Chown, E., Jones, R.M., & Henninger, A.
Emotions as Adaptive Behavior: A Model
(Very) Rough Draft

Introduction

People are suspicious of emotions. Decisions made under emotional circumstances are discounted as being less than sound. Prominent researchers in artificial intelligence give talks on why AI systems should never contain emotions. Emotions are not rational after all, and good judgement comes from rationality, or so the logic goes. In recent years, however, another view has emerged, one that accounts for the limitations of rationality and the adaptive advantages of emotional behavior.

In an influential book, *Descartes' Error*, Antonio Damsio (1994) recounted cases of patients apparently devoid of emotions after serious brain injuries. In some cases these patients were perfectly capable of "normal" reasoning and logic but still made very poor decisions. Damasio characterized one such patient's decision making landscape as "flat." The patient could see the alternatives, but the consequences of the various outcomes had no meaning to him. Damasio's "somatic marker hypothesis" postulates a connection between what we are thinking about and bodily states. For example, when thinking about an unpleasant experience one might get an unpleasant feeling in one's stomach. In patients such as the ones just described, this relationship apparently does not exist. In normally functioning people, on the other hand, these feelings can inform decision making without the need for elaborate deliberation.

The thesis of this article is that "gut feelings" on the whole are not capricious, but generally reflect a kind of compacting of experience that normally leads to fast, accurate, decisions. Organisms that rely on information rather than sharp claws must necessarily make high quality decision very quickly. Whereas a rational creature might analyze a situation in depth and rate possible outcomes on a scale from –10 to 10, an emotional creature's rating is a feeling rather than a number, and it is the feeling that translates a prediction into action. There is a sense in which the major difference between the two situations is that in emotional organisms a great deal of the processing work has already been done. Positive and negative feelings represent appraisals, after all, and can vary in magnitude on an arbitrary scale. The commitment to a reliance on preprocessed information reflects the importance of fast decision making in a dangerous world. It is the difference between quickly satisficing in order to address one's feelings and being lost in thought trying to figure out exactly what the optimal course of action is.

It is our contention that the three mechanisms of emotion that we focus upon in this paper provide fast, albeit heuristic, answers to three important questions about the state of an organism in the world. 1) How important is the current situation? 2) Is it good or bad? 3) Will I be able to deal with it effectively?

Cognition is often analyzed in artificial settings – comfortable adults passively making decisions that may not have any connection to their own lives. While this methodology can be extremely productive, especially for isolating specific cognitive factors, it may not be a true test of the cognitive system as a whole. For better or worse, emotions have a profound effect on human decision-making in real-life situations (Schwarz, 2000).

In this article we take the position that as the human capacity for processing information grew, so too did the need to process the information quickly and efficiently. The emotion system systematically expanded to fill these needs by providing a way to modulate the strength of responses, and by categorizing information by its potential usefulness as well as its quality. After examining how the emotion system serves human informational needs, we go on to look at the mechanisms of emotion in more detail. We then present an implemented computer model of the system, and detail the results of experiments that show the ways that emotions impact decision-making.

This is not intended to be a comprehensive theory of emotions. Such a theory would need to address a number of aspects of emotion, such as social and meta-cognitive aspects, that are beyond the scope of the article. We are also leaving out many of the physical aspects of emotions. Cognitive pain, for example might be accompanied by nausea, elevated heart rate, etc. Instead, the focus of this article is on what we view as the foundations of emotion, brain mechanisms specifically oriented towards automatically assessing and responding to the world. Notably, this leaves out more cognitive appraisals of information such as "am I in control of this situation?" This is in contrast to the Ortony, Clore and Collins (or OCC) model of emotions (1988) that essentially ignores mechanisms and subsequently includes a wider range of factors. We would argue that the inclusion of more purely cognitive factors is one reason why, for example, there is such little agreement on issues such as how many emotions there are and what their names are. It is simply very difficult to completely disentangle cognitive and emotional factors and we have chosen to do so by focussing on general purpose mechanisms. We feel that one benefit of beginning with mechanisms is that we have identified several very general mechanisms that cover a great deal of the emotional spectrum and provides a subsequent foundation for looking at how cognition and emotions interact.

Emotions as Adaptive Behavior

The beginnings of the human emotional system can be found in the limbic system. Our perspective is a direct descendant of the work of Kaplan (1991) who views the evolution of humans through the lens of their information processing needs. The critical idea is that each evolutionary step in the development of the emotional system further refines and develops the capacity of the system to serve the organism's information processing capabilities. At its most basic level any large-scale brain system needs to be general purpose with later refinements allowing for more speciality. There is a misconception often seen in popular literature (and in scientific literature as well) that emotions are a deterrent to rational behavior. This view is reinforced by the work of Kahnemen and Tversky and others (Tversky, & Kahneman, 1974; Nisbett & Ross, 1980) who have repeatedly found evidence that human decision making is decidedly not rational.

However, just as that data can be reanalyzed from an evolutionary perspective to show that rationality is not always what it seems at first blush, so too can human emotions.

From our perspective the primary role of emotions is as a fast way of detecting and reacting to important events. In a sense emotions represent a kind of complex chunking in which important triggers are tagged with strong reactions that can help to modify behavior in appropriate ways. In a dangerous world there is not always time for analysis and deliberation, so the emotional system provides a fast alternative to rationality. This is not to say that emotional responses are necessarily suboptimal however, on the contrary we will make the case that under normal circumstances they are appropriate and afford great advantages, mainly in speed, over a purely rational approach.

In this section we will provide an overview of the three systems that form the core of our model of emotions. We present the systems in order from oldest to newest and attempt to show how each subsequent system adds to the functionality of the emotional system as a whole.

Detecting Importance – The Arousal System

The most primitive building block of the emotion system appears to be the arousal system. This is based upon its seat in the oldest part of the brain, as well as the fact that many very simple organisms appear to have a kind of arousal system. In its most basic form the arousal system is a call to action. Simple organisms, for example, will move quickly when stimulated. As an evolutionary response, the heuristic seems to be that when things are happening it is better to do something, anything, than to do nothing at all. As we shall see, as other parts of the emotional system are put into place the effects of the arousal system will become increasingly more sophisticated. Nevertheless, the basic notion of arousal as a "strength of response" modulator will remain intact.

Valence – Pleasure and Pain

An obvious problem with the "do something in the face of stimulation" heuristic is that it does not differentiate between stimuli. The next piece added to the emotional system, the pleasure and pain system, helps make basic distinctions between beneficial and harmful stimuli. The adaptive benefits of a pleasure/pain system are virtually self-evident. At their most primitive levels painful stimuli are damaging to an organism while pleasurable stimuli are either replenishing or are oriented towards reproduction. Pleasure and pain, without the need for analysis, provides an organism with a strong message about its current state. Activities that bring pain need to be terminated quickly, while activities that bring pleasure should be continued. These signals are immediate and do not require any intermediate processing. By themselves pleasure and pain confer an adaptive advantage to organisms for the simple fact that pain should be avoided and pleasure extended. A creature that simply retracts a limb upon feeling pain has an adaptive advantage over one that must analyze the sensation and determine a rational course of action.

Pleasure and pain both denote important events, and therefore will increase an organism's arousal level. Whereas undifferentiated arousal might yield undifferentiated activity (for example agitation), pleasure and pain provide additional information for an improved general response. In the case of pleasure the response should be to *arrest* any new responses – taking a new action might cause the pleasure to stop. For pain, on the other hand, the proper response is to *excite* possible responses in order to do something to stop the painful stimulation. We will examine the details of how these responses work later in the article, for now we will merely assume that this is one of the things the arousal system does.

By themselves, pleasure, pain and arousal form a useful, general purpose, system. At this point they have little or nothing to do with information processing and accordingly they serve all manner of organisms well, not just higher animals. For information processing organisms, such as humans, pleasure and pain can serve in a greatly expanded role. With cognition comes the ability to make predictions. Among the predictions an organism can make is whether or not it will experience pleasure or pain based upon its actions. As with Damasio's somatic marker hypothesis, such predictions will be accompanied by bodily feelings commensurate with the results of the predictions. The ability to make predictions has been widely cited as the primary advantage of an information processing organism. Interestingly enough, whereas the direct sensation of pleasure stimulates arrest, and the direct sensation of pain stimulates excitement, for predicted pleasure and pain the results are fundamentally the opposite. An organism must not do what it thinks will bring it pain, and should do what it thinks will bring it pleasure. Again, in each case pleasure and pain should stimulate the arousal system, but this time it is in anticipation rather than in actual sensation. The fact that anticipated pain is handled differently by the brain than current pain has been shown by using brain imaging (Ploghaus, et al., 1999). The study used fMRI to show that pain and anticipated pain do activate some common regions (e.g. the medial frontal lobe), but that there were differences in activation as well, as pain and anticipated typically activated neighboring but distinct areas within a region. This, by the way, can be taken as evidence of the somatic marker hypothesis. Throughout this article there will be cases where we refer simply to the "pain" system or the "pleasure/pain" system. In such cases we implicitly mean both sensory and cognitive pain (and pleasure as the case may be).

