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 neuro-physiologically 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 preconditions into sensory pre-conditions that could be detected by reactive processes and named the set of processes to be activated or deactivated on such preconditions 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 ad-

vantage 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. Internal attentional focus is a further example of the use of constraints as a modelling device: in this case non-symbolic systems may constrain what goals are considered by the symbolic systems, the extent or type of memory retrieval that can be carried out, or the extent or type of actions that can be considered in planning. One could also allow the non-symbolic systems to exercise control over the pattern-recognition mechanism required to deliver motivations to the symbolic systems so that a greater or lesser number of motivations are handled.

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, allowing us to model emotional flooding.

4. What cognitive model?

This approach to integrating non-symbolic and symbolic systems requires a model of a different type from ACT-R or SOAR, though these both include mechanisms which can be applied within the symbolic systems. Meanwhile, neuro-physiological models of the brain are still too fragmentary and small-scale to be useful for this purpose.

One interesting and possibly more useful approach is offered by the PSI model of Dorner (Dorner and Hille 95). This focuses on emotional modulation of perception, action-selection, planning and memory access. Emotions are not defined as explicit states but rather emerge from modulation of information processing and action selection. These modulators include arousal level (speed of information processing), resolution level (carefulness and attentiveness of behavior) and selection threshold (how easy it is for another motive to take over), and thus provide the type of interface discussed in the last section to non-symbolic systems. The model also applies two built-in motivators - level of competence and level of uncertainty, which are thought of as the degree of capability of coping with differing perspectives and the degree of predictability of the environment.

It would be an overstatement to suggest that this model can be applied without alteration to the discussion of this paper – not least because aspects are specified too broadly for straightforward implementation (Bach 2003) – but the role of emotion it outs forward does correspond in part to the suggestions made here.

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.

Bach, J (2003) The MicroPsi Agent Architecture. Proceedings of ICCM-5, International Conference on Cognitive Modeling Bamberg, Germany (pp. 15-20), 2003

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.

Dorner, D., Hille, K.(1995): Artificial souls: Motivated emotional robots. In: Proceedings of the International Conference on Systems, Man and Cybernetics. (1995) 3828– 3832

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

Embodiment *vs.* Memetics: Is Building a Human getting Easier?

Joanna Bryson*
*Artificial models of natural Intelligence
University of Bath, United Kingdom
jjb@cs.bath.ac.uk

Abstract

This heretical article suggests that while embodiment was key to evolving human culture, and clearly affects our thinking and word choice now (as do many things in our environment), our culture may have evolved to such a point that a purely memetic AI beast could pass the Turing test. Though making something just like a human would clearly require both embodiment *and* memetics, if we were forced to choose one or the other, memetics might actually be easier. This short paper argues this point, and discusses what it would take to move beyond current semantic priming results to a human-like agent.

1 Embodiment

There is no doubt that embodiment is a key part of human and animal intelligence. Many of the behaviours attributed to intelligence are in fact a simple physical consequence of an animal's skeletal and muscular constraints (Port and van Gelder, 1995; Paul, 2004). Taking a learning or planning perspective, the body can be considered as bias, constraint or (in Bayesian terms) a prior for both perception and action which facilitates an animal's search for appropriate behaviour (Bryson, 2001).

This influence continues, arguably through all stages of reasoning (Chrisley and Ziemke, 2002; Lakoff and Johnson, 1999) but certainly at least sometimes to the level of semantics. For example, Glenberg and Kaschak (2002) have demonstrated the *action-sentence compatibility effect*. That is, subjects take longer to signal comprehension of a sentence with a gesture in the opposite direction as the motion indicated in the sentence than if the motion and sentence are compatible. For example, given a joystick to signal an understanding of 'open the drawer', it is easier to signal comprehension by pulling the joystick towards you than pushing it away. Boroditsky and Ramscar (2002) have shown that comprehension of ambiguous temporal events are strongly influenced by the hearer's physical situation with respect to current or imagined tasks and journeys.

These sorts of advances have lead some to suggest that the reason for the to-date rather unimpressive state of natural language comprehension and production in Artificially Intelligent (AI) systems is a consequence of their lack of embodiment (Harnad, 1990; Brooks and Stein, 2004; Roy and Reiter, 2005). The suggestion is that, in order to be meaningful, concepts must be grounded in the elements of intelligence that produce either action or perception salient to action.

The pursuit of embodied AI has lead us to understand resource-bounded reasoning which explains apparently suboptimal or inconsistent decision-making in humans (Chapman, 1987). It has also helped us to understand the extent to which agents can rely on the external world as a resource for cognition — that perception can replace or at least supplement long-term memory, reasoning and model building (Brooks, 1991; Clark, 1997; Ballard et al., 1997; Clark and Chalmers, 1998). However, despite impressive advances in the state of artificial embodiment (e.g. Chernova and Veloso, 2004; Schaal et al., 2003; Kortenkamp et al., 1998), there have been no clear examples of artificial natural language systems improved by embodiment.

I believe this is because embodiment, while necessary, is not a sufficient explanation of semantics. We *have* seen neat examples of the embodied acquisition of limited semantic systems (e.g Steels and Vogt, 1997; Steels and Kaplan, 1999; Roy, 1999; Billard and Dautenhahn, 2000; Sidnera et al., 2005). These systems show not only that semantics can be established between embodied agents, but also the relation between the developed lexicon and the agents' physical plants and perception. However, such examples give us little idea of how words like INFIN-

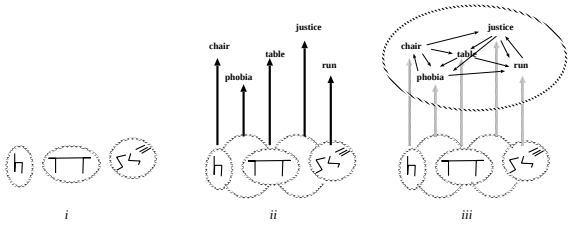
ITY, SOCIAL or REPRESENT might be represented. Further, they do not show the *necessity* of physical embodiment for a human-like level of comprehension of natural language semantics. On the other hand, it is possible that the semantic system underlying abstract words such as 'justice' may also be sufficient for terms originally referencing physical reality.

I do not contest the importance of understanding embodiment to understanding human intelligence as a whole. I *do* contest one of the prominent claims of the embodied intelligence movement — that embodiment is the only means of grounding semantics (Brooks and Stein, 2004). Roy and Reiter (2005) in fact *define* the term GROUNDED as 'embodied', which might be fine (compare with Harnad, 1990) if GROUNDED hadn't also come to be synonymous with MEANINGFUL. The central claim of this paper is that while embodiment may have been the origin of most semantic meaning, it is no longer the only source for accessing a great deal of it. Further, some words (including their meanings) may have evolved more or less *independently* of grounded experience.

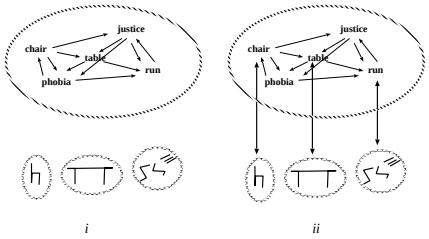
Figure 1: A two-dimensional projection of a semantic space, after Lowe (1997). The target words are taken from the experiments of Moss et al. (1995). Additional information on nearness is contained in the weights between locations in the 2-D space.

2 Memetics' role in development

We now know that humans could very well develop an interconnected web of words *independently* of the process of developing grounded concepts (See Figure 1). Grounding then becomes a process of associating *some* of these statistically acquired terms with embodiment-based concepts. Thus children can learn and even use the word JUSTICE without a referent. Gradually as they gain experience of complexity of conflicting social goals and notions of fairness develop a richer notion of both what the word means and how and when both the word and the grounded concept can be used in furthering their goals. But even before that, a relatively naive reference to the term could well accomplish something.



(a) Deacon's Theory



(b) Bryson's Hypothesis

Figure 2: In Deacon's theory, first concepts are learned *a(i)*, then labels for these concepts *a(ii)*, then a symbolic network somewhat like semantics *a(iii)*. I propose instead that grounded concepts and semantics are learned in parallel *b(i)*, then some semantic terms become understood *b(ii)*.

I want to be clear here: in my model, humans still acquire associations between these two representations, just as in the (Deacon, 1997) model that inspired it. What's different is the ordering. In my model, lexical semantics is learned in parallel with embodied categories and expressed behaviour. Subsequently, *some* words become grounded as connections are formed between the two representations (see Figure 2). Nevertheless, this model also leaves the door open to true memetics — perhaps *justice* is an evolved concept that has fundamental impact in our culture and institutions without anyone truly 'understanding' it in any deeply grounded way.

3 Building someone cheaply

The previous sections have talked about what composes current human intelligence. But let's change

the topic now to trying to build someone capable of a decent conversation, even of coming up with the occasional good idea apparently on their own. Someone that could pass the Turing test if you chatted to them at the bus stop for 20 minutes, assuming you couldn't see what they looked like.

