks relies on component skills that differ in their practice history. As a result, power functions with one term for practice could not be expected to provide accurate descriptions of learning curves. In particular, transfer situations that involve a mixture of old and new skills are likely to lead to perturbations in learning curves that require more than the simple version of the power law to describe. We explore the types of functions that are necessary in these situations and note the impact of transfer factors on learning rates.

Learning curves are used extensively in psychology for depicting how the accuracy or speed of performance on a task improves with practice. These curves, particularly those measuring speed, have characteristic shapes. Early in practice, speed improvements are dramatic, but taper off with continued practice - a case of diminishing returns. Snoddy (1926) was the first to note that when the logarithm of performance time is plotted against the logarithm of amount of practice, a straight line typically results. This indicates that performance time can be described as a power function of practice, as indicated by the following equation:

$$T = X + N P^c \qquad (1)$$

In Equation 1, T is the performance time on a task, P is the amount of practice on the task, $X + N$ is performance time on the first trial of the task, and X is the performance time after an infinite amount of practice. Power functions where $X = 0$ often provide very good fits to data, especially where large amounts of practice are involved. The parameter c in Equation 1 is the learning rate. The value of c is less than zero, to match the negative accelerated feature of the learning curve, and is also usually a value between 0 and -1. The closer the value of c is to -1, the faster the learning rate.

Crossman (1959) re-analysed the data of many experiments in varying domains, such as card sorting, addition of digits and cigar rolling, and noted that learning in all of these tasks conformed to what he referred to as "de Jong's law". De Jong (1957) had also noted the power function regularity in learning curves. Newell and Rosenbloom (1981) also examined learning curves in a wide range of domains. They examined the ability of power functions to fit learning curves in comparison to other functions that are also negatively accelerated, such as the exponential function and the hyperbolic function, and concluded that the power function regularly provided a closer fit to the data. The general form of the exponential function used to describe performance improvements with practice is:

$$T = X + N e^{cP} \qquad (2)$$

where the meaning of each parameter is the same as in the general form of the power function presented above (Equation 1). Newell and Rosenbloom, like Crossman, were impressed with the apparent lawfulness of the regularity in the learning data, so much so that they referred to the regularity as the power law of practice, and noted that it was one of the few laws in Psychology.

Indeed the presence of power functions in human learning data is so ubiquitous that the power law of practice has almost become an accepted fact in Psychology. The acceptance of the power law of practice within Psychology has grown to the extent that some researchers (e.g.,

Please address correspondence to: Craig Speelman, School of Psychology, Edith Cowan University, 100 Joondalup Drive, JOONDALUP 6027, WESTERN AUSTRALIA. Email: c.speelman@ecu.edu.au

Anderson, 1983; Logan, 1988) have suggested that for any theory of skill acquisition to be considered seriously, it must be able to provide an account of the law. Certainly anyone who has proposed a theory of skill acquisition in the last twenty-five years has promoted the ability of the theory to account for the power law as an important indicator of the theory's validity. It is important to note, however, that there have been suggestions recently that the power law may not be as lawful as has generally been believed. In addition, there is evidence that the shape of learning curves is affected by transfer. Both of these issues are considered in this paper, although it is the latter that is of main interest here. Given that the nature of learning curves is one of the important facts about skill acquisition that all theories strive to account for, it is imperative that a valid theory can account for factors that impact on transfer as well as their influence on learning curves.

The Power Law of Learning and its Alternatives

As would be expected with any law, once the power law of learning was suggested, many objections and challenges to its status as a law of human behaviour were mounted. In this paper we present a number of these challenges that demonstrate that describing the relationship between practice and improved performance with a power function is not a straightforward matter.

The apparent lawfulness of power function learning has been called into question recently, with strong arguments being presented for exponential function learning (e.g., Heathcote, Brown & Mewhort, 2000). When Newell and Rosenbloom (1981) first presented the case for the power law of learning, they explicitly compared the ability of the power function and the exponential function to account for a number of sets of learning data. Although Newell and Rosenbloom's comparisons found that learning data were consistently described better by power functions, doubt has been cast upon the validity of that conclusion.

At least three objections have been raised to the proposal that performance time is a power function of practice. The first objection concerns the quality of the evidence that supports the power law. Several researchers (e.g., Heathcote et al., 2000; Myung, Kim & Pitt, 2000) have pointed out that when power functions are most evident, the data being analysed represents average data from several individuals. This result is considered misleading because an artifact of the process of calculating a linear average of individual learning functions is that it will result in a learning curve that is described better by a power function than an exponential function, even if the individual functions were exponential (Myung et al., 2000). The second objection to the power law concerns evidence that favours exponential

learning functions. Following the first objection, when individual learning curves are analysed, exponential functions regularly provide a better fit to the data than power functions (Heathcote et al., 2000; Josephs, Silvera & Giesler, 1996; Rosenbloom & Newell, 1987). Finally, the third objection to the power law concerns the suggestion that there may not be one pure function that describes all learning curves. For instance, there is evidence that exponential functions are best for describing performance in some conditions, whereas other conditions result in learning curves that are best described by power functions (Carlson, 1997, pp. 59-60). Other researchers have suggested that learning curves on complex tasks may represent the operation of several underlying learning functions (Delaney, Reder, Staszewski & Ritter, 1998; Rickard, 1997, 1999). The combination of component learning functions could even involve a mixture of power and exponential functions (Heathcote et al., 2000).

Several years ago we also proposed that learning curves should be considered as summaries of many learning functions. This argument is detailed in the next section. It should be immediately obvious that, despite the foregoing, we have maintained the use of power functions as descriptions of learning. The main reason is that the model takes as its starting point Anderson's (1982) ACT* theory of skill acquisition, and in particular the ACT* account of the power law. Ultimately this feature of our argument for a componential view of the learning curve is of no consequence, as the final message applies no matter what form learning functions take, as long as they are negatively accelerated.

