A Computational Model Of Affect Theory: Simulations Of Reducer/Augmenter & Learned Helplessness Phenomena

Michael Sean Davis

Inflexxion, Inc.
320 Needham Street, Suite 100
Newton, MA 94025
mdavis@inflexxion.com

Overview

This paper outlines the initial results of an ongoing research effort to build and test a computational model that simulates key components of a theory of mind proposed by psychologist Silvan S. Tomkins in his four-volume series *Affect, Imagery, Consciousness* (1962, 1963, 1991, 1992). Tomkins' model of the mind is able to account for an impressive range of motivational phenomena, including the crucial developmental link between cognition and emotion, how conflicting motives are resolved and decisions about moment-to-moment behavior are made, and how individuals can be motivated simultaneously by highly idiosyncratic goals that emerge from experience, as well as the complete array of innate, instinctual drives.

To begin testing the explanatory power of Tomkin's model, key components of his theory have been implemented in computational terms and tested in experimental simulations. In particular, Tomkins' *affect theory*, a sub-theory of his overall view of the mind, has been targeted for primary research purposes. Affect theory has been selected as the focus of study because it is the most unique component of Tomkins' model, and many authors, e.g., Nathanson (1992) and Schore (1994), consider it to be his primary contribution to the understanding of the mind. Affect theory also has the potential to explain how "primary" emotions, as described by Damasio (1994), are generated and can evolve into cognitively generated "secondary" emotions over time.

Reducer/augmenter (Larsen & Zarate, 1991) and learned helplessness phenomena (Peterson, Maier, & Seligman, 1993) were selected for simulation purposes. These affective/behavioral phenomena have been repeatedly documented to occur in studies involving living subjects. Moreover, these phenomena are at the opposite ends of the nature/nurture continuum. For reducer/augmenter phenomena, differences in affectively motivated behavior are attributed to innate, constitutional factors. In contrast, learned helplessness effects are attributed to differences in environmental experience. The simulation of these causally diverse affective/behavioral tendencies, therefore, should provide a useful initial test of the range of phenomena for which Tomkins' theory can account.

PASSIO

The agent developed to embody Tomkins' affect theory is named PASSIO. The word "passio" is the Latin root for passion, which encompasses all the emotions, including joy, anger, distress, fear, and excitement, all of which have been incorporated into the system architecture. PASSIO is designed to function in an environment populated with abstract *objects*. Theoretically, objects can possess any number of features, but for the research described here objects have a location, size, and identifier (i.e., a unique value serving to distinguish each object from others in the environment).

Six system components constitute PASSIO's information processing and behavioral capacities. These include *feature detectors* (e.g., senses) designed to convert information about abstract features into a *message*; an *analyzer circuit* that uses all the messages produced by the feature detectors to calculate the system's arousal level; an *affect system* that tracks the system's arousal levels and triggers an affect message; a *cognitive system* that recommends the objects the system should pursue as a goal; *a motor system* that controls PASSIO's movement in the environment, and a *memory system* that records the affective experiences associated with objects PASSIO encounters.

The affect system produces five of the six primary affects conceptualized by Tomkins. These include *anger, fear, distress, interest, and enjoyment.* Based on Nathanson's interpretation of Tomkins' theory, a sixth affective state, *boredom* or no affect, is also included in the system. In accordance with Tomkins' model, affects are generated by patterns in arousal levels across time. Specifically, enjoyment is caused by decreases in arousal, interest and fear is caused by moderate and steep increases in arousal, respectively, and distress and anger are caused by persistently high levels of arousal, with the threshold for distress being lower than for anger. Boredom is generated when none of the triggering conditions for the five affects are satisfied.

The memory system records the affective experiences and familiarity data associated with the objects that PASSIO encounters. Whenever the distance between PASSIO and an object is less than a specified threshold, the updating procedures for the memory system are activated. PASSIO maintains data on the number of times it has encountered each object, as well as five different affective strength values (i.e., one for each affect). Affective strength values are updated using a modified version of the Widrow-Hoff technique outlined by Wilson (1995) based on the active affect at the time of each encounter.