Figure 2(b) implies that memetics can only give us half the story, but this is wrong on two counts. First, I do not think embodiment is necessary for concept formation. We develop concept for justice to go along with the label, and I expect this same process could go on for quite a lot of other words.

It is possible that we'd need to provide some preformed seed concepts to get the system rolling. This may be necessary for two reasons:

- Purely for bootstrapping the learning system. It's possible that all concepts formed from memetic experience *are* formed partially in relation to or contrast with established concepts, so our poor disembodied mind might need some good, rich precocial concepts to get started (see further Sloman and Chappell, 2005).

- Because our memetic culture might not carry knowledge *everyone* gets for free. Given that a huge amount of what it means to be human is embedded in our semantic assumptions, it is possible that the brain can fill in the gaps. Stroke and lesion patients sometimes recover enormous functionality deficits if they still have enough of their brain intact that they can use the existing bits. If sufficiently stimulated (the main point of therapy), these surviving parts can provide enough information about what the *missing* information should look like that the individual may recover some lost skills. However, it is possible that some concepts are so incredibly universal to human experience that there just isn't enough information in the culture to reconstruct them.

But in general, I still think it might be easier to program some concepts (or proto-concepts) by hand than to build and maintain a robot that is sufficiently robust and long-lived, and has a sufficiently rich motor and sensor capacities, that it could do a better job of learning such concepts from its embodied experience.

But the other reason Figure 2(b) is not showing us that memetics is half the story is because a very important part of the story is left out. Even if we had an agent with all the knowledge of a human (or say we had a search engine with all the knowledge any human has ever put on the web), if all that agent ever does is *learns*, it isn't very human-like. To build someone, we need not only basic capacities for perception and action (which in the meme machine's case is just language in and out) but also motivation and action selection (Bryson, 2001). Even the cheapest human-like agent would need to have a set of prioritised goals, probably some sort of emotional / temporally dependent state to oscillate appropriately between priorities, and a set of plans (in this case, syntax and dialog patterns) to order its actions in such a way that it can achieve those goals.

Fortunately, nearly everyone in AI who builds agents (even roboticists) builds this part of the system in software, so again, there is no driving reason to bring in embodiment. Of course, without a body these goals would have to be purely intellectual or social (find out about you, talk about me, figure out how to use new words appropriately) — many but not all human goals would be inaccessible to a disembodied meme machine.

4 Conclusion

This short paper argues that although embodiment is clearly involved in human thought and language usage, we have consequently evolved and developed a culture permeated with the knowledge we derive in our embodied existence, and as such a cheap but reasonably entertaining agent might be built with no embodiment at all. Of course AI has tried to do this for several decades, but I think they have come at it the wrong way, focusing on logic-based reasoning too much and case- or template-based reasoning too little. Humans however are imitation and case-learning machines — to such an extent that some of our wisdom / common sense may well have evolved memetically rather than ever having been fully understood or reasoned about by anyone.

Acknowledgements

I'd like to thank Push Singh for getting me to think about common sense, even though we never agreed what it meant. I'd like to thank Will Lowe for teaching me about statistical semantics, even though we still don't agree what the implications are.

References

Dana H. Ballard, Mary M. Hayhoe, Polly K. Pook, and Rajesh P. N. Rao. Deictic codes for the embodiment of cognition. *Brain and Behavioral Sciences*, 20(4), December 1997.

A. Billard and K. Dautenhahn. Experiments in social robotics: grounding and use of communication in autonomous agents. *Adap. Behav.*, 7(3/4), 2000.

Lera Boroditsky and Michael Ramscar. The roles of body and mind in abstract thought. *Psychological Science*, 13(2):185–188, 2002.

Rodney A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.

Rodney A. Brooks and Lynn A. Stein. Building brains for bodies. *Aut. Robots*, 1(1):7–25, 2004.

Joanna J. Bryson. *Intelligence by Design: Principles of Modularity and Coordination for Engineering Complex Adaptive Agents*. PhD thesis, MIT, Department of EECS, Cambridge, MA, June 2001. AI Technical Report 2001-003.

David Chapman. Planning for conjunctive goals. *Artificial Intelligence*, 32:333–378, 1987.

S. Chernova and M. Veloso. Learning and using models of kicking motions for legged robots. In *The Proc. of ICRA-2004*, New Orleans, May 2004.

Ronald Chrisley and Tom Ziemke. Embodiment. In *Encyclopedia of Cognitive Science*, pages 1102–1108. Macmillan, 2002.

Andy Clark. *Being There: Putting Brain, Body and World Together Again*. MIT Press, Cambridge, MA, 1997.

Andy Clark and David Chalmers. The extended mind. *Analysis*, 58(1):7–19, 1998.

Terrence Deacon. *The Symbolic Species: The coevolution of language and the human brain*. W. W. Norton & Company, New York, 1997.

Arthur M. Glenberg and Michael P. Kaschak. Grounding language in action. *Psychonomic Bulletin & Review*, 9(3):558–565, 2002.

Stevan Harnad. The symbol grounding problem. *Physica D*, 42(1–3):335–346, 1990.

David Kortenkamp, R. Peter Bonasso, and Robin Murphy, editors. *Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems*. MIT Press, Cambridge, MA, 1998.

George Lakoff and Mark Johnson. *Philosophy in the Flesh: The embodied mind and its challenge to Western thought*. Basic Books, New York, 1999.

Will Lowe. Semantic representation and priming in a self-organizing lexicon. In J. A. Bullinaria, D. W. Glasspool, and G. Houghton, editors, *Proc. of the Fourth Neural Computation and Psychology Workshop: Connectionist Representations (NCPW4)*, pages 227–239, London, 1997. Springer.

H. E. Moss, R. K. Ostrin, L. K. Tyler, and W. D. Marslen-Wilson. Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21:863–883, 1995.

Chandana Paul. *Investigation of Morphology and Control in Biped Locomotion*. PhD thesis, Department of Computer Science, University of Zurich, Switzerland, 2004.

Robert F. Port and Timothy van Gelder, editors. *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press, Cambridge, MA, 1995.

D. Roy and E. Reiter. Connecting language to the world. *Art. Intel.*, 167(1–2):1–12, 2005.

Deb Kumar Roy. *Learning from Sights and Sounds: A Computational Model*. PhD thesis, MIT, Media Laboratory, sep 1999.

Stefan Schaal, Auke Ijspeert, and Aude Billard. Computational approaches to motor learning by imitation. *Philosophical Transactions of the Royal Society of London, B*, 358(1431):537–547, 2003.

Candace L. Sidnera, Christopher Leea, Cory D. Kiddb, Neal Lesha, and Charles Richa. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1–2):140–164, 2005.

Aaron Sloman and Jackie Chappell. The altricial-precocial spectrum for robots. In L. Pack Kaelbling and A. Saffiotti, editor, *Proceedings of the Nineteenth International Conference on Artificial Intelligence (IJCAI-05)*, pages 1187–1194, Edinburgh, August 2005. Professional Book Center.

Luc Steels and Frederic Kaplan. Bootstrapping grounded word semantics. In T. Briscoe, editor, *Linguistic evolution through language acquisition: formal and computational models*. Cambridge University Press., 1999.

Luc Steels and Paul Vogt. Grounding adaptive language games in robotic agents. In C. Husbands and I. Harvey, editors, *Proceedings of the Fourth European Conference on Artificial Life (ECAL97)*, London, 1997. MIT Press.

Requirements & Designs: Asking Scientific Questions About Architectures

Nick Hawes, Aaron Sloman and Jeremy Wyatt
School of Computer Science
University of Birmingham
{n.a.hawes,a.sloman,j.l.wyatt}@cs.bham.ac.uk

Abstract

This paper discusses our views on the future of the field of cognitive architectures, and how the scientific questions that define it should be addressed. We also report on a set of requirements, and a related architecture design, that we are currently investigating as part of the CoSy project.

1 What Are Architectures?

The first problem we face as researchers in the field of cognitive architectures is defining exactly what we are studying. This is important because the term "architecture" is so widely used in modern technological fields. An agent's cognitive architecture defines the information-processing components within the "mind" of the agent, and how these components are structured in relation to each other. Also, there is a close link between architectures and the mechanisms and representations used within them (where representations can be of many kinds with many functions). Langley and Laird (2002) describe a cognitive architecture as including "those aspects of a cognitive agent that are constant over time and across different application domains". We extend this to explicitly allow architectures to change over time, either by changing connection patterns, or altering the components present. Excluding such changes from the study of architectures may prevent the discussion of the development of architectures for altricial information-processing systems (Sloman and Chappell, 2005).

2 Related Work

Historically, most research into cognitive architectures has been based around specific architectures such as ACT-R, SOAR, and ICARUS (for a summary see (Langley and Laird, 2002)). A lot of work has been devoted to developing iterations of, and extensions to, these architectures, but very little work has been done to compare them, either to each other, or to other possible design options for cognitive architectures. In other words, little work has been done on the general science of designing and building cognitive systems. Anderson and Lebiere (2003) have recently attempted to address this by comparing two different architectures for human cognition on a set of requirements.

