Emotion as an integrative process between non-symbolic and symbolic systems in intelligent agents

Ruth Aylett
MACS, Heriot-Watt University
Riccarton, Edinburgh EH10 4ET
ruth@macs.hw.ac.uk

Abstract

This paper briefly considers the story so far in AI on agent control architectures and the later equivalent debate between symbolic and situated cognition in cognitive science. It argues against the adoption of a reductionist position on symbolically-represented cognition but in favour of an account consistent with embodiment. Emotion is put forward as a possible integrative mechanism via its role in the management of interaction between processes and a number of views of emotion are considered. A sketch of how this interaction might be modelled is discussed.

1. Embodied cognition

Within AI, the problem of relating cognition and action has led to a well-known division of opinion since about the mid 1980s between the older symbolic computing position that classically saw action as a one-many decomposition of abstract planned actions from a symbolically-represented control level and the situated agent view, as in Brooks (1986), that saw action as a tight stimulus-response coupling that did not require any symbolic representation. This can be – and originally was – posed as an engineering question of how to produce systems that were able to act competently in the real world, hence the origin of the argument in robotics, where the real world cannot be wished away and where robot sensing systems do not deliver symbols.

At this engineering level, the 1990s saw a pragmatic reconciliation of these divergent positions in hybrid architectures, usually with three levels (Gat 97, Arkin and Balch 97, Barnes et al 97), in which the relationship between symbolic planning systems and non-symbolic reactive systems was resolved by giving the reactive systems ultimate control of execution but either allowing planning to be invoked as a resource when needed (for example as a conflict resolution mechanism, or to provide sequencing capabilities) or giving planning systems to ability to constrain and contextualise – but not determine - the reactive system in what is sometimes known as supervisory control.

However the argument that was carried on from an engineering perspective in robotics, and which to some extent is now a done deal, has subsequently continued at a more scientific level in cognitive science, a discipline within psychology that arguably owed its existence to ideas from classical AI and was heavily influenced by the symbolic world-view as seen in large-scale computational cognitive models such as SOAR (Rosenbloom et al 93) and ACT-R (Anderson et al 04). Just as in pre-Brooksian robotics, these models can be criticised for not providing any adequate account of the role of sensing or motor action, which are implicitly seen as peripheral to a cognitive model much as I-O capabilities are peripheral to a computer processor.

A deeper criticism arises from a view of agency which sees *embodiment* as a key starting point rather than an optional extra (Clark 98) and starts from a body in a specific environment that needs a mind to control it rather than a mind considering abstract problems. In the world-view of embodied cognition (Wilson 02), sensori-motor engagement is the ground from which cognitive abilities are constructed (as in Piaget's developmental psychology); thus neither cognitive abilities nor specific environments and interactions can be detached in the way that had been previously assumed, and a dynamic and process-based view supersedes a declarative and logical-inferencing based view. The Cartesian separation between mind and body which still seems to exist in multi-level architectures is abandoned in favour of brain-body integration, in which processes such as the endocrine system play a vital coordinating role.

This does not mean however that a symbolic account of cognition is a pointless activity either from a scientific or an engineering perspective (so we do not actually have to abandon the whole of cognitive science up to now as well as a large chunk of psychology). Just as computational neuro-physiology does not operate at the explanatory level of physics, there is no reason why cognition based on symbolic reference must be reduced to neuro-physiology, even though what is known about the way in which the brain works suggests that symbol manipulation is an emergent property of the dynamic system formed by interaction between neurons. Indeed, the very concept of emergence dictates that an emergent phenomenon is modelled at its own level of representation since it cannot be decomposed into any one of the components whose interaction produces it. Thus an account at the symbolic level may be a valid one as long as it does not incorporate incorrect assumptions about the relationship between this activity and sensori-motor engagement.

The power of symbolic reference to abstract out of the current sensory context, to project into imagined contexts, to discretise continuous experiences into conceptual aggregates, to communicate through language and to use the environment to scaffold engagement with the world and other humans (as through writing) has a substantial impact on both cognition and interaction for humans. Thus current work on social agents that aims to produce more human-like systems whether via robots or graphical characters must incorporate symbol-manipulation systems as well as the sensor-driven systems that allow them to act with some degree of competence in their physical or virtual world.

However an important characteristic of these human abilities that has not been replicated in computer-based systems is that unification of symbol-manipulation abilities with non-symbolic behaviours driven more directly by sensing that we observe in human activity. In the human case, the hypothesis that symbol manipulation is an emergent property of interaction between brain components suggests that the ability to move smoothly between cognition and reactive engagement with the world is probably a matter of adjusting the interaction between these components and does not therefore require explicit conversion between multiple representations in the way this is typically carried out in current hybrid agent architectures. Unfortunately the computational neuro-physiology account of specific brain subsystems is still fragmentary, and there is no short-term (or even medium-term) likelihood of producing the principled interactional account that this hypothesis requires. Arguably, solving the problem of emergent symbolic reference would not only deal with the origin of language but possibly also with consciousness. It is thus a non-trivial enterprise.

