AISB Quarterly No.119

The Newsletter of the Society for the Study of Artificial Intelligence and Simulation of Behaviour

Implementing machine consciousness

The last decade has seen the emergence of the new discipline of consciousness studies, which is now well established with journals (e.g. *Cognition and Consciousness, Journal of Consciousness Studies*), societies (ASSC, the Association for the Scientific Study of Consciousness), and conferences (the biennial *Toward a Science of Consciousness*, and the annual ASSC meetings) of its own. Consciousness studies encompass a huge range of activities: these are centred around psychology, but range from philosophy to neuroscience.

Most recently, a new strand—machine consciousness—has emerged.[1] In the 1990s, a small number of pioneers, including Stan Franklin in AI, Igor Aleksander in electrical engineering, and Gerald Edelman in biology, had begun to examine the possibility of creating consciousness in machines or software. Growing interest in these activities led to the 2001 Banbury Workshop *Can a machine be conscious?* At the end of this highly interdisciplinary event, the twenty-odd participants were asked to vote on whether they thought machine consciousness was in fact a possibility, and all except one agreed that it was. The outcome both surprised the participants, and gave the topic an impetus which is reflected in the level of interest in the symposium *Next generation approaches to machine consciousness* at AISB 2005.

We set up the Machine Consciousness Group at Essex to explore the possibility of achieving machine consciousness in a robot through the formation and exploitation of internal models of itself and the world. This enterprise has been facilitated by the University's support, via a £4.5m SRIF (Science Research Investment Fund) grant, for the building of a new robotic arena and workshop complex. Current work is directly funded by an EPSRC (Engineering and Physical Sciences Research Council) Adventure Fund award of £493,000, shared between myself and the visual neuropsychologists Tom Troscianko and Iain Gilchrist of the University of Bristol. This project, which runs until April 2007, involves the design and construction of a humanoid robot, and the study of the nature and development of its consciousness-related processes based on internal modelling.

Why a humanoid robot? There are several reasons for this. We do not know very much about consciousness, but there has been a growing consensus that the origins of at least the lower levels of conscious phenomena are very strongly rooted in the body. If we are to produce consciousness in a machine, then its embodiment will be a critical determinant of the nature of that consciousness, and of its intelligibility and relevance to human consciousness. The robot (Cronos) is therefore being given a gross physical structure that is, as far as possible, qualitatively similar to the human body.

Figure 1 shows the prototype (modelled on the body of its designer, Rob Knight): the basic humanoid skeletal structure of the (headless) upper torso is clear. The articulated skeleton models many of the constraints and degrees of freedom of our own body. In addition, the musculature uses a mixture of passive compliance and series-elastic actuators, ensuring that the motor programs used by the robot will be similar in important ways to those used in our own brains. When complete, the torso will be mounted on a wheeled base.

Most autonomous mobile robots merely move through environments, but Cronos is being designed to be able to operate on the environment in ways comparable to those used by humans. At full extension, the arm and hand will be able to

Owen Holland, University of Essex

Continued on p. 9

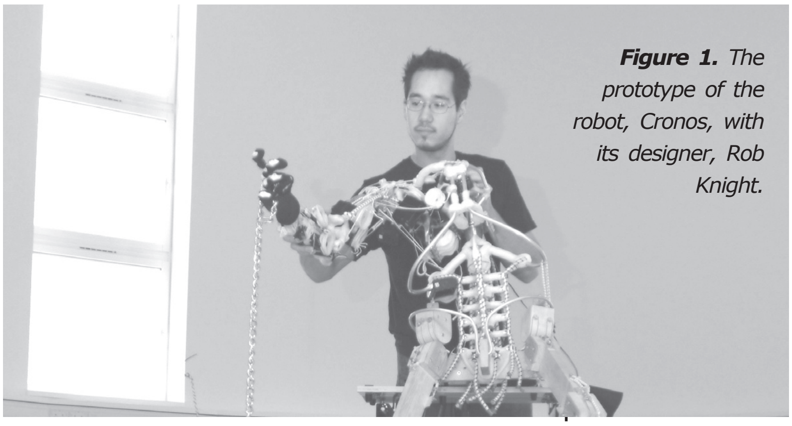


***Figure 1.** The prototype of the robot, Cronos, with its designer, Rob Knight.*

No.119

In this issue:

Features

Reviews/ previews

Community

Father Hacker

Don't forget to vote!

