Finding Aesthetic Pleasure on the Edge of Chaos: A Proposal for Robotic Creativity

Ron Chrisley
Centre for Research in Cognitive Science
Department of Informatics
University of Sussex
May 16th 2006

I have been developing a robot neurocontroller called the CNM since 1990. I propose to extend it in such a way that it allows a robot to exhibit creative behaviour, in the sense that human observers may find the products of the behaviour both of value, and novel. Thus, no direct modeling of human creativity is intended. However, I reserve the right to opportunistically and shamelessly exploit what is known about creativity in humans, when and as is expedient. Also, I suppose that there is always the possibility that insights into the human case may accrue from the artificial case by accident, as it were.

The CNM is, at root, a learned, forward model of the environment that allows a robot to anticipate or predict the sensory input it will receive if it performs a given action in a given context. This, in conjunction with built-in motivators (such as a "pain" signal when colliding with an object, or a "pleasure" signal when the robot's battery charge increases) can enable the robot to engage in adaptive behaviour by selecting actions that result in an expectation of pleasure or absence of pain.

The heart of the creativity-inducing extension to the CNM would be additional motivators, corresponding roughly to the two dimensions of creativity mentioned before: Value and novelty.

Value:

An assumption behind this approach is that a good way to make an artificial system be capable of creating works of value is to make it be capable of appreciating works of value. Thus, the ability of the CNM to evaluate its own output is crucial; but of at least equal importance is its ability for it to experience and evaluate the works of others, and the world in general. A way to avoid the representational bottleneck of having to "feed" a system digital encodings of others' work is to let it experience those works in the real world, as we do. Suppose, for example, that the CNM is operating within the domain of music. Then the proposal is that instead of giving the CNM a MIDI score of Beethoven to ponder, let it hear Beethoven coming out of your stereo, or better yet, out of your piano. But if it experiences others' works in the real world, then it should experience and evaluate its own works that way too. This double embodiment of appreciating and creating in the real world, in conjunction with a judicious choice of hard-wired prefrences akin to ours (e.g., for low integer relationships between frequencies), will go a long way to ensuring that what the CNM ends up valuing is the kind of output that we, too could value. It also allows for the possibility of serendipity, in the space between what the sensory inputs the CNM expected its actions to actually produce, and the sensory inputs the actions actually produced: the space we call "the world". These can augment more familiar ways of ensuring connections to human value, such as built-in preferences for human attention. For example, a preference for the sounds of human presence (especially applause!) would yield a preference for human proximity. Alternatively or additionally, one can formalise human approval by putting buttons on the robot that would allow listeners to provide approval or disapproval feedback.

Novelty:

Another kind of motivator will have the effect of encouraging novel, open-ended behaviour by pitting the creative and predictive capacities of the system against each other in a salutary co-evolutionary arms-race. Saunders and Gero have suggested motivating an agent in a way that can be characterised by a Wundt curve. The pleasure of a stimulus is a hump-shaped function of its novelty: the agent receives little pleasure out of non-novel experiences, maximal pleasure out of mildly novel experiences, and reduced, zero, or even negative pleasure for experiences that are "too novel". The CNM aims to improve on this idea with the inclusion of a motivator that is not dependent on novelty itself, but on something that often results in novelty: complexity. One can summarize the motivator as a preference for experiences that are on the subjective "edge of chaos": sensory inputs that almost, but do not quite, elude analysis or understanding, for that agent at that time. This motivator can be described with the same Wundt curve, but the x-axis of the "hump" curve is not novelty, but rather: effort (e.g., time) required to be able to understand. In the case of the CNM, understanding is a case of subsuming under prediction. If an experience is too simple/predictable, it will offer little pleasure from the complexity motivator (although it may result in pleasure because of one of the inbuilt motivators). If an experience is harder to predict (e.g., some time passes before prediction error can be held below a certain threshold), then more pleasure will be derived. However, if the error cannot be reduced, or only at great cost in time and/or resources, little or no pleasure will result. Thus, real-time dynamics will be crucial to the CNM's appreciation of a work.