It should be noted that many emotional models do not explicitly include pleasure and pain, but do divide emotions into those that have positive or negative valence. The precaution of not including pleasure and pain directly may be due to the differences between cognitive and sensory pain (and pleasure). Pain researchers do not appear to be as reticent about the link. The definition of pain, according to the International Association for the Study of Pain is "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" (Merskey, 1979, p. 250). The important point being that the emotional experience is intrinsic to the pain (Chapman, 1995; Craig, 1999).

It is at this stage that the basis for an emotional system can start to be seen. Having experienced pain in conjunction with some item, many creatures will begin to avoid the item as best they can. As Braitenberg points out with wonderful imagery (1986) this can be interpreted as a kind of hatred. Conversely if another item brings pleasure, creatures will try to be near the item. In more common terminology these behaviors are *avoid* and *approach*. From an adaptive perspective being able to predict pleasure and pain is even better than simply being able to detect pleasure and pain, but it must necessarily come after those simple abilities were established. Developmentally this is also true, as one must experience pain before one is able to predict it.

The step from simple pleasure/pain detection to the prediction of pleasure and pain is a large one and relies on several cognitive factors. First, there must be cognitive structure available that can be used to make predictions. Second, those predictions must be tagged somehow as being pleasurable or painful. Because pain and anticipated pain activate neighboring regions of the brain, Ploghaus et al. have speculated that this occurs through local interactions in those regions (1999). For example, someone putting their hand on a hot stove will experience pain. The combination of the cognitive structure active when the pain was experienced and the brain areas active because of the pain create a kind of associative link to other, neighboring, areas. Later when that same person thinks about putting their hand on the stove it is the neighboring areas that become active. This will recall the sensation of pain originally felt, but will not be equivalent to it. Presumably the difference is what allows the person to respond differentially in the two cases (e.g. reacting strongly in one case while doing nothing in the other).

The linkage between the cognition and the anticipation of pain is associative. Things that are experienced together in time are strongly associated cognitively. The strength of the association reflects factors such as the intensity of experience, as well as repetition. This association builds cognitive structure useful in prediction (Kaplan, et al., 1990). While the associative link to pain may not seem like a wonderful outcome, it serves a very useful purpose – namely providing a strong bridge between the cognitive system and the arousal system. In general information processing organisms do not need to "decide" to avoid painful things, they simply do because the combination of the pleasure/pain system, an arousal system, and predictive cognitive structure automatically makes it happen. In a sense the decision is left to evolution rather than to the individual and over the course of time this tradeoff has served individuals well. It is the strength of this response and its automaticity that makes emotions problematic to rationalists. If people were truly rational then all decisions would be made by weighing evidence and considering alternatives, but emotions dictate that many decisions are made on a different basis, one that favors fast action and safety. There are a number of flaws with the rational perspective. Among them is the fact that people rarely have the perfect information required for a true rational analysis. Probably even more important is the time required to make a "rational" decision. Emotional responses may be heuristic, but they are fast and the heuristics that they are based on have served innumerable generations.

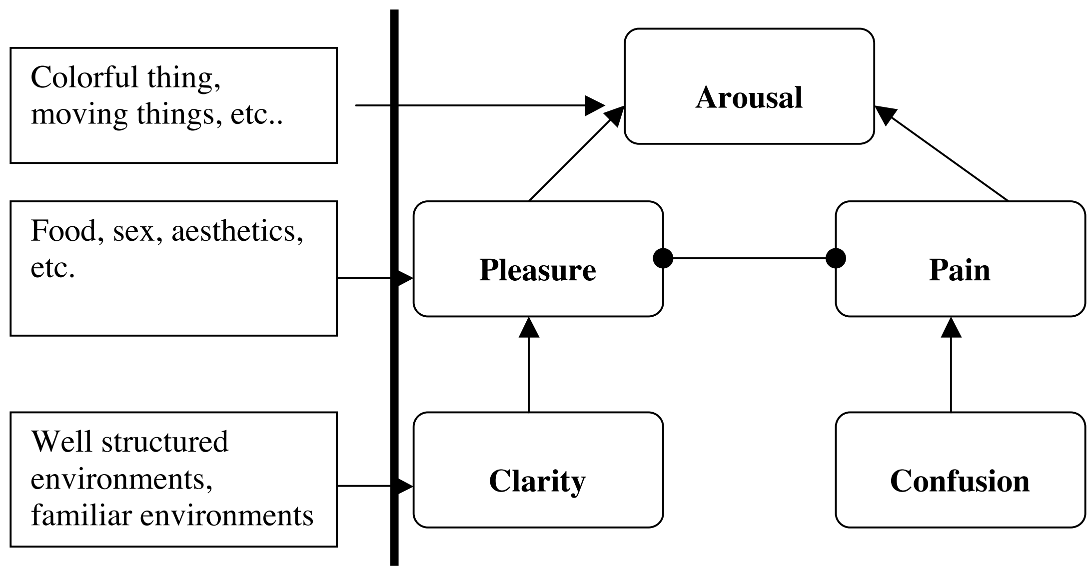
There is a further link between emotions and cognitive structure in that arousal fosters enhanced learning (refs). Again, from an adaptive perspective this is extremely sensible in that the things judged most important are probably the same things that should be learned best. "Important" in this case means "arousing" which can have some disconnect from things judged cognitively important (such as studying for a test). Again, judged by a modern perspective, this has the problem that we have organisms that are easily distracted by highly arousing stimuli – such as certain television programs – rather than "important" things – such as math homework. A more realistic perspective is that television programs do a (unfortunately) better job of taking advantage of stimulating people than many homework assignments do. Fortunately, however, the linkage between the emotional system and cognition, also means that cognitive factors can ultimately outweigh raw stimulation.

Aside from pleasure and pain (and the prediction of the same) there are other sources of arousal also grounded in evolution. Various stimuli, such as moving things, colorful things and large furry things, elicit arousal. While these responses can be considered to be emotional, they need not have positive or negative valence. One can be fascinated by snakes, for example, without being afraid of them. In such cases the arousal system will act in its original capacity, which is simply to provide undifferentiated stimulation of the organism. Interestingly, as Hebb pointed out (1972) the list of arousing stimuli grows larger with intelligence not smaller. This too can be seen as the result of evolution, which, over many generations, can sort out specific types of stimuli that impact an organism's survival chances. Hebb argued that we do not appear to be highly emotional creatures because we have structured our culture to minimize emotionally charged stimuli. So while it may seem childish or annoying to react strongly to otherwise harmless stimuli, such reactions probably served our ancestors well. And again, with learning, we are often able to overcome our instinctual reactions to such stimuli.

The Ability to Cope – Clarity and Confusion

Once the link between cognition and the emotional system had been established further refinements became possible. As Kaplan argues (1991) creatures that rely on information processing must necessarily be sensitive to the quality of information they have available and as a direct result of this clarity and confusion become important cognitive states. Clarity comes when one's internal model of the world is in accord with one's perception. Because of the importance of prediction, clarity will be particularly strong when a prediction has been made and subsequently confirmed. This is important cognitively because it is a strong signal that the cognitive structure involved in the prediction is sound. Confusion, on the other hand, comes when perception is at odds with one's internal model. Not surprisingly, a prime example of this comes when a prediction is made, but then is proven wrong. Confusion is important because it signals that the cognitive structure involved in the prediction is not adequate. From this perspective, clarity and confusion are fast measures of competence in the current environment. As we shall see, this can be done by a relatively simple, and automatic, mechanism. As such it is fast and does not require a cognitive process to perform the monitoring.

Since clarity is important and positive it is pleasurable. Since confusion is important and negative it is painful. Since these states trigger pleasure and pain respectively they will also naturally lead to higher levels of arousal. As in the previous scenarios, the pleasure/pain generated by the clarity or confusion provides a general signal of what should come next. In the case of clarity it is a signal that everything is working well and the major reaction of the cognitive system should simply be to reinforce the cognitive structure that generated the prediction. Because of the higher levels of arousal associated with the pleasure that clarity brings, this will tend to happen automatically. Because of the pleasure involved there will be additional association between the active cognitive structure and feeling good. A number of human foibles and predilections seem to arise from this including good luck charms and gambling obsessions. With confusion, on the other hand there will be a mismatch between what was expected and what happened. Learning is critical in this situation too, but it is not reinforcement of old structure, rather it is the fast creation of new structure that is the goal. Of course confusion and pain go hand in hand, the linkage of which helps explain why many students avoid difficult subjects in school. Rationally these reactions may seem absurd. In the context of survival, where a decision can mean life or death, they make far more sense.



