Three types of emotional effects that will occur in a cognitive architecture like Soar

Originally presented at:
Workshop on architectures underlying motivation and emotion
The University of Birmingham 11-12 August 1993

Frank E. Ritter
ritter@psyc.nott.ac.uk
Centre for Research in Development, Instruction and Training
Department of Psychology
U. of Nottingham
University Park, Nottingham
NG7 2RD

Introduction and abstract

ˆ

This short paper is offered as a foil to those produced by real thinkers about emotions.[1] It is also a chance for me to think about something new, for I haven't studied emotions before, although I have spent considerable time thinking about cognitive architectures. We will find that emotions will be and must be incorporated into architectures for cognition, that both areas of research touch each other, but that more work must be done improving and extending cognitive architectures before emotions arise. Basing a theory of emotion in terms of an architecture provides, in a very direct way, a new and necessary level of meaning to emotions. The agents we are studying perform information processing, and any theory of emotion must include that as well unless emotions are not influenced by the cognitive actions of the agent. As noted below, this appears not to be the case, an agent's emotions are directly dependent on the agent's knowledge, and the architecture will be influenced by its use. As a proposed unified theory of cognition (UTC), Soar (Newell, 1990), like any UTC, needs to be extended to cover emotions, so this extension is as important to Soar as it is for theories of emotion.

There are three main ideas that come to my mind when thinking about emotions within a cognitive architecture. The first is that many instances of emotions may not be resident explicitly within the architecture. Some of them may not be directly felt by the agent, but exist only as interpretations by observers or upon interpretation by the agent itself. The observational nature of emotions is explored below in the section on cognitively based emotions.

The second point is that there may be two or more mechanisms for a given emotional label.

[1] I would like to thank Peter Bibby and Sam Marshall for discussions on this topic. Any remaining naivete is mine alone.

This is important to keep in mind, that we do not need to find a single cause for each label, for this leads to silly arguments like the tastes great/less filling debate of reasons to like a brand of beer. This proposition is directly supported by two additional types of emotional effects that I propose can appear in an architecture: (a) direct chemical causes of emotions, such as those caused by endomorphine, as well as (b) changes to the architecture through use. These have been ignored by the cognitive architecture community, and need to be incorporated at some point, although I do not attempt to do so here. In the sections on wet emotions and those caused by exercising the architecture I sketch what they might look like, and the problems that they will cause.

Finally, a theory of emotions must incorporate learning. A UTC, nearly by definition, must incorporate learning, and learning once implemented appears to be ubiquitous. There are other roles and perhaps even other types of emotions, particularly as a way of publicly labelling an agent's state (Oatley, 1992), but they are not addressed here because they exist far above the level of architecture in the social band (Newell, 1990).

Grounding emotions in a process architecture

Upon reflection, what has most influenced the thoughts presented here about emotion is not the particulars of the implementation of Soar, the architecture that I have been working within for the last five years, so I need not go into its details. What grounds these thoughts are the fundamental assumptions behind Soar that allow it to be proposed as a unified theory of cognition. These assumptions are based on cognitive architectures (Newell, 1990) and the chunking learning mechanism (Laird, Rosenbloom, & Newell, 1986).

An architecture (in the case of Soar, basically a production system) represents the structure that does not change between tasks (Newell, 1990). Adding knowledge to the architecture to perform a particular task (in the case of Soar, productions) creates a process model of a performance in a particular domain. So I am basically sympathetic to Oatley's (1992) view of emotions as evaluations of plans, and think that it is headed in the right direction. I think that a theory of emotion, because it is partially based on plans, must be based on a full process model, and an architecture that can actually process information. As noted below, there are many potential emotional states, and the emotions we talk about represent only a smallest portion of those that occur, but the most visible to without process models.

The discussions here treat architecture in terms of how it processes information. Clearly, the brain remains the same physical brain upon changes such as ingestion of alcohol, and as such, retains much of the same physical architecture. What is interesting and allows us to make useful predictions of behaviour will be how does the information processing architecture change, and this is what is examined below.

A chunking learning mechanism (Laird, Rosenbloom, & Newell, 1986) is incorporated within the Soar architecture. How it works, to the level of approximation of this paper, does not matter. What does matter is when it occurs and the transfer of what is learned. The chunking mechanism in Soar applies automatically whenever a task is completed. However, it does not exist as a separate overseeing process. On the other hand, learning based on observing and reflecting on a model's own behaviour would have to be based on the knowledge of what to reflect upon, and an explicit decision to reflect.