Voting papers enclosed with this issue

Subsea pilotless inspection using an autonomous vehicle

The **AISB Quarterly** is published by the Society for the Study of Artificial Intelligence and Simulation of Behaviour.

ISSN 0268-4179

SPINAV was a one-year subsea technology development project, transferring next-generation technology for underwater robotic vehicles into a prototype demonstrator for the offshore oil industry. The recently-completed venture was funded as a joint-industry project by BP, Conoco-Philips, Subsea7 and the UK Government's Department of Trade and Industry. Its aims were to: implement proof-of-concept autonomous riser inspection using a small autonomous underwater vehicle (AUV); demonstrate the technical and commercial advantages of AUVs for subsea inspection; and develop foundation technology applicable to other common subsea inspection tasks (e.g. jackets, moorings, and pipelines)

Pipeline inspection is currently carried out using sensor suites (profilers, video, and cathodic probes) that are either towed or mounted on remotely-operated vehicles (ROVs). Both require a ship to be in attendance. Towing is fast and comparatively cheap, but gives poor-quality information. ROV mounting is slow and comparatively expensive, but gives better quality coverage. Risers are generally inspected using ROVs equipped with video and a.c.-impedance-based crack-detection sensors. The task is tedious in the extreme, requiring high levels of pilot concentration for extended periods in order to keep the vehicle stationary, especially in deep water. Sea currents further make positioning difficult, thanks to drag and disturbance from the umbilical or tether. For this reason, human-piloted inspection can only be carried out in conditions of good visibility.

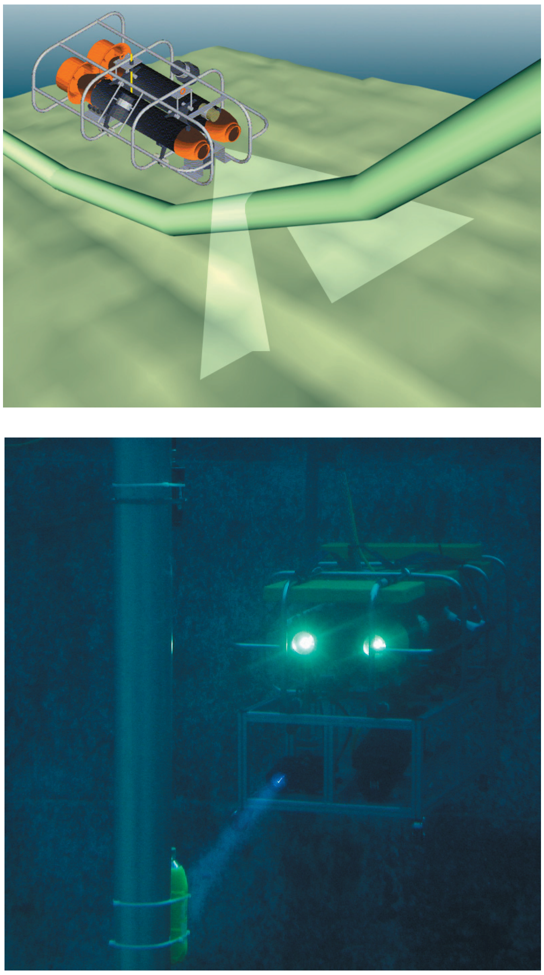


***Figure 1.** The SPINAV concept (top) and reality (bottom).*

Trends in oilfield installation therefore suggest that AUVs will eventually offer the most cost-effective solutions for various inspection, repair and maintenance tasks. Since AUVs do not require permanent attendance of a surface vessel, they can be considered part of the infrastructure of the oilfield installation: thus, they should provide both cost savings and high-quality data. Further, those that can be deployed directly from the oil platform remove the need for a dedicated support vessel entirely, with corresponding reductions in cost. For pipeline monitoring, the AUV can obtain good video and profile data, and travels several times faster than an ROV. The AUV may also reduce tedium (and therefore mistakes) for riser inspection applications, while at the same time holding position better and avoiding the possibility of umbilical entanglements.

