The First, Second and Third Person Emotions: Grounding Adaptation in a Biological and Social World

Chrystopher Nehaniv
Interactive Systems Engineering
University of Hertfordshire
Hatfield, Herts AL10 9AB, U.K.
c.l.nehaniv@herts.ac.uk
http://www.u-aizu.ac.jp/~nehaniv
fax: +44-1707-284303, phone: +44-1707-284470

Abstract

We consider the interpretation of emotions and similiar phenomena as support for survival and coping in the world. Grounded in the experience of an emotional agent, certain such emotions, drives or experiences are self-oriented (homeostasis, intake, outflow: hunger, pain, irritation), while others suggest a generalized or specific recognition of other agents or objects (curiosity, fear; or hatred, envy, yearning, greed). Other, more complex emotions are involved in relations to a second person (sympathy) or social regulation (shame, guilt, feelings of loyalty) or affective episodic structure (regret). Considering complex emotions in relation to other 'persons' yields insight into the roles and possible design of various emotions in behavioral regulation in biological, software, and social contexts.

Affective coloring of episodic memories of sequences of actions and experiences may suggest a mechanism for grounding of behavioral adaptation. We outline how such coloring of episodic memories might provide a mechanism for emergent spatial and social navigation.

Introduction

Human languages distinguish the self ("I" – the first person), and the other with whom the self directly interacts and whom the self addresses ("you" – the second person), and the other with whom one is not engaged directly ("he/she/they" - the third person). We argue that these grammatical categories of human natural language are useful in considering the adaptive role of emotions as they relate to, regulate, and control behaviors in a biological and social setting.[1] We shall use the term "emotion" rather more loosely than is common in the literature (e.g. (Harré & Parrott 1996)) to include phenomena often labelled as drives, feelings, emotions, and so on. Furthermore, we will not insist on distinction according to the manner in which such phenomena are experienced, whether consciously, as sensations with affective coloring, or as physiological drives and so on. Nevertheless, such notions as "drive", "feeling", "emotion", etc. could possibly emerge from the considerations in the domain of person and related factors. Furthermore, as suggested by an anonymous reviewer, this domain of person may turn out to correlate with different areas of nervous system and brain function.

We consider emotions or emotion-like phenomena in evolved organisms (whether or not they possess a nervous system) with a view to applications in artificial physical and possibly software agents, but also to understanding the role emotions play in the regulating the behavior of organisms embedded in their biological and social environments.

While we speak of the 'person' of emotions, this 'person' is understood here as any organism or agent, rather than a human person. No doubt in the case of the virus, bacterium, vegetable or fungal agent the reader may feel that the use of this term or the use of the term 'emotion' is too anthropomorphic, since it is likely (but impossible to determine) that such creatures do not have experiences in the sense that humans and other animals do. Yet we believe that there is a certain degree of universality here to justify the use of these terms, whether or not the corresponding emotions could arise in such a 'person', and, all the more so, that the notions will prove useful in the design and understanding of artificial agents (physical or software).

A (Generalized) Phenomenological Perspective

I feel, therefore I am.

The question of emotions has defied scientific analysis at least since the days of Aristotle, because emotions are of a subjective nature (Harré & Parrott 1996;

[1] Although it is important, it should be hardly necessary to say, however, that gaining insight from such considerations of the emotional agent's self and its relatedness to others does not in any way endorse or entail a linguistic, symbolic representational or logicist view of behavior and intelligence.

James 1884; Arnold 1968). The situation is similar to the problem of studying consciousness: each of us can be sure he or she is conscious via direct experience, but cannot *know* directly whether another human also experiences consciousness. Our experience of consciousness and emotion is *first-person.*

You are like me, therefore you feel.

We can observe that beings, who we hypothesize are like ourselves, have similar experiences. The assumption that other humans are *second-persons*, that they experience the world in a way similar to the way that I do, turns them from mysterious objects into beings whose actions make sense. We can understand their actions by attributing intent, motivations, and emotions to them similar to the ones we ourselves experience. Treating others as second-persons, empirically, has turned out (for most of us humans) to be a successful basis for us in dealing with our social world (but see the discussion of empathy and autism in (Dautenhahn 1997)). Thus, it seems that assumption that other beings have minds like our own is a good working hypothesis for a social being (at least for mammals living in individualized societies – although perhaps not for social insects (Wilson 1974; Dautenhahn 1998). Indeed, it may be an involuntary assumption, but the tendency to make it is in itself revealing.