Learning Rate

Snoddy (1926) was the first to report that learning data, when plotted on log-log axes, approximated a straight line. Snoddy also reported that the learning rate of his subjects, as indicated by the slopes of their learning curves on log-log axes, were identical:

> We have already indicated that the facilitation lines (learning curves) for all types of practice are parallel on log paper, which means that the slopes of the log equations are equal. From this it would be a good guess that all individuals, whether high or low in efficiency, might have the same slope for their facilitation lines. This turns out to be the case and n, which is equal to 0.25, is the universal slope. We have studied about 300 children, 500 insane, and over 1000 low, normal and superior adults among college students, and have invariably found the same average slope of the log equation in facilitation practice. It is of interest that many of our insane

subjects are well below the normal distribution curve for adults, and yet these have the same slope *n*. In our study of children of all age levels the slope of the facilitation lines is seen to be the same as for normal adults. This constancy of the slope of the log equations for all subjects is not only a thing of much interest mathematically and physiologically, but it simplifies enormously the mathematics involved in the treatment of clinical data. (pp. 22-23, parentheses inserted)

This finding of identical learning rates amongst such a disparate sample of people is indeed remarkable, for at least two reasons. First, it is unlikely that measurement of any other feature of human performance would yield an identical value across a group of people. Second, no mention is ever made of this remarkable observation in modern discussions of learning curves. The reason for this omission is clearly because no one observes any degree of consistency in learning rates amongst people. This may be due to the greater precision in curve fitting afforded by today's computers compared to the analytical techniques available to Snoddy. Certainly, our experience is that learning rates are just as variable as any other measurable characteristic of human performance.

Although the learning rate parameter of power functions is usually found to be between -1 and 0 there appears to be substantial variation within this range between tasks and subjects (e.g., Newell & Rosenbloom, 1981). However a systematic relationship between type of task and subjects' learning rate has yet to be established. This paper presents an account of this relationship, an account that shows how transfer and learning rate are intimately connected. This model was originally developed as part of Speelman's (1991) doctoral research, and elements of it have been presented and evaluated in Speelman and Kirsner (1993), Speelman (1995), Kirsner and Speelman (1996), Speelman and Kirsner (2001), and Speelman and Kirsner (2005).

Anderson's (1982) ACT* theory of skill acquisition included an account of the power law of learning that is probably the only account to be explicit about the determinants of learning rate. This account basically suggests that, all things being equal, learning rate is a constant for each person. Given the fact that subjects exhibit varying learning rates dependent on task situation, this account is obviously not sufficient. However, Anderson's account can be used as a starting point for the development of a more precise model. This model assumes that learning rate is affected by the relative amounts of practice of particular task components and also the relative number of processing steps involved in these components. Thus this model, which we refer to hereon as the linear component model, is in fact designed to account for the shape of learning curves following transfer, but will account for changes in learning rate as a matter of course. In the next section, Anderson's account of learning rate is described, followed by a description of the linear component model. Predictions that follow from the linear component model are compared with previously published data. It should be noted that development of the linear component model pre-dated Anderson's (1993, Anderson & Lebiere, 1998) revision of the ACT* theory, and so does not accurately reflect some of the details of the ACT-R account of the power law of learning. The ultimate message of the linear component model, however, is unaffected by the choice of theoretical starting point.

ACT* and Learning Rate

ACT* predicts that performance will become faster with practice as a result of both algorithmic improvement and strengthening. ACT* states that the process of compilation improves the procedure for performing a task by reducing the number of steps involved in the procedure. It is assumed that the number of steps is reduced by a constant fraction with each improvement. Anderson (1982) suggests that this algorithmic improvement follows a standard power function for improvement to an asymptote:

$$N = N^* + N_0 P^{-f} \qquad (3)$$

In this equation N represents the number of processing steps performed on Trial P (where P represents the amount of practice), $N^* + N_0$ represents the number of steps performed on Trial 1, and N^*, the asymptote of this function, represents the minimum number of steps that constitute the optimal procedure for performing the task. The exponent f represents the constant fraction by which the number of steps is reduced with each improvement.

The other contributing factor to performance time reduction in the ACT* theory is that of strengthening. In ACT* it is assumed that the representation of task-specific information is strengthened in memory with practice. The strength of a memory element then determines the speed at which it can be accessed and applied. Anderson (1982) shows that this reduction in application time with increased strength can also be described by a power function:

$$T = C + A P^{-g} \qquad (4)$$

In this equation T represents the time to execute a series of production rules, or productions, P represents the amount of practice, A is a constant that represents the time it takes to execute a certain number of productions and is therefore proportional to the number of productions involved in performing a task, the asymptote C represents the minimum time for execution of a certain number of productions, $C + A$ represents the maximum time for execution, and the

exponent *g* is a constant that represents the rate at which the strength of a memory element decays, and is a value greater than zero.

In order to derive a function that describes total time to perform a task Anderson (1982) combined Equation 3, which represents the number of productions involved in performing a task, and Equation 4, which represents the time per production. This combination resulted in Equation 5:

$$TT = (N^* + N_0\, P^{-f}) \times (C + A\, P^{-g}) \qquad (5)$$

This equation simplifies to Equation 6:

$$TT = N_0\, A\, P^{-(f+g)} \qquad (6)$$

if *N** and *C* are assumed to be zero. Anderson makes this assumption because Equation 5 is not a perfect power function, although it is a good approximation to one. Eliminating *N** and *C* results in a pure power function that can be further simplified to the equation introduced above as the general form of a power function that describes learning:

$$T = N\, P^c \qquad (7)$$

As described above, *N* is a constant related to the initial number of productions involved in performing the task and *P* is the amount of practice. This equation has a zero asymptote.

The step from Equation 5 to Equation 7 suggests that the learning rate for any task (*c*) is a negative constant that is determined by two other constants: the fraction by which the number of productions is reduced with compilation (*f*), and the decay rate of production strength (*g*). The latter two constants, according to Anderson (1982), are parameters of the cognitive system. Therefore learning rate is itself a parameter of this system. This implies that learning rate is determined by particular characteristics of a subject's mental functioning, that it is a constant which is 'built in' to each subject. Thus, in Anderson's account, there is no apparent facility for type of task to affect learning rate.

If learning rate is some form of constant for each subject, then presumably the range of learning rates that are observed between experiments is a result of different combinations of results from subjects with different learning rates. However this suggestion is not supported by experiments where learning rate was observed to vary from task to task within individuals and groups (e.g., Fitts, 1964; Grose & Damos, 1988; MacKay, 1982; Newell & Rosenbloom, 1981; Smith & Sussman, 1969; Snyder & Pronko, 1952). Therefore it does not appear that variation in observed learning rates results only from a sampling distribution of combinations of various constant learning rates. Variation in learning rate is also likely to result from an interaction between subject characteristics (i.e., their intrinsic learning rate) and task features. The form of this interaction is considered below.

Old and New Task Components

The form of the interaction between a constant learning rate and task features could involve the fact that, except for infants, most tasks have components that involve previously learned skills as well as components that are peculiar to the new task (Fitts & Posner, 1967, p. 19). The latter components will include both components for performing the new part of the task and components that integrate the functioning of the old and new components. This conception suggests that for any task there will be some components that have had more practice than others. The obvious consequences of such a suggestion are that (1) the older components will be faster than the new components (if the number of steps involved in the two sets of components is equivalent), and (2) the older components will have less room for improvement than the new components. These predictions in turn suggest further consequences. However, they also rely on a number of assumptions.