As we have shown, this system (Figure 1) is automatic and beneficial to survival. This also means that it can lead to aberrant behavior, particularly in environments that are not similar to the environments that we evolved in. This is not an article, however, about what to do about emotions, it is about understanding why we have them and what they are. The critical point is that the mechanisms described so far are oriented towards detecting various kinds of important events and responding to those events quickly and appropriately. At its core this is a relatively simple system, but it becomes considerably more complex as cognitive structure itself becomes more complex and different structures interact with each other.

It is our contention that the pleasure/pain system, the arousal system, and the clarity/confusion system form the foundation of the human emotional system. Further, various human emotional descriptors are simply labels that have been attached to combinations of these states and other cognitive factors. Rather than trying to postulate a large number of brain states or mechanisms for individual emotions, we view emotions as an emergent property of these mechanisms. In this system pleasure and pain differentiate positive emotions from negative ones and arousal is effectively a measure of emotional intensity. Just as time differentiates the appropriate reactions to pleasure and pain (e.g. arrest vs. excite) so too is it important in differentiating emotions. For example, the anticipation of pain is associated with the emotional label “fear.” Current cognitive pain, on the other hand, is associated with the emotional label “sadness.” It is precisely because there aren’t specific systems for each emotion that has lead to the wide disparity in emotional states that researchers are willing to agree upon (Damasio, 1994). This may be because further cognitive appraisals are useful in further subdividing emotional categories. For example, it is generally useful to distinguish the source of pleasure. If the source is personal, the emotion is labeled pride; if it is someone else there is a different label.

Despite the difficulty in agreeing on emotional labels there is evidence from work on isolating brain activity that four different areas of the brain have general emotional types associated with them (Robinson & Coyle, 1980; Davidson, et al. 1990). When the right frontal and right pareital areas of cortex are dominant, the mood is anxiety and fear. Dominant activity in the right frontal and left pareital areas result in sadness. When the activity is centered in left frontal and left pareital the mood is calm. Finally, activity in left frontal and right pareital is associated with joy. It is easy to see that positive affect is associated with the left frontal area while negative affect is associated with the right frontal area. The differentiation between left and right pareital, on the other hand, may reflect the intensity of arousal.

The brain mapping research reflects the relative agreement within the emotions community that emotions can be categorized by positive or negative valence (which we are taking to be pleasure or pain) and that arousal is an indicator of emotional intensity. Although our model includes an additional factor, namely clarity, many other emotional models include a similar construct usually labeled “certainty.” This is common, for example, in appraisal theories of emotions (Smith & Ellsworth, 1985; Lerner & Keltner, 2000) which also include a number of other, more cognitive, factors not included directly in our model. These factors include *attentional activity*, *anticipated effort*, *control*, and *responsibility*. Attentional activity refers to the degree that something attracts versus repels one’s attention. Anticipated effort refers to the degree that physical or mental exertion appears to be required. Control refers to the degree that events have been brought about by situations versus agents. Responsibility refers to whether the agent is responsible or whether something else is. Our position is that these factors are useful in many situations, but by themselves do not afford any immediate adaptive advantage. For example, being confused is a clear signal of trouble for an organism regardless of what it is doing. Knowing something will take effort, on the other hand, is useful knowledge, but is not necessarily cause for alarm by itself. One would not necessarily expect such knowledge to increase arousal, which is our standard for emotions. Instead, it is our position that such appraisals are more properly thought of in the realm of cognitive responses to emotions.

The mechanisms of emotion

The three components of the emotional system provide three measures of an organism's current state. The arousal system provides a rough measure of importance. The pleasure/pain system frames things in terms of positive or negative experiences. Finally the clarity/confusion system provides a relative measure of how well the organism should be able to cope with the current state of the world. The emotional system does not just inform the organism of its state, however, it also shapes decision-making and directly impacts cognition in other ways. We have already addressed some of these issues in passing and will now examine them in more detail.

In our view the emotional system provides a general framework such that an organism can quickly assess and respond to all manner of environments and situations. In turn cognition sharpens those responses and in some cases can even override basic emotional tendencies. It is in that context that we present our system. Later, when we discuss our implementation the interplay between cognition and emotion will be shown in more detail. For now, however, our focus will be on the general-purpose aspects of emotion.

As we have indicated, pleasure, pain, anticipated pleasure, and anticipated pain are all general signals that should elicit different responses. These responses are summarized in Table 1. "Arrest" in this case means that the organism will tend to avoid doing different things, while "Excitation" means that the organism will tend to take actions. For example, when experiencing pain, an organism will need to take action to stop whatever is causing the pain. When anticipating pain, on the other hand, organisms will tend to be more passive so as not to bring the pain on themselves. There is a large literature, for example, that shows that people are relatively unwilling to make choices that might lead to regret. For example some parents do not vaccinate their children because of potentially negative side effects, even when the chances of the side effects is only a fraction of the death rate of the disease (Ritov & Baron, 1990). This is a good example of when a general-purpose mechanism does not necessarily produce a good result and why cognition must sometimes be able to override the basic emotional impulse. It may be true that "do not do anything that will lead to a negative result" is a good heuristic, it is only a heuristic.

	Pleasure	*Pain*
Experiencing	Arrest	Excitation
Anticipating	Excitation	Arrest

Arousal is sometimes equated with optimum functioning. This is probably because graphing the relationship between arousal and performance forms an inverted-U curve (Hebb, 1972). Hebb frames this relationship in terms of how stimuli guide behavior. At very low levels of arousal, such as during sleep, stimuli do not affect behavior at all. At high levels of arousal, stimuli tend to overwhelm cognition. Hebb also points out that it is wrong to equate arousal with motivation, since highly aroused people can be so emotionally excited as to be incapable of any action at all.

A better analogy for arousal is emotional intensity. At median levels of arousal this intensity can result in cognitive focus, but when arousal is too high emotions run towards the panic side. The focus and intensity brought by arousal appears to come as the result of several neurotransmitters, notably norepinephrine and dopamine. One result of the release of these transmitters is a reduction in background cortical noise and a corresponding increase in activity in already active cortical areas (refs). In information theoretic terms, this is an increase in the signal-to-noise ratio. In behavioral terms the focus of attention is narrowed and perception is heightened. Given the idea that arousal increases after something important has been detected what we see is that attention narrows upon what was just perceived (the importance trigger, e.g. pain (Eccleston & Crombez, 1999)) and the response to that item is strengthened. When arousal is too high, then attention is dominated by perception and appropriate responses may not be available because they may be suppressed as background noise before they can be considered. Ultimately this can lead to a kind of arousal feedback loop and panic.

The increase in the signal-to-noise ratio of neural firing has another important side effect – an increase in learning. Cognitive scientists dating back to William James (1892) and D.O. Hebb (1949) have theorized that learning occurs as the result of correlated neural firing. A natural result of an increase in arousal and the corresponding increase in signal-to-noise is that learning will be more intense. Naturally this is exactly what a system designer would want if they were designing a subsystem that detected important events, such times are exactly when learning becomes especially important. Not surprisingly there is substantial behavioral evidence linking high levels of arousal to increases in learning (Revelle & Loftus, 1990).

It is at this point that we can begin to see the elegance of the emotional system. The cognitive system has a number of ways to detect important events. These range from the direct – physical pleasure and pain – to the evolutionarily based – such as a fascination with snakes – to the systemic – such as clarity and confusion – to the purely learned. Even though the range of these events is huge, one basic response is the same in every case – an increase in arousal. In turn the increase in arousal automatically changes the way information is processed in the brain in several important and effective ways, directly impacting attention and learning. It is a system that is general purpose and yet can be continually refined through learning.