Thus there will be three types of emotional effects within Soar. The most common type of emotional state within a Soar model will be strictly cognitive. These are dry, implicit emotions that represent evaluations of cognitive states, rather like Oatley (1992) proposes. Reporting these states as emotions will require reflection by the agent, or observation by an informed observer. They will not require changes to what we might think of as a rational cognitive architecture, but will require us to specify the knowledge necessary to reflect and to add emotional labels.

The second type of emotion in Soar will be caused by chemical changes to the brain, such as injections of dopamine or the natural occurrence of endomorphines. These types of effects have not been modelled in Soar. When they are, they will have to be implemented as fundamental but short term changes to the cognitive architecture. These effects may be directly perceived as emotional states (which I doubt), or as a tendency to end up with a particular type of evaluation. These emotions would be relatively straightforward to model compared with the third kind of emotions that I propose are caused by cognitive actions using up or creating physiological changes in the brain, the physical system that realises the cognitive architecture, and thus in the cognitive architecture itself. For example, repeated failure of cognitive operators might deplete brain chemicals used to implement them in the architecture. Sadly, this physiological response caused by cognitive actions would add an additional level of indeterminacy to modelling. Fortunately, this appears to have less effect than the knowledge to perform the task, and by definition only occurs after several cognitive actions have been performed. These types of emotions are taken up in turn below.

Cognitive, knowledge level emotions: plain, dry, and labelled only upon reflection

Some emotions can be modelled as partial state information, operator preferences, or the history of one or both these that are part of a process model for performing a task. When Oatley talks about emotions as mental states that can be defined and talked about (Oatley, 1992, p. 17), and as describing emotions as evaluations of plans (p. 55), I believe that he is talking about this kind of emotions.

Several of Oatley's emotions correspond directly to features within the Soar architecture. Cognitively based happiness and sadness would occur often as successful (or unsuccessful) operator application. Being bored may be the repeated application of the default Wait operator (Laird, Congdon, Altmann, & Swedlow, 1990) over a period of time long enough for reflection to indicate that there were not likely to be other operators proposed for a while. Abelson's (1963) plain attitudes as well can be interpreted as preferences for operations or states.

The states and situations just discussed occur repeatedly in Soar or any other architecture more often than and vary more frequently than emotions are typically reported. What, then, makes these states emotions? I suggest that these routine states of a problem solver are promoted to emotions when an observer attempts to summarise the initial agents performance. This summary can occur from an outside observer, or by the agent itself. This makes several predictions, some perhaps new: that emotions are computed, that the agent must set aside its task and be able to note its progress in order to have an emotion, that emotions are knowledge based, that self-computed emotions may differ drastically from an observer's.

Anger — reflection on an impasse. As a more extended example, consider anger. A possible cause of an emotion one might label anger, would be reflection upon an impasse in applying an operator and finding that another agent was related to the impasse and to the inability to proceed. The conflicting agent could have removed a resource, it could have modified a resource, or it could even be itself a recalcitrant resource. This description suggests that for someone to have this type of anger, they must have enough knowledge of the world to simulate it, and time to reflect upon the causes of their problems. As anecdotal evidence, consider the lack of anger in combat, where less is known and less time to reflect is available, compared to anger on the motorway. This knowledge-based view of anger also suggests that with additional knowledge, what was anger, can be changed with additional knowledge: "I did park in your garage, but I was told it was the one that I rented."

Conclusions. Within a cognitive architecture these emotions will be emergent, implicitly defined mental states (or histories of states) that are labelled only upon reflection by self or others. Situations that map to a given external label, such as happiness, will be similar, but can be quite different in their internal definition, in terms of what type of state or what operator actually indicated success. We only know that we are having them when we observe ourselves or someone else does. Because these are descriptions of states, and not states themselves, there can be mixed emotions. You can love the sinner and hate the sin, like the operator but not the state, or have had some successful operations and some unsuccessful operators, and be a long way from the goal. Finally, much of these emotions must come from either reflection, which our current models do little of, of from observation. Because we can perform that task, the first emotions will occur through this mechanism.