3 Architectures & Science

To advance the science of cognitive systems we need two related things: clear, testable questions to ask, and a methodology for asking these questions. The methodology we support is one of studying the space of possible *niches* and *designs* for architectures, rather than single, isolated, designs (Sloman, 1998b). Within such a framework, scientific questions can be asked about how a range of architecture designs relate to sets of requirements, and the manner in which particular designs satisfy particular niches. Without reference to the requirements they were designed to satisfy, architectures can only be evaluated in a conceptual vacuum.

The scientific questions we choose to ask about the space of possible architecture designs should ideally provide information on general capabilities of architectures given a set of requirements. This information may not be particularly useful if it is just a laundry list of instructions for developing a particular architecture for a particular application domain. It will be more useful if we can characterise the space of design options related to a set of requirements, so that future designers can be aware of how the choices they make will affect the overall behaviour of an agent. The questions asked about architectures can be motivated by many sources of information, including competing architecture designs intended for similar niches.

In order for questions about architectures, and their answers, to be interpreted in the same way

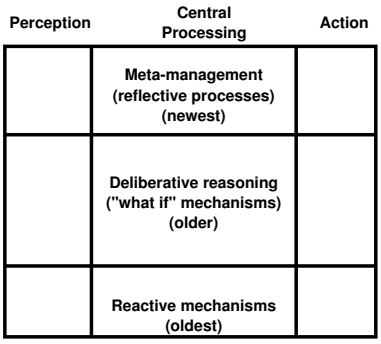


Figure 1: The CogAff Architecture Schema.

by researchers across many disciplines, we need to establish a common vocabulary for the design of information-processing architectures. As a step towards this, we use the CogAff schema, depicted in Figure 1, as an incomplete first draft of an ontology for comparing architectures. (Sloman, 2001). The schema is intended to support broad, two-dimensional, design- and implementation-neutral characterisations of architectural components, based on information-processing style and purpose. If an architecture is described using the schema, then it becomes easier to compare it directly to other architectures described in this way. This will allow differing architectures to be compared along similar lines, even if they initially appear to have little in common.

4 A Minimal Scenario

For our current research as part of the CoSy project[1], we are working from requirements for a pre-linguistic robot that has basic manipulative abilities, and is able to explore both its world and its own functionality. At a later date we will extend this to add requirements for linguistic abilities. We are approaching the problem in this way because we believe that a foundation of action competence is necessary to provide semantics for language. These requirements come from the CoSy *PlayMate scenario*, in which a robot and a human interact with a tabletop of objects to perform various tasks[2].

In our initial work on this scenario we will focus on the requirements related to the architectural elements necessary to support the integration of simple manipulative abilities with a visual system that supports the recognition of basic physical affordances from 3D structure. We see this as the absolute minimum system for the start of an exploration of PlayMate-like issues in an implemented system [3].

Our requirements analysis has led to the design of a prototype architecture which we believe will satisfy the niche they specify. Space restrictions do not permit a full description of the architecture, but in brief the architecture features multiple concurrently active components, including: a motive generator; information stores for currently active motives, general concepts, and instances of the general concepts; a general-purpose deliberative system; a fast global alarm system; a plan execution system; management and meta-management components; a spreading activation substrate; and closely coupled vision and manipulation sub-architectures.

The high-level design for this architecture is presented in Figure 2, and is in part inspired by our previous work on information-processing architectures (e.g. (Beaudoin, 1994; Sloman, 1998a; Hawes, 2004)). Although this design clearly separates functionality into components, these components will be tightly integrated at various levels of abstraction. For example, to enable visual servoing for manipulation (e.g. (Kragic and Christensen, 2003)), visual and proprioceptive perception of the movement of the robot's arm in space must be closely coupled with the instructions sent to the arm's movement controller.

The information-processing behaviour of the architecture is driven by motives, which are generated in response to environmental or informational events. We will allow humans to generate environmental events using a pointing device. The agent will interpret the gestures made with this device as direct indications of desired future states, rather than intentional acts (thus temporarily side-stepping some of the problems of situated human-robot interaction). Generated motives will be added to a collection of current motives, and further reasoning may be necessary if conflicts occur between motives. The deliberative system will produce action plans from motives, and these plans will be turned into arm commands by the plan execution system. This process will be observed at a high level by a meta-management system, and at a low level by an alarm system. The meta-management system may reconfigure the agent's processing strategies if the situation requires it (e.g. by altering the priorities associated with motives). The global alarm system will provide fast changes in behaviour to handle sudden, or particularly critical, situations.

[1] See http://www.cognitivesystems.org for more information.

[2] More information about the PlayMate is available at http://www.cs.bham.ac.uk/research/projects/cosy/pm.html.

[3] Our work on requirements from the PlayMate scenario is presented roughly at http://snipurl.com/cosy_playmate.

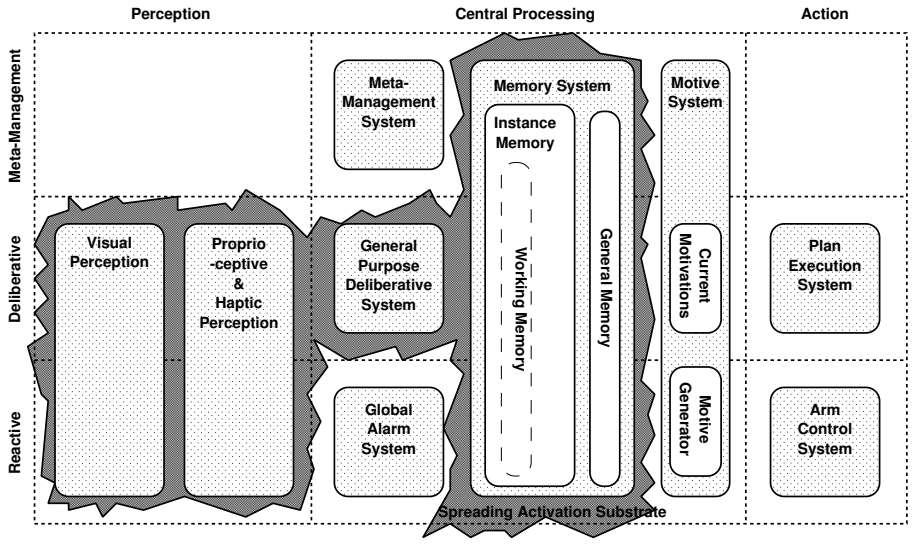


Figure 2: The Proposed Architecture.

Although a spreading activation substrate is featured in the design, we are not currently committed to its inclusion in the final system. Instead we are interested in the kinds of behaviour that such a design choice will facilitate. Information across the architecture may need to be connected to related information, and such an approach may allow the agent to exploit the structure of such connections by spreading activation, which may be based on co-occurrence, recency or saliency. We are also interested in investigating how to combine distributed approaches to representing and processing information with more localised approaches, and what design options this provides. Such a combination of processing approaches can be seen in the work on the MicroPsi agent architecture (Bach, 2005).

5 Architecture Evaluation

The question of whether, and why, our proposed architecture is appropriate for an agent with PlayMate-like requirements is quite hard to formulate in a way that is directly answerable. Instead, we must use our requirements analysis, and our previous experiences of designing architectures for such agents, to derive precise questions and suggest testable hypotheses from this. The following paragraphs present specific questions we could ask about the architecture, and many other architectures.

How can information exchange between architectural components be controlled, and what trade-offs are apparent? For example, should information from visual perception be pushed into a central repository, or should task appropriate information be pulled from vision when necessary, or should this vary depending on the system's information state, goals, the performance characteristics of subsystems, etc.?

What are the relative merits of symbolic and subsymbolic (e.g. spreading activation) processing methods when applied to collating information across the entire architecture? The proposed spreading activation substrate could interact with various processes and ontologies, and record how information is manipulated. Alternatively, this could be implemented as a central process that must be notified by other processes when certain operations occur. These different approaches could be compared on their proficiency at managing large volumes of multi-modal information, their ability to identify changes of context across the architecture, the difficulty of integrating them with other processes, or the ease with which they facilitate other operations (such as attentional control).

To what degree should the architecture encapsulate modality-specific and process-specific information within the components that are directly concerned with it? Cross-modal application of the early processing results can increase accuracy and efficiency in some processes (c.f. (Roy and Mukherjee, 2005)). In other cases information may be irrelevant, and attempts to apply it across modalities may have the opposite effect whilst increasing the computational load on an architecture. We could explore this notion more generally by asking what types of information should, and should not, be made available by architectural components whilst they are processing

it, and what use other architectural components could make of such information.

Given the types of information the architecture will be processing, what are the advantages and disadvantages of having a single central representation into which all information is translated? How do these advantages and disadvantages change when additional processes are added into the architecture?

What role does a global alarm mechanism have in PlayMate-like domains, how much information should it have access to, and how much control should it have? For example, an alarm mechanism may have access to all the information in the architecture and risk being swamped by data, or it may have access to limited information streams and risk being irrelevant in many situations.

