Applying Systemic Design to the study of ‘emotion’

(To be presented at the AICS’94 conference in Dublin, Ireland)

Tim Read

School of Computer Science
University of Birmingham

Emotion has proved a difficult concept for researchers to explain. This is principally due to both terminological and methodological problems. Systemic Design is a methodology which has been developed and used for studying emotion in an attempt to resolve these difficulties, providing a step toward a complete understanding of ‘emotional phenomena’. This paper discusses the application of this methodology to study the three mammalian behavioural control systems proposed by Gray (1990). The computer simulation presented here models a rat in the Kamin (1957) avoidance experiment for two reasons: firstly, to demonstrate how Gray’s systems can form a large part of the explanation of what is happening in this experiment (which has proved difficult for researchers to do so far), and secondly, as avoidance behaviour and its associated architectural concomitance are related to many so called ‘emotional states’.

Introduction

The concept of ‘emotion’ occupies considerable space in the literature. ‘Emotional phenomena’ (often expressed using folk psychological terms: general ones like ‘moods’, ‘desires’, ‘dispositions’, and specific ones like ‘angry’, ‘curious’, ‘happy’) play an important role in human behaviour, both at a personal and cultural level. There are many different theories of ‘emotion’, couched in different research paradigms, although little agreement between them. As the author pointed out in Read (1993), major problems in this area are the following: *Terminological*: namely using non-scientific folk psychological terminology in a scientific context, the result being that the same terminology is used to describe different parts of a given concept. Such language is circular in nature, and must be relegated to part of the boot-strapping process of generating an architecturally-grounded theory, and not be used in an aprioristic way to specify it (Read and Sloman, 1993). *Methodological*: as ‘emotion’ is a multifaceted concept, individual narrow research paradigms encounter difficulties both covering the entire concept, and addressing research from different fields Read (1993).

Systemic Design

Systemic Design (Read, 1993) represents a variant of the design-based approach (Sloman, 1993) to provide a methodology for studying ‘emotion’ and other allied phenomena in biological organisms. This approach stresses three aspects: firstly, the application of the design-based approach to work toward a theory of architectural functionality at different levels, secondly, the incorporation of data from different research disciplines to assist the different functional levels with the application of (explicit and implicit) empirical constraints, and thirdly, the incorporation of phylogenetic considerations to ensure that any previous problems that influenced the development of the mechanisms (and are no longer part of the current environment) still play a role in the design.

The term ‘level’ is being used here in two distinct senses: firstly, to refer to the degree of abstraction of process and representation above the system hardware, and secondly, to refer to the domain of functionality. Process and Representational Abstraction: the notion that the mind/brain transition should not be considered in a dualistic sense, but as a hierarchy of virtual machines, has been put forward by many people, including Sloman (1993). Central to this thesis is the idea of architectural abstraction, both for the computational process and representational formalisms that make up the various virtual machines. The transition takes place from processes that are closely integrated with the hardware using essentially non-symbolic forms of representation, through to higher levels of progressively more abstract and symbolic forms of representation.

Functional Domain: This concept has been proposed by Gray (1990). He defines three functional levels: firstly *Behavioural*, i.e. the way in which the organism interacts with the world, secondly *Computational*, i.e. the information processing occurring in the mental architecture, and thirdly *Neural*, i.e. the actual brain systems that make up that architecture. Gray then argues that, in order for a theory of such an architecture to be developed, we need to explain its functionality at the three levels, and their interconnection. Gray’s three levels provide a framework in which Systemic Design can be used. Furthermore we can now make the tenets of Systemic Design stronger by stating that the architecturally-grounded theory developed will be defined purely in terms of those levels, and that any terminology emerging from such a design process must map across them. There may not be a direct mapping between any two of the three levels, as a concept in one level may map to several concepts in another one (such as the way a high level computer languages relate to their implementation).

The aim of the design exercise is to produce a ‘broad and shallow’ (Bates, Loyall and Scott, 1991) computer model (one which attempts to model some aspect of most systems present in the agent to some degree of detail), as opposed to a ‘narrow and deep’ (one which attempts to model a ‘key’ system in great detail). The construction of the model and its subsequent testing give rise to extra theory, data, and emergent phenomena that can be used to refine the original theory of architectural operation, and any associated terminology.