PASSIO's cognitive system is limited to one primary task: selecting goals for the system to pursue in the form of objects. The cognitive system consists of four simple stochastic modules, each one named for the type of behavioral activity it is intended to produce. These modules are *withdraw*, *engage*, *familiarize*, and *explore*. All the modules have specific affective triggers: enjoyment activates familiarize, interest activates explore, any of the negative affects activates withdrawal, and boredom activates engage. Each module uses distinctive criteria for selecting goals designed to maximize the system's experienced ratio of positive to negative affect based on the inherent logic of Tomkin's model. That is, withdraw biases the system towards objects associated with enjoyment, engage towards objects associated with interest, explore towards new objects, and familiarize toward nearby objects.

Once the cognitive system selects a goal or object to pursue, the motor system moves PASSIO from its current location toward the goal in the most efficient manner possible (i.e., in a straight line). The speed with which PASSIO moves towards the goal is proportional to the system's arousal level, with higher arousal levels generating higher speeds. This feature reflects Tomkins' *principle of correlation*, which states that the physical response generated by an affect generally resembles or "correlates" with the affect's triggering stimulus.

Reducer/Augmentor Simulation

Reducer/augmentor phenomena deals with the relationship between innate neurological differences, levels of Central Nervous System (CNS) stimulation, emotional responsiveness, and behavioral tendencies. Researchers that study the phenomena contend that there are powerful individual differences in the way people subjectively experience and respond to the same sensory stimulation. At the opposite ends of this continuum, there are two types of individuals referred to as *augmenters* and *reducers*. At the physiological level, augmenters amplify or increase incoming sensory stimulation, whereas reducers dampen or decrease the same sensory data. In order to explain these differences, researchers postulate the existence of a centrally located stimulus intensity modulation mechanism. Theoretically, this mechanism augments sensory stimulation for augmenters and attenuates it for reducers. As a result, reducer/augmenter theory generates specific predictions about how augmenters and reducers will experience and respond to different levels of environmental stimuli. Stemming from the dampening effects of the modulation mechanism, it is assumed that reducers experience low-stimulation conditions as generally more dull and monotonous than augmenters. It is also presumed that reducers pursue stronger and more intense forms of stimulation than augmenters in an effort to counteract this effect. Over time, various methodologies have been developed to differentiate between augmenters and reducers (i.e., Petrie, 1967 and Vando, 1974). Using these methods, decades of empirical research generally support the predictions of reducer/augmenter theory (i.e., Buchsbaum & Silverman 1968 and Mishara & Baker, 1981). Reducer/augmenter has also been extended into the domain of emotion. Larsen & Zarate (1991) have confirmed that reducers find low-stimulation tasks more boring than augmenters, and that reducers are more likely to engage in activities that have a high probability of evoking emotion in their daily lives. Indeed, Larsen concludes that reducers "seek out" forms of CNS stimulation, including "strong emotional responses," in order to "regulate their arousal upward" (Larsen & Zarate, 1991, p.721).

Environment

All reducer/augmenter simulations were conducted in a Gradated Object (GO) environment, referred to as GO. This version of GO was rectangular in shape, contained seven rows and seven columns, and was populated with oval objects. Each row was constituted by seven uniformly sized and distributed fixed objects, which essentially serve as landmarks, as well as a singular moving object traveling back and forth at constant speeds along the x axis. The size and speed of moving objects were gradated according to their position on the y axis, with higher y positions being associated with proportionally greater size and speed. Thus, rows closer to the top of the environment contained moving objects that were faster and larger than those in rows near the bottom. See Figure 1 below for a graphic representation of GO.

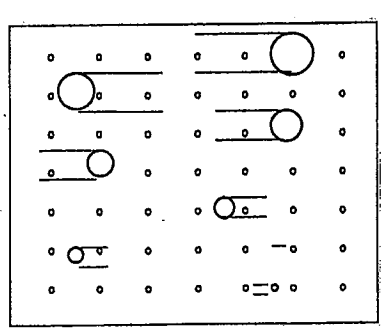

Figure 1: GO Environment

Finally, all objects, fixed and moving, generated a form of constant "radiant energy" directly proportional to their size, with larger objects generating more energy than smaller ones. The purpose of using the GO environment was to simplify the measurement of sensation-seeking, as higher positions on the *y* axis are explicitly associated with more stimulation.

Experiment 1

A population of 150 PASSIO subjects were placed in the GO environment, and each run for a total of 20,000 time steps. All subjects had a *visual detector* that processed the radiant energy generated by an object, so that larger, closer objects caused greater arousal levels than smaller objects further away. To determine their place along the reducer-augmenter continuum, each subject was randomly assigned a *reducer-augmenter coefficient* ranging from 0.75 to 1.25 using a uniform distribution function. This coefficient was multiplied by the system's arousal level each time step, so subjects with coefficient values greater than 1 were analogous to augmenters and those with values less than 1 were analogous to reducers. Given the structure of the GO environment, a subject's average position on the *y* axis over time was used as a measure for sensation seeking behavior.