Just as we cannot study the consciousness of others directly, we can also not experience the affective states of other humans or other living things, nor can we be sure that anything is experienced at all. The empirical phenomenological perspective asks not for external proof of such experience, but adopts as a working hypothesis "others are like me". We shall adopt this hypothesis also in the case of non-human mammals and even other animals – who (it seems) are certainly not experiencing complex human emotions like shame or pride. But, as a metaphor, an engineering assumption, or design principle, we use notion of experience of emotion to gain insight into how systems surviving in biological and social worlds might function.

Emotional Grounding

Masanao Toda's Fungus Eaters (Toda 1982) carry out certain behavioral programs for survival when they have certain set of "urges"; this results in adaptive behavior. Such urges serve to regulate behavior, yielding appropriate actions in appropriate contexts (e.g. eating when hungry). Steve Grand (Grand, Cliff, & Malhorta 1997) has used such notions in implementing CREATURES, a successful product based on artificial life technology in which so-called Norns grow-up in an environment learning and responding to stimuli while being governed by a set of urges (hunger, sex-drive, etc.) that may take on dynamically varying numerical values.

Elsewhere the author and K. Dautenhahn (Nehaniv & Dautenhahn 1998) have argued that "meaning" only arises for an agent historically embedded in its environment (*No agent, no meaning*). Emotion systems involved in feedback control of situated agents may serve to provide the grounding for embodied agents in the body/environment coupling. Moreover, affect may play an important role in memory and historical grounding. The psychologist R. Zajonc has shown in the 1980s at the University of Michigan (Zajonc 1980) that humans prefer stimuli that they have experience before to new stimuli (***Familiarity breeds content[ment]***), whereas rational humans who have lost some affective capacities due to aphasia are unable to function "rationally" (Damasio 1994).

First Person Emotions

First person emotions may be viewed as useful to regulators of the individual's own state. An organism maintains its state within certain acceptable ranges for various parameters and tends to act to restore its state to within these ranges if disturbed (homeostasis). Irritability or pain are experienced when there is a deviation away from the acceptable range of relevant parameters in or outside the body (acidity, glucose-level, pressure, suitability of environment for respiration – oxygen concentration, air pressure, etc.). Many emotional terms and terms for feelings are names for the experience of such deviations or the corresponding restorations (hunger, nausea, itching, dizziness, feelings of cold/heat/burning). Pleasure and pain are the most general terms for such first person emotions.

The point is that such first person emotions play a role in regulation of behavior involving only the organism itself in its environment (e.g. a snake moving onto a sun-baked stone when feeling cold). In the case of organisms or agents without nervous systems, various forms of taxis can also be construed as resulting from generalized emotions ("internal signals" or rather aspects of state in organismal dynamics) arising from impinging sensory data leading to behaviors.

Second Person and Social Emotions

To support complex social interactions, including grounding and adaptive behavior in individualized (and possibly anonymous) societies, other types of emotions are useful. Empathic experience and the recognition of other (possibly of non-conspecific) individuals as having an experience of the world analogous to one's own may be requisite for the experience of second person emotions. It is possible to imagine that such mechanisms do operate without necessarily requiring representation of other minds, i.e. in individuals without a theory of other minds (cf. the relation of this notion to possible mechanisms implicated in autism, e.g. (Baron-Cohen 1995)). But do dogs growling at someone approaching their food have a theory of mind? The perception of *intent* on the part of another conspecific or other animal seems likely not to require the representation of a mind constructing that intent.

Emotions concerning the behavior of others with whom one is not (potentially) engaged could be termed third person emotions. An example might be disdain, but not indignation since that latter results from considering consequences to the self (or second persons onto whom one empathizes). It is unclear whether third person considerations are genuinely emotional or even whether they could enter with any phenomenological relevance for a situated agent.

Dimensions of Emotion

Emotions are always experienced in the self. They may be present or absent, and they are valenced (positive or negative). They may have varying levels of intensity. An emotion may focus on the physical self, objects (or others to whom intent is not attributed in the emotion) in relation to the self, or engagement with another agents (second person emotions) or the perception by others – including real and imagined persons – of the self (second person emotions). They may involve valuation of motives of others (other social emotions). In another dimension, consequences for that agent or actions of this agent are important in emotion. Emotions may focus on temporal episodes or just on the current state of the world.

Tangled Knots

Blending of the first person and the other gives rise to the second person, while relations to the self complicate matters further (in addition to the consideration of other dimensions mentioned above).

Interaction of biological agents is grounded in first and – in those agents that perceive others — also second person emotional experience.