Clearly wet emotions: externally caused physiological changes

The implementation of this effect on emotion is relatively straightforward. We know that exercise, certain foods, and other mediating chemicals perhaps associated with various mental illnesses will influence various facets of the brain and thus, the cognitive or information processing architecture. General arousal caused by exogenous factors is included here, as well as gross effects on cognition due to such physical changes as extremes in temperature, the bends while diving, and drugs with known cognitive effects such as alcohol and LSD. There appear to be three ways these changes would lead to emotions. The most awkward is that these changes lead directly to reports of emotions. This requires postulating chemical sensors. The second explanation is that they are simply labeled by any ongoing cognitive or social activity (Schacter & Singer, 1962). The third explanation is slightly more complicated and novel, but I think more parsimonious – these architectural changes might include changes to the basic processing rate, the amount of working memory (and its decay rate if you believe in that), and perhaps even the operator creation and selection scheme may be influenced in subtle but important ways. When it comes time to reflect and assign an emotional label, these changes will have lead the processing into a state typically associated with that change.

Because some of these emotions are directly contradicted or modified by another emotion's brain chemicals (or represent the opposite extremes of concentration), less of these emotions can co-exist. They can, however, coexist with cognitive, reflection based emotions, so there can be mixed emotions based on the cognitive state supporting one conclusion, and the physical state of the brain suggesting another. As far as I know the size and type of these effects on the cognitive architecture remain to be specified, although work has started in labeling these effects (e.g., biases in anxiety disorders, Eysenck, 1993).

Interactive emotions: Wet emotions caused by exercising an architecture

Finally, because the architecture on a cognitive level must be realised on a physical system, there may be interactions between the architecture and it use that lead to states or processing styles that would be given an emotion as a label. These types of interactions may perhaps occur through particular repetitive cognitive actions depleting a brain chemical used in their performance. In this interaction the physical architecture would shine through, although not in a classical cognitive sense of limits on reaction time, but as changes to the types of actions that can be performed or that are preferred by the decision process.

While I find these changes unanticipated by cognitive science, there is at least one analogous situation in the physical world. (These are few, for there must first be an architecture that processes information, knowledge for it to use, and for it to be implemented in a non-perfect medium, which rules out most computer based analogies.) Consider, for example, a football team (it doesn't matter which kind). The members will tire differentially based on their position and the initial strategies used. Over time, different play emerges in the team as various members tire and new strategies are selected based on this. Understanding its play will involve understanding its players, their strategies, and how implementing the strategies influence the players physically.

Frustration, for example, appears to be based on or at least require repeated cognitive actions. That it is also influenced by the body's state as well (e.g., hunger), suggests that this emotion may be related to changes to the architecture. Frustration is thus not a purely reflection based emotion, but one where actions on the cognitive level interact with the underlying physical architecture. In this case, the mental state is not only being influenced by the physical state of the agent, but by an interaction of what happens on the cognitive level with the physical state of the architecture.

There may also be a type of anger that is non-cognitive caused by reflection upon hate. In this case modifications to the architecture would result in a context independent anger caused chemically by the previous negative evaluations. Anecdotally this is supported by individual differences in how easy people become upset or stay that way. Within this view, moods are longer term modifications to the architecture caused by modifications to the physical aspect of the architecture that also modify the information processing architecture. The physical changes, and thus the moods, may be caused by physical changes, such as exercise, as well as actions on the cognitive level. Specifying these changes in detail is beyond the scope of this paper, but should be possible.

Preferences involving the self as agent. In particular, preferences and evaluations regarding the welfare and safety of the agent appear to be particularly susceptible to causing these types of changes to the architecture, such as fear, shame, excitement before success, love and hate This suggests that there may be an architectural bias to self, something which no architecture that I am aware of incorporates.

By definition these emotions represent changes to the architecture. How they interact with reflective emotions and other effects is not clear. These changes, whether labelled emotions or not, will prove the hardest to model and understand because they change the definition of architecture. Admitting these allows that in humans the architecture and knowledge are not separable as proposed by Newell (1990), but are linked through the history of their mutual interaction.

What about learning?

Let's now consider what influences learning will have on the three types of emotions sketched out above. The dry, cognitively based emotions will be directly and largely influenced by learning. What states occur, when they occur, their order, and the probability of success should increase with learning. Given learning, more operations will be successful, or at least not as often attempted. This suggests that given time, if cognitive happiness is merely getting on with it or performing plans successfully (Oatley, 1992, p. 55), happiness, if not complacency, should increase in a fixed environment.   As a person ages, any changes to an architecture through repeated and unsuccessful application (what could perhaps be labelled as child-like behaviour) should also decrease, leading to less interactive emotions. With better, or at least smoother performance on the cognitive level, the supporting physical architecture should get more stable.