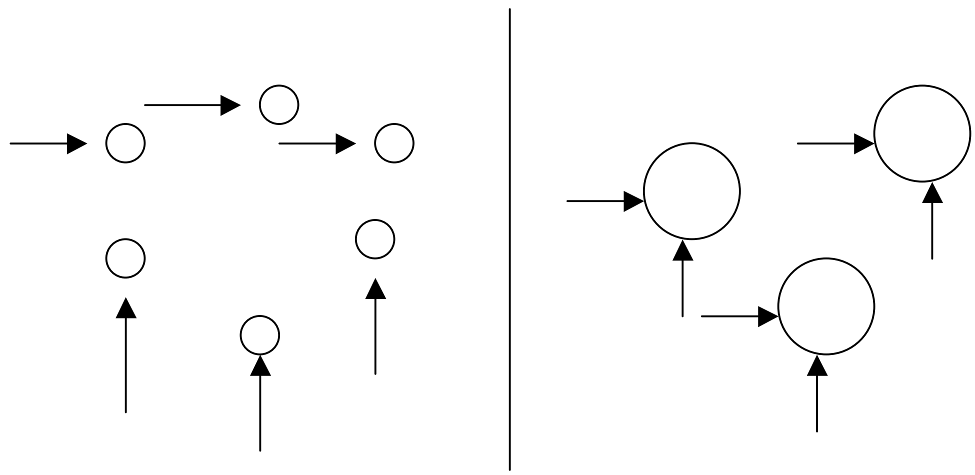
The final major mechanisms in the emotional system detect clarity and confusion and represent one way that learning refines the system. True intelligence is often linked with self awareness and the clarity/confusion mechanism is a good example of why self awareness is important. Essentially organisms that rely on information processing need a metric of how well they are doing and a clarity mechanism fills that role in an automatic, general purpose, way. To see how a clarity mechanism might work it is necessary to consider what is happening cognitively in each case. Clarity comes at moments when previously nebulous cognitive structure comes into focus. This would typically happen when an event in the world provides the perfect example to reinforce a cognitive structure still in its early stages – a tentative prediction has been made and has come true. What is critical is that the world and the cognitive structure are in accord. This is important because it signals that the structure should be strengthened since the world is supporting it; since importance leads to arousal and arousal leads to learning, this will happen automatically.

Confusion, by contrast, comes when cognitive structure is at odds with perception. This too is a case where learning must be strong, and the resultant arousal has the required effect. In the case of confusion, cognitive structure has generated an expectation and that expectation has been violated. The result is unfocussed neural activity as one set of things is perceived while another set has been active as part of cognition (Figure X). Clarity and confusion, therefore, are associated with markedly different brain states. With clarity, neural activity will be relatively focussed. With confusion, on the other hand, it will be more diffuse. Kaplan and Ivancich (ref) have proposed a clarity mechanism that responds differentially to these general conditions.

It is important to note at this point that, unlike pleasure and pain, clarity and confusion are not at opposite ends of a continuum. The opposite of confusion is expertise, which can even lead to boredom. In completely novel environment there is no knowledge to limit the possibilities of what might happen next. One aspect of learning is a sort of pruning away of such possibilities based upon experience. Whereas a novice chess player might try to examine every possible move, for example, an expert will only look at two or three (ref). Between the novice and the expert are stages where there is still enough novelty for uncertainty. Clarity comes with success during those times. Experts will not have the same experiences of clarity because they will not have the same level of uncertainty. On the contrary, it is more likely that experts will have to cope with boredom as their cognitive structure becomes so compact and efficient that it generates very little activity. Expert chess players, for example, can play multiple simultaneous games because no single game will require any processing beyond what is already habitual.

Each stage of cognitive development within an environment will engender different types of behaviors. In this article we are focussing on the earliest stages, corresponding to clarity and confusion, because they are intrinsic to emotions. Expertise can generate a different kind of emotion, however, namely the lack of it which we generally refer to as boredom. An environment that was exciting and dangerous the first time may seem old and stale by the tenth time. In such environments there is nothing arousing because the familiarity and competence in them implies safety. Hebb (1972) has framed this in terms of the need for excitement and play. The low levels of arousal associated with boredom can actually spur an organism to try new and different environments. This is a sound survival strategy as expertise in multiple environments is useful if environments change or disappear. Meanwhile, the confusion inherent in new environments will tend to lead the organism to occasionally retreat back to the familiar, and therefore the safe. Not all environments become boring. Environments with natural sources of pleasure and pain, for example, will not become boring; those experiences will generate their own arousal.

Other environments, particularly natural environments, have features that are interesting even in the absence of arousal. Kaplan and Kaplan (ref) in researching environmental preferences, have categorized such environments along such dimensions as mystery, coherence, and legibility.



As with the other components of the emotional system, the nature of clarity and confusion fosters the need for specific responses. In this case the responses have to do with learning. With clarity the match between the world and the internal model of the world should be reinforced. With confusion, on the other hand, the internal model, or lack thereof, has generated invalid expectations. Fortunately, the increased arousal associated with both of these states will tend to lead to the correct behaviors automatically. Recall that one effect of arousal is to increase signal-to-noise and suppress cortical background noise in favor of what has been perceived. In a clarity state what has been perceived and what has been expected are the same. This means that the cognitive structure supporting the prediction will be automatically reinforced. With confusion, however, the incorrect expectations will essentially serve as the cortical noise in the system since they are not being reinforced by perception. This means that those same predictions will tend to be inhibited while what is perceived will become strongly active due to the increase in arousal. The active cognitive elements will automatically benefit from the increased level of learning, so the cognitive structure will automatically tend to adjust itself in the proper direction towards what was just experienced and away from what was previously believed.

The clarity mechanism, as described, is a case where a general purpose, automatic mechanism has a clear advantage over a purely rational, or cognitive, model. Making a rational determination of confusion or clarity would require constant monitoring and processing. Even once a decision was made about whether or not the system was clear or confused, then additional decisions would have to be made about what to do about it and how to proceed. By contrast we have proposed a simple mechanism that naturally responds appropriately in the vast majority of cases. Once again it is also easy to see why this is a source of distress in the modern world. People do not like feeling confused and many will tend to shy away from environments where there is a danger of being confused.

Individual Emotional Differences and Behavior

The emotion system is inexorably linked to cognition. The arousal system directly impacts how information is processed, the pleasure/pain system impacts decisions on what to do or not do, and the clarity system informs as to the likely quality of decisions. By extension, individual differences in how emotions are processed and generated will lead to differences in the kinds of decisions people are likely to make. These differences are partially expressed as personality types. It would be foolish to claim that we can explain personality theory in terms of our emotional model. First, the emotional system is one part of personality, not the whole of it. And second, the personality literature is still in relative disagreement over a number of basic issues. Nevertheless it is instructive to look at how small changes in the emotional system could realistically generate certain personality or temperamental traits. In this section we will begin by deriving the basic sources of individual emotional differences in our model. We will then go on to analyze them in the context of some of the most widely accepted models of personality and temperament. Later we will use these traits in testing how differing emotional types can impact decision making in a cognitive model.

Individual differences

Returning to Figure 1, we can see that the various parts of our model are linked together. Clarity brings pleasure, which in turn brings arousal, etc. Each of these major pieces represents a source of individual differences and these differences can be explained by using the links between the pieces. If the figure were to be viewed as a neural network (realistically it is a vast simplification of a network) the links would be the connections between nodes. It is our position that the links between the components are relatively static – that is to say they are not subject to learning. Individual differences in emotional type in this case would arise from differently weighted links in the model. For example, some people may be highly susceptible to experiencing pain. In the model this could be expressed by having strong links into the pain portion of the model. To simplify our discussion we will consider three basic link differences and go on to show how those differences might be expressed in behavior.

The first difference has to do with the susceptibility to becoming aroused. Individual differences in this case mean that one person might remain calm in the face of horrific circumstances and even personal pain while another might become highly agitated at the slightest disturbance.

The second difference has to do with pleasure and pain. We do not rule out the possibility that a single individual might be differentially sensitive to pleasure versus pain, but given the lack of direct evidence for it and the nature of this discussion, we will focus on general sensitivity to both. What we are claiming is that two people who receive identical stimulation will actually perceive different levels of pain because they will be differentially sensitive to pain. There is even some evidence that links such different sensitivies to eye color (Rosenberg & Kagan, 1987). A person more sensitive to pain might be more likely to focus on negative events because such events have an unusually strong effect.

The final difference has to do with clarity and confusion. The individual differences marking clarity and confusion are slightly trickier than in the other cases. This is because they are confounded with the kinds of cognitive structure that an individual generates. Fast learners may not become confused as often as other people due more to how they learn, than to how prone they are to being confused. Nevertheless, it is reasonable to assume that different people have different thresholds in becoming confused or in experiencing clarity. On the one hand, a person who is prone to feeling confused might shy away from the novel or the complex because of the negative feelings associated with being confused, on the other, some people may be able to find clarity where others see only chaos.

Personality theory

The dominant current theory in personality research is generally referred to as the "Big 5" theory (Digman, 1990; Costa and McCrae, 1995). Statistical analysis has generally shown that there are five orthogonal personality dimensions. In practice there is general agreement about three of the dimensions, and a little less about the final two dimensions (and there is another line proposing that there are only three dimensions (Eysenck, 1991)). We will focus on three of the Big 5 dimensions, and attempt to show that each these can be understood as an individual difference in the emotional model we are presenting. The three factors we will focus upon are generally called *extraversion*, *negative emotionality*, and *openness*. The two we will spend less time with are agreeableness and conscientiousness.