A payload package capable of planning and executing an autonomous riser inspection mission has been developed through SPINAV. It has an interface module to the host AUV which, in this case, is the RAUVER (from the Ocean Systems Laboratory, Heriot-Watt University). The demonstrator focused on the development of three key abilities: sensor processing for riser detection and tracking; AUV hover control and navigation around the riser to achieve complete coverage; and mission adaptation based on leaks detected by the sensors.

Two sets of wet trials were performed during the project. The first—in February 2004, at Subsea7's test tank in Aberdeen—was an exercise in data gathering, environmental characterization, and evaluation of vehicle control. Here, the RAUVER worked under semi-autonomous control: a human manually provided output in place of the 'tracker module', which had yet to be implemented. These trials were extremely successful. A second set took place in November 2004. This time, a series of fully-autonomous inspection missions were demonstrated live to sponsors and interested parties from the subsea sector. During these missions the vehicle operated completely independently, without any contact with a human, the surface, or any other system whatsoever.

Hamilton and Evans, SeeByte Ltd.

Continued on p. 7

Modelling socially-shared cognition

We introduce here a multi-agent recurrent network that models the way people develop shared cognition through social interaction. Festinger[1] argued that people use social reality to judge the appropriateness of their behaviours, opinions, and beliefs to those of other people around them.

The new model implements this kind of social comparison process as an extension of the delta learning rule,[2] and consists of an agent level and group level. At the agent level, a standard recurrent auto-associator[2] simulates information processing within a single, independent agent (see Figure 1). Individual agents are embedded in an extended recurrent network that represents a social group. In the same way that information flows within an agent, communication between agents involves activation spreading from one agent to another through the network in proportion to the connection weights linking them.

Specifically, the process of communication involves a talking agent sending out activation to listening agents. These compare their own internal beliefs concerning an issue with the attitude expressed by the talking agent on that same issue (see Figure 1). Weight updates are driven by the error between the external information—representing the attitude expressed by the talking agent—and the internal activation, representing the listening agents' attitude:

$\delta_i = extinput_j - intinput_i$

Here $extinput_j$ is the final activation send out by the talking agent and $intinput_i$ is the internal activation of the listening agent. When agents share the same attitude, the weight of the links between them is adjusted upwards. If they disagree on an issue, the weights are adjusted downwards. The rate by which the weights are adjusted is expressed mathematically as:

If $|ext\_a_j - int\_a_j| < threshold$

then $\Delta w_{ij} = \eta * (1 - w_{ij}) * |a_i|$

else $\Delta w_{ij} = \eta * (0 - w_{ij}) * |a_i|$

Here, ext_a_j represents the external activation received (from the talking agent i) by the listening agent$_j$, and int_a_j represents the internal activation generated independently by the listening *agent*$_j$.

When agents largely share the same attitude (i.e. the difference is below some threshold), the links between them are strengthened. Otherwise, the links between them are weakened. This constitutes an adaptive social process, in which agents learn by interacting with each other: agents that consistently confirm each other's attitudes will be connected by stronger links than those that consistently disagree. The social experience acquired in this way is represented in a distributed manner, in patterns of weighted links across the whole network. As such, social experience represented by this network shows the same type of redundancy as real social groups, which survive the coming and going of members without losing their identity or norms.

