A Theory of the Mind as a Complex System

[1]Craig P. Speelman, [2]Kim Kirsner and [3]Stephan Halloy

[1]School of Psychology and Social Science, Edith Cowan University, Australia
[2]School of Biology, University of Western, Australia
[3]The Nature Conservancy, Santiago, Chile

ABSTRACT

Five principles of skill acquisition are presented based on a review of research on human learning and expertise. Essentially these principles state that practice leads to faster and more efficient uses of knowledge. This enables faster performance and results in less demand on mental resources. In turn these outcomes enable higher level behaviours to be attempted. Ultimately skills are developed through refinement of many component processes. A theory of the mind is proposed that borrows from theories of complex adaptive systems. In this theory, the mind is conceived of as consisting of agents that compete for resources associated with processing information. The nature of this competition is similar to that observed in physical and biological systems in that agents survive or disappear depending on their usefulness. This theory is shown to be capable of explaining the five principles of skill acquisition, without these principles being explicitly built into the theory. Implications for other theories of skill acquisition are considered.

Keywords: Skill Acquisition; Complex Systems; Cognition, Mind; Mental Resources, Biological Systems

1. INTRODUCTION

In the field of Cognitive Psychology most explanations of cognitive phenomena involve the proposal of systems without any consideration for the origins of these systems. We consider this to be a problem for the field because such attempts at explanations can be considered to be merely re-descriptions of the phenomena. An obvious solution to this problem, then, is that the origins of systems should always be considered and in the case of human cognition, we suggest that learning should play a central role in the development of cognitive systems. This paper represents a proposal for how a complex systems approach can provide an account for how humans learn and how cognition can develop through experience with the world.

The paper is divided into several sections. In the first section we describe five principles of learning that we have identified from research into human skill acquisition. We then propose a theory of how the mind develops-the Component Theory of Skill Acquisition-that adopts complex systems principles proposed by Halloy (1998) and Halloy and Whigham (2004). Finally we show that such a theory can account for the five principles of learning and importantly, without building those principles into the theory.

1.1. Principles of Skill Acquisition
1.2. Principle 1: Practice Leads to Faster Performance

Principle 1 is the most obvious feature of learning. When something has been learned from a previous experience and it can be utilised at some later moment in time, performance at that later moment is typically faster than previous performance. This principle applies to all aspects of behaviour, not just overt behaviour such as performance of a task. For instance recognition of experiences as familiar is faster as the number of recognition attempts increases (Pirolli and Anderson, 1985). Perception of objects is faster with increased experience (Crovitz *et al.*, 1981). It is probably not unreasonable to suspect that this principle reflects a basic characteristic of neural functioning (Altmann *et al.*, 2004; Barnes, 1979; Bolger and Schneider, 2002).

A common explanation for the effect that practice has on the speed of performance is that practice leads to faster,

Corresponding Author: Craig P. Speelman, School of Psychology and Social Science, Edith Cowan University, Australia

more reliable activation of knowledge structures. This effect of practice is often referred to as strengthening in theories of skill acquisition and memory (Anderson, 1982). That is, as access to an item in memory increases, the representation is in some way strengthened, which means that it becomes easier to access (i.e., access is faster and more reliable) and more resistant to forgetting (Anderson, 1983; Pirolli and Anderson, 1985). So, another version of Principle 1 is that practice leads to strengthened knowledge structures. When performance relies on access to memory representations, the greater the strength of those representations, the faster will be the performance.

An important feature of the way in which performance speed improves with learning is that it is negatively accelerated. That is, performance improvements are typically dramatic early in practice and taper off as practice increases until some asymptote is reached. There is a great deal of debate as to the particular mathematical function that best describes this pattern of improvement. For instance, many people claim that learning curves are best described by power functions-the so-called power law of practice (Anderson, 1982; Logan, 1988; Newell and Rosenbloom, 1981). Others suggest that exponential functions provide a better description of learning curves (Heathcote *et al*., 2000; Josephs *et al*., 1996; Rosenbloom and Newell, 1987). There are even suggestions that learning curves represent summaries of several component learning curves that each vary in their form (Heathcote *et al*., 2000). For our current purposes it is not important which function provides the best description of a learning curve. Only the negatively accelerated feature of learning curves is crucial for our argument.

1.3. Principle 2: Practice Leads to Efficiencies in Knowledge Access

According to Principle 1, practice leads to faster performance. As mentioned above, one of the reasons for this is that repeated access to memory representations strengthens these representations and this facilitates further access, which in turn facilitates performance that relies on this access. Another reason why practice leads to improved performance speed is what we identify as Principle 2: When people practice a task, the way in which they perform the task changes. Typically practice leads to a more efficient form of processing. This gain in efficiency can be characterised as a move from beginner level performance, which involves some deliberation about what responses are required, to mastery level performance, which is marked by immediate recognition of a situation and knowing the appropriate responses.

There are a number of different theories as to how practice leads to this sort of improvement in the efficiency of processing. According to the ACT theory

(Anderson, 1982; Anderson and Lebiere, 1998), improvement in processing is a result of practice leading to a reduction in the number of processing steps through either deleting unnecessary processing steps or collapsing a number of simple processing steps into fewer more complicated processing steps that have the same effect as the original steps. The proponents of the SOAR theory (Newell, 1990) make similar claims. According to the Instance theory (Logan, 1988), the efficiency of processing is improved by moving from a situation where processing steps are executed in a serial manner to another situation where stimulus conditions trigger the appropriate response without any intervening deliberation. In other words, where people may originally engage in a process of generating a solution to a problem, eventually with practice, when the same problem is presented the appropriate solution is retrieved directly from memory.

Although the various theories propose different means by which practice leads to more efficient processing, all of the theories lead to the same prediction: With sufficient practice of a task where the stimulus-response relationship is consistent, performance will eventually reach the stage where perception of a known stimulus will trigger an automatic response (i.e., seeing "3 x 4 = ?" will automatically lead to a response of "12"). It is clear that moving from a situation where several processing steps are required before a response is generated to a situation where a stimulus invokes a response automatically will result in considerable savings in the amount of time to perform the task. But this process does not only save time. One view of the result of this process is that people are able to make use of consistencies in the world to set up short cuts in the ways in which they deal with the world. A problem and all of its associated stimuli, goals and processing steps necessary to lead to a response, can be represented mentally in a compressed form, much like a shorthand version of a word.

Research on expertise is full of examples where the acquisition of expert knowledge is accompanied by a change in the way the domain of expertise is perceived. Experts perceive particular configurations of stimuli like most people recognise words. That is, the configurations are recognised automatically as meaningful and depending on the goal of processing, may be relevant to decisions about appropriate responses. In contrast, novices behave like someone who is learning to read and is just able to recognise that certain visual patterns (e.g., letters and letter combinations) are relevant to this task. That is, novices are typically barely able to recognise as relevant components of those configurations that experts automatically appreciate as a whole. The clearest example of this feature of expertise comes from the domain of chess. A common memory task that is used to

examine expertise differences in chess involves presenting people with a chess board and a configuration of pieces that corresponds to some point in the middle of a game. Exposure to this configuration is usually restricted to a few seconds. The board is then covered up and the subject's task is to reproduce the configuration on another board. Accuracy in this task is a direct reflection of chess expertise. Novices can typically reproduce 33% of a configuration. Players of an intermediate standard can reproduce 49% and players of the Grand Master level can reproduce 81% (Chase and Simon, 1973). Importantly, these differences in accuracy are not observed when players are presented with random configurations of pieces. The usual explanation for these results is that, as players gain expertise, they learn to associate certain configurations of pieces with the state of the game. For instance, there are attack configurations that pose danger for the opposition's queen and there are defence configurations designed to protect the king. So, the acquisition of chess expertise is associated with the ability to automatically recognize particular meaningful configurations of chess pieces. Evidence for this view comes from the fact that when players are observed attempting to reproduce mid-game configurations in the memory task, novices tend to position pieces in groups that typically are related in superficial ways, such as they all appeared in the same region of the board, or they all had the same colour. In contrast, chess masters positioned pieces in groups that corresponded to meaningful configurations (e.g., attack, defence). These configurations were not restricted to pieces of the same colour or the same region of the board but would involve pieces of both colours that may have been positioned some distance from each other. Thus chess experts appear to be "reading" a chessboard in terms of groups of pieces that correspond to meaningful configurations-indeed they may even have names for these configurations (e.g., "The Nimzo-Indian Defence", "The King's Gambit" and "The Giuco Piano- see Kasparov, 1985).

Many other domains (e.g., viewing X-Ray slides of the human body: Lesgold *et al.*, 1988; physics problems: Larkin, 1983; photographs of basketball games: Allard and Burnett, 1985; memory for plays in sport: Allard and Stakes, 1991) reveal the same sorts of perceptual differences between experts and novices. Such research on expertise has a number of features in common. Experts 'see' things differently to novices. This difference in perception appears to be related to the ability to identify configurations of stimuli as representing meaningful wholes rather than as groups of individual stimuli. These configurations are meaningful because they are related to the purpose of the processing they perform in their domain of expertise. That is, after

years of experience in a domain and years of practice performing particular tasks, certain configurations of stimuli automatically trigger responses and these responses could be in the form of actions, decisions, thoughts, or production of a label. Thus the automatic response to a stimulus configuration is a large component of expert performance. That this feature appears to be a characteristic of all areas of expertise illustrates the pervasiveness of Principle 2.