Valenced responses to stimuli (positive or negative) by the body have been implicated both the source of emotional expression and prior to cognitive aspects (James 1884; Darwin 1892; Zajonc 1984). For many and certainly for more complex emotions, cognitive factors (Ortony, Clore, & Collins 1988) conditioned by cultural factors (Landman 1996) play a role in the genesis and experience of emotion. Temporal aspects are also evident in emotions such as hope and regret.

Implications for Agents and Artifacts

These admittedly preliminary and unpolished considerations, while not leading to any full classification of emotion and related phenomena, already do lead us to some consequences for the phenomenology of emotion in the design of social agents and the understanding of biological ones.

How can emotion improve decision-making?

Obviously, pain and pleasure are important in reinforcement learning. But emotion can serve as a contextualization cue for appropriate action. Doing something that doesn't feel right leads an organism to another approach – to use a computational metaphor – to switch to another "program". The intensity of emotion providing a measure for the degree to which this is desirable.

An emotion such as regret expresses pain about actions or lost opportunities in the past, and thus leads to evaluation of prior actions and reflections on their appropriateness for the past context — potentially, results of such analysis can guide future behavior and facilitate an escape from the immediacy of merely reactive reinforcement learning.

How is emotion useful in social behavior?

Social emotions such as gratitude, guilt, regret, sympathy, or being proud of another all require the existence (or perception) of a second person. Clearly such emotions have social implications, since they are likely to influence behavior toward others. The fact that they exist suggests that they play may play a role in group selection – that is, the selective advantage for the individual gained by virtue of belonging to a successful group. Such emotions as well as the recognition of the social standing of others may lead to natural ethics in a community of agents (Sigmund 1998; Ridley 1996).

Primates spend a lot of time in social relations, in grooming, forming and breaking alliances in complex hierarchies of power relations. It may be true that social creatures with more complex social structures need larger, more complex brains. Concern about the relations to and the intentions of others seems a major focus of primate social behavior (Dunbar 1997). The social intelligence hypothesis (Byrne & Whiten 1988) states that abstract and more generalized reasoning skills evolved in humankind based on the preadaptation of existing social reasoning skills such as those seen in other higher primates. Thus, socially situated emotion may well have been a prerequisite for the development of humanlike intelligence. If this is the case, the implication is that to achieve such a level of artificial intelligence may well require not only computational power, and not only affective computation, but support and grounding for second person and social emotions.

How can emotional and episodic memory be integrated?

The effect of emotion, especially intense emotion, on memory is well-known (e.g. (Gregor 1997)). Experiments of (Zajonc 1980) shows that humans prefer stimuli to which they have been exposed before rather than other stimuli (the so-called *mere exposure effect.*) This could be adaptive since it would tend to lead an agent toward the familiar (as more pleasant) rather than the unknown. Going toward the familiar has at least two advantages:

(1) the agent tends to stay in contexts of which it has more experience and in which it hence has been able to survive or otherwise act appropriately

(2) this could serve as an aid to navigation without resort to maps – going toward the familiar is likely to lead one home or to a destination the path to which the agent has traversed before.

Indeed, this later mechanism suggests an manner for navigation to arise as an emergent phenomenon.[2] Many of us can find our way home or to the store following the familiar – even on the second time the trip is made – while being totally incapable of drawing or visualizing a geometric or in any way accurate map of our route (not to mention the space in which this trajectory is followed).

Such considerations suggest that emotion may play a role for cognitive maps, in which affective features may mark episodic sequences resulting in an 'emotional coloring' not only of actions but of sequences of actions and experiences.

A natural supposition then is that second person and social emotions could play a role in the affective coloring of episodes of social interaction and thus serve to lay down a cognitive, non-representational map emerging during social interaction and useful in navigating the social world.

Expression of affectively colored historical memories (episodes) appropriate for the agent situated in its world could be approached via the algebraic framework of (Nehaniv & Dautenhahn 1998) to provide historical grounding for an affective system.

What could affective implementations teach us about 'wet' biological systems?

Computer and robotic simulations of emotional systems might contribute to biological research by providing frameworks and testbeds in which purported mechanisms for various behaviors could be studied. Success or failure of an artificial implementation of someone's explanatory mechanism could never prove that a biological system has the same hypothesized structure or uses the hypothesized mechanism, but successful implementations could show that a proposed mechanism may be sufficient and feasible for realizing the phenomenon displayed by an emotional organism embedded in its environment.[3] Emotionally intelligent artificial agent implementations could be used to gauge the explanatory power of affective mechanisms in explaining successful social behavior, possibly even at the level of non-individualized societies like ants, termites, even *Dictyostelium discoideum*, as well as social mammals).