The Work of Jeffrey Gray

The work of Gray in the general area of ‘emotion’ provides a suitable opportunity to test and apply Systemic Design. Gray (1975) has proposed that the mammalian brain contains three basic systems that control behaviour: the Behavioural Inhibition System (BIS), the Behavioural Approach System (BAS), and the Fight Flight System (FFS). He suggests that faults in these systems underlie pathologies like anxiety disorders and schizophrenia, and that for a complete understanding of these brain systems they must be defined at the three levels of operation stated above. His research to date has provided detailed accounts of the Neural and Behavioural level, but little of the Computational level and the information processing that occurs within and between these systems. When outlining the basics of Gray’s model, it is necessary to use his terminology, despite the problems mentioned above.

The BIS responds to perceptual input corresponding to ‘signals of punishment’, ‘non-reward’, ‘novelty’ and innate fear stimuli. On encountering such an input, the BIS interrupts the current motor sequence, increments ‘arousal’, and increases the organism’s ‘attention’ to the relevant stimulus. The BIS can operate in one of two modes: firstly, as a monitor, where predictions of future states of the world are matched against results of the current one as it unfolds, and secondly, as a controller, when an error in the prediction occurs, the BIS interrupts the current motor sequence, and increments both ‘arousal’ and ‘attention’.

The BAS responds to perceptual input corresponding to ‘signals of reward’ and ‘non-punishment’, and initiates motor responses to approach said stimuli. The FFS represents a much simpler system. Upon the receipt of actual punishment (e.g., an electric shock) or ‘non-reward’ (e.g., the non-appearance of expected food), the FFS has the tendency to interrupt the other two systems and generate an unconditioned escape response or some form of defensive aggressive behaviour.

The operation of these systems is further complicated by their interaction. Part of my research aims at specifying the conditions under which a particular system will take control from its peers. Examples of such situations are the approach/avoidance conflict and the escape/avoidance transition.

Gray’s work (Gray, 1987) is founded on experimental work done on animals, and on human psychopharmacology. The systems proposed essentially map onto the limbic system and parts of the prefrontal cortex. To be consistent with both Gray's experimental work and with the aims of Systemic Design of keeping close to the phylogenetic considerations, the architecture I have designed is for the rat. The justification for working with a model of a rat, and its applicability to human 'emotional phenomena' comes from the following research areas: differential pharmacological studies of the effects of drugs that work on these systems, lesion and brain damage studies, experiments demonstrating similar behaviour, and brain imaging techniques.

The objectives of my work are four-fold: firstly, to produce a specification of the functionality of Gray's three systems (both individually and the way they interact) at a Computational level, secondly, to build a computer simulation model which simulates psychological experimental data, and demonstrates the effectiveness of Gray's three systems, thirdly, to demonstrate the efficacy of the Systemic Design methodology, and fourthly, to refine the terminology that Gray uses for his model.

Conceptual architecture

The conceptual architecture which has resulted from the design process I have undertaken is illustrated in Figure 1. It is termed a conceptual architecture as its implementation details differ slightly from the form illustrated here for software design reasons and efficiency.

The model of the rat exists as a computational architecture contained within a notional body which is located in the simulated experimental environments to differentiate between this simulated rat and the real animal during the following discussions, this model will henceforth be referred to as Simrat.

Interactions between Simrat and the environment occur via sensory input (to evaluate the state of the world), internal information processing (as dictated by the computational architecture), and motor output (for relocation and interaction purposes).

Attempting to model sensory processing and representation in any degree of detail (as taken from current theoretical ideas of the same) is beyond the scope of this research, and unnecessary for the goals outlined above. Hence the form and representation of perceptual stimuli follow a simplified form of deictic representation (Chapman, 1990) (each different modality provides different kinds of 'affordances' to the architecture). The four sensory modalities simulated within Simrat are intended to correspond to vision (each stimulus has both a location within space, and a reference key), sound (each stimulus has only a reference key), smell (each stimulus has only a reference key), and body damage (represented as a numerical level).

Simrat moves in the the world via three motor actuators, one for its head and one each for its front and rear limbs. The current implementation of this model uses high level motor commands of the form: start-of-action end-of-action, and doesn't refer to the motor actuators directly, although they have been included in the model to cater for any future extensions that may be required.