1.4. Principle 3: Learning Leads to Less Demand on Working Memory

The combination of Principles 1 and 2 can be considered as comprising a third principle. The ability to recognise groups of stimuli as meaningful and to do so automatically, results in a freeing up of mental resources. In particular, working memory is often described as being limited in capacity such that it can only hold a certain number of items at any time (e.g., 7 ± 2 items according to Miller, 1956). Principle 3 states that learning will lead to a situation whereby this capacity limit can be circumvented. For example, if an English speaker is presented briefly with the following set of letters:

uremfolta

and then asked to recall them, they might have difficulty remembering all of the letters, particularly in the correct order. If the letters were rearranged, however, into the following order:

formulate

memory for the set of letters is likely to be far more accurate. With the first set of letters, there are nine separate pieces of information, a number that may exceed most people's working memory capacity. In the second set, however, because we recognise the letters as together forming a word, the nine pieces of information can be processed as one piece of information. In this way, working memory will only contain one piece of information-a pointer to an item in our long term memory for words-which means that working memory will have spare capacity for any other information that is to be held there. Presumably, then, another six or so sets of letters that are similarly arranged into words could be held in working memory without seriously affecting the ability to recall them. In this way we could conceivably recall 63 letters or more (i.e., 7 words at 9 letters each) without too much trouble. Thus, we are able to make use of our knowledge of words to apparently circumvent the normal capacity limits on working memory. This phenomenon is often referred to as 'chunking' in the

memory literature (Baddeley, 1990; Miller, 1956). It refers to the ability to interpret information in meaningful chunks and it is the chunks that make up the limited number of items of information that can be held in working memory.

According to Principle 3, then, learning within a particular domain often leads to the ability to automatically process information in that domain in ways that result in fewer demands on working memory. An everyday example of this principle in operation is the use of acronyms, such as ASIO, CIA, STM and RAM. Acronyms represent small sets of letters that correspond to the first letter of each word in a set of words (i.e., ASIO for Australian Security Intelligence Organisation). Thus a smaller number of letters can be used in place of a larger set. After sufficient experience with an acronym, the acronym can take on the meaning of the set of words it represents. Thus seeing "ASIO" will invoke the same associations as "Australian Security Intelligence Organisation" but without the necessity to process four words. Thus small numbers of letters can represent much larger sets of letters. In this way less information is processed and more cognitive resources are available to process other information. Further motivation for adopting this linguistic convention is that it becomes a more efficient mode of communication (i.e., less time is spent on the same concept) and saves cognitive effort (Anderson *et al.*, 1998).

The study of expertise has highlighted many areas where chunking occurs and where the nature of the chunks (size, complexity) is related to the degree of expertise attained by an individual. For example, in the memory task used to examine the cognitive processing associated with playing chess, novices and masters place pieces on the board in distinct groups, suggesting they have processed particular pieces together and remember them as chunks. In addition, novices and masters recall the same number of such chunks, suggesting that the two groups of players are subject to the same capacity constraint on working memory. Importantly, though, the number of pieces in each chunk is greater for the masters than for the novices (3.8 pieces vs. 2.4 pieces, Chase and Simon, 1973). Hence masters can remember a greater total number of pieces than novices. A similar observation has been made with expert waiters (Ericsson and Polson, 1988).

Principle 3 has important implications beyond the fact that expertise in a domain can result in sometimes extraordinary memory skills for information in that domain. One of these implications is that, by developing a strategy whereby large amounts of information can be processed with only a small amount of working memory resources, the expert has

considerably more working memory capacity available for other forms of processing than is the case with a novice. As a result, the expert is capable of a greater level of complexity in their behaviour than the novice. This phenomenon we label as Principle 4.

1.5. Principle 4: As Expertise Increases, Fewer Mental Resources are Required to Perform a Particular Task, Enabling the Development of a Hierarchy Of Skills

The first three principles, taken in combination, characterise learning as leading to a situation whereby more and more of the knowledge that underlies performance can be retrieved faster and more reliably as expertise increases. As a result an expert has more knowledge at their mental fingertips that can be accessed quicker than the novice. Furthermore, this increased accessibility of expert knowledge frees up mental resources for other forms of processing. Certainly this characterisation of the attainment of expertise matches the common experience that when embarking on a new task (e.g., driving a car) we can often feel so overwhelmed by the various elements of the task that require our attention that we feel as if we cannot do the task at all. Eventually, though, with increased experience, the task seems to get easier. The task is not changing, of course, we are. We slowly gain more of the knowledge that is required about how to perform the task and our ability to use this knowledge increases. Ultimately we reach the stage where the knowledge is executed automatically and we can feel as if performing the task requires no effort whatsoever. Thus someone who has been driving a car for ten years or more probably engages so few mental resources for the actual operation of the car that they have plenty of resources available for increased vigilance on the road (and so are involved in fewer accidents than novice drivers, Adams, 2003) and are capable of performing other tasks while driving (e.g., singing along to the radio, conducting conversations, planning a new route to avoid a traffic jam) that have little impact on the driving task itself. Principle 4, then, identifies the fact that as expertise is acquired in a domain, more and more mental resources become available and so further development of behaviours becomes possible.

Everyday life is full of examples where increased experience with a particular domain or task leads to a transition through a hierarchy of skills. Infants who can barely comprehend or produce language, or orient themselves in three-dimensional space eventually learn to communicate with speech and text and may even learn to pilot a plane. The distance between novice and expert performance in these domains is clearly great, but so is

the amount of time and opportunity for practice. The amount of improvement on a novel task that can be observed in one hour in the laboratory can be anywhere from 50-90% (e.g., dropping from 41 seconds per trial to 15 seconds, Speelman and Kirsner, 2001, Exp.1). Given years of practice to master the many components of a complex task, vast leaps in performance levels are possible. Principle 4 suggests that the acquisition of adult-level skills is a matter of learning component skills to a level of performance that enables sufficient mental resources to be made available for the development of a new set of component skills. According to this principle, then, complex behaviours should develop in a stage-like manner. When component skills are new, resources will be used to cope with the demands of the task. As component skills improve with practice, a level of mastery of the task is reached such that fewer resources are required to perform the task. The freeing up of mental resources makes possible the performance of higher level behaviours, which may require the development of a new set of component skills. By proposing Principle 4 as a principle of learning, we are making the strong claim that the acquisition of skills, such as language comprehension and production and performance of mathematical operations, from infancy to adulthood, should be characterized by clear stages in development, not necessarily related to biological maturation, where the trajectory is through a hierarchy of behaviours, from low level to higher level behaviours, where mastery of some behaviours must always precede development of other behaviours and performance within a stage will be marked by improvement of component skills without necessarily any improvement in the overall target behaviour. Thus the degree of improvement apparent at any point in time will depend on the level of granularity of the analysis of behaviour. At a high level, improvement may appear discontinuous but at a lower level improvement may be gradual but continuous.

1.6. Principle 5: Mastery in a Domain Involves the Application of an Array of Component Processes, with Varying Degrees of Specificity to Tasks and Contexts. These Processes are Recruited in a Manner that Allows for Consistent Performance under Stereotypical Situations and Flexible Performance Under Unusual Circumstances

Principle 5 expresses an assumption that underlies the previous four principles. That is, many behaviours reflect the execution of a vast array of component processes. Component processes range from those that are developed specifically for the particular behaviour being performed,

to those that are useful across a broad spectrum of behaviours. The extent to which skills are specific to a particular context is determined by several factors, but in essence, people adapt to a task situation and their skill reflects the nature of this situation. According to Principle 5, then, all behaviours involve a transfer situation, where the level of performance is determined by the extent to which existing component processes can be recruited and new component processes need to be developed for the task at hand. Furthermore, the time to perform a task is a sum of the time to execute the component processes necessary for performance of the task. When performance of the task commences, old component processes will be some way along their own learning curve and new component processes will be at the beginning of their particular learning curves. The learning curve exhibited for performance of the task, then, will reflect a combination of the component learning curves.

Some behaviours will involve component processes that are applicable across a wide range of domains. Reading skills, for instance, are recruited by a vast array of tasks facing adults. As a result of at least twenty years of reading in a large number of contexts, most adults' reading skills would be just as applicable to reading on a computer screen as part of learning document editing skills as reading recipes in acquiring cooking skills. Hence performance improvements in these behaviours are unlikely to be a result of improvements in component reading processes (although benefits may well accrue for jargon words associated with the particular skill domain). Instead, improvements in performance of these behaviours are more likely to be the result of refinement of component processes that are specific to the particular behaviour. That is, the amount of performance improvement observed with a task will be a function of the amount of improvement that occurs on component processes and the relative contribution of well-practiced and new component processes to the overall performance.