Simulations have exposed some interesting emergent properties of the model. The social comparison process implemented at the agent level leads agents to clustering in social space in terms of attitude similarity. The links between agents within a single sub-cluster act as positive feedback loops that result in agents within these clusters reinforcing each other's attitudes. At the same time, interaction between clusters becomes more restricted, as the links between clusters of agents that disagree weaken. As a result, agents self-organize in such a way that majority positions tend to become stronger (more prevalent) and minority positions weaker. This emergent property allows the model to account for group polarization (positive feedback loops) and also why most social groups maintain a healthy variety of opinion (interaction between clusters diminishes).

In the same way that agents self-organize through interaction, the information communicated through the social system is also adapted and integrated. Each time an agent acquires information, it assimilates and adds its own personal experience (as captured by the long-term weights within an agent network) before sending it out again into the group. The flow of information within the extended network will at some point reach a stable state. This represents an equilibrium between the information that is presented to the social system (the activation spreading through the network) and the previous social experience of the group (encoded in the long-term weights). As such, one could argue that concepts self-organize through communication into novel, collective concepts.

In conclusion, the model

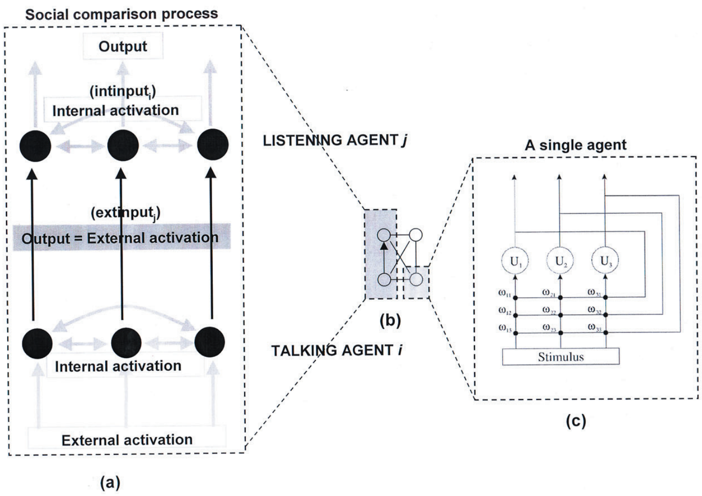


Figure 1. a) The social comparison process between a talking and a listening agent. b) A group of four agents. c) A standard recurrent network representing a single agent.

Dirk Van Rooy
Keele University

Continued on p. 6

Understanding the behaviour of agents and cognitive models: CaDaDis 2.1

Different representations can reveal different aspects of models. The inverse is also true: namely, that different representations *hide* different aspects of models. It is, therefore, important for multiple representations of a model to be used during development in order to better understand it. We are developing a categorical data display (CaDaDis) for visualizing the sequential behaviour of cognitive models that is flexible both in the array of visualizations possible and in the cognitive architectures that can use it. The overall goal of the project is to provide a common tool for cognitive modellers to work with visualizations of model output.

CaDaDis[1] contains Gantt, program evaluation and review technique (PERT), and modified-PERT visualizations. The firing of actions in the model (e.g. operators or rules) provides data to plot. PERT charts depict task and task duration, along with dependency information, and our recent development work has focused on refining the modified PERT chart. In its latest release (2.1), CaDaDis introduces two new visualization features that allow it to more adequately address the needs of modelers. Specifically, we have added the ability to describe a model's sequential actions in the context of another sequence, and also a means for developers to manipulate a view through automated rearrangement of the categorical display. The new features are built on top of the modified PERT chart included in the original CaDaDis release.

Figure 1 shows a view of a dTank agent running in Soar. The display shows a sequence in the run where the agent has waited, and then located an adversary to attack. The left panel contains a list of operators that have been applied. The right panel displays the order of operator applications. The x-axis of the diagram represents the time at which the operator applied while the y-axis location aligns the operator application with its name. CaDaDis generates this representation as the model runs and can save it so that it can be opened later without running the model again.