The first assumption underlying the above predictions is that for any one person the rate at which new skills are learned is a constant. This is as suggested by Anderson's (1982) conception of the power law of learning. A 'new' skill here is conceived of as a skill that involves no components that have had previous practice. This includes those skills that are necessary for integrating the functioning of old and new skills into the same goal structure. The second assumption is that all components underlying the performance of a task will improve according to the power law of learning and, with practice, will continue to do so at the same learning rate. It should be noted that this is a convenient starting assumption, but because of the objections raised about the validity of the power law, there is cause to doubt the validity of this assumption.

In summary then, the above conception suggests that, in most cases, learning a new task involves continued practice on old skills. These are skills that have been practised in the context of some other task. Learning a new task will also involve the development of new skills. These are skills that are required to fill the gap between the repertoire of old skills possessed by the trainee and the skills necessary to perform the new task. These new skills will involve both task-related skills and skills for integrating the functioning of old and new task-related skills. In order to evaluate the predictions based upon this concept of task learning it will be informative to contrast this concept with an idealised situation where all components of a task are learned from scratch.

The idealised situation is a simple one - improvement is a function of practice and follows the power law of learning.

The initial time to perform the task is a function of the number of components or steps involved in the task. Thus the whole situation can be described by one power function. In contrast, the more realistic situation is unlikely to be accurately described by one power function. The simple reason is that in this situation each component will not have had equal amounts of practice. As a result a power function with one term that describes amount of practice is not sufficient. This then raises the question of how varying amounts of practice can be incorporated into a function that describes improvement on a task with old and new components.

One possibility is suggested by Anderson, Conrad and Corbett (1989) who propose that "acquisition of ... skill can be predicted by composing simple learning functions for (the) units (p. 503)" underlying tasks. One interpretation of this proposal is that it is suggesting that components of a task have their own learning functions and that the learning function for the task as a whole is a combination of these separate functions. With respect to the current discussion, this would suggest that old and new components of a task improve with practice according to their own learning functions. These separate functions would then include the fact that the components have had unequal amounts of practice. The learning function for the task would then be a combination of these "old" and "new" learning functions.

The form of this combination needs to be considered before the implications of this suggestion can be examined. Underlying a great deal of the research into skill acquisition is the assumption that the more steps involved in a task the longer it takes to perform (e.g., Anderson, 1982; Carlson, Sullivan & Schneider, 1989; Staszewski, 1988). The assumption implies a serial process where each step contributes a particular amount of time to the total task time. Following this logic with the combination of old and new skills requires the combination to be a serial one. That is, the processing of one set of components should not impinge upon the processing of the other set except to provide input information. If this is the case then the learning function for the whole task should be a simple linear combination of power functions describing improvement in each of the underlying components. If the components can be separated into old and new then this function will have the following form (Speelman, 1991):

$$T_{task} = T_{old} + T_{new}$$
$$= N_0 P_0^c + N_n P_n^c \qquad (8)$$

This equation represents the linear combination of two power functions of the form described in Equation 7. Terms with the subscript o represent parameters of the old components of the task, and terms with the subscript n represent parameters of the new components. Equation 8,

then, is the mathematical expression of the model we refer to as the linear component model.

There are a number of implications of the linear component model that should be made explicit. The first is that the contribution of each set of components to the total task time is weighted by the number of steps involved in each set. That is, the greater the number of steps in a set of components, the greater will be the contribution of this set. The second implication is that this weighted combination will be qualified by the amount of practice that the sets of components have had prior to the combination. This qualification has two related forms: (1) the more practice a set of components have had, the faster they will be, and so practice serves to reduce the contribution of a set of components to the overall performance time of a task; (2) as the amount of practice of a set of components increases the room for improvement decreases.

The most important implication of the linear component model concerns the rate at which improvement will occur in the total task. In this model the learning rate (c) of the two separate power functions is the same in each function. This represents the assumption described above that the learning of all components of a task for any one person is a constant. Incorporating this assumption into the model results in a power function describing improvement in the overall task that has a different learning rate to that of each of the components. This difference is always in the direction of a reduction: the learning rate of the total task will be slower than the learning rate of its underlying components. The amount by which the learning rate will be reduced is a function of the relative number of steps in the old and new components, and of the relative amount of practice each set of components had prior to combination.

For example, consider the case of a subject who has practiced a task for 6 sessions. Let the task have 100 steps ($N = 100$) and the learning rate be -0.8 ($c = -0.8$). The improvement in the time to perform the task can now be described by the equation $T = 100 \, P^{-0.8}$ (this is only a loose description as N in Equation 7 is only proportional to the number of processing steps/productions involved in a task, not equal to this number). Now suppose the subject is given a new task to practice that includes all of the steps in the old task plus a new set of steps that number 20. The subject will be able to perform the old steps quickly but will be starting from scratch with the new steps. The time to perform such a task that includes old and new components can be described by the combination of the power functions that would describe improvement on the separate components. Thus,

$$T = T_{old} + T_{new}$$
$$= 100 \, P_{old}^{-0.8} + 20 \, P_{new}^{-0.8}$$
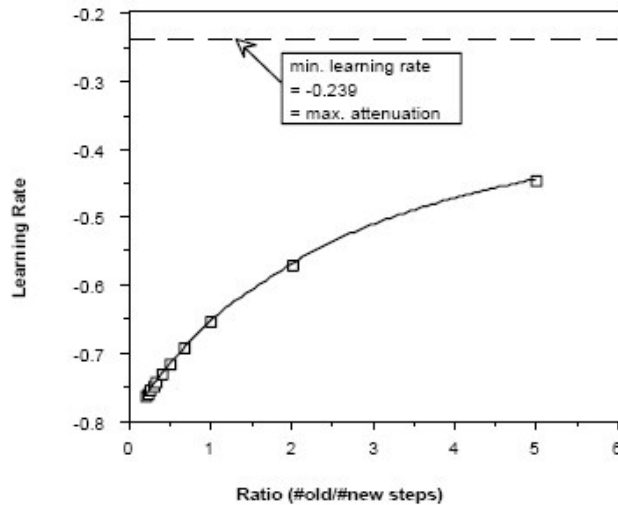
where $P_{old} = P_{new} + 6$.

Figure 1: Learning rate as a function of the ratio of the number of steps in old skills vs. the number of steps in new skills. The data points were generated from a learning example described in the text. The line represents an interpolation of the data points.