This is enough for the CNM to be an appreciator of novel works. But where there is evaluation, there can be creation: In the case of sensory experiences that were the result of perceiving the CNM's own creations, the reward signal from the motivator will have the result of changing the action selection policy in such a way that both overly simplistic, and also "uncompressable", incomprehensible works are less likely to be produced in future. What will be left are works on the edge of chaos, where valuable novelty lives.

As a result, the CNM will be aiming at a moving target. What once was pleasurable because of its elusiveness will become boring, because now easily understood. However, there does remain the theoretical possibility that conceptual development is not monotonic. Especially in connectionist systems, learning to understand something new may make one forget something previously learned, potentially making it of interest again. Preventing such limit cycles may require the introduction of some kind of explicit autobiographical memory as a more conventional means of calculating novelty, if the cycles are not prevented by some other means (such as the bored faces of other agents that do have a memory of the CNM's earlier, and now repeated, output).

A further extension of this approach can be achieved by making the CNM's own states, particularly the states of its motivators, the subject of "perception" and prediction, and therefore a potential source of meta-motivation. Experiences that have an effect on the CNM's motivators such that a pattern is produced that the CNM's model of its own motivator activity initially fails to predict, but eventually succeeds in predicting, will be preferred (and will be therefore more likely to be produced). Perhaps there could be hard-wired preferences for certain motivator patterns as well. But one must be careful to not give the CNM too much access to its own states: If the self becomes easily predicted, any object produced by that self will tend to be rendered easily predictable, reducing its subjective aesthetic value. The only option then would be for the system to seek out regimes of activity that evade the predictive capacities of its own self model. Perhaps this could even serve as a way to ground what is meant by "too complex" in this context: Whatever can be predicted only by modelling oneself. However, such prediction will also require modelling of the environment, so this bound will not be inconveniently low. On the other hand, being able to predict one's own perceived bodily movements (toe-tapping?) may itself be a source of pleasure via various motivators, including the complexity motivator.

The intended robotic platform for this system is a pair of AIBO ERS-7 robots. Having more than one CNM agent will dramatically foster the co-evolution of the generative and evaluative capacities. But the AIBOs are motorically challenged; they can't draw, for example. They can produce sound, but it would hardly be in the embodied spirit of this proposal if the rich source of creativity that is the gap between conception and realisation were not made available to the CNM. So although the initial stages of the project may involve simple auditory output, a further stage will explore the following possibility: Impose a mapping between joint angles and sounds produced, at least when the CNM AIBO is in "creative mode". Thus, every slight move the AIBO makes (or is imposed upon it by the world, or another AIBO) will result in a change in its sound output. In order to make more pleasurable sound outputs, the AIBO will have to move. The resulting music improvisation through dance, and observation of such by humans and the other AIBO, should open up dramatic new possibilities for creative output.

I look forward to hearing from others about the undoubtedly enormous amount of work along these lines that has already been done!

References:

Saunders, R., Gero, J.S. (2001) Designing for Interest and Novelty: Motivating Design Agents. In de Vries, B., van Leeuwen, J., Achten, H. (eds.), CAADFutures 2001. Dordrecht: Kluwer, pp 725-738.

On the CNM architecture:

Chrisley, R. and A. Holland (1995) "Connectionist synthetic epistemology: Requirements for the development of objectivity," Niklasson, L. and Boden, M. (eds.) Current Trends in Connectionism: Proceedings of the 1995 Swedish Conference on Connectionism. Hillsdale, NJ: Lawrence Erlbaum, pp 283-309; ISBN 0 8058 1997 5. http://www.cogs.susx.ac.uk/users/ronc/papers/scc95.pdf

Chrisley, R. (1993) "Connectionism, Cognitive Maps, and the Development of Objectivity," Artificial Intelligence Review 7, pp 329-354. http://www.cogs.susx.ac.uk/users/ronc/papers/sncc92.pdf

Chrisley, R. (1990) "Cognitive Map Construction and Use: A Parallel Distributed Processing Approach," in Touretzky, D., Elman, J., Hinton, G., and Sejnowski, T. (eds.) Connectionist Models: Proceedings of the 1990 Summer School. San Mateo, CA: Morgan Kaufman. pp 287-302. http://www.cogs.susx.ac.uk/users/ronc/papers/Chrisley-Cognitive_Map_Construction_and_Use.pdf