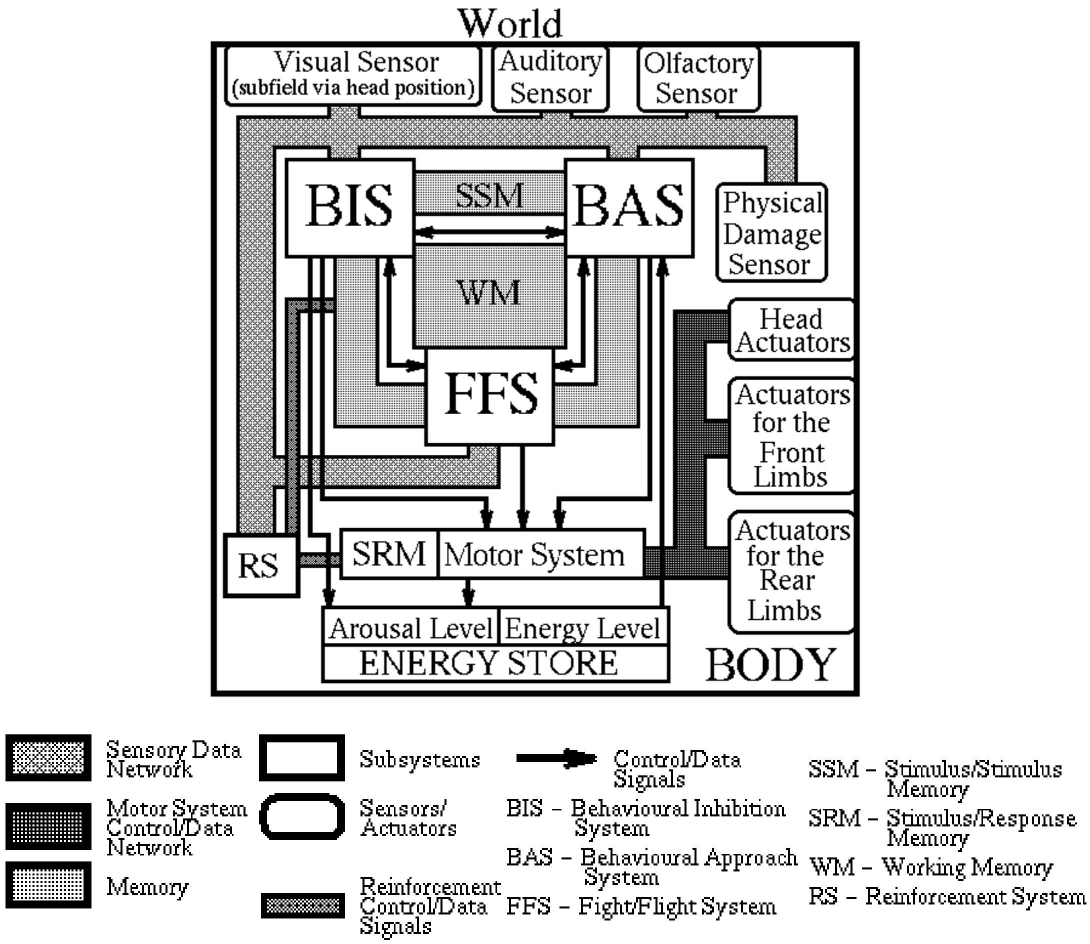

Figure 1: Conceptual architecture diagram

There are two types of memory system in Simrat, a long term store of different types of associative connections between different representations, and a short term store of the current contents of perceptual processing, together with activated stimulus-stimulus (SS) associations, and any stimulus-response (SR) sequences that have been established.

In order to discuss the structure of the long term memory it is necessary to mention briefly some of the terms connected to associative learning and animal conditioning (a form of associative learning). Associative learning is said to take place when an association is formed between two types of representation. The implication is that the occurrence of the first representation will activate a link to the second. In this model there are three types of association that can be formed:

Conditioned Stimulus - Unconditioned Stimulus (CS - US). An unconditioned stimulus is defined to be one that is intrinsically reinforcing to Simrat, i.e. on encountering a US, motor responses will be initiated to either increase its interactions with the stimulus (longer duration of interaction, closer proximity, etc. [in which case the US is termed appetitive]), or to decrease its interactions with the stimulus (shorter duration, larger distance etc. [in which case the US is termed aversive]). When a US is encountered, perceptual stimuli that are present at the time of its occurrence will be associated with it, by virtue of the process of associative learning. The stimuli in question are called conditioned stimuli, as their presence activates a conditioned association with the US. The strength of the CS - US association grows with repeated encounters by virtue of reinforcement.