Extraversion

Extraverts are typically categorized as people who prefer to be with other people. Extraverts are outgoing and assertive and sometimes are described as craving excitement. Introverts on the other hand prefer to be by themselves and are often described as reserved. Introverts tend not to seek excitement. Within this category there is some research that indicates that extraverts also tend to have more positive affect than introverts (Gross, et al., 1998; Carver, et al., 2000).

Arousal theorists have long made the case that extraverts are more difficult to arouse than introverts. The excitement needed by extraverts is to generate arousal. Going back to the notion that graphing performance against arousal generates an inverted U, extraverts tend to be on the lower left end of the U and need excitement to push their arousal to the middle. The stimulation provided by people and novelty can provide such excitement. Introverts, on the other hand, will tend to be pushed to the lower right of the U when in stimulating environments. Therefore they will tend to try and control for this by seeking quieter places and by being alone.

It is tempting to postulate that the source of excitement for extraverts is pleasure. This would tend to be supported by the studies linking extraversion and positive affect (Watson & Tellegen, 1985). However, our model offers another explanation. Over the course of their lifetimes, extraverts will learn what environments they do well in, and what environments they do not do well in. In novel environments, especially those that are stimulus intensive, extraverts will expect to perform well and accordingly will develop a confident attitude. Introverts will not tend to have such positive experiences. Because they become over-stimulated, they will not perform as well and will generally come to associate those negative feelings with such environments. The argument here is simply that extraverts will tend to have more positive associations with many kinds of environments because of their experiences. On their own and free of stimulation, they should not be any more prone to feeling good than anyone else. This conclusion is supported by studies that show exactly that extraversion correlates only very weakly to the magnitude of changes in positive affect for positive stimuli (Gross, et al., 1998).

Negative emotionality

This dimension is differentiated by the strength of a stimuli required to elicit a negative response. At one end of the spectrum are people who are “resilient.” They are described as being calm, slow to discourage, and handling stress well. At the other end are people who are “reactive.” These people are uneasy, quick to anger and embarrass, and do not handle stress well.

Since this factor is framed in terms of negative emotions, it is tempting to simply link it to pain. This is probably a mistake since pleasure and pain are so closely linked. Rather it is more likely the case that we simply pay more attention to pain, partly because pain is more important from an evolutionary perspective (since pain is equated to danger). In any regard, the central point is that people who experience pain more easily are more likely to focus on it in their daily lives. This factor is sometimes labeled “neuroticism” and is often related to anxiety. Our contention is that this anxiety comes as the result of low levels of pain experienced to a nearly constant degree. In many people this pain is easily ignored, but for people with a lower threshold the pain will be very real and they will come to associate it with most aspects of their lives. Viewed from this perspective neurotics may be responding to life in a very reasonable way. The same studies that showed only a weak link between extraversion and changes in positive affect, showed very strong correlations between neuroticism and changes in negative affect (Gross, et al., 1998).

Openness

This dimension is more controversial than the previous two and has been left out in three-dimensional models such as Eysenck's (1981; 1991). This trait divides people into explorers and preservers. The names of the traits suggest the linkage to boredom and play. Preservers are described as being interested in the here and now, preferring the familiar and being conservative about change. Explorers, by contrast, daydream, are open-minded and prefer variety.

It should not be surprising that this trait is more controversial than the others as it is probably muddied somewhat by the interplay of clarity and confusion with how people learn. Nevertheless, the clarity/confusion mechanism provides insight into the different types. Presumably preservers do not like new things because they abhor becoming confused. It is simpler to fit something into existing structure than to deal with learning new structure. Explorers, on the other hand, may be addicted to clarity and easily prone to boredom. Too much of the same thing will cause them pain, so they will seek the novel.

Agreeableness and Conscientiousness

The final two dimensions of the Big 5 model are agreeableness and conscientiousness. Agreeableness is a category generally included in personality models and reflects the sources that people use to define correct behavior. This would appear to be a social factor and is not relevant to our model of emotions. Conscientiousness is defined in terms of goals. Again this does not appear to relate to emotions as discussed in this article.

Temperament

In studying temperament, Kagan (1998) has taken a more cautious approach to assigning categories than most personality researchers. The one major category that he is willing to commit to, distinguishing fearful and bold people, is similar to the extraversion/introversion dimension in personality theory. This work is notable with regard to our emotional model for several reasons. First, most of Kagan's work deals with children. Kagan has shown that temperament is essentially set in a child after about two weeks, and tends to be stable for a lifetime. This is a strong argument that temperament is genetically based and essentially a parameter for a given individual. Second, Kagan identifies norepenephrine as the single critical factor that seems to distinguish fearful people from bold people; based upon the quantities of norepenephrine in the brain it is possible to determine if someone will be generally fearful (high levels) or bold (low levels).

Summary

We do not believe that the parameters in our model cover the gamut of individual differences in personality or decision making. We do, however, believe that examining such differences in the context of a model such as this can be enlightening. As with everything in this article we stress that cognitive structure and experience will have a major impact on how these differences are expressed. Introverted people may learn effective strategies to cope with highly arousing environments for example. The emotional system provides a kind of template for how a person is likely to react, but cognitive structure is able to shape those responses within certain boundaries.

In the rest of the article we show how this is the case by describing an implementation of our model in an agent model and then by showing how differences in emotional parameters change the decision making processes of the agent.

Implementing Emotions in a Symbolic Architecture

Although we have described the emotional model in connectionist terms, we have implemented it in the context of a rule-based system, specifically in the Soar architecture. The reason for this is fairly simple: rule-based agents are simply more sophisticated and have a wider range of capabilities than their current neural network counterparts. Soar agents in particular have been built to do cognitively demanding tasks such as flying airplanes (refs). The task we have chosen – that of a special forces team operating in enemy territory – is also appropriate because it is highly emotionally charged and fast decisions often have life or death consequences, thus emphasizing the adaptive nature of emotions. The difficulty of implementing our emotional model in Soar, on the other hand, is that in some ways the model does not map cleanly to a rule-based system. We will describe our implementation in two phases: first will be how we implement the model in general architectural terms. Second, will be implementation details specific to the special forces task.

Soar Agents

A key component of all Soar models is that all activity is cast as a succession of decisions involving operators and goals. The decisions are based on an internal representation of the current situation, which is built up based on realistic simulated sensors. To make a decision, a Soar system performs a parallel retrieval from long-term associative memory (implemented as a very fast rule-based system) to get preferences for selecting the next "operator". An operator might represent an action as simple as "put block A on block B", or as complex as "build structure C". The retrieved preferences are analyzed, and a decision is made for the current best operator.

Once the current operator is selected, long-term memory is again consulted (via rapid, parallel, retrieval rules) to carry out the operator. If it is a simple operator, this will result in either a new output command being sent (which involves the agent's effectors), or some changes to the internal state of the system (such as classifying input). If the selected operator is complex it will become a goal to be achieved through decomposition into one or more sub-operators. This activity can recurse, leading to the dynamic construction of an active goal hierarchy.

An agent's knowledge in Soar is encoded in the rule-base. This long-term memory, together with the goal hierarchy, provides a smooth integration of reactive control and goal-driven behavior. Thus, the system quickly responds to changes in its environment, as it also selects new operators based on active goals.

Implementing the emotion system in rules

We have opted to implement the pieces of the emotional system directly in Soar rules. This decision was made as much of a matter of programming convenience as anything else. An alternative would be to implement the emotional system outside of the Soar architecture and develop special input and output links specifically to communicate between the two systems. Instead we just use the existing input and output channels and create internal structures corresponding to the various emotional building blocks. We will describe this in some detail when we get into our example. First, however, we will discuss the effects that having emotional states have on the rules that already exist in our Soar agents.

The area of greatest impact will have to do with arousal. As we have noted, arousal tends to focus cognition. Further, highly aroused people will tend to revert to doing the things that they are most familiar with. In terms of a rule-base this means that agents will have differential access to rules based upon their arousal level. The simplest way to code this is to put arousal thresholds on rules. At the bottom end, some rules will not fire because they require a certain level of motivation that only occurs with at least moderate levels of arousal. Procrastinators, for example, need the arousal jump caused by a looming deadline before they start working on projects, even though they possess the tools necessary to do the work. At the high end, some rules will drop out at high levels of arousal as the cognitive system essentially relies more heavily on the familiar and what has worked in the past. In theory these rules will tend to come in two varieties: first are those rules that are not well learned, especially for things that have only been recently learned. Second, are rules that require more deliberate cognition. In Soar terms, for example, these would be rules that have a large number of conditions. The idea being that the increased focus of arousal makes matching all of the conditions less likely. The inverted U performance curves associated with arousal can be viewed as on the one hand not putting enough knowledge to use when arousal is very low, and on the other not having access to the complete knowledge base when arousal is high.