This function now has an overall learning rate of $c = -0.44$ (i.e., plot values for T against P_{new} on log-log axes and the gradient of the resulting straight line is -0.44). Therefore, learning rate has been attenuated as a result of combining two skills that differ in the amount of practice they have had and the number of steps involved with their execution. The rate of improvement in the overall task is slower than in the components underlying performance in the task. However the attenuation will not always be as dramatic as in this example. As mentioned above, the amount of attenuation is moderated by two factors.

The first moderating factor on the amount of attenuation in learning rate is the relative number of steps in the old and new components of a task. The effect of this factor on the overall learning rate is depicted in Figure 1. The data points were generated from the above example, where $P_{old} = P_{new} + 6$ and $c = -0.8$. The number of steps in the new component was varied from 0 to 500 and the number of steps involved in the old component was kept constant at 100. It is clear from this figure that an increase in the ratio of old to new steps increases the attenuation of the learning rate. The function depicted in Figure 1 has minimum and maximum boundaries. The minimum boundary corresponds to the situation where there are no old steps involved in the task. The learning rate in this situation corresponds to the intrinsic learning rate of the system, which is -0.80. The maximum boundary represents the maximum attenuation effect. This corresponds to the situation where the old steps outnumber the new steps to the extent that the new steps have no effect on the overall learning rate.

The point of maximum attenuation is equivalent to measuring the learning rate of a task as if it were a new task

and ignoring the fact that it has been practiced for 6 sessions. The subsequent sessions of practice would elicit performance times that improved at what appeared to be a slower rate than the earlier sessions. However this attenuation is simply a result of using an inappropriate point to represent 'session one'. Figure 2 illustrates that in this situation a practice function with a 'slow' learning rate is in fact the tail end of a practice function with a faster learning rate. This phenomenon can result in inaccurate measures of learning rate when prior experience with a task is not taken into account. Newell and Rosenbloom (1981) suggest that using a more general form of the power function that incorporates the amount of prior practice can improve the accuracy of power function descriptions of practice data. Such a function would have the following form:

$$T = N (P + E)^c \tag{9}$$

This equation is the same as Equation 7 except that the term which represents the amount of practice on a task is now divided between practice that is observed (P) and practice prior to observation (E). This form of the power function has been shown to provide a better fit to some practice data but is also no better than simpler functions (Equation 7) with other data (Singley & Anderson, 1985). The reason why Equation 9 is no better than Equation 7 for some practice data is that the assumption of prior experience with a task is too general an assumption. As demonstrated above, prior experience may only apply to some components of a task. Other components will not have had any practice. Hence the linear component model (i.e., Equation 8) may be a more accurate depiction of some situations.

The second moderating factor on the attenuation of learning rate is the relative amount of practice that the old and new skills underlying a task had prior to their combination. Figure 3 illustrates the effect on learning rate of increasing the amount of prior practice of old skills in the example introduced above. All of the data points in this figure were derived from the example situation with the number of old steps constant at 100 and the number of new steps constant at 20. The learning rates were calculated for the learning functions that resulted from varying the amount of practice of old skills in Equation 8. Figure 3 shows the result of varying P_{old} from 0 to 100 sessions. When the old skills have had no practice prior to combination the learning rate is simply the intrinsic learning rate of the system (-0.80). This corresponds to the situation where all components of the task are new. As the amount of practice of old skills increases the attenuating effect on learning rate increases until the combination of old and new skills has its maximum effect. In this situation, when old skills have been practised for 8 sessions prior to the combination of old and
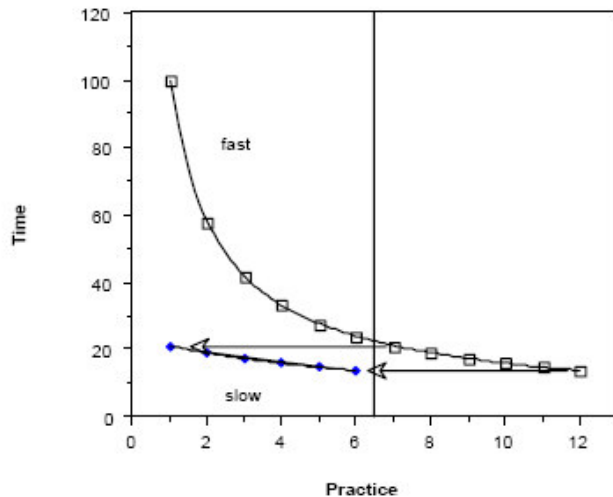
Figure 2: Demonstration that a practice function with a 'slow' learning rate may be the tail-end of a function with a faster rate. The last six data points of the fast curve (learning rate = -0.80) have been displaced 6 practice units. The new curve now has the slower learning rate of -0.239.

new skills, the resulting learning rate is at its slowest at -0.439. Beyond this point, increasing the amount of practice old skills have prior to combination has a diminishing effect on learning rate. This diminishing effect continues until the old skills have had so much practice that any further practice results in only negligible improvement. At this point the combination of old and new skills has no effect on the overall learning rate. The learning rate for the overall task will now be completely determined by the rate at which performance on the new skills improves, and this will be at the intrinsic learning rate of the system (-0.80 in this situation).

In summary, the linear component model leads to the prediction that when a task involves old and new components, this task will be learned at a slower rate than that at which each of the two sets of components improves (i.e., the constant intrinsic rate of each person). The amount by which this learning rate will be attenuated will be moderated by the relative number of steps between old and new components of the task, and by the amount of practice that the old skills had prior to learning the new task. Some evidence exists in the research literature to support these predictions. This evidence is presented below.

Evidence in Support of the Linear Component Model

In order to evaluate the linear component model, data is required from a very specific type of situation. Firstly, there needs to be performance data from both a training phase and a transfer phase in order to compare learning rates in both phases. Secondly, sufficient information about the tasks performed in the two phases is required in order to make judgements about the relative number of old and new skill components utilised in the training and transfer tasks. We found three studies in the research literature that provided sufficient information to enable a comparison between the data from these studies and predictions derived from the linear component model.

