Representing and Processing Emotions in a Cognitive Architecture

Pat Langley

Center for Design Research
Stanford University

Thanks to John Laird, Christian Lebiere, David Nicholas, and Paul Rosenbloom for useful discussions. This research was supported by Grant No. FA9550-20-1-0130 from AFOSR, which is not responsible for its contents.

The Ubiquity of Emotions

Emotions play a central role in most aspects of human life; they color and modulate our activities, both physical and mental.

This raises an important and interesting scientific question:

- *How are emotions related to cognition?*

More broadly, what function do emotions serve in an integrated cognitive system?

Science fiction often depicts human-level AI systems as devoid of emotion, but does this really make sense?

Emotions and Rationality

AI's default view is that emotions are 'irrational' evolutionary holdovers which are *detrimental* to intelligence.

- We can build systems that – *to some extent* – reason, plan, and communicate without emotions.
- But Simon (1967) has argued affect and emotion are crucial for controlling cognitive attention.
- And Damasio (1994) reports brain-damaged humans – with little or no emotion – who cannot make decisions.

This suggests that human-level cognitive systems may actually *require* emotions.

Some Distinctions

Both academic papers and everyday language often confound some distinct concepts:

- *Affect.* The positive or negative aspect of some experience.
- *Mood.* A global variant of affect for an entire cognitive system.
- *Emotion.* A mental structure related to goals and beliefs about an event, agent, or object.
- *Feeling.* An affective or hormonal response that is associated with an emotion.

This talk will focus on emotions, which are the most complex and interesting from an AI perspective.

Examples of Emotions

We view many emotions as important enough to name, such as:

Happy	Sad	Angry	Afraid
Worried	Despairing	Love	Proud
Courageous	Disappointed	Relieved	Pleasantly surprised
Frustrated	Satisfied	Helpless	Annoyed
Irritated	Disgusted	Resentful	Envious
Jealous	Embarrassed	Guilty	Ashamed
Regretful	Offended	Self righteous	Sympathetic
Pitying	Amused	Wonder / Awe	Schadenfreude

Other mammals have emotions, but not as complex as those in humans, which suggests a strong cognitive component.

Representing Emotions

Before we can discuss emotional processes, we must consider how to *represent* them.

Marsella, Gratch, and Petta (2011) distinguish three ways to encode emotional content:

- *Dimensional* models – points in N-dimensional space
- *Anatomical* models – activations in neural circuits
- *Appraisal* models – relations among cognitive structures

These frameworks suggest radically different ways to represent and process emotions.

Drawbacks of Dimensional Models

Many dimensional models characterize emotions as *points* in a three-dimensional space:

- *Pleasure* – measure of valence
- *Arousal* – level of affective activation
- *Dominance* – measure of control

Synthetic characters often use such 'PAD' models (Wachsmuth, 2008), but they ignore key facts:

- Emotions are *about* some event, person, or object;
- We can have *mixed* emotions about the same target.

This suggests that they involve much richer cognitive structures.

Emotions as Cognitive Structures

Appraisal models view emotions as inferred relations among mental structures based on situations.

Ortony et al. (1988) describe 22 configurations for emotions organized around events, objects, and other agents.

These patterns serve as 'elicitation' conditions on emotions that involve relations among:

- Goals, intentions, expectations, and beliefs
- Inferences about others' goals, intentions, and beliefs

This suggests that such emotional structures are *abstract* and *domain independent*, much like rules of dialogue.

The PUG Architecture

The PUG architecture (Langley et al., 2016) supports embodied agents with four core ideas:

- *Qualitative beliefs are grounded in quantitative descriptions*
- *Symbolic goals have numeric utilities that reflect tradeoffs*
- *Discrete skills have associated control equations*
- *Mental simulation guides task and motion planning*

Research on PUG has focused on simulated mobile robots but can be used for many intelligent agents.

Moreover, the architecture has potential for explaining links between cognition and emotion.

PUG's Knowledge Structures

PUG incorporates four distinct types of generic, long-term knowledge structures:

- ***Concepts*** – Define relational categories, attributes, and *veracity*
- ***Motives*** – Indicate *utility* of relations conditioned on situation
- ***Skills*** – Specify *control values* based on match to *target concepts*
- ***Processes*** – Predict *changes* in attributes given current values

PUG uses these elements for conceptual inference, reactive control, heuristic evaluation, and plan generation.

Two Facets of Emotions

We should distinguish between two aspects of emotions that seldom appear in the literature:

- *Long-term* generic structures that *generate* these instances
- *Short-term* structures that encode emotional *instances*:
  - *(disappointed John (failed John CompSci101))*
  - *(resents John (passed Sam CompSci101))*

This dichotomy maps directly onto PUG's distinction between generic ***concepts*** and specific ***beliefs***.