To take a general example, consider a rule having to do with fleeing. Rather than having elaborate conditions specifying all of the possible times that one might run from a threatening situation it is simpler to have a single rule. This rule might not have any conditions beyond a minimum arousal threshold. Essentially the rule is"if I get too excited, then run away." General rules of this type might be thought of as "fight or flight" defaults provided by evolution. Although the rule might apply a great deal of the time, it could be a low preference rule and would therefore not typically fire. On the other hand, as arousal gets higher the number of available alternatives will begin to dwindle because some of them may have high arousal thresholds. This will act to make fleeing an increasingly attractive alternative that is more, or less, likely depending on the agent's experience in that domain (with additional experience fewer rules will drop out at high arousal levels). At the same time there is no reason to flee even in a novel environment if there is nothing threatening happening. This is taken care of by the minimum threshold on the rule. In this case arousal acts as a kind of tag that enables or disables portions of the overall rule-base.

In more connectionist terms it is possible to think of each rule as a distinct cognitive element. Arousal can affect virtually any rule because its effects are general. Rules will respond differentially to arousal due to a large number of factors such as how well-learned they are.

Pleasure and pain have a differential impact on rules as previously noted by the arrest/excite distinctions in Table 1. There is a question of whether these distinctions need to be explicitly modeled in a rule-based system. Rules, after all, already implicitly carry the assumption that actions are to be performed. Presumably such rules are designed to maximize pleasure and minimize pain by their very nature. Actions that might bring pain, for example, would generally not be coded as rules. Whereas an associative system would need some sort of inhibitory mechanism to prevent certain thoughts from becoming actions, rule-based systems essentially prune the thoughts. On the one hand the associative system thinks "if I go into the cave I'll run into the bear" on the other the rule-based system simply has a rule specifying specifically what to do. Indeed expressing knowledge of the form "do not go into the cave because a bear lives there" is difficult for many rule-based approaches.

The effects of clarity and confusion are expressed in the rule base mainly by their impact on pleasure and pain and subsequently arousal. It is the case, however, that people learn to recognize when they are confused and can develop behaviors based upon that recognition such as avoiding difficult subjects in school. It is probable, however, that this is a meta-cognitive effect and one that need not be modeled directly.

There is one other way that clarity and confusion impact the rule-base. Since Soar systems do not typically think ahead the same way a chess playing system might, for example, there needs to be a way to simulate the expectations the rules implicitly contain. To this end we have created an expectation data structure such that the effects of rules can be explicitly enumerated. When the expectations are violated it results in an increase in confusion; when expectations are met the result is an increase in clarity. We will cover this in more detail when we discuss our implementation.

Activating emotions

Inputs to the emotional system come from both the perceptual system and from the cognitive system (in the form of rules). In the case of the perceptual system we simply code some inputs as being painful, arousing, etc. In addition, it is important to tag each stimulus with the object that generated it if available. For example, a snake is an arousing stimulus. By tagging the stimulus, the extra information can be used to inform the attention system. Items that are highly arousing, pleasurable, etc. are likely to dominate attention. In Soar this also impacts which percepts are stored in working memory and how long they should be kept there. We are actually blurring the distinction between pure perception and cognition by handling some percepts (such as seeing enemy soldiers) as though they were naturally arousing or painful. While these percepts may indeed immediately generate arousal it is through associations built up over time as background knowledge.

Some inputs may be more purely cognitive. Some actions, for example, might be pleasurable (e.g. helping a team member). In such cases the corresponding Soar rules can directly generate inputs to any of the pieces of the emotional system.

Emotional equations

The implementation of emotional equations in our system is directly inspired by the TRACE model of Kaplan et. al. (1992). TRACE is used to model populations of neurons that comprise cell assemblies, the connectionist equivalent of a symbol. Rather than trying to model the behavior of individual neurons, the choice was made to model the gross behavior of large groups of neurons. This has the advantage that it does not require perfect knowledge of how the individual elements behave. In turn, the TRACE model was inspired by the population models used by ecologists to study large groups of animals.

With all of our quantities we are going to build in a range of values representing the fraction of their maximum activation. E.g. there is some maximum amount of pain that a person can experience and were they experiencing it, it would register a value of 1 (for 100%). Half of that would be 0.5, etc. We will start with simple forms of the equation and successively refine them into their final forms. These equations will be expressed as difference equations, meaning the value of a quantity at time T, will often depend on its value at time T – 1. Since the time steps are discrete these are not differential equations.

We will begin with arousal. Every person will have some baseline, or typical, level of arousal that they will tend to be around in the absence of stimulation. In truth people are constantly being stimulated so it may not even possible to measure such a thing, but we can assume some basic norm. We will call that level of arousal *AB*.

At any given time there will be a number of factors working to push arousal up and down. These can be thought of as inputs to arousal. These inputs include stimuli that are naturally arousing (e.g. snakes), as well as pleasure and pain. In addition to pleasure and pain themselves, people are also sensitive to changes in pleasure and pain. We can collect these together in a single equation. It is possible that people are differentially sensitive to each of these factors individually, but that introduces an extra level of complexity beyond the scope of this article.

Arousal Input(T+1) = Stimulation(T) + Pain(T) + Pleasure(T) + ΔPain(T) + ΔPleasure(T)
or, rewriting for compactness:

AI(T+1) = S(T) + P(T) + PL(T) + □P(T) + □PL(T)

Since we are dealing with proportional values, it is necessary to normalize these inputs between –1 and +1. Essentially the most arousing thing possible would be to max out the Stimulation dimension along with either Pleasure or Pain, while incurring a large change in that dimension. Assuming that each of these factors are themselves already normalized, the maximum Arousal Input level would be 3. Therefore in our final equation we will divide by 3.

We must further modulate the increase (or decrease) in arousal by the amount of change possible. For example, if arousal is already at 0.9 it can only go up another 0.1. To simplify, we will use the inverse distance from the baseline as this modulator. Finally, individuals will respond differentially to these inputs depending upon how easily aroused they are. This arousal sensitivity is included as *Asens* in the equation. Putting everything together we get:

Move(T+1) = Asens * (AB – ABS(AB – A(T)) * (S(T) + PL(T) + P(T) + □P(T) + □PL(T)) / 3

In addition to the arousal inputs, there is the tendency of arousal to retreat to baseline levels in the absence of stimuli. This recovery rate will vary for individuals and will be given a constant value of AR. The maximum amount of recovery that can take place, of course, is the difference between the current arousal (**A**), and the baseline level (*AB*).

Recovery Max(T+1) = A(T) – AB.

The recovery rate is influenced not only by the individual's recovery rate, but also by how aroused they currently are. For example, highly aroused individuals have a tendency to stay aroused (ref). This suggests that recovery should happen slower at extreme levels. This effect can be achieved by taking the absolute value of the difference between the current arousal level and the baseline level and subtracting that quantity from the baseline.

Recovery Magnitude(T+1) = (AB - ABS(AB – Arousal(T)))

By multiplying these two equations together we moderate how much recovery can happen, by how fast it should happen. Finally, we also include the individual's recovery rate and get:

Arousal Recovery(T+1) = AR * (A(T) – AB) * (AB – ABS(AB – A(T)))

Putting the input and recovery equations together we get:

A(T+1) = A(T) + Move(T+1) - Arousal Recovery(T + 1)

or

A(T+1) = A(T) + Asens * (AB – ABS(AB – A(T)) * (S(T) + P(T) + + PL(T) + □P(T) + □PL(T)) / 3 - AR * (A(T) – AB) * (AB – ABS(AB – A(T)))

where *Asens*, *AB*, and *AR* are constants. For ease of reading purposes, we'll drop the time tags from here on out with the understanding that they are implied. E.g.