Conditioned Stimulus - Conditioned Stimulus (CS - CS). Once an initially neutral stimulus becomes a CS it too can form other associations with stimuli that are present when it is encountered. Hence stimuli can exist as chains of associations from the initial CS down to the final US, e.g.: CS3 - CS2 - CS1 - US.

Conditioned Stimulus - Conditioned Response (CS - CR). Motor sequences that are used to either approach appetitive CSs or avoid aversive ones come themselves to be associated with the CS that triggered the response, and are called conditioned responses, CRs.

Now the basics of associative learning and conditioning have been introduced, it is possible to consider the structure of the long term memory. This is subdivided into two sections: an SS memory and an SR memory. The former contains associative stimulus pairs of two types: CS - US and CS - CS, while the latter contains associative pairs of one type: CS - CR. The Motor System takes the SRs from the short term memory when signalled to do so by the BAS, and converts them into motor responses which are manifested as changes in the world.

The Reinforcement System is the heart of the associative learning mechanism in Simrat. It is an attempt to extend the Rescorla-Wagner (Pearce, 1987) model of associative learning to include a notion of ‘attention’ as perceptual filtering, to take account of the Mackintosh (1975) attentional theory. The Rescorla-Wagner model of classical conditioning is notable, as despite its many shortcomings, it expresses mathematically the change ΔV_A in association strength between a stimulus A in terms of the asymptote of conditioning λ (dependent on the strength of the US), the sum of the associative strengths of all CSs present V_{sum}, and a generic learning rate parameter α, i.e.:

$$\Delta V_A = \alpha(\lambda - V_{sum}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

Mackintosh offers a different equation to account for the process of the change in associative strength, whereby such change depends only on the stimulus in question:

$$\Delta V_A = \alpha_A(\lambda - V_A) \quad (2)$$

In this case α_A is a stimulus specific learning rate parameter, dependent on the stimulus's physical properties and the rat's experience of it. Such an idea can be seen to be equivalent to part of existing two-process attention theories (Mackintosh, 1975) as the likelihood that a stimulus will be attended to by the rat determines whether the stimulus will be learned about and may itself change with experience. The associative learning mechanism in Simrat is based in part on the following equations which use parts of the above two equations:

$$\Delta V_S = \alpha_{SS}(\lambda_S - V_S) \quad (3)$$

$$\Delta V_S = \alpha_{SS}(0 - V_S) \quad (4)$$

Equations 3 and 4 combine features of equations 1 and 2 above to state the change in the associative weight of an SS relation (either CS - US or CS - CS). A point to note is that criteria dictating whether a relation is rewarded (3) or punished (4), are not implicit in the equations and come instead from the reinforcement system itself.

$$\Delta V_R = \alpha_{SR}(\lambda_{mod^{\delta K}}\lambda_R - V_R) \quad (5)$$

Equation 5 is similar to equations 3 and 4 in stating the change in the associative weight of an SR relation. This application of essentially the same learning equation to SR conditioning represents an extension to the ideas of researchers like Rescorla, Wagner and Macintosh who focus on SS conditioning. Equation 5 differs from 3 and 4 in its incorporation of the variable $\lambda_{mod^{\delta K}}$. This modifies the asymptotic strength of the SR conditioning to reflect changes in the degree of aversiveness (δK = δAv), non-aversiveness (δK = δA$\overline{v}$), appetitiveness (δK = δAp) or non-appetitiveness (δK = δA$\overline{p}$) (see (Gray, 1975) for details of these four terms) between the start and finish conditions of the motor response, as illustrated in figure 2.

Start state	Motor Response	Finish state
Environment_1	$\longrightarrow$	Environment_2

Figure 2 Environmental transition due to motor response

The four types of λ_{mod^K} are illustrated in equations 6 - 9 below:

Aversive	$\lambda_{mod^{Av}} = 1 - e^{-\delta Av - \xi^{Av}}$	$\delta Av = Av_1 - Av_2$	(6)
Non-aversive	$\lambda_{mod^{A\overline{v}}} = 1 - e^{-\delta A\overline{v} - \xi^{A\overline{v}}}$	$\delta A\overline{v} = A\overline{v}_2 - A\overline{v}_1$	(7)
Appetitive	$\lambda_{mod^{Ap}} = 1 - e^{-\delta Ap - \xi^{Ap}}$	$\delta Ap = Ap_2 - Ap_1$	(8)
Non-appetitive	$\lambda_{mod^{A\overline{p}}} = 1 - e^{-\delta A\overline{p} - \xi^{A\overline{p}}}$	$\delta A\overline{p} = A\overline{p}_1 - A\overline{p}_2$	(9)