Singley and Anderson (1985)

Singley and Anderson (1985) examined the extent of transfer between different computer text-editing programs. Two basic types of editors were used: (1) line editors, where only one line of text in a file can be viewed at a time and editing is on a line-by-line basis; and (2) screen editors, where the screen is filled with the contents of a file and users are able to designate the location to be edited by moving around the screen with a cursor. Subjects in this study were trained to operate two line editors (ED and EDT) and one screen editor (EMACS). Singley and Anderson were interested in the extent to which training on one versus two line editors would transfer to performance on a screen editor. They found that there was positive transfer both between the line editors and from the line editors to the screen editor. These results were interpreted as suggesting that all of the editors shared a certain number of productions necessary for their performance. Thus training with one editor provided the subjects with a set of productions, of which some were useful for operating another editor. In more detail, there was almost complete transfer between the line editors, which suggests that these editors share a large number of productions. In contrast there was only partial

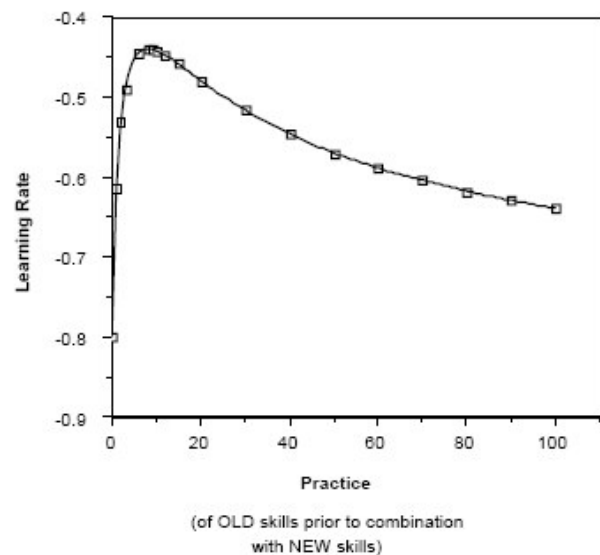(of OLD skills prior to combination with NEW skills)

Figure 3: Learning rate as a function of the amount of extra practice of old skills in comparison to new skills. The maximum attenuation of learning rate occurs when $P_{old} = 8$ and results in a learning rate = -0.439. The data points were generated from a learning example described in the text. The line represents an interpolation of the data points.

Table 1: Power functions describing performance with three text-editors during Training and Transfer phases of an experiment reported in Singley and Anderson (1985).

Editor		Performance	
Training	Transfer	Training	Transfer
EDT	ED	$T = 4.3\ P^{-0.53}$	$T = 3.7\ P^{-0.20}$
ED	EDT	$T = 4.8\ P^{-0.79}$	$T = 3.9\ P^{-0.34}$
ED/EDT	EMACS	$T = 3.9\ P^{-0.55}$	$T = 3.4\ P^{-0.41}$

transfer between the line editors and the screen editor, suggesting that the productions developed with the line editors were not totally sufficient for operating the screen editor. The development of more productions would have been necessary for efficient performance with this editor.

The most interesting result of this experiment for the current discussion concerns the differences in the shape of the learning curves during training and transfer. Singley and Anderson fitted simple power functions to the data and the equations of these functions are presented in Table 1. The design of the experiment was such that the subjects who performed in the Transfer phase with ED were trained with EDT and vice-versa. The subjects who performed in the Transfer phase with EMACS were subjects who had either been trained with one of the line editors or both, the total amount of training being equal for both groups of subjects. The equations in the Training column of Table 1 were derived from data from control subjects who practiced with only one editor.

It is clear from the equations in Table 1 that in all cases the learning rate during Transfer was slower than the rate during Training. Furthermore, the amount of attenuation was greater when transfer was between line editors (-0.53 to -0.20 and -0.79 to -0.34) than when transfer was from line editors to the screen editor (-0.55 to -0.41). These results are as would be predicted on the basis of the linear component model. First consider the line editors. The fact that there was almost total transfer between these editors suggests that they share a large number of productions. Therefore when subjects switch to one line editor after operating with the other line editor, very few new productions need to be developed. In other words, when these subjects operate the new line editor, underlying their performance will be a large number of old skills and a small number of new skills. As was shown in Figure 1, this is the type of situation where attenuation of learning rate is very large. In contrast, transfer from line editors to screen editors involves fewer shared productions. Thus subjects who trained with line editors and then switched to the screen editor would be able to use some of their previously developed productions but would also need to develop a relatively large set of new productions. As a result the ratio of old to new skills in this situation would be smaller than in the situation where transfer was between line editors. As shown in Figure 1, this predicts a smaller attenuation of learning rate in the former

than in the latter situation.

A similar interpretation of the Transfer results was suggested by Singley and Anderson. Initially they attempted to account for the different practice functions that were observed in Training and Transfer with Newell and Rosenbloom's (1981) general power function (i.e., Equation 9 above). However they concluded that the simple notion of including prior experience into the power function was not sufficient to account for the Transfer data. Singley and Anderson then suggested that in such transfer situations it was necessary to identify components of tasks that were general and specific. General components are those that are shared between tasks. Specific components are those that are peculiar to the particular task. Singley and Anderson then proposed that a more appropriate account of their transfer data would involve a power function that included two separate power functions, one for the general components and one for the specific components:

$$T = X + N_g\ P_g{}^c + N_s\ P_s{}^c \qquad (10)$$

This equation is equivalent in form to Equation 8 with an asymptote. Furthermore equating general components with old components and specific components with new components results in the two equations being equivalent in function as well. Unfortunately Singley and Anderson did not evaluate whether Equation 10 provided a better account of their transfer data than Equation 9. Furthermore, there was no explicit discussion of the implications of this conceptualisation of transfer on learning rates.

MacKay (1982)

MacKay (1982) reported data combined from two earlier studies (MacKay & Bowman, 1969; MacKay, 1981) in which the speed of reading sentences aloud was measured. Three types of sentences were examined: normal, scrambled and nonsense. The scrambled sentences were derived from the normal sentences by re-arranging the order of the words. The nonsense sentences were derived in turn from the scrambled sentences by substituting or rearranging letters in words to form pronounceable non-words.

MacKay reported two main results. The first was that normal sentences were read faster than scrambled sentences, which were read faster than nonsense strings. The second result was that the rate at which subjects improved their reading speed with practice was a function of the type of sentence being read. Subjects improved at the fastest rate

with nonsense strings, at a slower rate with scrambled sentences, and at the slowest rate with normal sentences.

The difference in learning rates can be accounted for by considering, at a fairly superficial level, the skills that underlie reading aloud the three types of sentences. First consider the normal sentences. Competent readers rely on skills that (1) convert familiar words into sounds, and (2) use syntactically based meaning to increase fluency (e.g., noun phrases such as "the dog" are read faster than "the" and "dog" read separately, Just & Carpenter, 1980). With adult readers these skills are unlikely to improve much with practice. Hence any improvement observed with this type of sentence will be at a very slow rate.

Now consider the scrambled sentences. Subjects who read such sentences aloud now perform without the benefit of the second type of skill involved in normal reading. That is, because the scrambled sentences did not contain familiar noun or verb phrases, words can only be processed as individual meaning units rather than as part of higher level concepts. Therefore with this type of sentence, subjects would be required to develop some other strategy to increase fluency beyond the level at which they can read a list of unrelated words. This suggests that reading these sentences results in the combination of old and new skills. The ratio of old to new skills with these sentences will be smaller than with the normal sentences and so the attenuation of learning rate will be less than with the normal sentences. Thus learning will be faster with the scrambled sentences than with the normal sentences.