Does the architecture need some global method for producing serial behaviour from its many concurrently active components, or will such behaviour just emerge from appropriate inter-component interactions? Approaches to component control include a single central component activating other components, a control cycle in which activity is passed between a small number of components, and other variations on this. Are there particular behaviours that are not achievable by an agent with this kind of control, and only achievable by an agent we decentralised control, or vice versa? If such trade-offs exist, how are the relevant to PlayMate-like scenarios?

Given the range of possible goals that will need to be present in the whole system, how should these goals be distributed across its architecture, and how does this distribution affect the range of behaviours that the system can display?

Obviously there are many other questions we could ask about the architecture, such as whether it will facilitate the implementation of mechanisms for acquiring and using orthogonal recombinable competences[4]. The process of designing and implementing architectures to meet a set of requirements involves the regular re-evaluation of the requirements in light of new developments. Inevitably, this means that other questions will be considered, and the above ones reconsidered, as the research progresses.

Acknowledgements

The research reported on in this paper was supported by the EU FP6 IST Cognitive Systems Integrated project Cognitive Systems for Cognitive Assistants "CoSy" FP6-004250-IP.

[4] http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0601

References

John R. Anderson and Christian Lebiere. The Newell test for a theory of cognition. *Behavioral & Brain Science*, (26):587–637, 2003.

Joscha Bach. Representations for a complex world: Combining distributed and localist representations for learning and planning. In Stefan Wermter, Günther Palm, and Mark Elshaw, editors, *Biomimetic Neural Learning for Intelligent Robots*, volume 3575 of *Lecture Notes in Computer Science*, pages 265–280. Springer, 2005.

Luc P. Beaudoin. *Goal Processing In Autonomous Agents*. PhD thesis, School of Computer Science, The University of Birmingham, 1994.

Nick Hawes. *Anytime Deliberation for Computer Game Agents*. PhD thesis, School of Computer Science, University of Birmingham, 2004.

Danica Kragic and Henrik I. Christensen. A framework for visual servoing. In Markus Vincze and James L. Crowley, editors, *ICVS-03*, volume 2626 of *LNCS*. Springer Verlag, March 2003.

Pat Langley and John E. Laird. Cognitive architectures: Research issues and challenges. Technical report, Institute for the Study of Learning and Expertise, Palo Alto, CA, 2002.

Deb Roy and Niloy Mukherjee. Towards situated speech understanding: Visual context priming of language models. *Computer Speech and Language*, 19(2):227–248, 2005.

Aaron Sloman. Varieties of affect and the cogaff architecture schema. In *Proceedings of the AISB'01 Symposium on Emotion, Cognition and Affective Computing*, pages 1–10, 2001.

Aaron Sloman. Damasio, descartes, alarms and meta-management. In *Proceedings of IEEE Conference on Systems, Man, and Cybernetics*, pages 2652–2657, 1998a.

Aaron Sloman. The "semantics" of evolution: Trajectories and trade-offs in design space and niche space. In Helder Coelho, editor, *Progress in Artificial Intelligence, 6th Iberoamerican Conference on AI (IBERAMIA)*, pages 27–38. Springer, Lecture Notes in Artificial Intelligence, Lisbon, October 1998b.

Aaron Sloman and Jackie Chappell. The Altricial-Precocial Spectrum for Robots. In *Proceedings IJCAI'05*, pages 1187–1194, Edinburgh, 2005.

Integration and Decomposition in Cognitive Architecture

John Knapman*
*School of Computer Science
The University of Birmingham, UK
J.M.Knapman@cs.bham.ac.uk

Abstract

Given the limitations of human researchers' minds, it is necessary to decompose systems and then address the problem of how to integrate at some level of abstraction. Connectionism and numerical methods need to be combined with symbolic processing, with the emphasis on scaling to large numbers of competencies and knowledge sources and to large state spaces. A proposal is briefly outlined that uses overlapping oscillations in a 3-D grid to address disparate problems. Two selected problems are the use of analogy in commercial software evolution and the analysis of medical images.

1 Introduction

Debates about cognitive architecture often deal with choices between rival techniques and modalities. By contrast, Minsky, Singh and Sloman (2004) emphasise the importance of integrating components of differing kinds.

A standard method in designing complex systems is to decompose into components and then connect them, simply because it is impossible for individual team members to hold all the detail in their heads. Because of such limitations, there are those who think that it is not possible for human beings to design systems that exhibit general intelligence comparable to their own (Rees, 2003). Even after fifty years of research, such counsels of despair are premature until we have explored more of the possibilities. The prizes are great, in part for a better understanding of ourselves and the enlightenment it will bring in the tradition of the dazzling human progress since the Renaissance in many fields, including medicine and technology. There are also strong commercial benefits in being able to build smarter systems that can undertake dangerous or unpopular work.

To make substantial progress, there are lessons to learn from work on the architecture of computer systems generally (Bass, Clements and Kazman, 2003), which informs us that a key responsibility of the architect is to define how the components fit together and interface with each other. This assumes that there is a degree of uniformity in the style of the components, so that their interfaces can be defined in a form that is commonly understood by the people working on them. Interfaces are then specified in terms, firstly of timing and flow of control (serial or parallel, hierarchical or autonomous, event driven or scheduled, for example), and secondly in terms of the format of data flowing between components.

A less well documented but nevertheless well known experience is that a small team of dedicated experts can achieve wonders compared with large teams of people with mixed ability. A small team of four highly experienced people can often produce efficient, reliable and timely systems that solve problems most effectively. By contrast, many research (and development) teams consist of one experienced person, who is very busy, and several bright but inexpert assistants. To get an expert team together to work full time and hands on for a period of years is expensive and disrupts other activities, but the results can be very exciting.

Even with such a promising kind of team organisation, interfaces have to be defined. In cognitive architecture, components will be of differing kinds. Connectionist methods that emphasise learning and convergence must somehow be combined effectively with symbolic processing. Sun (2002) shows how symbolic rules may be inferred from state transitions in a connectionist network, but that is only one of several ways in which interactions could take place. There are other promising numerical methods, such as KDDA, which has been applied successfully to face recognition (Wu, Kittler, Yang, Messer and Wang, 2004).

The strong emphasis on learning and learnability in connectionist methods is carried over to symbolic rules in Sun's method, but there are other benefits

from connectionism, such as approximate matching and graceful degradation, that need to be exploited in a combined system. Benefits like these accrue not just from connectionism but from other soft computing paradigms (see Barnden (1994) for a good discussion in the context of the study of analogy).

One method for linking components that use quite different representations from differing standpoints is to use numerical factors, whether these are interpreted as probabilities, fuzzy or other uncertainty values, ad hoc weights, or arrays of affective measures (e.g., motivation (Coddington and Luck, 2003), duty, elegance). Such techniques can also help to reduce search spaces and large state spaces, although they may sometimes smack of heuristics of the kind favoured by researchers in the 1970s.

2 Abstraction

Many different kinds of abstraction have been identified. In the field of computer system design, there are methods with well-defined levels of abstraction, the most concrete being a set of programs that implements a design. In the B and Z methods (Bert, Bowen, King and Waldén, 2003) there are formal proof procedures to show that a more concrete specification is a refinement of one that is more abstract.

Less formal methods, such as the Unified Software Development Method (Jacobson, Booch and Rumbaugh, 1999), based on the Unified Modeling Language (UML) still have the idea that a high-level design (as an object diagram) can be refined progressively until implementation. Beyond the writing of the programs, a more complete analysis sees the programs and their specifications as abstractions of their actual performance, as recorded in traces while they execute. The diagrams are found to be better for communicating between designers and developers than either formal-language statements or natural-language descriptions.

Even though there is a clear definition of abstraction in these examples, users often feel intuitively that the diagrams are less abstract than the text. Such an intuition throws up the difficulty of defining abstraction in a general way. It seems to depend on fitness for purpose as well as brevity and omission of detail.

In the case of a story told in a natural language, a synopsis is more abstract in the latter sense. In the case of scientific papers, the "abstract" is intended to help readers decide whether to read the main paper. The "management summary" of a business report allows busy senior people to understand enough to be able to trust and defend their subordinates' recommendations.

A reference to "the report", "the story" or "the paper" is clearly more abstract than having to repeat the content. Generally, referring to something verbally or symbolically is brief, enables it to be dealt with in its absence, and permits economy of thought, i.e., it leaves mental room, as it were, for other concepts to be introduced and related to it.

Uncertainty representations provide another form of abstraction, and one that is particularly easy to formalise (Baldwin, Martin and Pilsworth, 1995). A fuzzy value can stand for "the hand-written letter that is probably a k", or "the car that looked like a Jaguar". Such representations are the clearest candidate for a form of abstraction that can be used to enable disparate sources of knowledge to be combined effectively and rapidly. For example, a moving picture, some sounds, previous experience of steam trains and expectation that the Flying Scotsman will pass through the station at 11:00 a.m. combine so as to interpret the distant approach of the train while ignoring most other details.