How then are we to proceed with an integrated account that is not merely a pragmatic engineering kludge needed to produce competent social agents? The argument of this paper is that one should see process regulation as the key to the enterprise since even if the detailed mechanisms adopted may be invalidated by more extensive models of the brain, the basic approach is likely to be compatible with such models. The specific hypothesis of this paper is that affective systems should be considered as a key component of such regulation because of their role in attentional focus, in relation both to perception and cognition, as well as the management of goals, the allocation of internal resources, and the balance between thinking and acting.

2. Accounts of emotion

Just as accounts of action split into two camps in AI from the mid 1980s, there are two corresponding accounts of emotion and its role with respect to agency, one more related to models of the brain and nervous system, and process-oriented, and one related to symbolic models of cognition dealing with goal management and inferencing.

The first of these views, which aims at neurophysiologically plausibility, models emotion as part of a homeostatic control mechanism. Often incorporating a model of the endocrine system (Canamero 98) it suggests that emotion should be viewed as the set of brain-body changes resulting from the movement of the current active point of a brain-body process outside of an organism-specific ‘comfort zone’. It does not therefore require a single meter-like component in an agent architecture to represent *an* emotion, but offers a distributed representation interpretable in terms of the internal process states and external expressive behaviour as an emotion.

As well as an independent system state, one can also regard emotion in this framework as modifying the impact of an incoming stimulus. Thus emotion can be incorporated into action selection both indirectly and directly in much the same way as perception, and indeed can be thought of as functioning rather like an internal sensing process concerned with all the other running processes.

The second view of emotion is usually known as *appraisal theory* since it assumes a cognitive appraisal associated with perception that assesses the relationship between symbolic categories established via a model hierarchy using perceptual input and the cognitive-level goals of the agent.

Specific appraisal-based theories that have proved highly influential in the construction of graphical characters are those of Ortony et al (1988), which was based on a taxonomy of 22 emotion types, each with an associated appraisal rule, and that of Lazarus (1984), which links appraisal to action via the concept of coping behaviour. Sherer

(01) decomposes appraisal into a sequence made up of Relevance Detection, Implication Assessment, Coping Potential Determination and Normative Significance Evaluation, but remains tied to a top-down view of emotion in which cognitive processing results in later physiological changes.

Interestingly however one can interpret appraisal as an abstraction on a process of the same type as the first view in which a goal takes the place of a comfort zone. This does not mean that goals could never be independently determined at the cognitive level, but it offers the possibility of propagating the state of the homeostatically-regulated non-symbolic systems into the more abstract representational space of symbolic cognition.

In contrast to both of the views discussed above, Izard (1993) takes a heterogeneous approach which does not rule out appraisal but argues that emotion can also be generated directly by the nervous system, as in the first account, by empathy and by highly intense states of physiological drives such as hunger or lust. Such an approach is consistent with the integration between both views of emotion as part of an integration between different types of process within an agent.

3. A sketch of interaction

It is very tempting to view the integrative functions we are seeking as a way of linking different *levels* within a multi-level hierarchy. Within robotics this is exactly how these issues are discussed: symbolic cognition is a high-level system while non-symbolic reactive systems are low-level. We have avoided this terminology so far because it is highly ambiguous in other fields. Within psychology and cognitive science, high-level could also mean evolutionarily more recent or conscious as distinct from subconscious.

However we have argued above that it does make sense to think of a representational level for symbolic processing even if the way in which we implement it in a computer is very different from the way it is implemented in the brain. Once time is taken into account it is also clear that the processes on which symbol manipulation depends run with fewer real time constraints in terms of delivering motor commands, are able conceptually to stay at what we would call a higher level of abstraction and as a result provide discrete categories covering what at a non-symbolic level would be seen as dynamic processes.

It is however misleading to think of reflection as an activity that runs wholly as symbol manipulation and reaction as an activity that does not use symbol manipulation at all. As Wilson (02) argues, reflection is grounded in the mechanisms of sensory processing and motor control that evolved for interaction with the environment even when it is being applied for purposes that do not require immediate activity in the specific environment of the current moment; what she describes as *offline cognition*. At the same time, appraisal is an example of online cognition which may be quite reactive.

In fact, emotion seems to be closely tied to the distribution of internal resource between the processes producing symbol manipulation and others with tighter connections to motor action, as witness emotional flooding, in which cognitive activity seems to substantially shut down. We can think of this as a type of internal attentional focus which may be more closely tied to the attentional focus proposed by perception for a tight loop with the current environment or less tightly coupled to it when more offline cognition is taking place.

As a sketch of interaction, we finally consider a possible relationship from symbolic to non-symbolic (what might be called top-down processing if we adopt the levels vocabulary) and then from non-symbolic to symbolic.