Subjects' reducer-augmenter coefficients were then correlated with their sensation-seeking score, i.e., there average location on the *y* axis across all time steps. As hypothesized, reducer-augmenter coefficients were negatively correlated with overall sensation-seeking ($r = -0.54$, $p < .001$), so that subjects with lower reducer-augmenter coefficients generally tended to prefer environmental regions closer to the top of the environment than subjects with higher reducer-augmenter coefficients.

Although the experiment confirmed the basic hypothesis concerning general sensation-seeking tendencies of reducers and augmenters, there was evidence that reducers and augmenters may actually have more complex sensation-seeking profiles than expected. When viewing sample runs, it appeared that augmenter subjects initially sought out more stimulation, i.e., preferred locations higher on the *y* axis, than reducer subjects. Over time, however, reducers and augmenters appeared to settle into their expected patterns.

Experiment 2

To explore the possibility that reducers and augmenters may actually have more intricate sensation-seeking profiles than anticipated, another experiment was conducted. Intended to facilitate assessment of sensation-seeking profiles of reducers and augmenters over time, the design for this experiment was a repeated measures MANOVA with 12 assessment periods, one factor (reducer or augmenter), and one dependent variable (sensation-seeking). In this model, each assessment period represented the average *y* location over 1667 time steps, so again each run lasted for approximately 20, 000 time steps. Each subject (N = 150) was assigned a reducer-augmenter coefficient with procedures identical to those used for the correlation study described above. Subjects with coefficient values greater or equal to 1.0 were considered augmenters, and those with values less than 1.0 were considered reducers.

An analysis of the results revealed a significant within subjects interaction effect [$F(11, 1639) = 7.29$; $p < .001$]. A plot of the mean sensation-seeking scores for reducers and augmenters at the 12 assessment periods (shown below in Figure 2) revealed the nature of this finding. The plot suggests the existence of a cross-over effect in regard to the sensation-seeking tendencies of reducers and augmenters. That is, in an unfamiliar environment, augmenters initially appear to engage in more sensation-seeking activity than reducers. As subjects became more familiar with the environment over time, however, both reducers and augmenters appear to discover the environmental regions in which they feel the most comfortable. As predicted by the research literature, for reducers these environmental regions ultimately involve higher levels of stimulation than those regions preferred by augmenters.

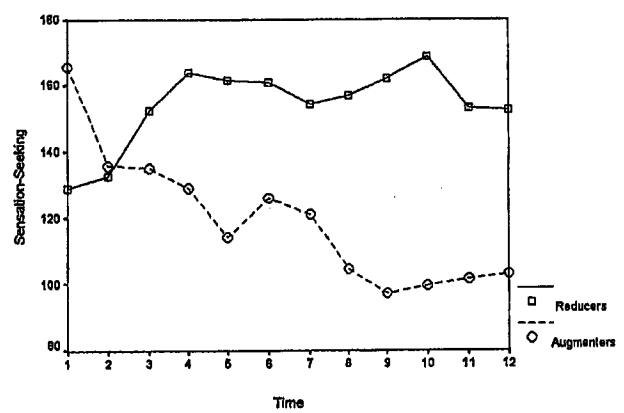

Figure 2: Sensation-seeking scores across time

Learned Helplessness Simulation

Learned helplessness theory postulates a fundamental association among three variables: *contingency*, *cognition*, and *behavior* (Peterson, Maier, & Seligman, 1993). Contingency refers to the objective relationship between actions and their outcomes. Cognition refers to the way in which contingencies are understood and explained, with the content of the explanation creating future expectations about how particular actions and outcomes are related. Finally, behavior refers to the observable consequences of the cognitions that have developed as a result of experience.

Given the nature of the relationship between these three psychological variables, learned helplessness theory stipulates that one of the most fundamental contingencies for both animals and people alike is controllability versus uncontrollability. In particular, when either animals or people or exposed persistently to contingencies characterized by uncontrollability, a set of expectations evolve that produces a constellation of maladaptive behaviors. One of the most problematic consequences is the development of a chronic passivity that precludes the discovery of more favorable circumstances. Other consequences include cognitive retardation, low self-esteem, and heightened negative affectivity. As a whole, these consequences are referred to as "learned helplessness effects."