3 Large State Spaces

Problems often entail large numbers of possibilities. Although both abstractions and numerical methods can help to reduce the possibilities, there are frequently cases where many possible states must be carried forward before higher abstractions can be used to eliminate some. Sometimes called the "AI-complete" problem, there have been hopes that quantum information processing could address it. However, the breakthrough has not so far come. Apart from special cases where the data has cyclic properties (as in modal arithmetic for code breaking), the main benefit is that large state spaces (having N states) can be searched in time proportional to $\sqrt{N}$ instead of $N/2$. This can be worthwhile in some cases (e.g., reducing 500 billion steps to one million), but must await the availability of suitable hardware. The programming skills required are rather daunting.

Some people have suggested that the unstable periodic orbits (UPOs) of chaotic oscillators (Crook and Scheper, 2001) can represent potentially infinitely many things. It has been observed that the signals in a brain appear to be either random or chaotic before settling rapidly to a coherent state (Tsui and Jones, 1999). In theory, a random signal contains all possible frequencies, but a practical random signal is limited to the bandwidth of the channel and takes a long time to distinguish from a sum of many overlapping signals of different frequencies, an idea taken up in the Proposal section below. A chaotic signal is somewhere in between, and is also indistinguishable in practice (Gammaitoni, Hänggi, Jung and Marchesoni, 1998) from the other two.

4 Requirements

A test bed for the exploration of ideas is needed that supports the following requirements:

1. Combining modalities, especially connectionist, symbolic and affective
2. Combining competencies, including (but not limited to) analogy and structure matching, vision and formal language interpretation
3. Abstraction, with emphasis on combining knowledge sources
4. Scaling to large state spaces, particularly exploring efficient forms of parallelism
5. Scaling to large numbers of knowledge sources

5 Proposal

To explore these issues and requirements, one approach is to design and simulate a programmable signal-processing network capable of both symbolic and connectionist processing, with commitment to as few preconceptions as possible.

A flexible structure is proposed that will allow for the interplay of several loose decompositions. We will allow an element to belong to several groupings, which can be nested. It must identify with which grouping any particular communication is associated. With such a protocol, elements are not confined to a hierarchical organisation, but hierarchies are still possible, and communication channels are more manageable than in a complete free for all.

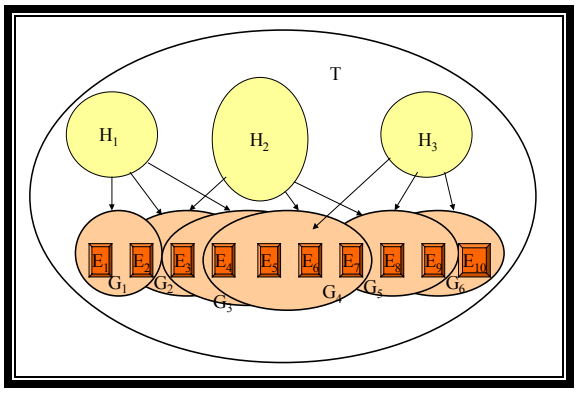


Figure 1: Illustration of Flexible Structuring – containment is shown by nesting or with arrows

A message between two elements has to conform to the representational format of their common grouping at some level of nesting. Then, if the prime method of communication is broadcasting, for example, broadcasting (both sending and receiving) would only take place within a grouping and would follow the representational convention for that grouping. In the illustration of Figure 1, E_1 and E_2 can communicate by the conventions of their containing groupings G_1, H_1 and T. E_1 can communicate with E_3, E_4, E_5 and E_6 by the conventions of H_1 and T, because these four are in G_3, which is in H_1. E_1 can only communicate with E_{10} by the conventions of T or by some form of relay through intermediate subordinate groupings.

Within such a general framework, it is proposed to exploit the parallelism inherent in modulated overlapping signals of many frequencies embedded in a 3-D grid. One such model is described by Coward (2004) as an attempt to emulate aspects of the architecture of the brain. Uncertainty models, including Bayesian nets (Pearl, 1988) and fuzzy logic, are to be accommodated. It must be suitable for both learning and programmed behaviour. Programming provides the flexibility to explore challenging applications, and it allows certain abilities to be built in. For other capabilities, there should always be at least the possibility that the symbols and mappings defined could be acquired through experience or by an indirect process such as analogy, deduction or abstraction.

There must be support for widely differing data types, particularly for image processing and formal language processing. The particular problems under consideration are in two domains:

1. The application of analogical reasoning to commercial software evolution
2. Analysis of medical images

It must be reasonably clear how the framework may be extended to other modalities, e.g., movement control for vehicles or robots.

A particularly elegant form of programming is functional programming, where every program is a transformation from the input parameters to an output value. A function is defined in mathematics as a set of ordered pairs of input and output. Most programmers think of procedures as behaving algorithmically, but the alternative definition sees a function as a transformation from input to output via memory lookup. The effect of a procedure giving multiple results can be achieved by multiple functions that take the same parameters. Functions of several variables can be decomposed into (single valued) functions of pairs of variables.

This view of a function is particularly convenient for the parallel processing of overlapping signals of many frequencies. Such signals can encode ranges or uncertainty in data but can also represent patterns of input, such as image intensities or characters in text.

Frequency can be used to encode data values, with amplitude representing strength. A transformation converts from an input frequency to an output frequency and may adjust the weight. Thus it may transform one pattern to another or may perform simultaneous logical operations on parallel data. The transformation of patterns using weight adjustment

can be equivalent to that performed in connectionist networks, with a natural mechanism for incorporating learning. Logical operations may not need to perform weight adjustment, but conjunction and disjunction between two sets of signals are needed. Disjunction can be achieved by summing. Conjunction requires frequency matching.

Some programming models require a global state, for example the query and subgoals in PROLOG, whereas object-oriented (OO) programming localises the state information in separate objects.

A 3-D grid can contain many channels, and the signals can persist for some time, somewhat in the manner of objects in OO programming, though with different dynamics. Together they may encode a very large state space, even though there is a limit to the number of signals that can be carried on one channel because of the constraints of bandwidth. They are well suited to representing data structures such as parse trees or image region classifications.

The interactions are different from those in quantum computing; there, each state is completely integrated but is processed in isolation from all others. In the 3-D grid, a state is distributed, but information from many states can be mixed.

6 Conclusion

After half a century's work, there remain many ideas that can be explored, particularly in the arena of integration. It would be most exciting to see what a dedicated team of four or five experts able to work full time for a period of years could achieve. However, the challenge remains of discovering a small enough set of representations that are general enough for interfacing between the kinds and styles of components identified but are nevertheless succinct enough to be computationally tractable.

Acknowledgements

I'm grateful for valuable discussions with Aaron Sloman.

References

J.F. Baldwin, T.P. Martin and B.W. Pilsworth. *Fril – Fuzzy and Evidential Reasoning in Artificial Intelligence*, Research Studies Press, Taunton, UK and Wiley, New York, 1995, p.54

John Barnden. On Using Analogy to Reconcile Connections and Symbols, in Levine, D.S. and Aparicio, M. (eds.) *Neural Networks for Knowledge Representation and Inference*, Hillsdale, New Jersey: Erlbaum 27-64, 1994

Didier Bert, Jonathan Bowen, Steve King, Marina Waldén (eds.) *ZB 2003: Formal Specification and Development in Z and B: Third International Conference of B and Z Users*, Turku, Finland, June 4-6, LNCS 2651, Springer, 2003

Len Bass, Paul Clements and Rick Kazman. *Software Architecture in Practice (2nd edition)*, Addison-Wesley, 2003

Alexandra Coddington and Michael Luck. Towards Motivation-based Plan Evaluation, in *Proceedings of the Sixteenth International FLAIRS Conference (FLAIRS '03)*, Russell, I. and Haller, S. (eds.), AAAI Press, 298-302, 2003

L. Andrew Coward. The Recommendation Architecture Model for Human Cognition. *Proceedings of the Conference on Brain Inspired Cognitive Systems*, University of Stirling, Scotland, 2004

Nigel Crook and Tjeerd olde Scheper. A Novel Chaotic Neural Network Architecture, *ESANN'2001 proceedings – European Symposium on Artificial Neural Networks*, Bruges, Belgium, 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 295-300, 2001

Luca Gammaitoni, Peter Hänggi, Peter Jung and Fabio Marchesoni. Stochastic resonance, *Reviews of Modern Physics*, **70**(1), 270-4, 1998

Ivar Jacobson, Grady Booch and James Rumbaugh. *The Unified Software Development Process*, Addison Wesley Longman, 1999

Marvin Minsky, Push Singh and Aaron Sloman. The St. Thomas Common Sense Symposium: Designing Architectures for Human-Level Intelligence, *AI Magazine*, **25**(2), 113-124, 2004

Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, California, 1988

Martin Rees. *Our Final Century*, William Heinemann, London, p.134, 2003

Ron Sun. *Duality of the Mind: A Bottom Up Approach Toward Cognition*, Mahwah, New Jersey: Erlbaum, 2002