Acknowledgements

The author gratefully acknowledges stimulating discussions with Joseph Goguen, Kerstin Dautenhahn and Meurig Beynon as well as other participants at the International Workshop on Computation for Metaphors, Analogy and Agents (see (Nehaniv 1998)), on many of the topics touched on here.

[2] Such emergence of navigational skill would offer a radical alternative to what traditional roboticists refer to as "localization", i.e. determining one's location on an internal, external or constructed map (possibly using external beacons and landmarks) despite drift, freak perceptions, sensor noise, etc.

[3] In this connection of computationally-inspired biology, a good example are the insect-like robots of Barbara Webb (Webb 1994), who showed that a much simpler mechanism than several proposed elaborate ones was sufficient to account for phonotaxis in mate-seeking crickets and produced as side-effects other cricket-like aspects of movement and behavior.

References

Arnold, M. B., ed. 1968. *Nature of Emotion : Selected Readings.* Harmondsworth: Penguin.

Baron-Cohen, S. 1995. *Mindblindness: An Essay on Autism and Theory of Mind.* MIT Press.

Byrne, R., and Whiten, A. 1988. *Machiavellian Intelligence.* Clarendon Press.

Damasio, A. R. 1994. *Descartes' Error: Emotion, Reason and the Human Brain.* New York: G.P. Putnam & Sons.

Darwin, C. 1892. *The Expression of Emotion in Man and Animals.* Reprinted by the University of Chicago Press, 1965.

Dautenhahn, K. 1997. I could be you - the phenomenological dimension of social understanding. *Cybernetics and Systems Journal: Special Issue on Epistemological Aspects of Embodied AI* 28(5):417–453.

Dautenhahn, K. 1998. Meaning and embodiment in life-like agents. In *(Nehaniv 1998).* Preliminary version in: *Plenary Working Papers in Computation for Metaphors, Analogy and Agents (6-10 April 1998 - Aizu, Japan).* University of Aizu Technical Report 98-1-005. 24–33.

Dunbar, R. 1997. *Grooming, Gossip, and the Evolution of Language.* Harvard University Press.

Grand, S.; Cliff, D.; and Malhorta, A. 1997. CREATURES: Artificial life autonomous agents for home entertainment. In *Proc. First International Conference on Autonomous Agents (AGENTS'97 – Marina del Rey).* Association for Computing Machinery.

Gregor, J. A. 1997. *Memory & Remembering: Everyday Memory in Context.* Addison Wesley Longman Limited.

Harré, R., and Parrott, W. G., eds. 1996. *The Emotions: Social, Cultural and Biological Dimensions.*

James, W. 1884. What is emotion? *Mind* 9:118–205.

Landman, J. 1996. Social control of 'negative' emotions: The case of regret. In *(Harré & Parrott 1996).*

Nehaniv, C., and Dautenhahn, K. 1998. Embodiment and memories – algebras of time and history for autobiographic agents. In Trappl, R., ed., *Cybernetics and Systems '98*, Vol.2, 651–656. Austrian Society for Cybernetics.

Nehaniv, C., ed. 1998 (to appear). *Computation for Metaphors, Analogy and Agents.* Lecture Notes in Artificial Intelligence Series, Springer Verlag.

Ortony, A.; Clore, G. L.; and Collins, A. 1988. *The Cognitive Structure of Emotions.* Cambridge University Press.

Ridley, M. 1996. *The Origins of Virtue: Human Instincts and the Evolution of Cooperation.* Viking Books.

Sigmund, K. 1998 (to appear). The social life of automata. In Nehaniv, C., ed., *Mathematical and Computational Biology.* Providence, Rhode Island: Lectures on Mathematics in the Life Sciences Series, American Mathematical Society.

Toda, M. 1982. *Man, Robot, and Society.* The Hague: Nijhoff.

Webb, B. 1994. Robotic experiments in cricket phonotaxis. In Cliff, D.; Husbands, P.; Meyer, J.-A.; and Wilson, S. W., eds., *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior (August 8-12, 1994, Brighton, England)*, 45–54.

Wilson, E. O. 1974. *The Insect Societies.* Harvard University Press.

Zajonc, R. 1980. Feeling and thinking: Preferences need no inferences. *American Psychologist* 35:151–175.

Zajonc, R. 1984. On the primacy of affect. *American Psychologist* 39:117–123.