1.7. Summary

The five principles of learning describe a number of features that are general to all forms of skill acquisition. Essentially these principles state that practice leads to faster and more efficient uses of knowledge. This enables faster performance and results in less demand on mental resources. In turn these outcomes enable higher level behaviours to be attempted. Ultimately skills are developed through refinement of many component processes.

2. COMPLEX SYSTEMS IN HUMAN COGNITION

As they have been described above, the five principles of learning do not necessarily imply the

operations of a complex system. It is other evidence of human cognitive performance that suggests to us that complex systems underlie human cognition. As has been noted in a large number of contexts, within distributions of words in individual vocabularies (Zipf, 1949) and in print sources (Le *et al.*, 2002), there is a power function relationship between a word's frequency of occurrence and the rank of this frequency value in comparison to other words. This relationship, known as Zipf's Law (Adamic, 2000), refers to the size of occurrence of an event relative to its rank i. The law asserts that the size of the i'th largest occurrence of an event is inversely proportional to its rank and that this relationship is a power function. In English, or any other language for that matter, it follows that the terms with the highest frequency will occupy the smallest classes, whereas the terms with the lowest frequency will occupy the largest classes. Quantitatively, the law asserts that:

$$P_i = P_1 i^{-\alpha}$$

where, $\alpha \approx 1$ and P_1 is the probability of the most frequent word. The relationship can also be expressed in terms of the number of words in each word frequency class. Thus, for English, there might be just one word in the range 1,000-10,000 occurrences per million, but 50,000 words in the range 1-10 occurrences per million.

Zipf-like power functions have been reported for a variety of inanimate as well as animate phenomena. An illustrative list includes the size of earthquakes, the scale of forest fires and the height of sand-hills. Zipf-like functions have also been reported for complex social phenomena such as city size, the distribution of professions and the magnitude of stock market crashes. In addition, the list can be extended to include income distribution and visitors to internet sites. The ubiquity of these functions (which includes Pareto's law) introduces a particularly challenging question. Do they reflect some artifactual process involving the law of probabilities, or do they reflect an as yet unidentified principle that stands above the distinctions between the organisation of animate, inanimate and social systems? The most parsimonious approach to this problem involves the assumption that several descriptive and modeling levels must be involved and that convergence can be expected at only the highest level, if at all (Halloy and Barratt, 2007). Consider for example recent work on genomic properties. Luscombe *et al.* (2002) noted that frequency of occurrence of the generalized molecular parts associated with genomes followed the power law with a few parts occurring many times and most parts occurring only a few times. Luscombe *et al.* (2002) attributed these

patterns to a DNA duplication process as genomes evolved to their present state. It might be appropriate to hypothesize that while this explanation involves the same general principle as that which applies to the Zipf-like functions for words, the relevant explanations must involve different physical material. They involve different domains and the critical question concerns the presence or otherwise of a single over-arching principle, a principle that could be applied to both the animate and inanimate domains while protecting the assumption that they enjoy distinct physical mechanisms.

Halloy advanced a model that reflects this point of view (Halloy, 1998). Halloy's argument includes reference to evidence involving the evolution of both animate and inanimate systems and employs an over-arching principle to account for the ubiquity of Zipf-like functions. Complex adaptive systems, according to Halloy, consist of agents that are made up of particles. Agents compete with each other for resources. Particles are the basic unit of resource for which agents compete. Agents can grow in size because they have been able to attract more particles, or they can split to create two or more smaller agents. According to Halloy (1998, p.5), "The abundance distributions of agents tends to a power function with increasing slope toward the right in a log-log rank abundance relation or a lognormal." But the term 'tend' is critical. Halloy adopts the further assumption that "natural systems will approximate to log-normal models when left to their internal mechanisms, while distancing themselves from the log-normal when pressured by external forces" (Halloy, 1998, p.3), an assumption that "circumvents the debate on the appropriate mathematical distributions to fit to natural systems" (Halloy, 1998, p.3). Nonetheless, the lognormal distribution in a frequency-abundance context is "…a signature of complex systems." (Halloy, 1998, p.2).

It is on the basis that features of distributions of words in individual and cultural lexica are consistent with abundance distributions associated with the operation of complex adaptive systems that we feel inspired to suggest that the human mind is a complex system and to develop a theory of the mind that is a human instantiation of Halloy's theory (Halloy, 1998; Halloy and Whigham, 2004). Thus the mechanisms underlying our theory honour Halloy's description of the behaviour of agents and particles in complex systems. We describe Halloy's theory in detail and then follow this with a presentation of The Component Theory of Skill Acquisition that assumes the mind is a complex system.

2.1. Halloy's Resource Attraction Theory

According to Halloy, all complex systems possess a range of characteristics that give rise to similar statistical

relationships. Four important concepts in Halloy's resource attraction theory are those of resources, particles, agents and boundaries.

"Resources (are) anything for which agents may compete…Particles are the minimum units of resources. From an agent's viewpoint they are discrete packages of resources of variable size or 'mass'. Particles are analogous to individuals in a biological population, to quanta of light or space in a plant community, to particles of dust in the cosmos, or to economic elements…Agents arise when an initial undifferentiated mass of particles breaks up or coalesces (i.e., boundaries are formed) into a number of parts. Each agent contains or controls a number of particles. Agents are analogous to species or companies…Boundaries are formed where interactions are proportionally more important between the particles inside the agent than they are between them and particles outside. The same applies for boundaries between systems at a higher level. Boundaries fluctuate and have a certain degree of permeability." (Halloy, 1998, p.3).

"Complex adaptive systems have been characterized as systems made up of interacting agents which use rules to maximize their survival…A unifying feature of such systems is that agents are "greedy", i.e. they attract resources, as much resources as they can grab. However, in evolved systems this attraction may become remarkably subtle, with time delays and complicated strategic decisions to forego a resource here and now for one in the future and somewhere else. Since agents are greedy, they necessarily compete for resources. Hence possibly the primal feature of complex systems is greed (or more euphemistically, resource attraction) and competition as its secondary outcome. It is this resource attraction and competition which in turn determines the primary interactions between agents, as well as the adaptive nature of agents changing rules to outcompete others." (Halloy and Whigham, 2004, p.4).

The competition for resources between agents in a complex system can result in the development of clusters of agents into larger agents or the splitting of large agents into smaller ones. The conditions under which these two outcomes occur are determined, in part, by the level of attraction between and amongst agents and particles in the system. This attraction is proportional to the existing resources of the agents (i.e., larger agents are more "attractive") and inversely proportional to the difficulty in obtaining resources (i.e., agents are more likely to attract particles or to combine with other agents when it is easier to do so).

Two important outcomes of the competition for resources that Halloy and Whigham (2004) highlights are differentiation and adaptation of agents.

"(A)ll agents eventually reach a size where their growth is not practical within their infrastructure. At this point they split…into sibling or parent and offspring agents (e.g. bacteria splitting, plants sprouting new shoots…). Initially, as they are informationally almost identical, these siblings may be considered part of the growing agent. However, this split has set the stage for the drifting of information which leads to differentiation and diversification. As differentiation (inevitably) proceeds, the siblings become different agents separated by informational barriers and competing with each other. In biology this is known as speciation." (p.5)

"(E)volving agents typically explore new pathways and opportunities to attract resources. This is a consequence of resource attraction as modified by differentiation. As they explore new state space and rules, some agents find more efficient ways to capture resources and survive, while others die off. This is the process of evolution and adaptation." (p.6)

Halloy (1998) has demonstrated that complex systems with the features described above can evolve into systems with abundance distributions that tend to lognormal (i.e., a class of distributions of which Zipf's law is a subset). That is, complex systems possess a small number of large agents, a large number of small agents and a smooth transition between these two extremes.

Lognormal abundance distributions have been observed throughout nature and indeed throughout a range of human affairs. For instance, Halloy and Whigham (2004) report that such distributions exist in "planet sizes, earthquakes, animal and plant sizes and abundances, sizes of firms, behaviour of the stock market, (and) traffic congestion." (p.7). All of the situations in which such distributions have been observed constitute "networks of interacting things under non-equilibrium conditions" (Buchanan, 2000, p.16) and such systems are known as complex systems. Halloy's theory represents an explanation for how such systems evolve and is general to all complex systems. Below we extend this ubiquity to human cognition and thereby suggest that the functioning of the mind follows universal laws of nature.

3. THE COMPONENT THEORY OF SKILL ACQUISITION: THE MIND AS A COMPLEX SYSTEM

Up until now, we have referred to skilled behaviour as comprising the actions of many component processes. We have not been precise about the nature of these processes, but it has been sufficient to simply consider these as properties of the brain that carry out some form of information processing. Ultimately these component processes must be related to the functioning of neurons

in the brain, which essentially are processors of information. In order to develop a theory of skill acquisition that honours Halloy's specification of a complex system we will define our component processes as consisting of agents. Certainly this is not a precise specification of component processes, but this lack of specificity actually reflects part of our argument, that the specification of component processes will depend on the level of analysis (more on this later).