The implications of the λ_{mod^K} variable are crucial to understanding the nature of the associative learning a rat undergoes in establishing conditioned motor responses in given environmental conditions. The key factor is the 'degree of change' between the start (e.g. Av_1) and finish (e.g. Av_2) states of the motor response, as preset by the expectancy of the finish state as given by the type of conditions prevalent in the start state. Hence, for example, if an aversive stimulus is present in the environment and the rat responds to it, to either remove the stimulus or in some way decrease its effective valence, λ_{mod^K} will depend on the degree of change of aversiveness due to the motor response, as per equation 6.

The four ξ constants in equations 6 - 9 reflect the difference between start and finish states required for λ_{mod^K}s value to change for the four stimuli types. I.e. the rate of change, which will be different, for example, between aversive and appetitive start and finish states (i.e. a small appetitive increase may be more rewarding that a similar level of aversive decrease).

This model attempts to deal partially with the notion of 'attention' by recognising the importance of novelty for associative learning, or as it is modelled here, non-novelty. This model draws a connection between two views of 'selective attention' as proposed by Mackintosh (1975) and Pearce and Hall (1980). Mackintosh argues that the more reliably a CS predicts a US, the more it gains the animal's 'attention', while the implications of the Pearce and Hall model are the opposite, i.e. that stimuli which have not been encountered before gain more of the animal's 'attention' than good predictors, which can be acted on with little processing (Pearce, 1994).

This model makes the connection by falling back to Gray's three systems, in the context of the animal data, as follows: The different types of stimuli (well known CSs and unknown Ss) are processed in fundamentally different ways by the architecture. Well established stimuli will be processed by either the BIS or the BAS depending on their aversive/appetitive nature. If the CS in question is aversive it will (in a similar way to an unknown stimulus) tend to cause the BIS to interrupt any current motor responses, and orient toward the stimulus in order to gain more data about it. Whether the architecture then proceeds to monitor the stimulus will depend on the degree that previous encounters with it have established suitable motor responses. Hence the overall behaviour is as follows. If a stimulus is unknown (or a known CS with no correspondingly strong SR) the architecture will orient toward it and monitor it to gather more data. This will consequently increase its likelihood of forming part of any learned associations that arise from the encounter. Conversely, if the stimulus is a known CS with a strong SR, the SR will be initiated as soon as the CS appears, and no orienting response will take place.

Stimuli occurring in the environments Simrat encounters, that have no appetitive or aversive properties, become conditioned to what is called non-novelty. Stimuli with non-novelty values take longer to be conditioned subsequently by the presence of any reinforcers (i.e. as evident in latent inhibition (Mackintosh, 1974)).

Finally, it is important to mention the notional body of Simrat, and its tangible components, the energy store/level, arousal level, and damage level. The former is the central energy repository of the body, whose level (mediated by the ambient bodily energy) is held in the Energy Level indicator. The arousal level represents the current flow of energy from the store into the body, and directly affects the effectiveness of the motor system and processing time constraints. The arousal level is an important indicator of potential threat to Simrat, and (as will be illustrated in the next section) influences the control and functionality of Gray's three systems during certain situations. There are arguably three categories of action open to a rat in an experimental situation:

$$\text{US (e.g. shock)} \longrightarrow \text{UR (e.g. running)}$$
$$\text{CS (e.g. buzzer)} \longrightarrow \text{SR (e.g. running)}$$
$$\text{CS (e.g. buzzer)} \longrightarrow \text{Arousal} \longrightarrow \text{SR (e.g. running)}$$

Arousal is measured experimentally in rats in terms of their heart rate, or galvanomic skin response (Gray, 1975). Initially when an aversive US arrives the BIS increases arousal and eventually the rat finds the correct UR to escape the CS. If the rat can learn (to some degree) which motor responses are effective to remove the aversive CS, then two possible situations can occur. If the rat can easily learn the correct SR, then as the SR is established, the ability of the CS to elicit arousal is reduced, until eventually the presence of the US causes the rat to produce the SR and then return to its behaviour prior to the stimulus occurrence with little or no change in its arousal level (Gray, 1975).