With the nonsense sentences subjects are at even less of an advantage in terms of being able to use already established skills. Because these sentences were derived from the scrambled sentences there is no syntactically based meaning to increase fluency. In addition the sentences contain no familiar words and so pronunciations are not as easily accessible as with real words. Instead subjects must rely on pronunciations derived from the pronunciations of real words that look similar. This derivation process is unlikely to be a well-practiced skill. Without many other old skills to influence performance with these sentences, it seems plausible that the ratio of old to new skills in this condition is lower than in both of the previous conditions. Therefore the combination of old and new skills in reading nonsense sentences will have little attenuating effect on learning rate and so subjects will improve with these sentences faster than with the other types.

The MacKay results support our claim that a consideration of the relative number of old and new skills used to perform a task according to the relationship depicted in the linear component model will enable predictions about the relative learning rates in tasks with different ratios of old to new skills.

Snyder and Pronko (1952)

It appears that the linear component model can also provide a good account of the acquisition of perceptual-motor skills. Snyder and Pronko (1952) investigated the effect of wearing inverting lenses (i.e., these make the world look upside down) on performance in a motor task. For 15 days subjects were trained with normal vision on the Purdue pegboard task - a task that requires precise visual control of motor responses. Following this training phase, the subjects were then given 27 days practice on the same task whilst wearing inverting lenses. This latter condition had the obvious effect of slowing performance compared to the normal vision condition.

A more interesting result concerns the rate at which performance improved in the two conditions. The practice data from this experiment has been re-plotted in Figure 4. The lines represent the best-fit power functions of the form described by Equation 7. It is clear from this figure that, although performance with inverted vision never attains the speed of that with normal vision, the performance in the former condition improves at a faster rate (-0.20) than in the latter condition (-0.10).

The differences in learning rate can be accounted for by assuming that performance in the inverted vision condition involves the same skills as in the normal vision condition plus a new set of skills that cope with the inverted visual information. Performance in the normal vision condition can be considered to involve both old and new skills. The old skills are concerned with co-ordinating visual information and motor responses. The new skills are more task related and concern rules specifying peg placement. In the same vein, the inverted vision condition involves the same old and new skills plus another set of skills concerned with translating the inverted visual information into normal orientation information. This means that subjects can adapt existing skills to operate in the new environment rather than having to develop a whole set of new skills. It also implies though that more new skills underlie performance in the inverted vision condition than in the normal vision condition. The fact that the performance on the first day of inverted vision was almost 60% slower than on the first day of normal vision suggests that the skills to be learned in order to cope with the inverted vision substantially outnumber the skills needed to place the pegs. Thus the ratio of old to new skills in the inverted vision condition is smaller than in the normal vision condition. The prediction can now be made on the basis of the linear component model that the combination of old and new skills in this task will result in a greater attenuation of learning rate in the
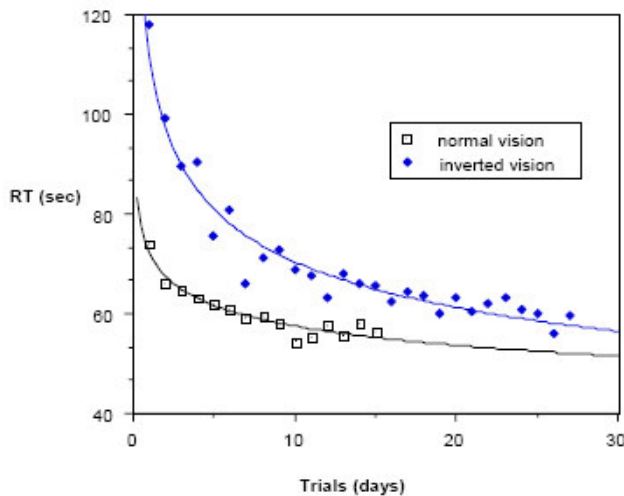
Figure 4: Data reported in Snyder and Pronko (1952). Data points have been plotted on linear axes. The lines represent the best-fit power functions with the following equations:

$$\text{Normal Vision: } T = 72.418 \, P^{-0.10}, \, r^2 = 0.913 \, ;$$
$$\text{Inverted Vision: } T = 111.98 \, P^{-0.20}, \, r^2 = 0.935.$$

normal vision condition (see Figure 1). This then accounts for Snyder and Pronko's finding that performance with inverted vision improved at a faster rate than with normal vision.

Summary

The studies discussed in this section, when considered as a whole, provide consistent support for the notion that the combination of old and new skills can affect the rate at which a task is learned. The way in which learning rate is affected depends on the extent to which old components have been practiced prior to combination with the new components, and the relative number of steps involved in the old and new components. One of the studies discussed above (Singley & Anderson, 1985) showed a slowing of learning rate from Training to Transfer. Another study (Snyder & Pronko, 1952) showed an increase in the learning rate from Training to Transfer. The MacKay (1982) data did not come from a transfer design but did show differences in learning rates between conditions which were accounted for by considering the relative amounts of practice and number of steps involved in the underlying components.

Task Learning Curves as Summary Functions

One of the major implications of the linear component model is that learning rate, as exhibited on a particular task, is by no means a constant of the cognitive systems of research participants. It is clear that overt learning rate can be affected by the combination of old and new skills. In addition, some research suggests that learning rate is also affected by factors such as motivation and anxiety

(Speelman & Kirsner, 2005). Thus, as suggested above, there is evidence that is counter to what is implied by Anderson's (1982) account of the power law of learning, that the rate at which performance on a task improves is a constant of the learning system. It may well be the case, though, and was an assumption of the argument developed above, that new component skills that contribute to overall performance on a task are acquired at a rate that is a constant for any individual.

The fact that the rate at which performance on a task improves can be a function of previous experience with components of the task is not surprising. Certainly it would be expected that learning a task would be facilitated if a trainee is familiar with parts of the task. However, what is important to realise is the apparent distinction between the learning of a task and the learning of components underlying performance of the task. The linear component model predicts that a combination of old and new skills will affect learning rate. This effect involves an attenuation of learning rate compared to the rate at which the old task components were originally learned. Thus the learning rate of a 'new' task will be different to that of a previously observed new task, that is, the task in which the old task components were learned. The significance of this finding is that the rate of improvement on a task is determined by the history of the components that underlie performance of the task, rather than experience with the task itself. Thus variability in task learning rates appears to result from the combination of underlying component skills with varying application histories that all improve at the same rate. Therefore, contrary to what is implied by Anderson's account of the power law of learning, the rate of improvement on a new task is not directly determined by parameters of the cognitive system. Instead it is the rate of improvement of component skills that underlie performance of the task that is determined by the parameters of the cognitive system. Only when performance on a task relies on the execution of completely new skills will the task learning rate be the same as the cognitive system's learning rate. However, given that for adults most tasks involve the execution of component skills that vary in the extent to which they have been practiced, improvement on a new task is unlikely to ever be at the same rate at which the underlying components improve.