The central tenet of the theory we are proposing - the Component Theory of Skill Acquisition - is that the human brain and therefore the human mind is a complex system. The agents in this complex system receive, process and transmit information. Depending on the level of analysis, these agents may be individual neurons, or networks of neurons, or even networks of networks. The degree to which consideration of these agents as networks of neurons will assist in understanding their function will also depend on the level of analysis. For example, developing an understanding of the processing of lines and edges in visual stimuli may rely on a focus on the performance of individual neurons, however it may be more sensible to explain the comprehension of written text by recourse to higher order agents that correspond to networks of neurons. Ultimately, though, an agent will only have utility by virtue of its input and output connections-that is, an agent must receive information and then pass it on once it has been processed. Thus all agents exist in networks with other agents. Furthermore, an agent will incorporate feedback mechanisms whereby the success or otherwise of the agent's processing will determine the likelihood of that agent performing that processing in the future. Details of these feedback mechanisms are provided below.

For the brain to be considered a complex system, its agents must compete for some resource. In our view this resource is information, because it is a *sine qua non* for adaptation to the world. The fundamental drive of agents in the mind is to be used to process information. This feature of our theory mimics that of neural systems where the survival of connections between neurons depends upon regular activation (Bruer and Greenough, 2001; Latham-Radocy and Radocy, 1996). In our model, agents compete to process information. If the outcome that results from the operation of an agent leads to success in achieving some goal, then the agent will be likely to be recruited for processing in the future. Thus success can lead to an increased potential for further success for an agent and hence continued survival. Failure, however, will lead to a reduced potential to be used in future and so possibly the demise of the agent.

Agents live or die on the basis of their usefulness. Firstly, they have to compete to be used. Secondly, if the result of their processing is successful, this then increases

the chance of continued survival for the winning agent. This principle, then, suggests some fundamental drives of the mind. Early in life these drives will relate to survival. For example, if the products of an agent's functioning in a particular situation lead to food, water, or warmth, then the feedback associated with these rewards ensures that the agent is used again in future when a similar situation is encountered and alternative agents are less likely to be recruited. Thus an infant can learn the utility of crying, calling for "mum" and saying "drink". Similarly, social rewards can act as drivers for learning (e.g., if a particular facial feature leads to positive attention from people, repeat it; if uttering a certain sound leads to food, a toy, or affection, repeat it).

In complex systems, agents are usually referred to as growing in some sense as they attract further resources. The sense of growing, however, is dependent on the situation in which a complex system is being considered. In our complex system version of the mind, agents grow with use in the sense that, with success agents come to be recruited more often to perform a particular task and so they come to dominate processing. Agents will increase their chances of being recruited to perform a task through a number of mechanisms, including performing the task more efficiently than competing agents (see below), or forging connections with other agents. In other words, there will be an increase in the number of situations in which an agent is useful. There will be limits, however, on the extent to which forging connections with other agents leads to an increased usefulness-some connections will prove to be fruitless because the processing that the agents can do is not relevant for some tasks (e.g., for some infants, every animal is a "dog", until that response regularly attracts no reinforcement). Thus, the nature of the environmental demands will shape the usefulness of an agent.

In addition to the success or otherwise of an agent's actions, the competition between agents for the right to process information is decided by the speed with which the agents complete processing. In a similar vein to Logan's (1988) Instance theory, the agent that completes the necessary processing in the shortest time will be the winner of the competition. That is, the fastest agent is most likely to be used in future. The products of the fastest agent are used and hence that agent receives "success" feedback. This feedback has the effect of making this agent more likely to be used in similar situations when they occur in the future.

The competition to be used that agents engage in can result in the combination of agents to form larger sized agents and also the splitting of large agents into smaller, more specific agents. The conditions under which these events occur will be associated with the particular task

presented to the complex system. The system will try to respond to any challenge that the environment presents and so the resulting agents will be a match for the environmental demands. If a challenge can be met by combining agents, then this will occur. If a particular environmental niche is detected that requires a more specific processing task than an existing general agent can complete, then a more specific agent may "break away" from the parent to exploit this opportunity to be used. Following Halloy and Whigham (2004), the competition to be used amongst agents results in a lognormal distribution of agents such that there will be a large number of small agents that have very specific purposes, a small number of general purpose agents and an inverse non-linear relationship between frequency and size for those agents between these extremes.

A concrete example of this type of system comes from issues surrounding word recognition. A general feature of skilled performance (Principle 5) is that sometimes it is useful to have skills that can solve problems in many situations and other times a more specific set of skills will be necessary. This feature is obvious in language skills. For instance, some English words are used in a broad range of situations (e.g., the) and others have a far more limited range of usefulness (e.g., hydrogen). That is, the word "the" can be used as the definite article for just about every noun in English and so is likely to appear in the majority of English sentences. The word "hydrogen", in contrast, is likely to appear in sentences relating to chemical contexts. Thus the agent for processing "the" is used in more situations than the agent responsible for processing "hydrogen". The "the" situations are not all going to be common-the situations in which "the" will occur are as varied as the topics of discourse, whereas the "hydrogen" situations are far more likely to have a common feature-that is, they will concern the element called hydrogen. The "the" agent is thus an extremely general agent that can be invoked in a number of situations whereas the "hydrogen" agent is only likely to be used in a small number of quite specific situations. There are very few words that are as ubiquitous as "the" ("a" and "I" are other examples), but there are thousands of words like "hydrogen" that appear in restricted contexts. In other words, there tend to be many more low frequency words than high frequency words. This distribution of words of course corresponds to Zipf's law, which is a member of the lognormal family of abundance distributions.

There are many ways in which agents can improve their competitiveness for resources, or in other words, improve the likelihood of being used. To achieve this, agents must complete their processing in less time than competing agents. In our view, these improvements are the same as those we have highlighted with respect to the effects of practice on performance. That is, practice can lead to faster and more efficient forms of processing. We propose that there are many mechanisms whereby these changes can occur. For instance, agents may process information faster with repeated usage, a reflection of changes to neural function that have been observed to result from practice (Altmann *et al.*, 2004; Barnes, 1979; Bolger and Schneider, 2002; Eccles, 1972). Alternatively, agents may combine with other agents in ways that reduce inefficient forms of processing (e.g., unnecessary processing steps can be skipped). Agents can also split if environmental demands suggest that a smaller agent with a more specific processing function will be more useful. The particular improvement strategy followed will be determined by the nature of the environmental challenge being tackled. Importantly, when a new challenge is encountered, there will be many potential means for improving the competitiveness of agents and so there will be greater potential for improved processing performance. As the challenge becomes more familiar, though, there will be less potential for improving further the competitiveness of agents and so performance improvements will be less likely. Thus performance changes over the course of practice will be negatively accelerated, which is evident in the characteristic shape of learning curves. Importantly, this negative acceleration characteristic is also a general feature of lognormal distributions in that it is relatively easy for an agent to shift from a rank of 1000 to a rank of 990 (i.e., the proportional difference is low), compared to shifting from a rank of 10 to 1 (i.e., the proportional difference is many times larger).

A fundamental feature of our theory is that the likelihood of an agent being used in the future depends on the success of its processing. An important question, then, is by what mechanism does current performance affect future likelihood of use? In our theory this occurs as an inherent feature of the feedback process. We adopt the Kirsner and Dunn (1985) idea that every instance of processing results in a record of that processing. We suggest, however, that the record gets stored as part of the agent that performed the original processing, as a means of recording feedback of the results of the agent's processing. When an agent completely fulfils the goals of its processing, a record is created that reflects the operations of the agent. This record then is a bit of information, a particle in Halloy's terms, that is attracted to the agent and is thus stored as part of the agent. The effect of this process is that the agent grows. By growing in this manner, an agent becomes more attractive to future resources. That is, the agent is more likely to be used when faced with the same situation again. It is more likely to be used in future similar situations because its greater mass as

a result of successful past performance means it will be faster than other agents. In addition, a history of successful performance in a range of situations will increase the scope of applicability of an agent (see below).

The concept of an agent growing with successful application leading to an increased attractiveness is analogous to gravitational attraction. That is, in a physical system bodies with large mass possess greater gravitational attractiveness compared to bodies of smaller mass and so other bodies in the physical vicinity will be more attracted to and hence move toward the larger bodies. Similarly, in a complex system, the larger the mass of an agent, the greater the likelihood that it will attract particles in its region. In the mind, particles in the 'region' of an agent represent particles relevant to an agent's processing. Therefore, when a demand is presented to the system, if two agents are of equal relevance to the demand, the agent with the greater mass will be more attractive to the particle that is up for grabs. That is, it will be more likely to complete the necessary processing and hence will collect the particle that represents the information about that processing episode. This particle will then add to the agent's ability to attract further particles in future.