Alban Pui Man Tsui and Antonia Jones. Periodic response to external stimulation of a chaotic neural network with delayed feedback, *International Journal of Bifurcation and Chaos*, 9(4), 713-722, 1999

Wu, X.J., Kittler, J., Yang, J.Y., Messer, K. and Wang, S.T. A New Kernel Discriminant Analysis (KDDA) Algorithm for Face Recognition, *Proceedings of the British Machine Vision Conference 2004*, Kingston, UK, 517-526

On the structure of the mind

Maria Petrou and Roberta Piroddi
Imperial College London
Department of Electrical and Electronic Engineering, Exhibition Road, London SW7 2AZ
{maria.petrou,r.piroddi}@imperial.ac.uk

Abstract

The focus of any attempt to create an artificial brain and mind should reside in the dynamic model of the network of information. The study of biological networks has progressed enormously in recent years. It is an intriguing possibility that the architecture of representation and exchange of information at high level closely resembles that of neurons. Taking this hypothesis into account, we designed an experiment, concerning the way ideas are organised according to human perception. The experiment is divided into two parts: a visual task and a verbal task. A network of ideas was constructed using the results of the experiment. Statistical analysis showed that the verbally invoked network has the same topological structure as the visually invoked one, but the two networks are distinct.

1 Introduction

The key to reproducing intelligence and consciousness lies in the understanding and modelling of the organisation of functional areas, the privileged pathways of communication and exchange of information, and their more likely evolution and growth, as in Aleksander (2005). The heart of a biomimetic robot is the pump that sustains the blood-flow of information. The information comes from a number of functional parts which do not necessarily need to be bioinspired: it is possible to produce depth maps without knowing how the visual system produces them. What is important to know is how the depth map is used for navigation and in how many other different tasks it is used. The organisation and evolution of the information, as well as the fusion of pieces of information of varied degrees of uncertainity, are characteristics of a living being and this is what needs to be emulated directly from a biological system.

The Project of the Human Genome Consortium (2003) culminated with the completion of the full human genome sequence in April 2003. Now the research moves forwards, after having named and counted genes. Understanding the organisation and interaction of genes is the new frontier of biotechnology. This endeavour is helped by the development by Barabasi (2002) of a new science of networks, which sheds light into the complexity of organisation of living and artificial mechanisms alike.

The focus of this century will be the brain. Many researchers have already moved to the race for mapping the human brain neurons and understanding their functionalities, for example in Koslow and Hyman (2000). However, neurons are just one element. What about that special products of the human brain, the ideas? Is it possible to gain insight into the world of the ideas? Is it possible to count them? Are ideas organised in a recognisable way?

In the article by Macer (2002) the Behaviourome project was first proposed. In analogy with the genome project aimed at mapping the human genome, here the aim was to count ideas and to find out whether the number of ideas is finite, uncountable or infinite. One of the proposed means of obtaining the final goal was to provide a mental mapping of ideas and their interrelationships.

The organisation of information may be modelled in mathematical terms as a dynamic network. We are intrigued by the possibility that from the lowest possible level of information sources, namely the neurons, to the highest possible level of information products, the ideas, the same structural network may be underling the architectural building of their organisation. Networks have been used for a long time to represent complex systems. They have been popularised recently by Barabasi (2002), who studied self-organising networks. Self-organising networks may be random or scale-free. There is evidence that the organising architecture of a scale-free network underpins the structure and evolution of biological systems (like cells), social systems (like one's circle of friends) and artificial networks (like the Internet).

We designed an experiment in order to study how higher level information invoked by different stimuli,

namely visual and verbal, is organised in the mind. The importance of such a speculative attempt lies in the fact that if a symmetry between lower and higher levels is found, then the same mathematical model may be used to bridge the gap between top-down and bottom-up approaches to create artificial brains and minds.

2 Ideas and networks

In order to proceed in our experiment we need to focus on two elements: ideas and networks. Ideas (and their relationships) are the subject of this research. We need to define them and to highlight some elements of the disciplines that study their relationships.

According to Macer (2002) ideas are mental conceptualisations of things, including physical objects, actions or sensory experiences, that may or may not be linguistically expressible.

Self-organising networks, occurring in the natural world and as a development of human activities, may be modelled as being random or scale-free. Both scale-free networks and random networks are *small world networks*, which means that, although the networks may contain billions of nodes, it takes only a few intermediate nodes to move from one node to any other. Let us indicate by k the number of incoming or outgoing links from any node in the network. The difference between a random and a scale-free network is quantified by the probability density function $P(k)$ of a node having k incoming or outgoing links. In the case of a random network, $P(k)$ has a Poisson distribution, $P_1(k) \sim \exp(-k)$. In the case of a scale-free network, the probability may be modelled by a power function, $P_2(k) \sim k^{-c}$, where c is a positive constant.

3 Experimental methodology

We often see something which triggers some thought, which triggers another thought and so on. We colloquially refer to this as a flow of thoughts. We may try to stimulate such a trail of thoughts by showing an image or mentioning a word to a person. Each person will generate their own path of linked thoughts. The collection of ideas that have been thought may be visualised as a network of ideas and analysed as such. So the questions we would like to answer are:

1. How many ideas on average exist between any two randomly chosen ideas?
2. Does this number depend on whether the stimuli for these ideas are presented to a person in a visual or in a linguistic way?
3. If a network of ideas can be generated, which is its topology?

There are serious difficulties in the design of an experiment in order to answer the above questions. The most important difficulties are:

1. How does one define an idea?
2. How does one deal with the enormous number of different ideas?
3. How can one expose a subject to a collection of ideas?
4. How does one count the intermediate ideas between any two ideas?

The brief answers to the above questions are as follows.

1. Since we wish to show the ideas in the form of an image in the visual experiment, we need to restrict ourselves to an idea being an object, e.g. umbrella, flower, car, sun, etc.
2. It is obvious that we need to restrict ourselves to a finite number of ideas. We decided to use 100 nouns. There were various reasons for that, mainly practical, related to the way the stimuli had to be presented to the subjects. The nouns had to be chosen in an objective way, so as not to reflect any prejudices or associations of the investigators. We chose them to be words uniformly spread in the pages of a dictionary.
3. We showed to each subject the full collection of ideas as a matrix of 10×10 images arranged in an $A0$ size poster. Each noun was represented by a clip-art from the collection of Microsoft Office. This ensured that effort had gone into the design of the images so that they were most expressive for the particular object they were supposed to depict. Before use, each image was converted into grey and it was modified so that the average grey value of all images was the same. This ensured that no image would be picked out because of its excessive brightness or darkness.
4. A picture from the collection was picked at random and then the subject was asked to pick the next most similar one, and then the next most similar one and so on. Every time an object was picked after another one, a link was recorded between these two objects.

We conducted the experiments on a sample of 20 subjects. The sequence of actions that constituted the visual experiment are listed below.

- Allow the subject at the beginning to see the whole table of objects for several minutes, to familiarise themselves with what is included in the restricted world.
- Pick up one of the objects at random and highlight it by putting a frame around it. Ask the subject which of the remaining objects is most similar to this. Once the subject has made their choice, cover the highlighted object, highlight the object they had picked and ask them to pick the next most relevant object. Stop when half of the objects are still visible. This ensures that the subject has still plenty of choice when they make their choice of the last object. The subject may be allowed to say that an object has no relation to any other object in the table. Then the particular experiment may stop. Another experiment may follow starting with a different initial object.
- The position of the objects should not be changed during the experiment, as we wish the person to know what is included in the restricted world in which they have to make their choice.

The verbal experiment was aimed at identifying how many intermediate ideas exist between any two ideas presented in a verbal form. This experiment took place several weeks after the visual experiment. The same subjects took part in both experiments. Each subject was presented with a table containing 100 words corresponding to the 100 pictures presented in the visual experiment. The experiment followed these steps:

- Allow the subject at the beginning to see the whole table of words for several minutes, to familiarise themselves with what is included in the restricted world.
- Pick up one of the words at random and highlight it by putting a frame around it. Ask the subject which of the remaining words is most similar to this. Once the subject has made their choice, cover the highlighted word and ask them to pick the next most relevant word. Stop when half of the words are still available. The subject may be allowed to say that a word has no relation to any other word in the table. Then the particular experiment may stop. Another experiment may follow starting with a different initial word.
- The position of the words should not be changed during the experiment, as we wish the person to know what is included in the restricted world in which they have to make their choice.

4 Experimental results

The first point to investigate is whether there are ideas that are selected more frequently than others, i.e. whether there are ideas that are more popular or spring into mind more often, or all ideas are generally selected with the same frequency.

ID	Idea	Frequency	# Connections
42	graduate	16	15
29	doctor	12	11
99	wedding	14	11
2	airport	12	9
33	electricity-mast	14	9
96	tree	14	9
90	teacher	16	8

Table 1: Ideas that manifest a *hub* behaviour in the visual network.