Conversely if the rat can only learn an SR that sometimes removes the aversive CS, then the CS will still cause the generation of arousal as it continues to be encountered. Such arousal is functional for the animal, both in preparing its body to resist the effects of the US, and in biasing the operation of its architecture to tend to continue initiating motor responses, beyond the point where previous learning has shown them to be ineffective; which may make the difference between survival and none. It is the reduction of arousal, or more accurately the response to arousal, that this model postulates as a central part of the notional 'drive' that existing two-process theories of learning require for learning to take place (Gray, 1975). The BIS reacts to aversive novel stimuli in order to increment arousal. To be able to do this it must store a correlation between a threatening stimuli and the occurrence of the threat in question. This correlation is termed the 'arousal correlate' within this model, and exists as a separate consequence of the conditioning process to the usual CS - US relation.

An Example Scenario

Simrat is being applied to several experimental scenarios. This paper discusses only the Kamin (1957) avoidance scenario, which explores the relative effectiveness of CS termination/US avoidance during avoidance learning. The apparatus for the experiment consists of a 'shuttle box' (a box divided into half by two projections at the centre of each side wall). Four groups of rats (eight rats in each) are tested (one rat at a time) in the shuttle box, each group under different conditions.

For all four groups an electric buzzer (CS) is sounded, followed some seconds later by the application of an electric shock (US) to the half of the box the rat was in when the buzzer sounded. They can all terminate the shock once it has started by moving to the other half of the box (UR, termed shuttling). The rat's motor responses influence the buzzer and shock as follows: for group one shuttling in advance of the shock prevents the occurrence of the shock (avoidance) and terminates the buzzer. For group two, shuttling prevents the shock but does not terminate the buzzer, which continues for several more seconds. For group three, shuttling terminates the buzzer but does not prevent the shock. Group four can neither prevent the shock in advance nor terminate the buzzer.

Figure 3 below illustrates the results obtained from both the experimental rat data and my simulation (an average of five runs). The two graphs 3A and 3B respectively show the percentage of correct shuttle responses (i.e. moving to the other half of the box on hearing the buzzer before the shock arrives) for each group of ten trials.

This experiment tried to demonstrate whether conditioning is reinforced by virtue of the termination of an aversive CS or the avoidance of an aversive US. As the results show, it is both, since either stimulus effect alone is only partly effective (groups 2 and 3 in figure 3). Several researchers (as well as Kamin himself) have tried to interpret these results, in order to argue about what is actually learned. However, such descriptions whilst being of great theoretical interest are of little practical value in understanding what is actually possible for a given set of postulated mechanisms to achieve (as was discovered by the author during testing of the present computer model Simrat).

Even a very convincing theoretical explanation of the phenomena such as that provided by Mackintosh (1974) is far from sufficient to explain what is happening, as references to generalities about certain types of stimuli miss the subtle types of interaction effects between these stimuli that make it so very difficult for one simplistic set of mechanisms to have all the properties necessary to reproduce the behaviour.