Although the gravitational analogy is a useful metaphor for understanding the nature of the competition for resources that occurs between agents, we can be more specific about the mechanism underlying the relationship between an agent's successful processing and a subsequent increased likelihood of future recruitment. As stated above, particles in our theory represent processing records. But particles also enable future performance. That is, because they are a record for what worked in the past, they can function as a blueprint for what to do when that previous situation re-occurs. As a result, an agent that has completed a processing task successfully on many occasions will have a collection of particles that represent records of each of these processing episodes. More precisely, an agent is really only a collection of particles and so this growing collection of particles represents the growing mass of the agent. The mass of an agent is basically a collection of records of what happened in the past when a particular demand occurred. Alternatively, these records can be seen as a series of instructions about what to do should that demand re-occur.

Describing agents as collections of particles raises the question of why a large collection of particles (i.e., an agent with large mass) gives rise to faster performance than a smaller collection of particles (i.e., an agent with smaller mass). To answer this question we again borrow from Logan's (1988) Instance theory. Within an agent particles will differ in the speed with which they can be utilized as processing instructions. To

explain why this is the case, consider **Fig. 1.** Both panels of this figure depict a finite number of particles (N = 20 in both cases). Depending on the current goals, these particles may or may not represent useful forms of processing. Imagine that in both cases the X symbol represents the most relevant particle for the current goal. The other symbols represent particles that are less relevant to the current goal (i.e., solutions that are sub-optimal). In the left panel there is not as extensive a history of X being useful as in the right panel. That is, there are more Xs on the right. If we then imagine that recruiting one of the particles as a guide for performing the next task is a random search through these spaces and that the search ends when an X is encountered, there is clearly a greater chance of finding an X in the right panel than in the left panel. Furthermore, an X will be located sooner in the right panel than in the left panel. In general, then, the speed with which a relevant particle can be located is going to be determined by the number of such particles present-the more particles there are, the sooner one can be found and recruited. The speed with which an agent can perform a task, then, will reflect the particle that enabled processing in the shortest time. As in the Instance theory, the distribution of processing speeds amongst the particles within an agent (i.e., the time to recruit a relevant particle) will be a Weibull distribution (which is also a member of the lognormal family of distributions) and so there will be a power function relationship between the number of particles making up an agent (i.e., the number of successful processing episodes) and the speed of the agent's processing. Thus an agent is more likely to complete processing before other agents with less mass because it is more likely to have a relevant particle that can enable appropriate processing in a shorter time. Thus although it might be convenient to think of large agents as attracting a processing episode in their direction, it is more accurate to think of such agents as being the fastest to complete processing. An agent will dominate processing in the sense that it always does the job not so much because at some point it attains privileged status and so demands to do the processing, but more because it is simply the fastest to provide a processing result in a never-ending competition with other agents.

It is important to note that no two situations in the world are ever identical. Even reading the same word on a computer screen in identical formats on two occasions does not involve exactly the same situation because the contextual content of the person reading the word is slightly different from one moment in time to the next. Thus the human information processing system must be capable of tolerating differences in ostensibly similar stimuli in order to be able to identify them as such. Of course, there must also be limits on

this tolerance so that differences can be perceived. Certainly some neural tissue appears to be sensitive to stimulus differences. For example, there are neurons in the visual cortex that appear to be geared to recognising lines of a particular orientation (e.g., vertical), but which are still active in response to lines that do not match the ideal orientation (i.e., not completely upright). The extent of activation, however, is proportional to the extent to which the lines approximate the ideal orientation for those neurons (Hubel and Wiesel, 1962).

In Halloy's (1998) and Halloy and Whigham (2004) theory, the probability of an agent attracting a resource is inversely related to the distance to the resource. We honor this principle in the cognitive context by proposing that the relevance of an agent to the current goal of processing, or the similarity between the current conditions and the normal conditions processed by the agent, determine the likelihood of the agent being used and thus attracting a particle reflecting success. In other words, similarity determines the likelihood of an agent gaining mass and as a result, the rate of change of performance speed (Shepard, 1987).

For instance, this issue arises when an agent produces results that do not completely match the goals of processing. This could happen, for example, when an agent is used in a slightly different situation to the one in which it was developed. This partial success will nonetheless be stored as a record of the agent's processing. The record of the partial match will be stored with the agent and this will have two effects. Just as the storing of a record of complete success means an increase in the mass of the agent, a record of partial success will also increase the agent's mass. The increase in mass in this situation, however, will be less than that following complete success. This recognises the fact that partial matches between old knowledge and new problems can result in partial transfer (Greig and Speelman, 1999; Palmeri, 1997). The mechanism underlying this effect borrows from an idea in Palmeri's (1997) modification to Logan's Instance theory. The speed with which a particle can be utilised in the processing of a new situation will be a function of the similarity between the particle and the processing demand-the greater the similarity, the faster the speed of processing. Thus, the increase in the effective mass of an agent-that is, that characteristic of an agent that reflects the speed with which it will complete processing-that comes from a partial match will be proportional to the degree to which the agent satisfies the environmental demand. The second effect of storing a record of partial success is that information about the different conditions in which the agent was at least partially successful is stored with the agent. This has the effect of expanding the conditions under which the agent is potentially useful. Thus partial success can increase the potential usefulness of an agent in two ways: It results in increasing the mass of an agent that is associated with faster and more reliable recruitment of the agent, although this will be tempered by the degree of similarity between the experiences embedded in the agent and any new situation and it also results in an increase in the conditions in which the agent is applicable and so broadens the range of situations in which the agent could be useful.

Dealing with environments in which there is stimulus variability will result in agents that record many instances of complete success in one circumstance and a series of partial successes in another circumstance. As a consequence such an agent will possess two different types of particles. Under these conditions, there may be sufficient advantage in this heterogeneous agent (i.e., an agent applicable to several situations) splitting so as to create a number of smaller homogeneous agents that are more specific to the particular environmental circumstances. The advantage that will provide the motivation to such a split will be that the more specific agents will more completely satisfy the environmental demands and hence will receive greater increases to their mass than under partial matching conditions.

There will, of course, be a trade-off between the extent to which an agent develops to match the environmental demands (i.e., maximum increases to mass) and the frequency with which the specific environmental demands occur. Sometimes an agent that can deal with many situations will remain heterogeneous because the various situations do not occur sufficiently regularly to warrant a splitting to create a more specific agent. That is, the partial increases to the agent's mass that comes from partial matching will justify the continued existence of the agent in its current form, whereas a smaller agent that is specific to the particular situation will not be useful sufficiently often to justify its existence. There will be times, however, when particular situations will occur sufficiently regularly that the increases to the mass of a smaller, more specific agent adapted to that situation will justify its separate existence. (It is worth noting that the processes described in this paragraph are analogous to the processes of speciation in biological systems, as described for example in Mayr, 1963; Laurent, 1972).

There will also be times when the operations of an agent do not result in success. This can arise when an agent does not win the right to be used (e.g., another agent does the job, or no agent does the job and so the word is not understood). Alternatively, the agent does get to complete processing but the result does not constitute a successful outcome.
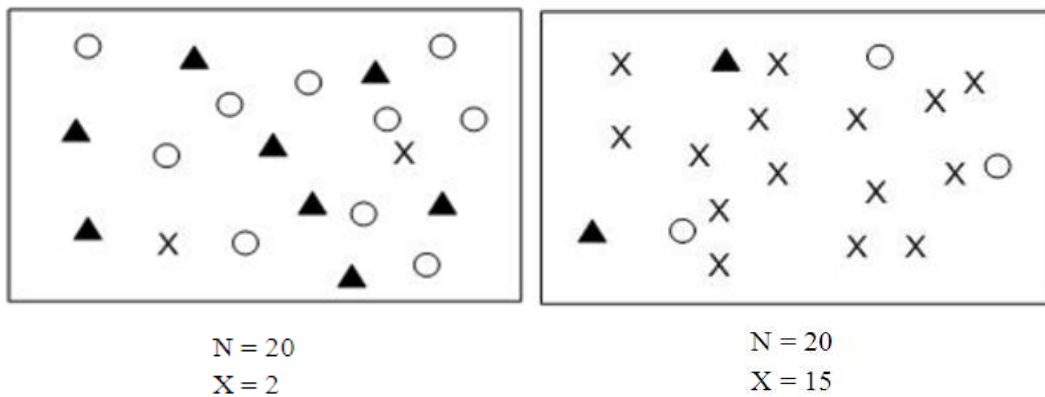
N = 20
X = 2

N = 20
X = 15

Fig. 1. Two hypothetical distributions of particles, where X represents the most relevant particle to the current processing goal

For example, the agent's output does not satisfy the goals of processing. In both cases the agent would not attract a particle that is the record of successful performance. As a result its performance goes unrewarded. In a sense this represents a situation of no change to the mass and hence status of the agent (although see below). However, this does not mean that the system remains the same and is unresponsive to failure. Such situations represent opportunities for other agents to prosper. The first situation is one where a competing agent wins the competition to perform because it completes processing faster than the other agent. If this winning agent continues to perform successfully then that agent will become the agent of choice in similar circumstances. In the second situation, where an agent produces a result that is ultimately unsuccessful, a demand on the system remains unsatisfied and so represents an opportunity for new agents to develop. Therefore, as a result of unsuccessful performance, due to performance being too slow or inappropriate, agents can effectively lose their preferred status as more successful agents take over performance of the task. It is important to note that competition between agents can result in successful agents emerging as dominant on the basis of success alone. There is no need to posit an explicit inhibition mechanism (although inhibition may or may not be necessary in neural implementations of a complex system of this sort).