To test the essential assertions of learned helplessness theory, the vast majority of learned helpless studies have utilized the triadic design, widely considered to be the gold standard in this area of research. The triadic design consists of three different conditions: a *controllable condition*, an *uncontrollable condition*, and a *no-pretreatment comparison condition*. In the controllable condition, subjects are exposed to a series of controllable aversive events. That is, each aversive event can be terminated if the subject discovers and engages in a particular response, such as pressing a lever or turning a wheel. Subjects in the uncontrollable condition are then paired with a member of the controllable condition, and exposed to a series of aversive events yoked to those previously delivered to the other member of the pair. As a result, an identical stimulus profile is administered to both groups of subjects. The critical difference is that subjects in the controllable condition can control their experience of the aversive stimuli via their behavior, whereas subjects in the uncontrollable condition are unable to influence their circumstances. Finally, subjects in the no-pretreatment condition are not exposed to any aversive stimuli, either controllable or uncontrollable. After these stimulus patterns have been administered to subjects in the different treatment conditions, all subjects are tested on a task reflecting disruptions attributed to learned helplessness. Typically, learned helplessness effects are inferred when subjects in the uncontrollable condition demonstrate difficulties with the task relative to subjects in the remaining two treatment conditions. Using this methodology, hundreds of studies have been conducted on both animal and human subjects. Consistently, the outcome is that only those subjects in the uncontrollable condition exhibit learned helplessness effects (Villanova & Peterson, 1992).

Feature Detectors

To test key theoretical constructs associated with learned helplessness phenomena, it was necessary to implement feature detectors for perceived efficacy, drive, and pain, as well as noise (e.g., a detector that consistently generated messages of random intensity levels). Perceived efficacy was viewed as the system's ability to attain a desirable outcome, it was set as inversely related to arousal levels, and was operationally associated with the affective strength values of the objects likely to be selected as the next system goal That is, efficacy levels were high and arousal levels low when likely goal objects were close and associated with high ratios of positive to negative affect. In contrast, efficacy levels were low and arousal high when likely goal objects were nearby and associated with high ratios of negative to positive affect. Drive, pain, and noise levels were positively and proportionally related to experimentally manipulated external stimuli. That is, arousal levels were increased when drive, pain, and noise levels were high, and decreased when these stimuli were low.

Environment

All learned helplessness simulations were conducted in a *shuttle box environment* that resembles those used in learned helplessness studies with laboratory animals. The basic shuttle box was rectangular in shape and contained 9 fixed oval objects that formed three rows and three columns. Unlike in the GO environment described previously, all the fixed objects were sufficiently large and pressed against one another so their perimeters were touching. Each object, therefore, constituted an environmental space. See Figure 3 below for a graphic representation of the shuttle box environment.

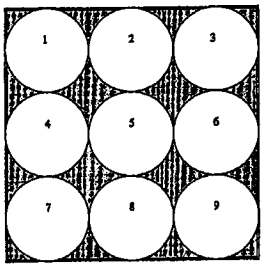

Figure 3: Shuttle Box Environment

Experiment

To simulate the classic triadic design, four-hundred and fifty PASSIO subjects were divided randomly into three groups: controllable, uncontrollable, and no-pretreatment. Subjects were run in groups of three, starting with a controllable condition subject, an uncontrollable condition subject, and then a no-pretreatment comparison subject. In addition, all subjects were randomly assigned an efficacy level condition: none, low, and high. For subjects in the no efficacy condition, perceived efficacy had no influence on the system's overall arousal level. For subjects in the low and high efficacy conditions, perceived efficacy inversely influenced arousal levels slightly and highly, respectively.

At the start of experimental procedures, subjects in the controllable condition were placed in a starting area (area

9) and exposed to a series of "shocks" (e.g., high levels of stimuli sufficient to generate negative affect) emitted randomly from the various objects in the environment. In the controllable condition, these shocks were temporarily terminated when subjects overlapped with a pre-assigned reward area (area 1) in the environment. After the reduction in the pain stimulus occurred, subjects were moved back to the starting area. Uncontrollable subjects were then placed in the same starting area, and exposed to identical "shocks" as the previous subject in the controllable condition. In the uncontrollable condition, however, subjects were unable to terminate the "shocks," regardless of their behavior. Also, uncontrollable subjects were allowed to roam freely, and were not moved back to the starting area when overlapping with the reward area. All controllable and uncontrollable conditions subjects were run for a total of 5000 time steps with these contingencies in effect. Subjects in the no-pretreatment condition were neither exposed to the treatment environment nor any form of aversive stimuli.