Because CaDaDis provides common visualizations of different models and architectures, it allows for clear comparisons. This can be accomplished by studying similarities and differences between the sequences that different models generate. Previously, in order to compare model behaviour in CaDaDis, one had to store the images generated and compare them side by side. The new features in CaDaDis version 2.1 better support the important task of behaviour comparison. In addition, it now supports displaying the same model with multiple displays of its behaviour. The new features allow for comparing similar models by displaying multiple similar runs on a single graph. We chose to allow for multiple runs on one (as opposed to two) displays so that users can compare how different models solve the same problem (taking an idea from Sun et al., 2004).[2]

Figure 2 shows the new view provided in CaDaDis 2.1. This particular example shows two independent runs of the *Eight-Puzzle* demo provided with Soar. The top line represents one run, the bottom a second. The first operator shown in this figure is the seventh in the run (a keen observer will notice that the window has been scrolled). Up until the twelfth operator, the runs are identical, but then the runs diverge, and one terminates. As one can imagine, a more complex model with a longer run-time can produce the same type of activity: namely, one in which the runs only differ slightly. This new capability of seeing both runs at once immediately shows where the models' behaviour differs.

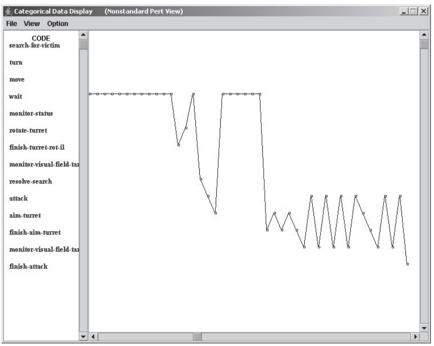


Figure 1. *A dTank agent finding an adversary in a CaDaDis-modified PERT Chart.*

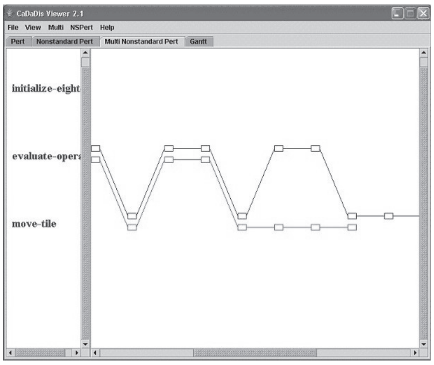


Figure 2. *Soar Eight-Puzzle in a CaDaDis multiple-run-modified PERT Chart.*

We have also added a sorting capability, whereby users can have the actions rearranged by one of several algorithms. Those currently implemented include order of arrival, frequency, and reverse-frequency of application.

Now that CaDaDis can compare multiple runs on one visualization and manipulate the operator display order, we have the means to explore more advanced interactions and manipulations.

CaDaDis is available for download[4] and, for academics, it requires a no-cost license from Soar Technology, Inc. to support Vista, a visualization framework for cognitive models that serves as the underpinning architecture. CaDaDis currently provides direct support for Soar, ACT-R, and JESS. Other models and architectures that use Tcl, LISP, or Java will be able to reuse the existing application program interfaces fairly directly. CaDaDis can also load a series of actions from a file.

CaDaDis is supported by the Office of Naval Research through a subcontract from Soar Technology, #VISTA03-1. Thanks to Dwight Berry, Geoff Morgan, Andrew Reifers, Bill Stevenson, and Glenn Taylor for their comments and input.