The suggestion that the mass of an agent does not change following unsuccessful performance may explain why, under some circumstances, people persist with inappropriate behaviour beyond the point at which they learn of the inappropriateness of their behaviour. For example, in the Luchins' water jar problem, following the development of a mental set to approach all problems with a particular solution, the majority of people persist with the unnecessarily complex solution after experiencing a problem that could only be solved with a simpler solution (Luchins, 1942). The complex system

explanation for this observation is that until an alternative agent is developed that can complete successful processing faster than the original agent can complete its unsuccessful processing, the unsuccessful agent will persist in producing inappropriate behaviour.

As stated in Principle 1, practice on any task typically leads to better performance. Another feature of the relationship between practice and performance that is just as commonly observed is that a lack of practice leads to poorer performance. That is, if someone practices a task for a period of time, performance typically improves in both accuracy and speed, but if the person ceases practice for some time, their performance upon resuming the task is never as good as it was at the end of the previous performance period. Skills seem to suffer a form of decay such that if they are not used, something is lost and this results in poorer performance. Some skill acquisition theories build in a decay parameter to account for this observation (e.g., Anderson, 1982). In our view there are several reasons why such apparent decay of skills occurs. Firstly, the complex system that is the mind is implemented in a biological system. It will therefore require some form of neural resource to maintain any form of mental representation over time. There may then be a limit on the ability to maintain representations that have not recently been of use. Thus agents that are currently "top of the pops" as far as usefulness is concerned grow in mass and this may occur at the expense of other not currently useful agents. This suggests a principle of conservation of mass whereby a constant mass is shared amongst all agents, such that any growth in the mass of some agents that reflects current usefulness is matched by a distributed reduction in mass of all other, not currently useful, agents (an example of how this can be modelled by the resource attraction theory is shown in Halloy, 2001). Another means whereby the apparent decay of skills can come about is associated with the idea that higher level skills require

the co-ordination of many agents that each perform some sub-component of the task. If a task is not performed for some time, the agents underlying performance of the task may be useful in some other task. So, although the agents themselves may suffer no loss of mass due to inactivity, the connections between the agents that enables their co-ordination to perform the original task may fade with lack of use. For example, a guitarist may work up a solo comprised of various riffs and licks for a particular song. After many years of performing the song the guitarist gets tired of the song and drops it from his repertoire. Following several years of playing other songs, in which all the riffs and licks from the deleted song appear, but in different orders and across different songs, he receives a request to play the old song. He will find that despite all of the riffs and licks remaining in his repertoire, co-coordinating them smoothly into the solo of the original song will not come easy. The first new performance of this old song is likely to be "clunky", or at least not as effortless and elegant as the final performance in the original tenure of the song.

Elsewhere (Speelman and Kirsner, 2005) we introduced a Fluency Threshold as part of the Component Theory of Skill Acquisition. This threshold corresponds to the point at which someone has the ability to attempt a new task. Prior to this point, the component processes necessary to perform the task are not fluent enough to fit within the person's resource constraints. Thus the demands of the task outweighed the available resources. At the Fluency Threshold, however, the component processes necessary to perform the task have been practiced to the extent that the resources required to perform the task do not exceed those available. In the complex system version of the component model, this Fluency Threshold can be understood in terms of the competition between agents to be used. As described already, agents with greater mass (i.e., more successful experience) have a greater chance of being recruited to perform a task than agents with less mass. There are several reasons for this. If an agent is too slow, another faster agent may win the right to be used (i.e., it produces a solution before the slower agent completes processing). In addition, the complex systems of the human mind exist in a dynamic world, where task demands include time constraints. A slow agent may not complete processing in time for the demands of the task and so the benefits of successful processing are not realised. As a result, no reinforcement for performance will be received. Thus, such small, 'young' agents are not reliably applied in certain circumstances to enable consistent performance. Reliability of application comes only with sufficient successful past application. When a novice attempts a task that requires the application of several component processes, that is, several agents, they will only be able to complete the overall task

successfully when the necessary agents have grown to a sufficient extent that they can do their job reliably. If any agent is insufficiently large to be reliable, then a link in the chain of processing will be inconsistently performed and the overall task will not be completed successfully.

As someone gains experience in a particular domain, the agents responsible for performing components of a task will become faster and more reliable. That is, when necessary, they will more consistently do the required job successfully. Eventually the agents will meet the Fluency Threshold conditions for successful task performance. That is, the person will have sufficient mental resources available, in the form of a set of reliable agents, to attempt the new task.

With further successful practice on this task, the agents responsible for the task components are rewarded for acting in concert by being recruited as a team in future. Indeed, if several agents consistently operate in succession to complete a task, there may come a time when agents that occur later in the chain come to 'anticipate' the point of their own application. Initially agents may only be sensitive to outputs of the agent that immediately precede them in the chain. It is possible, however, that with experience, agents later in a chain can become sensitive to outputs from agents earlier in the chain than those that immediately precede them. Eventually these later agents may become sensitive to the initiating conditions of the task so that these lead directly to the results of the final agents in the chain and so unnecessary processing steps can be eliminated. This form of learned anticipation is characteristic of all forms of learning, such as chains of associated conditioned stimuli leading to a conditioned response in classical conditioning, or the development of complex behaviour in operant conditioning and also the chunking of information that facilitates comprehension and memory in domains such as language and chess. Ultimately, then, a network of agents that enable performance of a particular task could potentially create a new, higher order agent that is adapted for performing this particular task. Furthermore, this higher order agent could then serve as a component agent on some other, even higher order task. Thus the processes involved at one level of the system occur at all levels of the system.

4. HOW THE MIND AS A COMPLEX SYSTEM GIVES RISE TO THE FIVE PRINCIPLES OF SKILL ACQUISITION

In this section we consider the extent to which the Component Theory of Skill Acquisition can account for the five principles of skill acquisition without explicitly building them into the theory. In doing so we show how the principles can be explained in terms of the operation

of agents and particles. We show too that each principle can be understood as a by-product of the adaptations of the complex system that is the mind. When the system adapts to the environment, it will do so according to the characteristics of a complex system and the adaptations will then exhibit particular features that are consistent with the five principles. Hence the principles are emergent features of the adaptations of agents.

4.1. Principle 1: Practice Leads to Faster Performance

Performance of most measurable behavioral tasks will involve the operation of several agents. Practice on such tasks leads to faster performance for several reasons. One is that individual agents process information faster with practice. As the number of times an agent completes processing increases, more particles representing records of these episodes will be stored with the agent. These particles enable future performance and so as the collection of particles increases in size, so too does the chance of recruiting a useful particle in less time. As a result, the speed with which an agent can complete its processing can increase with practice and this can lead to faster performance on a task. As mentioned in the previous section, though, the benefit to performance time of increased numbers of relevant particles diminishes as a power function of the number of particles (i.e., practice). Thus performance improvements on a task will be a negatively accelerated function of practice. In addition, performance of a task that involves the operation of several agents can get faster with practice as a result of changes to the particular agents involved in performing the task. That is, practice can lead to more efficient forms of processing as a result of redundant agents being dropped from processing. The opportunities for such improvements in efficiency are likely to be much greater early in practice compared to later and so improvements in performance time that result from this mechanism will also be a negatively accelerated function of practice.

4.2. Principle 2: Practice Leads to Efficiencies in Knowledge Access

When completion of a task involves the operation of several agents, practice can lead to the individual agents processing information faster and redundant agents being dropped from processing. As a result, "super" agents can develop that are responsible for performing the task in fewer steps than the original set of agents. That is, one agent can do the job of several agents. Thus, with reading experience, several agents that are separately responsible for recognizing the individual letters of a word can be superseded by an agent that recognizes the whole word.

4.3. Principle 3: Learning Leads to Less Demand on Working Memory

The idea that there are working memory constraints on the performance of a task is usually invoked when a task is attempted that seems to require more than someone is capable of performing. For instance a task may require someone to pay attention to more information than is apparently possible. An example of this would be someone who is learning a language and they are required to comprehend a number of sentences that include many unfamiliar words and that are spoken very quickly. Initially their ability to comprehend each word may be non existent or too slow to enable all of the information about each word to be integrated into some realization of the meaning of each sentence. With growing expertise with the particular language (i.e., they become familiar with more words and the speed with which they can access their knowledge of these words increases), their ability to process such sentences increases. That is, they can comprehend a sentence soon after it is uttered. In this type of situation, our theory would propose that initially the person does not possess agents for word recognition that are sufficiently reliable and fast as to enable comprehension of the utterances. Words keep being uttered without comprehension keeping up. Thus each sentence bypasses the listener. With practice, however, agents become very fast and reliable in their processing and so enable almost instantaneous processing of language. So, rather than a representation of a sentence needing to be retained in working memory for long periods until 'young' agents can process the words, 'old' agents are able to process the sentence quickly and so free up space in working memory.