Subjects in all three conditions then underwent identical assessment procedures during which they were evaluated for learned helplessness effects. The shuttle box environment used for these assessments was identical to the one used during the treatment procedures. That is, all objects used in the original treatment environment remained unchanged in the assessment environment. As part of the assessment protocol, however, modifications to key environmental contingencies were made. First, no pain stimuli were administered to subjects as part of the procedures. Second, the starting and reward areas were moved to area 7 and area 3, respectively. These changes made it necessary for controllable and uncontrollable subjects to discover new contingencies as part of their task. Third, the effect of overlapping with the reward area was modified. Rather than initiating a reprieve from pain stimuli, which were not included in the assessment protocol, a reduction in drive levels typically sufficient enough to trigger enjoyment occurred. Finally, when subjects overlapped with the reward area they were automatically moved to the starting area. All subjects were run for a total of 2500 time steps with these contingencies in effect. During these runs, subjects were evaluated for *task performance* (number of times overlapped with reward area), *activity levels* (mean standard deviations of locations on the *x* and *y* axis), and *negative affectivity* (total amount of negative affect).

Treatment condition main effects, efficacy level main effects, and interaction effects were significant for all three variables (e.g., $P < .001$ in all nine cases). Given these effects, post hoc pair wise comparisons were run separately for subjects in each efficacy condition. For subjects in the no efficacy and low efficacy conditions, these analyses revealed pair wise patterns inconsistent with learned helplessness phenomena. For subjects in the high efficacy condition, however, the analysis revealed pair wise patterns consistent with learned helplessness phenomena for all three variables (i.e., in comparison to controllable and no-pretreatment scores, mean uncontrollable condition scores were significantly lower on task performance, significantly lower in activity levels, and significantly higher on negative affectivity).

As depicted in Figures 4 and 5 below, an especially striking finding was the differences in the task performance and activity scores across the three treatment conditions. In the low efficacy condition, for example, the mean task score for subjects in the uncontrollable condition was near optimal (>70), and even exceeded the scores of subjects in the controllable condition. In the high efficacy condition, however, the mean task score for the uncontrollable subjects plunged to almost 0, while the scores for both the controllable and no-pretreatment groups remained at near optimal levels. These poor performance scores were the result of uncontrollable condition subjects becoming "frozen" in the starting area – a phenomena widely reported to occur in learned helplessness studies involving animals.

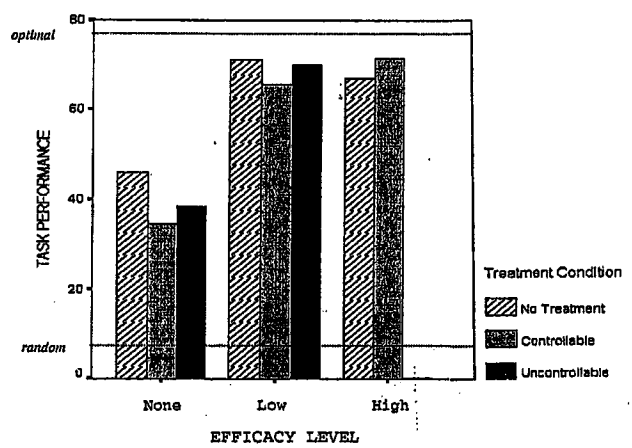

Figure 4: Mean task performance scores for all conditions

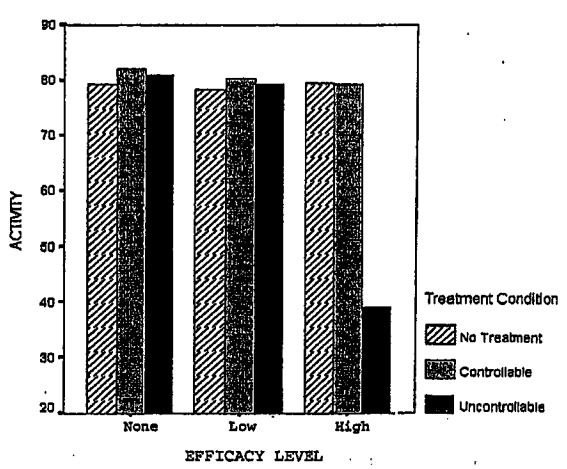

Figure 5: Mean activity scores for all conditions

Discussion

This research has implications for a range of theoretical and methodological issues, including those related to Tomkins' theory of affect and overall model of the mind, the relationship between cognition and emotion, and the development of truly autonomous agents. First, that a system motivated almost exclusively by simulated affect behaved consonantly with well-replicated phenomena from studies involving living subjects strengthens Tomkins' basic view of the affect system. In particular, the simulation results are consistent with Tomkins' claim that affects, as he conceptualizes them, represent a primary form of motivation, and that their motivational power is activated by patterns in arousal levels across time.