It should be noted that the argument about the attenuation of learning rates that results from the combination of old and new skills should not be construed as a claim that transfer leads to someone becoming a slower learner. Certainly part of the claim is that the rate of improvement on a task in which some of the requirements have been mastered before will be slower than if the task

was entirely new. But this slower improvement rate is on the overall task. New task components will be learned at the rate that someone normally learns new things. That is, someone's capacity to learn new things is unaffected by the transfer situation. It is only when measurement of performance takes into account performance on both old and new task components is a slower overall learning rate observed. In other words, learning rate on the overall task is attenuated, but learning rate on the new task components is unaffected.

The picture of skill acquisition that emerges from this discussion of learning rates is one where skilled performance involves the execution of a number of component skills. These component skills may be represented cognitively as production rules (although there is nothing in the above account that implies this constraint). When a number of new productions are executed together to perform a task, performance will improve according to a power function that has a learning rate determined by the parameters of the cognitive system. If the task conditions applicable to the successful execution of these productions continue to be present, then performance on the task will continue to improve according to the same power function. This will be the case even if task conditions change, as long as the appropriate conditions for the execution of the productions remain in the stimulus environment or are produced by the execution of other productions. Thus collections of productions can develop the appearance of skill modules, where changes in task conditions do not affect their execution, nor the pattern of their improvement. New productions may develop alongside these already well-practiced productions as task conditions dictate. These new productions will improve according to a different power function, although one with a learning rate again determined by the parameters of the cognitive system. In other words, new productions will improve at the same rate as the one that describes the improvement of old productions. Thus old and new productions will improve together according to their own learning curves. Although these learning curves will have the same basic learning rate, the momentary learning rates of the two sets of productions will be different because these productions are at different points along their learning curves. Performance on the task as a whole will improve at a rate that is not the same as the learning rate of new productions. In fact, the learning rate on the total task will be a function of the combination of the separate learning curves that describe the improvement of the different sets of productions that underlie performance on this task. Thus the power function that describes improvement on the total task can be seen as an aggregate of the learning curves of the components that underlie

performance of this task. In other words, just as performance time on the total task is a function of the time to execute all of the productions that contribute to the overall performance, so is the learning curve for performance on the total task a function of the learning curves describing improvement in all of these productions.

Although the simple addition of two power functions that describe improvement in old and new skills provides a reasonable account of changes in learning rate, more complicated tasks than those examined so far will no doubt involve more complicated combinations of functions. The large variation in learning rates for different tasks that is typically observed, both between and within subjects, may be a result of examining performance on tasks that involve various combinations of skills with varying practice histories. This apparently simple explanation for variations in learning rate belies the complexity that underlies the learning of a task. For example, it would be an immensely difficult exercise to determine the relative contributions to the overall pattern of improvement in a task that are made by, on one hand, the number of skills involved in various task components and, on the other hand, the practice histories of these skills.

The fact that new tasks rarely involve the execution of only new skills calls into question the ability of simple power functions to account for improvement with practice. Although such functions may be able to describe the improvement that is observed, the parameters of such functions will not be accurate reflections of features of the components underlying performance. For instance, in Anderson's (1982) account of the power law of learning, the intercept value of a power function (i.e., N in $T = N P^c$) was said to be proportional to the number of productions executed in the performance of a task. However, this will only be the case if all of the productions being executed are new. If some of the productions are old, and therefore are further along their learning curves than the new productions, then this coefficient will underestimate the number of productions being executed. Similarly, Newell and Rosenbloom's (1981) revised version of the simple power function (i.e., $T = N (P + E)^c$) may be able to account for improvement on a task that relies on the execution of old productions only, but all of these productions must have been practiced to the same extent for the parameters of this function to be informative (i.e., E must be the same for all of the productions). Therefore when a task involves the execution of productions with varying practice histories, as most tasks will, the best way to describe improvement on the task is to combine separate power functions that describe the improvement of the various sets of productions. The difficulty however is in estimating a trainee's experience

with particular task components and the number of productions that are executed to perform these components. The degree to which such an exercise is undertaken will depend on the accuracy that is desired from a function designed to describe improvement on the task.

The argument presented above, that learning curves reflect summaries of learning curves of component skills, was derived from a discussion of learning rates. Others have made similar claims about learning curves, but have reached these conclusions from different directions. As mentioned earlier, Heathcote et al. (2000) argued that power function learning curves were only found in averaged group data, and that individual learning curves were best described by functions that represented the combination of exponential and power functions. Also mentioned earlier was Singley and Anderson's (1985) suggestion that the type of transfer observed in some situations could only be understood by considering the skills that were specific to particular tasks and skills that were general to several tasks. As a result, they proposed that performance improvements on a transfer task should be described by a function that combined separate learning functions for the specific and the general skills. Some researchers (e.g., Delaney, Reder, Staszewski & Ritter, 1998; Rickard, 1997, 1999) have reported that, in some tasks, people utilise different strategies at different times. These researchers found that the resultant learning curves were best described by functions that included separate functions for each strategy. There is evidence, then, from various sources to support the claim that when people perform new tasks, they utilise old skills and develop new skills. The combined improvement on all of these skills that occurs with practice on the new task is the improvement that is observed on the new task.

It is important to note that the implications of the linear component model still apply regardless of the form of the learning functions that describe improvement of component skills. That is, if learning proceeds according to a power function, an exponential function, or some combination of these (Heathcote et al., 2000), combining learning functions that differ in their intercept values (i.e., starting times) and/or amounts of practice will result in an overall function with a learning rate that differs from the learning rates of the component functions. This result is due to the combination of negatively accelerated functions. The momentary learning rate at any point along such curves is different to any other point along the curve. When a task learning curve is a summary of the learning curves of several component skills, and the components are at varying points along their learning curves (i.e., different practice histories), then the task learning curve will reflect the combination of several momentary learning rates. Furthermore, the difference

between task learning rate and the learning rate of components will occur regardless of the form of the underlying learning curves, as long as they are negatively accelerated.