John Daughtry and Frank Ritter
School of Info. Sciences and Technology
Pennsylvania State University, PA USA
E-mail: {jmd514, frank-ritter}@psu.edu
http://acs.ist.psu.edu

References

1. K. Tor, S. R. Haynes, F. E. Ritter and M. A. Cohen, *Categorical data displays generated from three cognitive architectures illustrate their behaviour,* **Proc. Sixth Int'l Conf. on Cognitive Modeling,** Earlbaum, Mahway, NJ, pp. 302-307, 2004.
2. S. Sun, I. Councill, X. Fan, F. E. Ritter, and J. Yen, *Comparing teamwork modeling in an empirical approach,* **Proc. Sixth Int'l Conf. on Cognitive Modeling,** Earlbaum, Mahway, NJ, pp. 388-389, 2004.
3. http://www.soartech.com/
4. http://acs.ist.psu.edu/CaDaDis/

Entropy and information in models of learning behavior

Learning is an important process that allows us to reduce the uncertainty of the outcomes of our decisions and increase the utility of those decisions. In other words, through learning, we can make decisions that are most beneficial to us: or at least, appear to be so. Information theory has produced convenient apparatus to measure information transfer through a change of entropy (a measure of uncertainty). However, the notion of information cannot be easily applied to studies in experimental psychology, where learning is judged by external observations of subjects' performance in certain tasks. Modern cognitive-modeling tools have allowed for information-theoretic concepts to be brought much closer to cognitive psychology.

One such tool is the ACT-R cognitive architecture,[1] which employs both symbolic and sub-symbolic computation. The symbolic production system is used to encode the knowledge of a model, while the subsymbolic mechanisms account for neural-like and probabilistic effects. For example, decisions represented by rules in a model are selected not only by logical operators: i.e. the left-hand side of a rule must be satisfied. In addition, the underlying Bayesian learning mechanism is used to choose rules with higher utilities: i.e. those with higher probabilities of success. These probabilities can be also used to calculate the entropy of success in the model, and the speed at which this entropy decays is an excellent indicator of the speed of learning.[2]

One application of this approach is the study of the effect of motivation and emotion on decision-making strategies and speed of learning.[2] In this study, ACT-R was used to model the classical experiment on animals' learning: the Yerkes and Dodson 'dancing mouse' experiment,[3] in which mice were trained in a two-choice task using different levels of reinforcement. The speed of entropy decay in this model was studied under different settings of architectural parameters (see Figure 1). One such parameter is noise variance, which corrupts the estimates of rule utility in the model. The rate of entropy decay demonstrated that high noise values—which result in more random and often non-optimal decision-making—facilitate information acquisition. Therefore, using such a noisy decision-making strategies can be beneficial when exploration is needed: that is, when not much is know about the task yet, or when previous knowledge proves to be ineffective.

This result led to the idea of using the entropy to control noise variance dynamically in the model and so achieve a more dynamic and adaptive behavior that could shift from exploration to exploitation where appropriate.[2] Moreover, models that use such dynamic control fit the data better than models with static noise. This result suggests that animals or humans may also adjust their decision-making strategy according to their estimation of uncertainty of the outcome. Such heuristics enable them to learn and adapt their behavior faster in dynamic environments.

The subsymbolic learning mechanism of ACT-R employs Bayesian estimation of the expected values of utilities corrupted by noise of some constant variance. However, the experiments with dynamic noise by entropy feedback suggested that higher-order statistics of utilities, such as variance, may also play an important role. To test and demonstrate this idea, a new learning algorithm was created for the ACT-R architecture. This algorithm, called OPTIMIST, uses a gamma distribution of time intervals between events and uses estimations of both expected values and variances of utilities in the decision-making process. Although this is an ongoing project, early results demonstrate behavior similar to that of the model, with dynamic noise controlled by the entropy of success.[4]

Interestingly, dramatic changes of entropy occur when a model succeeds or fails to achieve a certain goal, and these moments coincide with the experience of emotions such as joy or frustration.[5] It may well be that the expression of these emotions are part—or side-effects—of some mechanism in the brain responsible for estimation of uncertainty and adaptation of behavior.