Second, the results also suggest, however, that Tomkins' model would be improved if it included the supposition that arousal levels are inversely related to moment-to-moment assessments of perceived efficacy. Indeed, with the efficacy concept included, Tomkins' theory is able to provide the crucial theoretical link between cognition and emotion often glossed over in other models. Although numerous other models, such as the OCC model (Ortony, Clore, & Collins, 1988), recognize the critical connection between cognition and emotion, they often fail to explicate the underlying *physiological* and *developmental* basis of this relationship. In contrast, when efficacy levels are linked to arousal levels within the context of an affect system as conceptualized by Tomkins, it becomes possible to explain the basic physiological processes by which cognitions and emotions merge and become fused over time and with experience.

Finally, the results of this research may also have some implications for the development of autonomous adaptive systems (i.e.,agents that can "cope or survive in an incompletely known and uncertain environment independently of external control" (Wilson, 1996, p.325). As illustrated by PASSIO – which was not designed to attain a predetermined goal or accomplish a predefined task – agents motivated by an affect system similar to the one proposed by Tomkins are indeed able to function autonomously. These results suggest, therefore, that researchers attempting to develop truly autonomous agents may find Tomkins' work an especially useful resource.

References

Buchsbaum, M. & Silverman, J. (1968). Stimulus intensity control and the cortical evoked response. *Psychosomatic Medicine, 30,* 12-21.

Damasio, A.R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain.* Gosset/Putman Press, New York, NY.

Larsen, R.J. & Zarate, M.A. (1991). Extending reducer/augmenter theory into the emotion domain: The role of affect in regulating stimulation level. *Personality & Individual Differences, 12, 713-723.*

Mishara, B.L. & Baker, A.H. (1981). Individual differences in stimulus intensity modulation in the elderly. *International Journal of Aging and Human Development, 13,* 285-295.

Nathanson, D.L. (1992). *Shame and Pride: Affect, Sex and the Birth of the Self.* New York: W.W. Norton & Company.

Ortony, A., Clore, G.L., Collins, A. (1988). *The Cognitive Structure of Emotions.* Cambridge University Press, Cambridge, MA.

Peterson, C., Maier, S., Seligman, M. (1993). Learned Helplessness: A Theory for the Age of Personal Control. New York: Oxford University Press.

Petrie, A. (1967). *Individuality in pain and suffering.* Chicago: University of Chicago Press.

Schore, A.N. (1994). *Affect Regulation and the Origin of the Self: The Neurobiology of Emotional Development.* Hillsdale, NJ: Elbaum.

Tomkins, S.S. (1962). *Affect, Imagery, Consciousness: Vol. 1. The Positive Affects.* New York: Springer Publishing Company.

Tomkins, S.S. (1963). *Affect, Imagery, Consciousness: Vol. 2. The Negative Affects.* New York: Springer Publishing Company.

Tomkins, S.S. (1991). *Affect, Imagery, Consciousness: Vol. 3. The Negative Affects: Anger and Fear.* New York: Springer Publishing Company.

Tomkins, S.S. (1992). *Affect, Imagery, Consciousness: Vol. 4. Cognition: Duplication and Transformation of Information.* New York: Springer Publishing Company.

Vando, A.A.. (1974). The development of the R-A scale: A paper and pencil measure of pain tolerance. *Personality and Social Psychology Bulletin, 1,* 28-29.

Wilson, S.W. (1995). Classifier fitness based on accuracy. *Evolutionary Computation,* 3, 149-175.

Wilson, S.W. (1996). Explore/Exploit strategies in autonomy. In P. Maies, M. Mataric, J. Polloack, J.A. Meyer and S. Wilson (Eds). From Animals to Animats (pp. 325-332). Cambridge, MA: The MIT Press/Bradford Books.