Knowledge that is learned can also be transferred. Cognitive dissonance, for example, is based upon reflection, and any conclusions created will also be learned, and as such, can transfer to other domains. Admitting changes to the cognitive architecture caused by changes to its implementation has implications for transfer. Evidence of state dependent learning suggests that learning within an architecture may be dependent on the state of the architecture at the time of learning.

Conclusions

Architectures provide a language for operationally defining emotions. Considering emotions in terms of a process architecture suggests several changes to the theories of emotion, and provides an additional, although terribly difficult direction for extending a cognitive architecture such as Soar. The notes presented here suggest that a theory of emotion not incorporating or based on a process model will be incomplete and miss the point of many emotions (many of you at the workshop presumably already agree with this but I thought I had to say it). Examining emotions with respect to a process architecture also makes some testable predictions. For example, that anger, requires specific knowledge and reflection.

The most basic emotions occur often but require reflection to be noticed, reported, or remembered. The naive theory presented above suggests that the first type of emotions are easy to model. They arise within the architecture as an agent goes about its business. Most emotional states go unnoticed or commented upon. These emotions will not be evident to the agent without reflection, that is, the agent will not say "I'm happy", they may just be happy because of success. What makes these states emotional labels is another observer, even the same agent, reflecting and observing the behaviour and interpreting the actions as an affective state. This observer/observee difference offers another dimension creating further differences in labelling the causes of emotions. Upon reflection or communication with other agents, they are labelled, shared, and remembered. This suggests that we must first get our cognitive models in order before we can understand emotions, and that reflection and learning look to be important as well for understanding emotion, and where the action will be in these models.

Some emotions look like changes to the architecture, or architecture changes based on processing resources (wow!). These types of emotions will cause us problems when modelling. They will be difficult to model for at least four reasons: (a) They require a complete model of the knowledge and architecture used to perform the task; (b) They represent interactions between processes and the architecture, and these interactions have not been specified; (c) They represent changes in the midst of behaviour, adding an additional level of indeterminacy; and (d) There will probably be large individual differences on this level. These emotional effects appear daunting, and quite frankly, I don't look forward to modelling them just yet.

There will be multiple causes of some of the most common emotional labels. Some emotions appear to have dual causes, that is, be two (or more) separate phenomena. Happiness may have several strictly cognitive causes, and at least one brain based cause. We will have to move away from labels from folk psychology to something based on the causes in the architecture and the knowledge level. This is not to say that there are an infinite number of emotions, but that within a single term there may be multiple, coherent sets of causes of that emotion, including the possibility that the causes of that emotion are in the observer. This suggests that categorical statements about emotions, or even individual emotional labels, are likely to be wrong. Not all emotion will be irrational, not all emotion will be biologically affected, and you will also see differences within a single emotional state, such as happiness, based on the actual reasons for applying that label.

Several reasons can now be seen why cognitive science has paid little heed to emotions. Our models have lacked the ability or need to model the fine levels of evaluation on which reflection and thus emotions are based. The lack of attention may also reflect the relatively rarity of emotions in basic cognition, for it is proposed here that it is only upon reflection that states are seen as emotions. Other influences on cognition outside of the knowledge level but connected with emotions are problematic as well. The relatively modest changes to the architecture that these physiological changes represent require understanding the unmodified cognitive architecture first, which we are still struggling with.

References

Abelson, R. P. (1963) Computer simulation of "hot" cognition", in Tomkins, S. S., & Messick (eds.), *Computer simulation of personality: Frontier of psychological theory*, New York, New York: Wiley & Sons.Eysenck, M. W. (1993). Cognitive factors in generalized anxiety disorder, Invited lecture presented at the III European Congress of Psychology, also published in *Acta Psychologica Fennica*, XIII, 99-112.

Laird, J. E., Congdon, C. B., Altmann, E. & Swedlow, K. (1990). Soar User's Manual: Version 5.2 (Tech. Rep.) CSE-TR-72-90. Electrical Engineering and Computer Science Department, University of Michigan. Also available from The Soar Project, School of Computer Science, Carnegie-Mellon University, as technical report CMU-CS-90-179.

Laird, J. E., Rosenbloom, P. S.., and Newell, A. (1986). Chunking in Soar: The anatomy of a general learning mechanism. *Machine Learning*, 1(1), 11-46.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Oatley, K. (1992). *Best laid schemes: The psychology of emotion*, Cambridge, England: Cambridge University Press.

Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. Psychology Review, 69(5), 379-399.