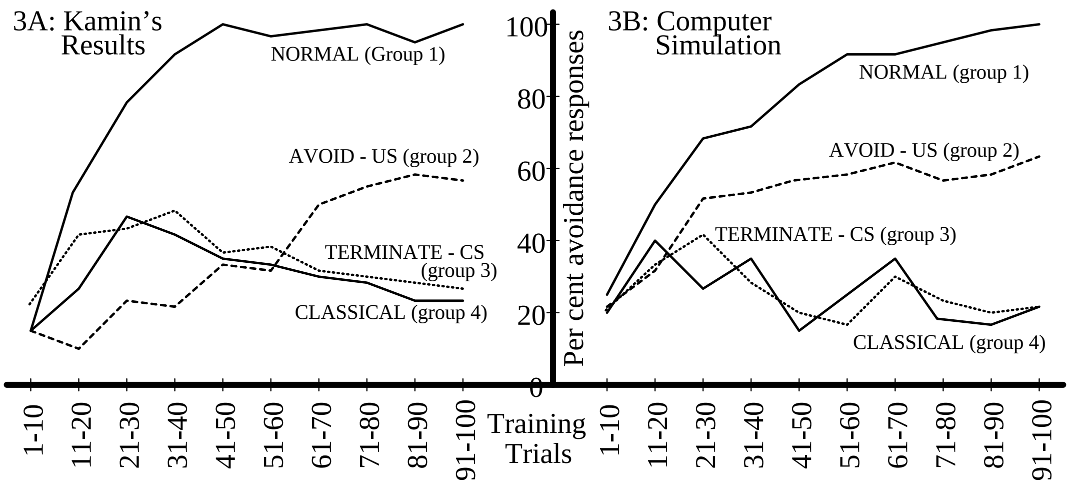

Figure 3 Avoidance per cent SRs for shuttle box experiment

The key to the success of the model (success is here defined to refer to the general structure [position, slope, value] of the behavioural responses illustrated in the graphs) in reproducing the experimental data (in this particular scenario) lies in three main areas: firstly, Gray's three systems which provide structure to the processing and interaction of environmental stimuli; secondly, the reinforcement system, and thirdly, the effects of arousal on Gray's systems. We now discuss how these effects in the model produce the behaviours illustrated in figure 3. The model represents the environment at each end of the shuttle box, together with the buzzer. So in total five stimuli combinations are possible. Table 1 below illustrates the final weights of these stimuli (only three of the five are shown in the table as the weights for each end of the box are very similar due to the behaviour being symmetrical), together with the weights of the corresponding shuttle and non-shuttle responses for each of the above four groups.

	box ends			buzzer			box ends + buzzer			SRs	
Groups	Av	A$\overline{v}$	Ar	Av	A$\overline{v}$	Ar	Av	A$\overline{v}$	Ar	sh	$\overline{sh}$
1	18.7	66.7	58.9	97.6	0	2.3	98.5	0	10.9	98.7	10
2	22.5	88.2	53.7	28.1	15	100	22.9	79	93.5	65.7	10
3	99.9	0	99.9	99.9	0	99.9	99.9	0	99.9	10	10
4	99.9	0	99.9	99.9	0	99.9	99.9	0	99.9	10	10

Table 1 Resulting SS and SR weights after 100 trials of the computer simulation (Groups: 1 = normal, 2 = avoid-US, 3 = terminate-CS, 4 = classical. Av = Aversive, A$\overline{v}$ = non-aversive, Ar = arousal correlate, sh = shuttle response, $\overline{sh}$ = non-shuttle response).

The first group are able to avoid the electric shock as well as terminate the buzzer, so as the first row in the above table illustrates, the box ends come to be non-aversive stimuli, and the buzzer combinations aversive. The shuttle response hence comes to be conditioned to both the aversive and non-aversive stimuli, and is strongly reinforced since the results are two fold (shock avoidance and buzzer termination). The arousal correlates are low as the correct performance of the shuttling response has prevented a correlation from forming between the aversive stimulus and shock. Hence when the buzzer sounds it is tagged as being an aversive stimulus by the BIS, which triggers the BAS both to tag non-aversive stimuli, and to initiate the correct shuttling response.

The second group respond initially to both the box ends and the buzzer as aversive stimuli, but once the shuttling response is used, the presence of the buzzer without shock, causes it and its combinations to gain non-aversive status. However the change in status of the buzzer combinations delay the learning of the shuttling response, so at the end of the trials whilst both box ends and buzzer combinations have become non-aversive stimuli, the shuttling response is still only partially reinforced, and the arousal correlate remains moderately high for the buzzer combinations. Hence when the buzzer sounds it is tagged as being non-aversive by the BAS. If the level of the arousal correlate is above an innately defined minimum level, then it will cause the BIS to raise the arousal level of the rat. This in turn causes the FFS to respond to the concomitant increase in the activation level (the probability that some motor response will be selected [via a transitory weight increase]) to attempt to initiate a response. If however the arousal correlate level is below the minimum defined level, then the BAS will initiate the shuttling response. Analysis of Simrat during the simulation of the group 2 conditions, shows that the initial difference between the slope of the graph for this group and the experimental results (compare figures 3A and 3B) is due to the way in which aversive and non-aversive stimuli complement each other. This effect may not occur for the actual animals, as the establishment of opposing CS types (e.g. non-aversive and aversive) for existing conditioned stimuli may be delayed. Further investigation is needed.

The third and fourth groups have similar behaviour, in both cases the shuttling response doesn't change the relative aversiveness between the start and finish states, so whilst buzzer and box combinations soon become aversive stimuli, the weights of the shuttle responses soon attenuate to an innate minimum level. It is the quick establishment of a high value for the arousal correlate for all stimuli that raises the level of the responses from an otherwise practically non-existent level. Hence when the buzzer sounds, the BIS establishes some combination as being strongly aversive, and the correspondingly high arousal correlate causes the BIS to increment the arousal level, at which point the FFS reacts to attempt to initiate a motor response as appropriate. The very low levels of the SRs weights make such reaction far from productive, as illustrated in figure 3.