4.4. Principle 4: As expertise Increases, Fewer Mental Resources are Required to Perform a Particular Task, Enabling the Development of a Hierarchy of Skills.

Two things can happen when a set of agents are used consistently in the performance of a task: (1) each agent completes its specific task in less time; and (2) some agents may no longer contribute to the overall performance of the task because other agents take over their processing. Thus as the history of successful performance grows, agents develop in such a way as to perform the task in a faster and more efficient manner. Being able to complete processing quickly is an advantage particularly when performance is in the context of a dynamic task where time constraints exist. These time constraints will include things like the existence of an environmental threat (e.g., an oncoming car) that requires some evasive action be taken in some

minimum time, or two people engaged in conversation where boredom could result if the conversation does not proceed at some minimum rate, or a complex task that requires intermediate products of processing be stored (e.g., double digit multiplication) but storage of these products in memory is subject to decay over time. Thus there is often considerable motivation to perform a task faster, not the least of which is to overcome the constraints of a basic level of performance in the domain. This motivation will provide the impetus to develop agents that are specific to the particular task at hand rather than utilise agents that have been useful in previous contexts. Agents adapted to specific contexts will be more likely to perform a task in fewer steps than agents cobbled together from previous relevant experiences. Ultimately, specific agents may develop to the extent that they enable automatic performance of a task rather than the slow and ponderous performance associated with more general agents. Thus a certain environmental challenge will trigger an automatic response rather than a chain of processing steps that may or may not produce an appropriate response. Typically, however, there is also motivation to perform at greater than basic level performance. Developing agents that enable fast and automatic performance of the basic task will mean that there may now be time enough available to start attempting more complex forms of the task. Thus, as someone becomes more fluent at evading an environmental threat, such as getting off a road in time to avoid an oncoming car, then one may be able to attempt another desirable behaviour, such as learning to cross a highway.

4.5. Principle 5: Mastery in a Domain Involves the Application of an Array of Component Processes, with Varying Degrees of Specificity to Tasks and Contexts

Sometimes agents will develop that are specific to a task and cannot be used in the performance of any other task. At other times, agents will develop that can be recruited in the performance of several tasks. The nature of a domain will determine the relative mix of these types of agents and therefore skills. That is, if a task environment is such that a particular job has to be completed in a particular way, then agents will develop that are highly specialized to perform that task. The existence of such highly specialized agents will be ensured by the continued demand from the environment for such processing. In contrast, a task environment that requires many different performance types in varying contexts demands a flexible set of skills. As a result, agents will develop that are smaller in scope, specific to finer grained details of the task, but be capable of being recruited by other agents in order to complete the overall task. Thus performance in such a varying domain is unlikely to reach

the automatic level of the more constrained environment, but is likely to be more flexible.

Another way to express Principle 5 is that people are sensitive to regularities in a task environment. The skills they develop to perform the task and their ability to transfer these skills, are a reflection of their adaptation to these regularities. Expressed in terms of the agent theory, this principle arises because the task environment determines the potential for particular types of agents to be used. Agents will develop to exploit opportunities and will do so in a manner that matches the peculiar requirements of that domain. As a result, the nature of the agents, in terms of whether or not they can be recruited to perform in other task environments will be determined by the nature of the task environment to which they originally adapted.

In sum, we have demonstrated that the five principles of skill acquisition all emerge from the adaptations of agents. Thus we have not had to build them into the fabric of the system, unlike other theories of skill acquisition (Speelman and Kirsner, 2005).

5. CONCLUSION AND IMPLICATIONS

We have outlined here a theory of the mind as a complex system. This theory describes how experience with the world leads to the development of agents that enable performance of tasks necessary for dealing with challenges posed by the world. The theory is not a fully-realised theory in the sense that computer simulations are possible based on the details we have sketched above (although Halloy (1998, 2001) has developed computer simulations of his model). The development of such a version of the theory is a task we have set ourselves for the future. Our main aim here, though, is to convince others of the importance of this task. In essence the theory we have presented here represents a claim that the contents of the mind are entirely a product of its interactions with the world. If we begin with this assumption, there are some enlightening implications for many areas of Psychology and for the entire discipline of Psychology. We outline one of these implications below (for other implications see Speelman and Kirsner, 2005).

5.1. Skill Acquisition

For many years the one great constant about research in skill acquisition was that practice on a task led to performance improvements that followed a power function. This feature of skill acquisition is known as the power law of learning (Newell and Rosenbloom, 1981). As mentioned earlier, however, there has been controversy recently about whether or not learning curves are indeed best described by power functions and in fact whether or not the power law should actually have the status of a law. One conclusion that seems safe from recent discussions of this issue is that power function

learning curves are most often seen in group data, that is, data that is averaged over several individuals. Learning curves tend to be far less smooth in individual data, although there are instances where they do occur as smooth functions (Speelman, 1991). Therefore the power law seems only to apply in certain circumstances and hence begs the question about its lawfulness. How then can it be that a generalization applies under some circumstances, but under others it may or may not apply?

The Component Theory of Skill Acquisition implies a resolution of this issue. According to the theory, individual learning curves on tasks are a reflection of the improved performance of component processes (agents) throughout practice. Some agents will be newly created for the task and so will probably have a long way to improve. Other agents will be virtual modules in the sense that they are as good as they are ever going to get and so will not contribute to performance improvements on the task. The performance of some agents will improve with practice in a smooth manner, others will improve according to a step function and there are likely to be many variants in between these extremes. All agents, however, will need to improve to survive and there will be limits on the extent of improvement possible. Indeed some agents may have reached the extent of improvement (i.e., very old but useful agents) yet they will still need to continue to be useful to survive. The nature of the particular learning "curve" for each agent will be largely dependent on the particular processing engaged in. For instance, a task like counting is likely to be associated with a slow incremental improvement in performance that corresponds with a strengthening of number facts in memory (Aunola *et al.*, 2004). In contrast, a task such as Duncker's Candle Problem would show a dramatic improvement in performance once a workable solution has been provided or is discovered. Thus different forms of processing involve different potentials for improvement and will therefore determine the nature of the improvement that can occur. Nonetheless, a task that involves the collaboration of teams of agents will typically show learning improvements that approximate a power function (i.e., are negatively accelerated and monotonic). This is because the averaging of several learning functions to create one omnibus learning function will always result in a power function (Haider and Frensch, 2002; Heathcote *et al.*, 2000; Murre and Chessa, 2011; Myung *et al.*, 2000). For the same reasons, smooth power functions will be more likely to be observed in the learning curves of groups of individuals than in individuals' learning curves. Thus, rather than the power law being an all-encompassing law for all occasions, as envisioned by Newell and Rosenbloom (1981), it is probably more accurate to state that learning behaviour tends to a power function, with individual cases being highly subject to 'noise'. That is, performance time will be a power function of practice when performance time data represents an averaging of performance times collected from groups of people, or component processes, that each improves individually with practice but not necessarily as a power function. When there is not substantial averaging of separate learning functions, then the power law will not apply. It is important to note, then, that the lawful aspect of the power law comes from the mathematical property of averaging several functions rather than from some property of the brain. This then frees our theory and any other theory of skill acquisition from the constraint that it must contain a learning mechanism that not only obeys the power law, but explains it. Our theory, in fact, can explain why power function learning is observed in some circumstances and not in others, without in fact having a learning mechanism that follows a power function exclusively. Learning curves reflect the rise and demise of agents in a complex system competing to perform a task and it is the ubiquitous properties of such systems that give rise to the regularities observed in learning.

6. REFERENCES

Adamic, L.A., 2000. Zipf, Power-laws and Pareto-a ranking tutorial.

Adams, C., 2003. Passenger Injuries in Crashes in Western Australia, 1996-2000. Report to The Royal Automobile Club of WA (Inc), April 2003. (Summarised in *Road Patrol Magazine*, April/May 2003)

Allard, F. and J.L. Stakes, 1991. Motor-Skill Experts in Sports, Dance and Other Domains. In: Toward a theory of expertise: Prospects and limits, Ericsson, K.A. and J. Smith, (Eds.). Cambridge University Press, New York, pp: 126-152.

Allard, F. and N. Burnett, 1985. Skill in sport. Canadian J. Psychol., 39: 294-312.

Altmann, C.F., A. Deubelius and Z. Kourtzi, 2004. Shape saliency modulates contextual processing in the human lateral occipital complex. J. Cognitive Neurosci., 16: 794-804. DOI: 10.1162/089892904970825

Anderson, J.R. and C. Lebiere, 1998. Atomic Components of Thought. 1st Edn., Routledge, Mahwah, ISBN-10: 0805828176, pp: 490.