From sociological studies reported by Gladwell (2000), we know that such nodes act as *hubs* and play an important role in a network. We found that only 7% of the ideas are both very frequent and very well connected. This 7% of ideas are connected to 60% of the remaining ideas for the visual experiment, presented in table 1, and to 50% of the remaining ideas for the verbal experiment, presented in table 2. These numbers show evidence of a small world behaviour for the networks of ideas we have built.

ID	Idea	Frequency	# Connections
7	bicycle	8	10
22	circus	7	12
48	house	10	12
69	rain	7	11
71	road	7	9
83	shopping	7	9
99	wedding	8	11

Table 2: Ideas that manifest a *hub* behaviour in the verbal network.

The number of intermediate ideas between any two given ones may be investigated using two measures. The first measure is the *average path*, defined as the average length of the path between any two nodes, if such a path exists. To calculate it, one has first to calculate the *average* distance between any two given

nodes and then average all these distances over the whole network. We denote this measure by $\bar{a}$. The second measure is called *mean path*. First, one has to find the *shortest path* between any two given nodes, if there is any. Then the shortest paths are averaged over the entire network. We denote this measure by $\bar{m}$.

The second research question was to discover whether the number of intermediate ideas between any two given ones depends on the way the stimuli are presented to the subject. In table 3 we list the path length characteristics for the two experiments. Taking into consideration the standard deviations of these measures, one notices that the path lengths are statistically the same, irrespective of the method used for hint giving.

Experiment	$\bar{a}$	σ_a	$\bar{m}$	σ_m
Visual experiment	13.3	1.9	11.5	3.5
Verbal experiment	13.9	1.7	12.6	4.2

Table 3: Summary of path length measures.

However, the ideas which act as hubs in the two cases are different, as one may see by comparing tables 1 and 2. This strongly indicates that the ideas are organised in two different networks, one verbal and one visual, which, however, exhibit similar topologies.

To further examine this topology we use the degree density $P(k)$, as this is a measure that characterises a network independently from the number of its nodes. We test here for the following null hypotheses:

- H_1: The data have been drawn from a population with Poisson distribution $P_1(k)$.
- H_2: The data have been drawn from a population with power law distribution $P_2(k)$.

Our analysis showed that we may reject the H_1 null hypothesis at 95% confidence level of rejection, while hypothesis H_2 is compatible with our data.

5 Conclusions

The results show:

- a small world behaviour of the networks obtained both by visual and verbal cues, indicated by the low mean path and low clustering coefficient values;
- a correspondence between the visual and verbal network in the value of the mean path and the value of the possible number of hubs;
- a correspondence in the topology of the networks, the two networks being statistically equivalent in topology;
- a difference between the visual and verbal networks indicated by the concepts that acted as hubs in the two networks;
- evidence that the networks are organised as scale-free ones.

If the mind organises itself in the form of networks, it appears that these networks are strongly influenced by the hardware on which the mind resides, i.e. the brain, otherwise we should not have a different network for the flow of ideas when visual stimuli are used, from that created when verbal stimuli are used. It is possible, therefore, to hypothesise the existence of a hierarchy of organisation of information which flows from the lowest level of electro-chemical information of the neurons, to the highest conceptual abstractions of ideas. The hierarchy is supported by the same underling network framework, which is one of scale-free topology. The network represents the hardware of the organisation, dictating which dynamical modification is plausible in the evolution of the information. What we should look for, in order to create a link between top-down and bottom-up approaches, is a mapping between neural structures and their higher level correlates.

References

I. Aleksander, *The World in my Mind, My Mind in the World: Key Mechanisms of Consciousness in people, Animals and Machines*, Academic Press, 2005.

L. Barabasi, *Linked: The New Science of Networks*, Perseus Publishing, 2002.

M. Gladwell, *The tipping point. How little things can make a Big difference.*, Little, Brown and Company, 2000.

International Human Genome Sequencing Consortium, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–922, 2003.

S. Koslow and S. Hyman, "Human brain project: A program for the new millennium," *Einstein Quarterly Journal of Biology and Medicine*, vol. 17, pp. 7:15, 2000.

D.R.J. Macer, "The next challenge is to map the human mind," *Nature*, vol. 402, pp. 121, 2002.

Social Learning and the Brain

Mark A. Wood*
*Artificial models of natural Intelligence (AmonI)
University of Bath, Department of Computer Science
Bath, BA2 7AY, UK
cspmaw@cs.bath.ac.uk

Abstract

Social learning is an important source of human knowledge, and the degree to which we do it sets us apart from other animals. In this short paper, I examine the role of social learning as part of a complete agent, identify what makes it possible and what additional functionality is needed. I do this with reference to COIL, a working model of imitation learning.

1 Building a Brain

The problem of building a brain is one facing me at this very juncture in my research. I need a brain capable of controlling indefinitely a complete agent functioning in the virtual world of *Unreal Tournament* (*UT*) (Digital Extremes, 1999). As a game domain, clearly UT is not an exact replica of the real world, and much is simplified or omitted altogether. However, it does provide an opportunity to study a very broad range of human behavioural problems at a tractable level of complexity, as opposed to other more realistic platforms which allow only the study of narrow classes of problems.

My research thus far has chiefly been in the area of social learning (particularly imitation), as I believe this is key to survival in a world where there are unfortunate consequences if things are not learned quickly enough. We humans also dedicate a vast amount of brain space to learning, and social learning in particular, compared to other species. In the following section, I will explain what I think the role of social learning is and why it is important. I will then briefly overview COIL, a model extending CELL (Roy and Pentland, 2002) from language learning to social learning in general. I describe both what COIL requires to function and how it would be extended and complemented to form a complete brain system. I conclude with a discussion.

2 The Role of Social Learning

Human infants seem to be innately programmed to imitate from birth (Meltzoff and Moore, 1983). Many animals, particularly the great apes, benefit from similar kinds of social learning mechanisms (Byrne and Russon, 1998), but none to the extent that we do. The speed and accuracy with which we can assimilate goal-directed (ie. task-related) behaviour from others is unique. Of course, communicating via language and the ability to reproduce actions at fine temporal granularity are among the human-specific skills which facilitate this learning. Taking these things into consideration, it would be wise to consider including social learning capabilities in any system designed to function as a complete brain.

Furthermore, autonomous agents need skills: whether 'basic', low-level skills such as co-ordinating motor control, or 'complex', high-level skills such as navigation. To acquire task-related skills at any level, I believe there are at least four types of things which need to be learned (Bryson and Wood, 2004, see also Figure 1):

1. *perceptual classes:* What contexts are relevant to selecting appropriate actions.
2. *salient actions:* What sort of actions are likely to solve a problem.
3. *perception/action pairings:* Which actions are appropriate in which salient contexts.
4. *ordering of pairings:* It is possible that more than one salient perceptual class is present at the same time. In this case, an agent needs to know which one is most important to attend to in order to select the next appropriate action.

Some of these may be innate, but those which are not must be acquired using a combination of individual and social learning. For example, assume we

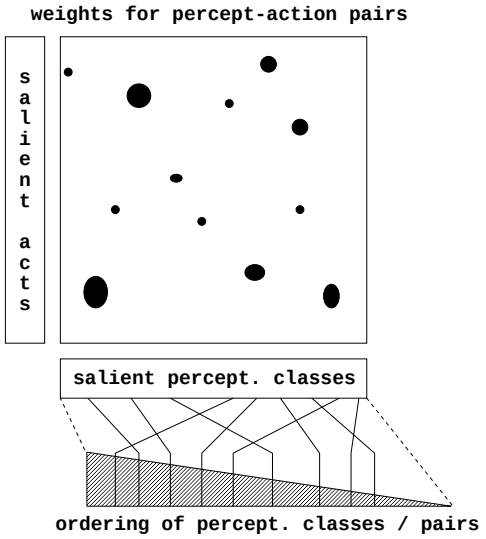


Figure 1: Task learning requires learning four types of things: relevant categories of actions, relevant categories of perceptual contexts, associations between these, and a prioritized ordering of the pairings. Assuming there is no more than one action per perceptual class, ordering the perceptual classes is sufficient to order the pairs.

have an agent which can issue motor commands, but does not initially know the results these commands will have on its effectors. Using visual and proprioceptive sensors (say) to measure these effects, and trial-and-error (individual) learning, a mapping between commands issued and effects produced can be created. This example is deliberately analagous to human infant 'body babbling' (Meltzoff and Moore, 1997). However, assuming a reasonable number of 'primitive' actions can be learned this way, the set of skills that can be built from these blocks is exponentially larger (and so on as more skills are acquired). To attempt to learn all skills through trial-and-error, then, would be to search randomly through these huge, unconstrained skill spaces — very inefficient.