Appraisal theory refers to emotion concepts as *appraisal frames*.

Emotional Concepts

We can specify emotional concepts in a PUG-like notation:

- An agent A is *disappointed* about event E if A wanted E, expected E, and believed E did not occur:

```
((disappointed ?A ?E)
  :conditions ((goal ?A ?E) (expect ?A ?E) (belief ?A not(?E)) )
```

- An agent A1 is *jealous* of agent A2 if A1 wants object O, believes he does not possess it, and believes A2 does possess it:

```
((jealous ?A1 ?A2 ?O)
  :conditions ((goal ?A1 (possess ?A1 ?O))
               (belief  ?A1 not(possess ?A1 ?O))
               (belief ?A1 (possess ?A2 ?O))) )
```

PUG can encode domain-level concepts, but we must extend it to handle these meta-level structures.

Two Forms of Emotional Processing

We can distinguish two types of cognitive tasks that make use of such emotional concepts:

- *Generation*, which produces emotions for a primary agent
  - *(disappointed John (failed John CompSci101))*
- *Understanding*, which infers the emotions of other agents
  - *(belief John (disappointed Jane (failed John CompSci101)))*

We need both for a full account of emotion's links to cognition.

Naturally, we should ask whether PUG's existing mechanisms for conceptual processing can handle them.

Conceptual Inference in PUG

The PUG module for *conceptual inference* derives all beliefs consistent with environmental perceptions by:

- Matching conceptual rules to infer beliefs like (*robot-at R1 O1*)
  - Computing values of numeric attributes associated with beliefs
  - Calculating *veracity* (degree of match) for each inferred belief
- Applying this recursively to generate the full deductive closure

This differs from Prolog's processing in that inference is driven by data (percepts) rather than queries.

PUG's current mechanisms should handle emotion *generation* but not emotion *understanding*, which requires abduction.

The Effects of Emotions

A complete computational theory of emotions must also explain their indirect effects on:

- Physical behavior
  - *E.g., crying about loss or punching someone in anger*
- Cognitive processing
  - *E.g., changing goal priorities or invoking planning*

This assumes that emotions are *not* evolutionary relics; instead, they serve as high-level *regulators* of cognition.

Emotions and Metacognition

Recall that emotional concepts specify abstract relations among goals, beliefs, and expectations.

Moreover, emotions influence domain-level decision making and planning. Together, these suggest that:

- *Emotions play a **metacognitive** role that operates over and influences base-level cognition.*

That is, emotional processing inspects traces of basic cognition and alters its course.

This postulate elaborates on Simon's (1967) idea that emotions serve as *regulators* of cognition.

Motivational Processing in PUG

PUG already includes structures – *motives* – that match against beliefs, generate goals, and compute utilities.

We can extend motives to match emotional beliefs and create goals to change others, as in the rule:

- ((wants ?P (disappointed ?Q _))
  :conditions ((disappointed ?P ?R) (believes ?P (cause ?Q ?R))))

This encodes an *eye for an eye* motive that leads a disappointed agent to want the responsible party to feel the same.

The goal's priority will depend on strength of the causal belief and level of agent disappointment.

Summary Remarks

This account unifies ideas from appraisal theory and cognitive architectures, the most important that:

- Emotions are symbolic *cognitive structures*
- Emotional *concepts* produce emotional *beliefs*
- These can both *generate* and *understand* emotions
- Emotions have a *metacognitive* influence on behavior
- *Motives* mediate this influence by altering *goals / utilities*

The PUG architecture holds promise for incorporating these theoretical postulates about emotion.

Sloman (2001), Gratch / Marsella (2004), and Marinier et al. (2009) have proposed similar accounts.

References

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

Gordon, A. S., & Hobbs, J. R. (2017). *A formal theory of commonsense psychology: How people think people think*, 513–544. New York: Cambridge University Press.

Gratch, J., & Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, *5*, 269–306.

Langley, P., Barley, M., Meadows, B., Choi, D., & Katz, E. P. (2016). Goals, utilities, and mental simulation in continuous planning. *Proceedings of the Fourth Annual Conference on Cognitive Systems*. Evanston, IL.

Marinier, R., Laird, J., & Lewis, R. (2009). A computational unification of cognitive behavior and emotion. *Cognitive Systems Research*, *10*, 48–69.

Minsky, M. (2007). *The Emotion Machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. New York: Simon and Schuster.

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

Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review*, *74*, 29–39.

Sloman, A. (2001). Varieties of affect and the CogAff architecture schema. *Proceedings of the AISB'01 Symposium on Emotion, Cognition, and Affective Computing*. York, UK.