Anderson, J.R., 1982. Acquisition of cognitive skill. Psychol. Rev., 89: 369-406. DOI: 10.1037/0033-295X.89.4.369

Anderson, J.R., 1983. Retrieval of information from long-term memory. Science, 220: 25-30. DOI: 10.1126/science.6828877

Anderson, J.R., C. Lebiere and M. Lovett, 1998. Performance. In: Atomic components of thought Anderson, J.R. and C. Lebiere (Eds.). Mahwah, NJ., Erlbaum.

Aunola, K., E. Leskinen, M-K. Lerkkanen and J-E. Nurmi, 2004. Developmental dynamics of math performance from preschool to grade 2. J. Educ. Psychol., 96: 699-713. DOI: 10.1037/0022-0663.96.4.699

Baddeley, A.D., 1990. Human Memory: Theory and Practice. 1st Edn., Psychology Press Ltd, ISBN-10: 0863771335, pp: 528.

Barnes, C.A., 1979. Memory deficits associated with senescence: A neurophysiological and behavioral study in the rat. J. Comparative Physiol., 93: 74-104. PMID: 221551

Bolger, D.J. and W. Schneider, 2002. Representation of visual word forms in fMRI habituation paradigms. Proceedigs of the 43rd Annual Meeting of the Psychonomics Society, (PS' 02), Kansas City, Missouri.

Bruer, J.T. and W.T. Greenough, 2001. The Subtle Science of How Experience Affects the Brain. In: Critical Thinking About Critical Periods, Bailey Jr., D.B., J.T. Bruer, F.J. Symons and J.W. Lichtman (Eds). Paul H. Brooks Publishing Co. Baltimore, Maryland, pp: 209-232.

Buchanan, M., 2000. Ubiquity. 1st Edn., Pheonix Orion Books, London.

Chase, W.G. and H.A. Simon, 1973. The Mind's Eye In Chess. In: Visual Information Processing, W.G. Chase (Ed.). Academic Press, New York.

Crovitz, H.F., M.T. Harvey and S. McClanahan, 1981. Hidden memory: A rapid method for the study of amnesia using perceptual learning. Cortex, 17, 273-278. PMID: 7285598

Eccles, J.C., 1972. Possible Synaptic Mechanisms Subserving Learning. In: Brain and Human Behaviour. A.G. Karyman and J.C. Eccles (Eds.). Springer-Verlag, New York.

Ericsson, K.A. and P.G. Polson, 1988. Memory for Restaurant Orders. In: The Nature of Expertise. M.T.H. Chi, R. Glaser and M.J. Farr (Eds.). Hillsdale, NJ., Erlbaum.

Greig, D. and C.P. Speelman, 1999. Is Skill Acquisition General or Specific? In: Perspectives on Cognitive Science: heories, Experiments and Foundations, Wiles, J. and T. Dartnall, (Eds.). Greenwood Publishing Group, Stamford, ISBN-10: 1567503829, pp: 171-190.

Haider, H. and P.A. Frensch, 2002. Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney *et al*. (1998) and Palmeri (1999). J. Exp. Psychol., Learn. Mem. Cogn., 28: 392-406. PMID: 11911396

Halloy, S.R.P. and B.I.P. Barratt, 2007. Patterns of abundance and morphology as indicators of ecosystem status: A meta-analysis. Ecol. Complexity, 4: 128-147. DOI: 10.1016/j.ecocom.2007.04.002

Halloy, S.R.P. and P.A. Whigham, 2004. The lognormal as universal descriptor of unconstrained complex systems: a unifying theory for complexity. Proceedings of the 7th Asia-Pacific Complex Systems Conference, Cairns Convention Centre, Cairns, QLD, Australia, pp: 309-320.

Halloy, S.R.P., 1998. A theoretical framework for abundance distributions in complex systems. Complexity Int., 6: 12-12.

Halloy, S.R.P., 2001. Effects on system structure, diversity and stability of the distance exponent in a resource attraction model. Complexity Int., 8: 18-18.

Heathcote, A.S., S. Brown and D.J.K. Mewhort, 2000. The power law repealed: The case for an exponential law of practice. Psychol. Bull. Rev., 7: 185-207. PMID: 10909131

Hubel, D.H. and T.N. Wiesel, 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol., 160: 106-154. PMID: 14449617

Josephs, R.A., D.H. Silvera and R.B. Giesler, 1996. The learning curve as a metacognitive tool. J. Experimental Psychol. Learn. Mem. Cogn., 22: 510-524.

Kasparov, G., 1985. Learn Chess. 1st Edn., Redwood Books, London.

Kirsner, K. and J.C. Dunn, 1985. The Perceptual Record: A Common Factor in Repetition Priming and Attribute Retention. In: Mechanisms of Attention: and Performance. M.I. Posner and O.S.M. Marin (Eds.). XI. Hillsdale, N.J., Lawrence Erlbaum. pp: 547-565.

Larkin, J.H., 1983. The role of problem representation in physics. In: Mental models. D. Gentner and A.L. Stevens (Eds.). Hillsdale, NJ., Erlbaum.

Latham-Radocy, W.B. and R.E. Radocy, 1996. Basic Physical and Psychoacoustical Processes. In: Handbook of Music Psychology, D.A. Hodges (Ed.). IMR Press, San Antonio.

Laurent, R.F., 1972. La Especiacion. Miscelanea, pp: 44-44.

Le, Q.H., E.I. Sicilia-Garcia, J. Ming and F.J. Smith, 2002. Extension of zipf's law to words and phrases. Proceeding of the 19th International Conference on Computational Linguistics, Aug. 24-Sept. 1, Howard International House and Academia Sinica, Taipei, Taiwan.

Lesgold, A., R. Glaser, H. Rubinson, D. Klopfer and P. Feltovich, 1988. Expertise in a Complex Skill: Distinguishing X-Ray Pictures. In: The Nature of Expertise M.T.H. Chi, R. Glaser and M.J. Farr (Eds.). Hillsdale, NJ., Erlbaum.

Logan, G.D., 1988. Toward an instance theory of automatization. Psychol. Rev., 95: 492-527.

Luchins, A.S., 1942. Mechanization in problem solving: The effect of Einstellung Psychol. Monographs, 54: 1-95.

Luscombe, N.M., J. Qian, Z. Zhang, T. Johnson and M. Gerstein, 2002. The dominance of the population by a selected few: Power-law behaviour applies to a wide variety of genomic properties. Genome Biol., 3: 0040.1-0040.7. PMID: 12186647

Mayr, E., 1963. Animal Species and Evolution. 1st Edn., Harvard University Press, Harvard.

Miller, G.A., 1956. The magical number seven, plus or minus two: Some limits on our capacity for. Psychol. Rev., 63: 81-97. DOI: 10.1037/h0043158

Murre, J.M.J. and A.G. Chessa, 2011. Power laws from individual differences in learning and forgetting: mathematical analyses. Psychon. Bull. Rev., 18: 592-597. DOI: 10.3758/s13423-011-0076-y

Myung, I.J., C. Kim and M.A. Pitt, 2000. Toward an explanation of the power law artifact: Insights from response surface analysis. Mem. Cogn., 28: 832-840. PMID: 10983457

Newell, A. and P.S. Rosenbloom, 1981. Mechanisms of Skill Acquisition and the Law of Practice. MIT Press Cambridge, MA, USA., ISBN-10: 0262680718, pp: 81-135.

Newell, A., 1990. Unified theories of cognition. 1st Edn., Harvard University Press, Cambridge, MA, ISBN-10: 0674921011, pp: 576.

Palmeri, T.J., 1997. Exemplar similarity and the development of automaticity. J. Experimental Psychol. Lear. Mem. Cogn., 23: 324-354. PMID: 9080007

Pirolli, P.L. and J.R. Anderson, 1985. The role of practice in fact retrieval. J. Exp. Psychol. Learning, Mem. Cogn., 11: 136-153. DOI: 10.1037/0278-7393.11.1.136

Rosenbloom, P.S. and A. Newell, 1987. An Integrated Computational Model of Stimulus-Response Compatibility and Practice. In: The Psychology of Learning and Motivation, G.H. Bower (Ed.). Academic Press, San Diego, pp: 1-52.

Shepard, R., 1987. Toward a universal law of generalization for psychological science. Science, 237: 1317-1323. PMID: 3629243

Speelman, C.P. and K. Kirsner, 2001. Predicting transfer from training performance. Acta Psychol., 108: 247-281. PMID: 11693252

Speelman, C.P. and K. Kirsner, 2005. Beyond the Learning Curve: The Construction of Mind. 1st Edn., Oxford University Press, Oxford, ISBN-10: 0198508859, pp: 288.

Speelman, C.P., 1991. Skill acquisition: Estimating the contribution of old skills to performance on new tasks. University of Western Australia, Nedlands, Australia.

Zipf, G.K., 1949. Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology. 1st Edn., Addison-Wesley Publishing, Cambridge, pp: 573.