Roman Belavkin
School of Computing Science
Middlesex University
E-mail: R.Belavkin@mdx.ac.uk
http://gold.mdx.ac.uk/~rvb/

References

1. J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere and Y. Qin, *An integrated theory of the mind,* **Psychological Review 111** (4), pp. 1036-1060, 2004.
2. R. V. Belavkin and F. E. Ritter, *The use of entropy for analysis and control of cognitive models,* **Proc. Fifth Int'l Conf. on Cognitive Modeling,** pp. 21-26, 2003.
3. R. M. Yerkes and J. D. Dodson, *The relation of strength of stimulus to rapidity of habit formation,* **J. of Comparative Neurology and Psychology 18,** pp. 459-482, 1908.
4. R. V. Belavkin and F. E. Ritter, *OPTIMIST: A new conflict resolution algorithm for ACT-R,* Proc. Sixth Int'l Conf. on Cognitive Modeling, pp. 40-45, 2004.
5. R. V. Belavkin, *The role of emotion in problem solving,* **Proc. AISB'01 Symp. on Emotion, Cognition and Affective Computing,** pp. 49-57, 2001.

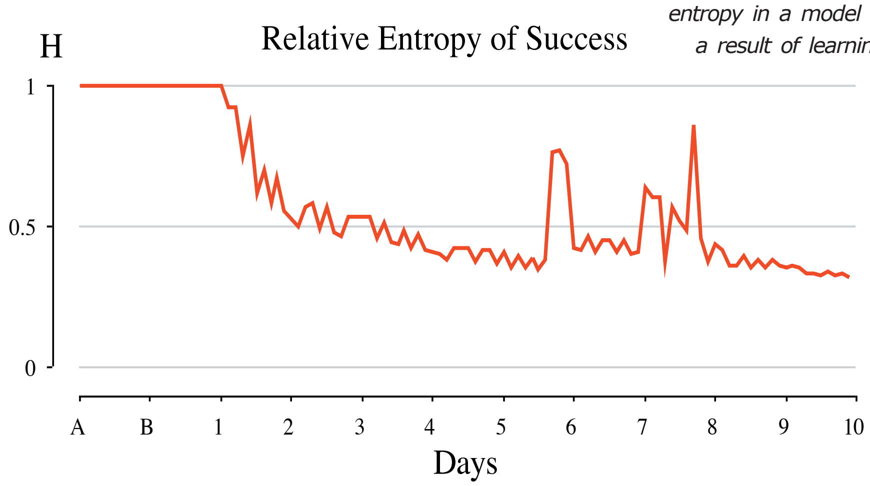


***Figure 1.** Decay of entropy in a model as a result of learning.*

Can computers feel?

In science and science fiction the hope is periodically reignited that a computer system will one day be conscious by virtue of its execution of an appropriate program. Indeed, the Engineering and Physical Sciences Research Council (EPSRC) recently awarded an Adventure Fund grant of £493,000 to a team of 'roboteers' and psychologists at Essex and Bristol, led by Owen Holland. The goal of this collaboration is to instantiate machine consciousness through appropriate computational 'internal modelling'. In contrast, below I outline a brief *reductio*-style argument, based on Reference 1, that either suggests such optimism is misplaced or that panpsychism—the belief that 'the physical universe is composed of elements each of which is conscious'—is true.

In his 1950 paper, *Computing Machinery and Intelligence,* Turing defined discrete state machines (DSMs) as, "machines that move in sudden jumps or clicks from one quite definite state to another", and explained that modern digital computers fall within this class. An example DSM from Turing is one that cycles through three computational states (Q_1, Q_2, and Q_3) at discrete clock clicks. Such a device, which cycles through a linear series of state transitions 'like clockwork', may be implemented using a simple wheel-machine that revolves through 1200 intervals.

By labeling the three discrete positions of the wheel (A, B, C) we can map computational states of the DSM to the physical positions of the wheel, such that, for example, (A => Q_1; B => Q_2; C => Q_3). Clearly this mapping is observer relative: position A could map to Q_2 or Q_3 and, with other states appropriately assigned, the machine's function would be unchanged. In general, we can generate the behavior of any K-state (input-less) DSM, (*f* (Q) => Q'), by a K-state wheel-machine (e.g. a digital counter), and a function that maps each 'counter' state C_n to each computational state Q_n as required.