3.1 From symbolic to non-symbolic

As mentioned in section 1 above, this is an issue that has been relatively extensively discussed in robotics, though in practice, the difficulties of creating a robot that is able to carry out more than a very narrow repertoire of actions has made most implementations either highly reactive or partially scripted.

Here, the issue is how to avoid the one-many expansion of discrete categories – for example planned actions – from symbolic form into non-symbolic form in a rigid mapping independent of sensing capabilities as in the classic Shakey-like approach. A view of reflection as a set of constraints on reactive systems allows this deterministic mapping to be avoided and replaced by contextual activation of groups of reactive processes. Thus in Barnes et al (1997) a planner mapped planning operator pre-conditions into sensory pre-conditions that could be detected by reactive processes and named the set of processes to be activated or deactivated on such pre-conditions being perceived.

It has the advantage that it allows reflection to overrule specific actions by an agent as well as to enable them. This supports a model of 'double appraisal' situations where for example hitting someone who has been offensive is overruled because of the way it will make the agent look to other people in a social group.

This approach does not require the initiative for reflection to come from an external task – it is equally compatible with the invocation of planning when reactive systems need it, whether to take advantage of sequencing capabilities or to deal with a situation in which reactive systems are not succeeding. However in this case it depends on the integration in the opposite direction, which is much more problematic and much less discussed.

3.2 From non-symbolic to symbolic

If constraints allow symbolic systems to impact non-symbolic ones without wholly determining them, pattern recognition is an obvious mechanism through which symbolic systems can discretise the dynamic variation of non-symbolic systems. Interpreting sub-symbolic configurations as either drives or emotions allows them to act as motivations within the symbolic systems and thus to initiate symbolic system activities such as planning.

This can be invoked from the symbolic process 'how do I feel?' 'what do I want to do?', but clearly can also, as just mentioned allow the non-symbolic processes to do the invoking, largely by associating high-levels of emotion with specific motivations. Here we think of a motivation as distinguished from goals by temporal scope and generality – thus dealing with hunger by eating is a motivation, while buying sandwiches or looking in the fridge would be examples of goals arising from this motivation. Within the symbolic systems then, emotion can be thought of as an integral part of goal management, and also as a heuristic weighting mechanism for large search spaces creating a search-oriented attentional focus.

These are aspects of the regulation of resources between symbolic and non-symbolic processes, and given that emotion is also heavily involved in sensory-motor coupling in non-symbolic systems, a very high level of emotion may truncate the symbolic process search space to the point where cognition almost halts.

Acknowledgements

The EU Humaine Network of Excellence on affective processing (http://emotion-research.net) provided an environment within which extended discussion has taken place on the topic of this paper: however it remains the view of the author not of the network as a whole. The European Commission partly funds Humaine but has no responsibility for the views advanced.

References

Anderson, J.R; D. Bothell, M. D. Byrne, S. Douglas, C. Lebiere, and Y. Qin. (2004) An integrated theory of the mind. Psychological Review, 111(4):1036–1060, 2004.

Arkin, R.A. and Balch,T. (1997) AuRA: Principles and Practice in Review. Journal of Experimental and Theoretical AI, 9(2), 1997.

Barnes, D.P; Ghanea-Hercock, R.A; Aylett, R.S. & Coddington, A.M. (1997) "Many hands make light work? An investigation into behaviourally controlled co-operant autonomous mobile robots". Proceedings, 1st International Conference on Autonomous Agents, Marina del Rey, Feb 1997 pp413-20

Brooks, R.A.(1986) A robust layered control system for a mobile robot. IEEE Journal of Robotics and Automation, RA-2(1):14–23, March 1986

Canamero, D. (1998) Modeling Motivations and Emotions as a Basis for Intelligent Behaviour, Proceedings of the First International Conference on Autonomous Agents eds. Johnson, W. L. and Hayes-Roth, B. ACM Press pp148—155, 1998

Clark, A. (1998). Embodied, situated, and distributed cognition. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (pp. 506-517). Malden, MA: Blackwell.

Gat E. (1997) On Three-Layer Architectures, in Kortenkamp D., Bonasso R.P., Murphy R. (eds.): Artificial Intelligence and Mobile Robots, MIT/AAAI Press, 1997.

Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. Psychological Review, 100, 68-90

Lazarus, R.S& Folkman, S. (1984). Stress, appraisal and coping. New York: Springer

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

Rosenbloom P., Laird J., and Newell A. eds. (1993) The Soar Papers: Research on Integrated Intelligence. MIT Press, Cambridge, Massachusetts, 1993.

Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), Appraisal processes in emotion. New York: Oxford University Press.

Velasquez, J.D. (1997) Modeling Emotions and Other Motivations in Synthetic Agents. Proceedings, 14th National Conference on AI, AAAI Press, pp10-16

Wilson, M. (2002) Six views of embodied cognition Psychonomic Bulletin & Review 2002, 9 (4), 625-636