Social learning can take many forms depending upon the nature of the agents in question: written or verbal instruction, explicit demonstration, implicit imitation, etc. An agent which is part of a society which facilitates such learning can take advantage of the knowledge acquired by previous generations. To do this an agent must be able to relate what it perceives to the actions it can execute; it must solve a correspondence problem between the instruction or demonstration and it's own embodiment (Nehaniv and Dautenhahn, 2002). For a learning agent in a society of conspecifics, this mapping is simple (although not trivial to learn), but for, say, a robot living among humans, solving this problem amounts to yet another skill that needs to be mastered. Socially-acquired skill-related knowledge can be used to significantly reduce the skill search space, allowing individual learning to merely 'fine-tune' new skills, taking into account individual variability within a society. The other alternative is that the 'instructions' acquired are coarse-grained enough to perfectly match existing segments of behaviour in the learner's repertoire.

3 Necessary Components

To better understand the components required for social learning in general, it makes sense to examine the information requirements of a model which is capable of such learning. The Cross-Channel Observation and Imitation Learning or COIL model of Wood and Bryson (2005) is suitable. This system is designed to observe via virtual sensors a conspecific agent executing a task, then in real-time output a self-executable representation of the behaviour needed to complete that task. It achieves this by matching the observed actions of this task *expert* with its observed perceptions of the environment. I now briefly explain the model and identify in general terms what is needed at each stage of processing (see also Figure 2).

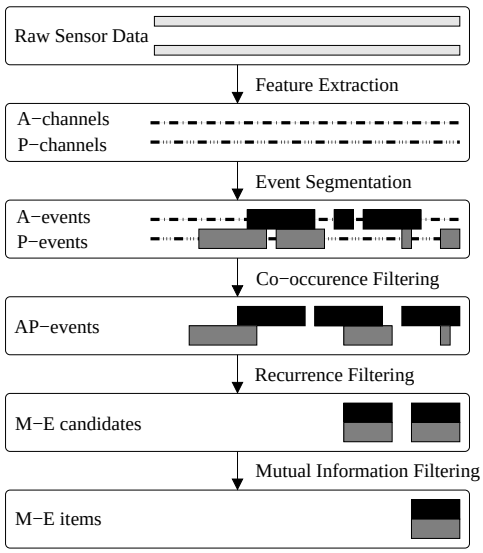


Figure 2: An overview of COIL.

Feature Extraction The inputs to this stage are raw sensory data. Depending upon the task, some of

these data are discarded and others are recoded or categorised. Remaining data are diverted into different channels ready for further processing, some specialising in action recognition and others in environmental parsing (perception). This stage therefore suggests a need for selective **attention**, compressed **representations** and **modularity** of processing.

Event Segmentation Following Feature Extraction, the channels containing the data are segmented into action and perception events depending upon the channel type. Events define high-level coarse-grained actions and perceptual classes, and are further divided into lower-level fine-grained subevents. This segmentation requires various **triggers** which are innate in the case of COIL, but could theoretically be learned.

Co-occurence Filtering Action and perception events which overlap in time are paired together and stored in a buffer (called Short Term Memory or STM). This requires **temporal reasoning** and **memory**.

Recurrence Filtering Co-occuring action and perception subevents which are repeated within the brief temporal window of STM are tagged. A chunk called a Motivation-Expectation or M-E Candidate, which represents the set of tagged pairs, is created and placed in Mid-Term Memory (MTM). Here we additionally use **statistical reasoning** and abstract judgments of **similarity**.

Mutual Information Filtering For each M-E Candidate, the maximum mutual information between its component action and perception subevents is calculated. Those which exceed some threshold are stored as M-E Items in Long Term Memory (LTM). COIL currently uses fixed **thresholds**, but again they could be acquired through experience. The LTM is the output of the system.

The innate skills which are necessary for social learning identified above can be provided by the hardware (memory, clock, etc.) and software (statistical algorithms, similarity metrics, etc.) of the agent.

4 Scaffolding COIL

I have looked at the basic components COIL needs in order to function as a social learning system. However, the extent of COIL's role within a complete agent, and the extra pieces which need to be added, remain in question.

There are a number of problems in assuming that a single monolithic COIL system can alone act as the 'brain' of our agent. Firstly the algorithm only learns – it has no capacity for making decisions or acting based upon what it has learned. Our most recent work demonstrates the addition of an extra module for exactly those purposes (Wood and Bryson, 2005). Secondly, a flat COIL system expected to carry out the high-level task of *life* would have to monitor every action and perception channel that could possibly be useful in achieving this task, or any of its subtasks. Even with the innate attentional capabilities of COIL's Feature Extraction stage, the algorithm's complexity is still exponential in the number of channels. Therefore, COIL seems more suited to learning local specialised tasks where the number of channels which need to be monitored can be reasonably constrained.

Let us assume instead that we have a number of COIL systems, each observing a localised task and its associated action / perception channels. We would need a method for discerning which of the following four scenarios is occuring:

1. A known task is being observed.
2. A known task is present[1].
3. An unknown task is being observed.
4. No known tasks are present or being observed.

It may be that scenarios 1 and 2 occur concurrently, in which case a decision would need to be made whether to observe and learn or join in with the execution of the task. Also, the presence of a task does not necessarily imply that the task should be executed. In scenario 4, social learning is impossible, and the product of previous social learning episodes is not applicable. This is where an individual learning module would come into play (see also Section 5).

A complete system that is capable of this arbitration is as yet unrealised. It may be possible to create a 'master' version of COIL which has high-level perception channels monitoring those environmental states which differentiate between local tasks, and action channels which cause lower-level task-specific COILs to be activated. On the other hand, a totally different system designed specifically to coordinate COILs (for social learning), RL (for independent learning) and acting may be more appropriate.

[1] A task is present if it is available in the environment for execution by the observer.

5 Discussion

In this final section, I highlight a number of research problems, some of which I will be investigating in relation to the COIL project, but all of which I believe will need to be studied before a complete working brain becomes a possibility. The balance between what is innate and what is learned, for both biological and theoretical robotic examples, has been discussed by Sloman and Chappell (2005). They claim that using a hybrid of the two may prove to be better than using either in isolation. We can further subdivide that which is learned into that which is learned socially, and that which is learned independently. Similarly, the best technique is probably to combine the two, and it is a thorough study of the balance and application of both that may result in significant progress toward constructing a complete agent. This task has at least the following component questions:

- What structures must be present to make independent learning possible? Presumably many of these will be innate, but how can social learning improve these structures / primitives and / or the efficiency of their usage (ie. learning how to learn better from another)?
- What primitives are needed to make social learning a possibility? Must they be acquired through trial-and-error learning, or can they be innate? If acquired, what is the cost of such acquisition? How do they differ from those required for individual learning?
- How does the embodiment of a given agent affect the structures / primitives best suited for both individual and social learning (both *what* is learned and the *way* it is learned)? How does this compare / interact with the affect the required tasks have on these primitives?
- How can individual and social learning be combined at both the practical task level and the abstract memory level? Do different combination strategies result in different levels of efficiency and / or goal accomplishment? Is this 'meta-skill' of hybrid learning itself innate, or somehow learned?
- What is the optimal trade-off between individual and social learning for a given task? How does this change with increasing task complexity? How is this affected by the nature of the task relative to, say, the embodiment of the executing agent?
- How can knowledge be consolidated to improve learning (of both kinds) next time? How are conflicts between what is learned socially and what is learned independently resolved? How easily applicable are social skills and their associated knowledge to individual learning situations, and vice versa?

I hope that these proposed research areas, and this paper as a whole, will in some way stimulate others into thinking about social learning in the context of a complete agent system.

References

Joanna J. Bryson and Mark A. Wood. Learning discretely: Behaviour and organisation in social learning. In *Proceedings of the Third International Symposium on Imitation in Animals and Artifacts*, pages 30–37, 2004.

Richard W. Byrne and Anne E. Russon. Learning by imitation: a hierarchical approach. *Behavioral and Brain Sciences*, 16(3), 1998.

Digital Extremes. *Unreal Tournament*, 1999. Epic Games, Inc.

Andrew N. Meltzoff and M. Keith Moore. Newborn infants imitate adult facial gestures. *Child Development*, 54:702–709, 1983.

Andrew N. Meltzoff and M. Keith Moore. Explaining facial imitation: A theoretical model. *Early Development and Parenting*, 6:179–192, 1997.

Chrystopher L. Nehaniv and Kerstin Dautenhahn. The correspondence problem. In Kerstin Dautenhahn and Chrystopher L. Nehaniv, editors, *Imitation in Animals and Artifacts*, Complex Adaptive Systems, chapter 2, pages 41–61. The MIT Press, 2002.

Deb K. Roy and Alex P. Pentland. Learning words from sights and sounds: a computational model. *Cognitive Science*, 26:113–146, 2002.

Aaron Sloman and Jackie Chappell. The altricial-precocial spectrum for robots. In *Proceedings IJCAI'05*, pages 1187–1192, Edinburgh, 2005.

Mark A. Wood and Joanna J. Bryson. Skill acquisition through program-level imitation in a real-time domain. Technical Report CSBU-2005-16, University of Bath Department of Computer Science, 2005. Submitted to IEEE Systems, Man and Cybernetics – Part B.