A = A + Asens * (AB – ABS(AB – A) * (S + P + + PL + □P + □PL) / 3 - AR * (A – BA) * (BA – ABS(BA – A))

Pleasure/Pain

The pleasure and pain equations are intertwined because pleasure and pain are mutually inhibitory. To simplify, we will focus on pleasure. The inputs to pleasure include things that are inherently pleasurable, cognitive pleasure (e.g. thinking of pleasurable experiences), and stimulation from the clarity system. At low levels these sources are additive. However, when one of the sources is particularly strong, for example at some threshold *Pthres*, it will tend to overwhelm the rest and dominate cognition. In reality there are probably further levels of distinction based upon factors such as how similar the sources are, etc. We will let the sources of pleasure be labeled P1 . . . PN. Then

```
if (MAX(P1. . . PN) > Pthres)
        Pleasure Input = MAX(P1. . . PN)
else
        Pleasure Input = MIN(SUM(P1 . . . PN), Pthres)
```

Pleasure and pain will be mutually inhibitory to some extent. This will be the case moreso at high levels. Call Pleasure Input *PlI* and Pain Input *PI*.

```
if (PlSens * PlI > Psens  * PI)
    if (PlSens * PlI > Pthres)
        Pleasure = PlSens * PlI
        Pain = 0
    else
        Pleasure = PlSens * PlI
        Pain = Psens * PI
else
    if (Psens * PI > Pthres)
        Pain = Psens * PI
        Pleasure = 0
    else
        Pain = Psens * PI
        Pleasure = PlSens * PlI
```

While pleasure and pain are somewhat in opposition, we postulate that clarity and confusion form a single system and therefore are even more directly in opposition.

Clarity Signal = ClSens * SUM(Cl1 . . . ClN)
Confusion Signal = Csens * SUM(C1 . . . CN)

Cl/C = Clarity Signal – Confusion Signal

Example: Special forces

The Soar behavioral model used to evaluate the emotion model was Special Operations Forces (SOF) Soar. This task involves a 6-man team inserted deep within enemy territory for reconnaissance purposes. Once inserted, they travel anywhere from 20-50 km to an *Objective Rally Point*, they split into three 2-man teams (i.e., two 2-man observation teams, and one 2-man radio team).

Seeking cover and concealment, the observation teams set up near the designated *Objective Observation Area* reports back to the radio team when an appropriate objective has been sighted. The radio team conveys the essential elements of the observations back to the base. This is done at separate *Transmit Sites*, which are away from the observers and change after each transmission. At the conclusion of the mission, they will make their way to a designated *Pickup Zone*.

The purpose of our simulations was to take existing Soar agents written for this environment and to add emotions to them in order to study how various emotional profiles will impact decision-making. To add to the emotional context of the environment we have added variations to the scenarios that include things such as unexpected gunfire, enemy attacks, etc. In principle, agents with different emotional profiles will react differently to identical situations – e.g. where one agent may continue to do its job, another might choose to flee, or become frozen with terror.

Individual Agents

Since our goal was to test the effects of emotions on decision-making we created agents with identical knowledge bases that only differed in their emotional parameters. This allowed us to isolate the effects of those parameters more effectively. For the purposes of our simulations, the important parameters were those directly associated with individual differences in an agent's emotional profile. These are the sensitivity to arousal, the sensitivity to pleasure and pain, and the sensitivity to clarity and confusion.

Having no a priori knowledge of what reasonable values of those parameters should be, we made several basic assumptions. First, we assumed a normal distribution of values around some mean. Since the parameter values all ranged from 0 to 1 the simplest way to proceed was to assume the mean to be 0.5 in each case. Once those assumptions were in place it is a simple matter to classify an agent's personality based upon the values of each of the parameters. For example an agent with a high arousal sensitivity (i.e. greater than 0.5) can be classified as an introvert (introverts make up less than 50% of the population, so this threshold is probably somewhat higher in practice). In a large-scale study we would next generate a population of agents at random, but our simulation runs in real-time so we could not do a large enough set of runs to make such a strategy worthwhile. Additionally, our main goal was simply to show the trends in decision-making associated with different types so we settled upon a strategy of generating clear-cut examples of each personality type (since there are three binary parameters, this amounts to generating eight separate agents). This is shown in Table X.

			Arousal	Pl/Pain	Cl/Conf.
Introvert	Neurotic	Explorer	0.75	0.75	0.25
Introvert	Neurotic	Preserver	0.75	0.75	0.75
Introvert	Stable	Explorer	0.75	0.25	0.25
Introvert	Stable	Preserver	0.75	0.25	0.75
Extravert	Neurotic	Explorer	0.25	0.75	0.25
Extravert	Neurotic	Preserver	0.25	0.75	0.75
Extravert	Stable	Explorer	0.25	0.25	0.25
Extravert	Stable	Preserver	0.25	0.25	0.75

Agent Rule-base

Our agents' rule-base consists of knowledge specific to the task as well as some general rules concerning highly emotional actions such as fleeing. At any given time a number of these rules might apply, so we also created a priority scheme to select amongst them. In our implementation the main effect of emotions are to provide differential access to the complete rule-base. In principle, when an agent has access to "rational" knowledge they should put it to use. Therefore in our scheme, deliberate actions have priority over emotional actions. In other words if the agent sees the enemy and two rules can fire, one corresponding to radioing a report as it is supposed to do, and another corresponding to fleeing, the agent will choose to radio the report.

There are several heuristics used to modify the existing knowledge base: 1) rules that have more conditions (and therefore are more deliberate) are less likely to fire when arousal is high. 2) Rules that are more active (e.g. running) require more arousal in order to fire. 3) Rules that are less well-learned are less likely to fire when arousal is high. Since the last heuristic is more subjective when modifying an existing knowledge base we did not use it in our simulation. The other heuristics can be applied virtually automatically.

An example of a general rule is a "flight" rule. Its conditions are based solely upon arousal. To simplify presentation, we will present our rules as simple **if-then** clauses rather than as pure Soar rules. This rule might look as follows:

```
IF (arousal > 0.85)
    run
```

This rule basically says that when things start getting emotionally charged a good response is to leave. Rather than trying to figure out all possible cases of when running is a good response the emotional system provides the context with a single value. Notice that agents that are quicker to arouse (e.g. introverts) will be much more likely to choose this action because their arousal is more likely to be in the target range.

Experience provides more deliberate, and specific, choices of actions.

```
IF (enemy*seen AND NOT enemy*sees*me AND arousal > 0.4 AND
                              arousal < 0.8)
       radio*report
```

This is a simplification of a rule that might be used by a radio team doing reconnaissance. The rule fires when the enemy is sighted and does not appear to have sighted the agent. It also only fires within a fixed arousal window. Essentially this covers cases where the agent cannot be bothered to report or is too emotionally excited or worried to remember to do it. The first case is extremely implausible since seeing an enemy should be an arousing event.

If these were the only two rules available to the agent, its range of behaviors upon seeing an enemy that does not see it would be either to run, radio a report, or do nothing. In this case nothing will be done if the agent's arousal level is below 0.4. A report will be radioed if arousal is between 0.4 and 0.8. Nothing will be done when arousal is between 0.8 and 0.85 (unlikely in a real agent since it will have many more responses available). The agent will run when its arousal is greater than 0.85.

Inputs

There are a number of sources of direct input to the emotional system in our scenarios. Some examples follow in Tables X through X+4. The values in these are subjective and not based upon any experimental evidence.

Arousal Inputs	Value
loud-noise yes	0.3
dangerous-location yes	0.3
warning-message yes	0.25
move-to-cover-and-concealment yes	-0.2
enemy-moves-away yes	-0.2
OK-message yes	-0.1

Pain Inputs	Value
im-hit yes	1
teammate-hit yes	0.5

danger yes	VAR
people-shooting yes	0.3
people-shooting-at-me yes	0.6
enemy-sees-me yes	0.5
moving-in-sight-of-enemy yes	0.25
mission-in-jeopardy yes	VAR
communication-effective no	0.25
teammate-killed yes	0.5
high-enemy-to-friendly-ratio yes	0.3

Pleasure Inputs	Value
teammate-hit no	0.2
mission-accomplished yes	VAR
subgoal-accomplished yes	0.15
communication-effective yes	0.1
danger-passed yes	0.1
enemy-disabled yes	0.15

Confusion Inputs	Value
Unknown value – benign	0.1
Unknown value – useful	0.25
Unknown value – dangerous	0.4
response-worked no	0.2
received-required-information no	0.2
input-overload yes	0.5
expectation-met no	0.1-1.0
leader-knows-im-alive no	0.4
enemy-exists yes	0.2
enemy-sees-me yes	0.2

Static Example

In this example the two-man SOF Observation team is stationed at the Observation Point and a high number of enemy have been sighted. For the purposes of the example we will assume that the agents have a starting arousal level of 0.5. The inputs to the emotional subsystem for this case may be seen in Table X+5.

Emotional Subsystem	Inputs	Value
Confusion	unknown-value useful	.25
Pain	high-enemy-to-friendly-ratio yes	.3
Arousal	dangerous-location yes	.3

Table X+5. Subsytem Inputs Considered for Static Example

These inputs are then processed by the various equations in the emotional system. Since different personality types will react differently to the different inputs, they will end up with differing levels of arousal. For this case, the results according to personality type are shown in Table X+6.