One important implication of the linear component model is that all human performance is the product of the operation of many component processes. The level of performance exhibited at any one moment is the result of the execution history of each component process, and the ability of the person to develop new processes when necessary. In this view every performance episode becomes a transfer situation (cf, Singley & Anderson, 1989) - in familiar situations old skills are utilised as is, whereas in new situations some degree of adaptation is necessary. Thus transfer determines the form of the linear component model that applies in any particular situation. That is, the degree to which transfer occurs in a particular situation and the nature of this transfer determines the components that comprise the specific version of the linear component model that applies to this situation. Clearly then, any factor that affects the nature and scope of transfer also imposes constraints on the nature of the linear component model. There are many prominent determiners of transfer and these have been considered elsewhere (Speelman & Kirsner, 2005), including context, which has important effects on the extent of transfer observed and hence the shape of learning curves.

Conclusions

This paper has demonstrated that the level of skill exhibited in any particular behaviour is fundamentally a matter of transfer. Just about all behaviour is composed of many component processes contributing to the overall performance. Skills developed in different contexts are recruited into service in the performance of new tasks. The absolute level of performance exhibited on a new task is determined by the extent to which old skills can be executed in the new context, and the extent of prior practice with these old skills. Furthermore, the rate at which performance of the new task improves is determined by several factors: (i) the performer's intrinsic learning rate; (ii) the relative proportion of old and new skills used to perform the task; and (iii) the extent of prior practice with old skills prior to the combination with new skills. The degree to which any of these transfer features can determine performance levels will be further complicated by: (i) the nature of old and new tasks; (ii) the nature of past and present training; and (iii) performance context. At present no theory attempts to take all of these factors into account in explaining performance, but we have commenced development of such a theory – the component theory of skill acquisition and transfer (Speelman & Kirsner, 2005). It seems reasonable to say,

however, that learning curve parameters, and in particular learning rate, are not constants of the cognitive system, but instead are variables that come under the influence of many intra and extra-individual factors. In any one situation, however, it may be safe to assume that a relatively constant intrinsic learning rate applies. Certainly though, Snoddy's (1926) belief of one learning rate that applied to every person and task is a fantasy.

References

Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369-406.

Anderson, J.R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J.R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J.R., Conrad, F.G., & Corbett, A.T. (1989). Skill acquisition and the LISP tutor. *Cognitive Science*, 13, 467-505.

Anderson, J.R. & Lebiere, C. (1998). (Eds.), *Atomic components of thought*. Mahwah, NJ:Erlbaum.

Carlson, R.A. (1997). *Experienced cognition*. Mahwah, NJ: Erlbaum.

Carlson, R.A., Sullivan, M.A. & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 517-526.

Crossman, E.R. (1959). A theory of the acquisition of speed-skill. *Ergonomics*, 2, 153-166.

De Jong, J.R. (1957). The effects of increasing skill on cycle-time and its consequences for time-standards. *Ergonomics*, 1, 51-60.

Delaney, P.F., Reder, L.M., Staszewski, J.J., & Ritter, F.E. (1998). The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science*, 9(1), 1-7.

Fitts, P.M. (1964). Perceptual-motor skill learning. In A.W. Melton (Ed.), *Categories of Human Learning* (pp. 243 - 285). New York: Academic Press.

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

Grose, E.M. & Damos, D.L. (1988). Automaticity and the transfer of mental rotation skills. *Proceedings of the Human Factors Society – 32nd Annual Meeting*.

Heathcote, A.S., Brown, S., & Mewhort, D.J.K. (2000). The power law repealed: The case for an exponential law of practice. *Psychological Bulletin & Review*, 7(2), 185-207.

Josephs, R.A., Silvera, D.H. & Giesler, R.B. (1996). The learning curve as a metacognitive tool. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 22, 510-524.

Just, M.A. & Carpenter, P.A. (1980). A theory of reading: From eye fixation to comprehension. *Psychological Review*, 87, 329-354.

Kirsner, K. & Speelman, C. (1996). Skill acquisition and repetition priming: One principle, many processes. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 22(3), 563-575.

Logan, G.D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492-527.

MacKay, D.G. (1981). The problem of rehearsal or mental practice. *Journal of Motor Behavior,* 13, 274-285.

MacKay, D.G. (1982). The problems of flexibility, fluency and speed-accuracy trade-off in skilled behavior. *Psychological Review*, 89, 483-506.

MacKay, D.G. & Bowman, R.W. (1969). On producing the meaning in sentences. *American Journal of Psychology*, 82, 23-39.

Myung, I.J., Kim, C. & Pitt, M.A. (2000). Toward an explanation of the power law artifact: Insights from response surface analysis. *Memory & Cognition*, 28(5), 832-840.

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

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.

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

Rosenbloom, P.S. & Newell, A. (1987). An integrated computational model of stimulus-response compatibility and practice. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol.21, pp. 1-52). San Diego: Academic Press.

Singley, M.K. & Anderson, J.R. (1985). The transfer of text-editing skill. *International Journal of Man-Machine Studies*, 22, 403-423.

Singley, M.K. & Anderson, J.R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Smith, K.V. & Sussman, H. (1969). Cybernetic theory and analysis of motor learning and memory. In E.A.Bilodeau & I.Mc.Bilodeau (Eds.), *Principles of skill acquisition*. New York: Academic Press.

Snoddy, G.S. (1926). Learning and stability. *Journal of Applied Psychology*, 10, 1-36.

Snyder, F.W. & Pronko, H.H. (1952). *Vision with spatial inversion*. Wichita, Kansas: University of Wichita Press.

Speelman, C.P. (1991). *Skill acquisition: Estimating the contributions of old skills to performance on new tasks*. Unpublished doctoral dissertation, University of Western Australia, Nedlands, Australia.

Speelman, C.P. (1995). The Shape of Learning Functions During Transfer. In P. Slezak, T. Caelli & R. Clark (Eds.), *Perspectives on Cognitive Science: Theories, Experiments & Foundations* Norwood, NJ:Ablex (pp 79-102).

Speelman, C.P. & Kirsner, K. (1993). New goals for HCI training: How to mix old and new skills in the trainee. *International Journal of Human-Computer Interaction*, 5, 41-69.

Speelman, C.P. & Kirsner, K. (2001). Predicting transfer from training performance. *Acta Psychologica*, 108, 247-281.

Speelman, C.P. & Kirsner, K. (2005) *Beyond the learning curve: The construction of mind*. Oxford: Oxford University Press.

Staszewski, J.J. (1988). Skilled memory and expert mental calculation. In M.T.H. Chi, R. Glaser & M.J. Farr (Eds.), *The nature of expertise*. Hillsdale, NJ: Erlbaum.