The above partial analysis of how Simrat produces the behaviour manifested by the rats in the Kamin (1957) experiment demonstrates how Gray's systems explain escape and avoidance behaviour. It is also important in illustrating how such systems underlie certain types of avoidance behaviour present in human behaviour (Gray, 1987) (although it is beyond the scope of this paper to deal with this issue in any detail). The implications of the explanation of this experiment are as follows: the presence of things that represent a threat to a human subject, can influence behaviour (either leading to immediate action or the initiation of longer term problem solving). The systems Gray propose that evolution has left us with, provide a mechanism for reacting to danger and threat in the environment. In situations where the threat is clear and a suitable response can be initiated, physiological arousal may remain low, and the effects of the BIS and BAS may be evident to the subject only in terms of the way their mental functioning changes in their presence.

A point of interest is the role of physiological arousal in this process. As for the rat in the shuttle box, the appearance of a threat can trigger arousal, which itself can lead a human subject to start to act to reduce that arousal even when no immediately clear response is at hand. It is the reflection on such arousal states that often form part of what are called 'emotional states' (James, 1884). The human mental architecture is far richer than that of the humble rat, and hence our capabilities to make use of such states are far more complex. Visual imagery (such as day dreaming) can invoke aversive images that cause the BIS to increment arousal. The salience with which this provides the image and associated mental processes, can lead to the aversive images being the subject of planning, problem solving, and general ruminative thought processes.

It is consequently not surprising that so much clinical effort has been devoted to developing behavioural and drug based therapies that act to prevent the BIS from affecting arousal and the concomitant mental processes (Gray, 1987). A final note to make on this subject is the author's intention that the future of this work on Gray's system is to extend the model to include separate systems that model the problem solving and planning processes often associated with 'cognition'. Such a model would hence be able to explore the effects of Gray's systems in influencing these processes.

Conclusion

This paper outlines the basis for the way in which Systemic Design has been applied to the three systems which underlie human emotional behaviour as proposed by Gray. While his claim is far from generally accepted in the research community, the systems provide a credible and interesting route for Systemic Design to study 'emotion'. The experiment detailed here has proved difficult for researchers to explain. The model presented here provides a considerably more rigourous explanation than previous purely psychological explanations, as it has been demonstrated that the systems it contains function together to meet the experimental behaviour.

The work outlined in this paper is approaching completion. When finished, the satisfied four research objectives outlined earlier in this paper are expected to give rise to a detailed analysis of the functional composition of Gray's systems in a way that could not have been achieved in conventional experimental or Neuropsychological paradigms.

References

Bates, J., Loyall, B., and Scott, R. W. (1991). Broad agents. *SIGART BULLETIN*, 2(4):38–40.

Chapman, D. (1990). Vision, instruction and action. Technical Report 1204, MIT Artificial Intelligence Laboratory.

Gray, J. A. (1975). *Elements of a Two-Process Theory of Learning*. Academic Press.

Gray, J. A. (1987). *The Psychology of Fear and Stress*. Cambridge University Press.

Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition and Emotion*, 4(3):269–288.

James, W. (1884). What is an emotion ? *Mind*, 9:188–205.

Kamin, L. J. (1957). The effects of termination of the cs and avoidance of the us on avoidance learning: an extention. *Canadian Journal of Psychology*, 11(1):48–56.

Mackintosh, N. J. (1974). *The Psychology of Animal Learning*. Academic Press.

Mackintosh, N. J. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. *Psychological Review*, 82(4):276–298.

Pearce, J. M. (1987). *An Intoduction To Animal Cognition*, chapter 4, pages 134–149. Lawrence Erlbaum Associates Ltd.

Pearce, J. M. (1994). Personal communication.

Pearce, J. M. and Hall, G. (1980). A model for pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87(6):532–552.

Read, T. (1993). Systemic design: A methodology for investigating emotional phenomena. In Rowles, C., Liu, H., and Foo, N. (Eds.), *Proceedings of AI '93, Melbourne, Australia*. Singapore: World Scientific.

Read, T. and Sloman, A. (1993). The terminological pitfalls of studying emotion. in Proceedings of WAUME '93, University of Birmingham, UK.

Sloman, A. (1993). Prospects for AI as the general science of intelligence. In *Prospects for Artificial Intelligence, proceedings of AISB93, the ninth biennial conference of the Society for the Study of Artificial Intelligence and the Simulation of Behaviour*, pages 1–10. IOS Press.