Personality Type			NEXT AROUSAL
Extraversion	Neuroticism	Explorer	0.53219249
Extraversion	Neuroticism	Preserver	0.53219249
Extraversion	Stability	Explorer	0.51971749
Extraversion	Stability	Preserver	0.51971749
Introversion	Neuroticism	Explorer	0.59456749
Introversion	Neuroticism	Preserver	0.59456749
Introversion	Stability	Explorer	0.55714249
Introversion	Stability	Preserver	0.55714249

Table X+6. Next Arousal Value Resulting from Scenario 2

From the example it is easily possible to see how behavior will start to diverge as the agent faces more and more emotional inputs. Essentially the differing levels of arousal will index different parts of the rule-base meaning that different personality types will make different kinds of decisions.

Results

Comparisons to Other Work

Our model is probably most similar to appraisal theories of emotions (refs). These theories are sometimes called "arousal/appraisal" theories because they postulate that emotions consist of arousal mediated by a series of cognitive appraisals. For example one model (Smith & Ellsworth, 1985; Lerner & Keltner, 2000) includes *certainty*, *pleasantness*, *attentional activity*, *anticipated effort*, *control*, and *responsibility*. It is easy to see that certainty correlates to clarity in our model while pleasantness corresponds to pleasure and pain. Attentional activity refers to the degree that something attracts versus repels one's attention. Anticipated effort refers to the degree that physical or mental exertion appears to be required. Control refers to the degree that events have been brought about by situations versus agents. Responsibility refers to whether the agent is responsible or whether something else is. The distinction we are making in our model is that in the continuum between purely deliberate and purely reactive responses, we are focussing on the reactive side. Specifically, the three components of our model can be put into completely mechanistic terms (of course these mechanism can be impacted by cognition). This is in keeping with our view of emotion as providing a quick response system to complex situations. Essentially what we are saying is that arousal, pleasure/pain and clarity/confusion provide an immediate, and general, initial framework for any situation. More deliberate cognition, and especially of the sorts included in appraisal theories, can then shape more specific responses.

It is important to stress that we are not claiming that appraisal factors not directly included in our model are not important to emotions. On the contrary, in many cases they are obviously important. On the other hand, we are claiming that the central emotional factors are the three that we have identified.

Another key theory of emotion, due to Ortony, Clore, and Collins (sometimes called "OCC") (1988) has probably had the greatest impact on the actual design of emotional agents. One reason for this is that the authors outline rules that might be used in an implementation. As with the appraisal models, the OCC model contains substantially more cognitive factors than our model. Indeed it is the interaction of cognition and emotions that is the authors primary concern. As a result the OCC model is considerably more specific than ours. One reason for the difference is how the authors of the OCC model view physiological evidence. They state "whereas the physiological concomitants of emotional experience are of indisputable importance, they throw little light on the cognitive components of emotion (p. 12-13)." We strongly disagree. For example, it is simple to use "emotional intensity" as a substitute for arousal. In many ways this can work well, but "emotional intensity" does not have the same set of attachments as arousal. Increased arousal implies increased learning, for example, while an increase in emotional intensity does not. The fact that intense arousal changes neural firing has other implications for cognition. While it is true that it may be possible to sort out these implications without resorting to physiological arguments, it is difficult to see why using such information is not useful.

Concluding Remarks

Acknowledgements

This work was sponsored by the U.S. Army Research Institute's Emotional Synthetics Forces STTR – Ph II, contract number: DASW01-99-C-0037 and by the National Science Foundation under Grant No. 0092605. We gratefully acknowledged that support. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation (NSF).

References

Braitenberg, V., (1986) *Vehicles: Experiments in Synthetic Psychology*, MIT Press, MIT.

Carver, C.S., Sutton, S.K. & Scheier M.F. (2000). Action, emotion and personality: Emerging conceptual integration. *Personality and Social Psychology Bulletin*, Vol 26(6). p. 741-751.

Chapman, C.R. (1995). The affective dimension of pain: A model. *Advances in Pain Research & Therapy*, **22**, p. 283-301.

Costa, P.T., and McCrae, R. R., (1995). Primary Traits of Eysenck's PEN system: three- and five- factor solutions. *Journal of Personality and Social Psychology*, 69, 308-317.

Craig, K.D. (1999). Emotions and psychobiology. In Wall, P.D. & Melzack, R. (eds.) *Textbook of Pain*. Fourth Edition. Churchill Livingstone, p. 331-343.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Avon.

Davidson, R.J. & ???? (1990). Approach-withdrawl and cerebral asymmetry. *Journal of Personality and Social Psychology*, 58, p. 330-341.

Digman, J. M., (1989). Five robust trait dimensions: Development, stability, and utility. *Journal of Personality*, vol. 57, no. 2, p. 195-214.

Digman, J. M., (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology 1990*, vol. 41, p. 417-440.

D'Ydewalle, G., Ferson, R. & Swerts, A. (1985). Expectancy, Arousal, and Individual Differences in Free Recall. *Journal of Memory and Language*, 519-525.

Eccleston, C., & Crombez, G. (1999). Pain demands attention: A cognitive – affective model of the interruptive function of pain. *Psychological Bulletin*, **125**, 356-366.

Eysenck, H.J., and Eysenck, M. (1985). *Personality and Individual Differences: A Natural Science Approach*. New York: Plenum Press.

Eysenck, H.J. (1981). A Model for Personality. New York: Springer.

Eysenck, H.J. (1991). Dimensions of Personality: 16, 5, or 3? Criteria for a Taxonomic Paradigm. Personality and Individual Differences, vol. 12, no., 8, p. 773-790.

Gross, J.J., Sutton, S.K. & Ketelaar, T. (1998). Relations between affect and personality: Support for the affect-level and affective reactivity views. *Personality and Social Psychology Bulletin*, Vol 24(3), p. 279-288.

Hebb, D.O. (1949). The organization of behavior. New York: Wiley.

Hebb, D.O. (1972). *Textbook of Psychology*, 3rd ed. Philadelphia: W.B. Saunders Co.

James, W. (1892).*Psychology: The briefer course* (1961 ed.). New York: Harper & Row.

Kagan, J. (1998). Galen's prophecy. Westview Press.

Kaplan, S. (1991). Beyond rationality: Clarity-based decision making. In T. Garling and G. Evans (Eds.), *Environment, Cognition, and Action: An Integrative Multidisciplinary Approach*, pp. 171-190. New York: Oxford University Press.

Kaplan, S., Sonntag, M. & Chown, E. (1991) Tracing Recurring Activity in Cognitive Elements (TRACE): A Model of Temporal Dynamics in a Cell Assembly. Connection Science 3, 179-206.

Kaplan, S., Weaver, M., & French, R. (1990). Active symbols and internal models: Towards a cognitive connectionism. *AI & Society*, **4**, p. 51-71.

Lerner, J.S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition and Emotion*, 14, p. 473-493.

Merskey, H. (1979). Pain terms: A list with definitions and a note on usage. Recommended by the International Association for the Study of Pain (IASP) Subcommittee on Taxonomy. *Pain*, **6**, pp 249-252.

Milner, P.M. (1991). Brain Stimulation Reward: A Review. *Canadian Journal of Psychology*, vol. 45, p. 1-36.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgement*. Englewood Cliffs, NJ: Prentice Hall.

Ortony, A., Clore, G.L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press.

Ploghaus, A., Tracey, I., Gati, J.S., Clare, S., Menon, R.S., Mathews, P.M., & Rawlins, N.P. (1999). Dissociating pain from its anticipation in the human brain. *Science*, **284**, 1979-1985.

Revelle, W. & Loftus, D. (1990) Individual differences and arousal: implications for the study of mood and memory. *Cognition and Emotion*, 4, 209-237.

Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making*, **3**, 263-277.

Robinson, R.G., & Coyle, J.T. (1980). The differential effect of right vs. left hemisphere cerebral infarction on catecholamines and behavior in the rat. *Brain Research*, 188, pp. 63-78

Rosenberg, A. & Kagan, J. (1987). Iris pigmentation and behavioral inhibition. *Developmental Psychobiology*, 20, p. 377-392.

Rusting, CL, and Larsen, R.J., (1996). Extraversion, Neuroticism, and Susceptibility to Positive and Negative Affect: A Test of Two Theoretical Models. *Journal of Personality and Individual Differences*., vol., 22, No. 5, p 607-612.

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition and Emotion*, 14, p. 433-440.

Smith, C.A., & Ellsworth, P.C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48, p. 813-838.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, p. 1124-1131.

Watson, D., & Tellegen,, A. (1985). Towards a consensual structure of mood. *Psychological Bulletin*, 98, p. 219-235.