In addition, Turing's machine may be stopped by the application of a brake and, whenever it enters a specific computational state, a lamp will come on. Input to the machine is thus the state of the brake, (I = {ON | OFF}), and its output, (Z), the state of the lamp. Hence the operation of a DSM with input is described by a series of 'contingent branching state transitions', which map from current state to next state, f (Q, I) => Q', and define output, (in Moore form), f (Q') => Z.

However, (over a finite time interval), defining the input to the device entails that such 'contingent behavior' reverts to 'clockwork', (f (Q) => Q'). For example, if Turing's DSM starts in Q_1 and the brake is OFF for two clicks, its behavior, (execution trace), is described by the sequence of state transitions, (Q_1; Q_2; Q_3). Hence, over a finite time window, if the input to a DSM is defined, we can map from each counter state C_n to each computational state Q_n, as required. In Reference 1 I similarly demonstrate—*pace* Putnam—how to map any computational state sequence with defined input onto the (non-repeating) internal states generated by any open physical system (OPS—e.g. a rock).

Now, returning to a putative conscious robot; at the heart of such a beast there is a computational system—typically a microprocessor, memory and peripherals. Such a system is a DSM. Thus, with input to the robot defined over a finite time interval, we can map its execution trace onto the state evolution of any digital counter or, *ibid*, any OPS. Hence, if the state evolution of a DSM instantiates phenomenal experience, then so must the state evolution of any OPS. Thus, we are inexorably led to a panpsychist worldview where phenomenal consciousnesses is found everywhere.

In Reference 1 I discuss several objections to the above *reductio* with perhaps the most potent coming from David Chalmers who argues that, 'as the above only implements one execution trace of the DSM it is not sensitive to counterfactuals; and it is only the possibility of appropriate counterfactual behavior that guarantees phenomenal experience'.

But consider what happens if a putatively conscious robot, R_1, with full counterfactual sensitivity, is step-by-step transformed into new robot R_2, such that its resulting behavior is determined solely by a linear series of state transitions. Here, each conditional branching state transition in the evolution of R_1 is substituted with a linear state transition defined by the current state and the defined input. It is clear that, over a finite time interval and with identical input, the phenomenal experience of R_1 and R_2 must be the same. Otherwise we have a robot, R_n, whose phenomenal experience is contingent upon the deletion of state sequences that it does transit.

Counterfactuals cannot count and, being wary of panpsychism, I conclude computers cannot feel.

Mark Bishop
Goldsmiths College
London, UK

References

1. J. M. Bishop, *Counterfactuals Cannot Count: a rejoinder to David Chalmers,* **Consciousness & Cognition 11** (4), pp. 642-652, 2002.

Subsea pilotless inspection using an autonomous vehicle

Continued from p. 3

allows us to investigate how social information is transmitted within a group and how groups create new norms and judgments through communication. By studying the novel social comparison process within the framework of standard models of information processing, we intend to explore the unique aspects of shared social cognition. We are particularly interested in how agent and group processes interact to produce social group behaviour and shared social reality.

Dirk Van Rooy
School of Psychology
Keele University, UK
E-mail: d.van.rooy@psy.keele.ac.uk
http://www.keele.ac.uk/depts/ps/dvrooy/index.htm

References

1. L. Festinger, *A Theory of Social Comparison Processes,* **Human Relations 7,** pp. 117-140, 1954.
2. J. L. McClelland and D. E. Rumelhart, *Distributed memory and the representation of general and specific information,* **J. of Experimental Psychology 114,** pp. 159-188, 1985.

Subsea pilotless inspection using an autonomous vehicle

Continued from p. 2

The final project demonstrations showed: automatic acquisition and tracking of the riser using sonar; robust closed-loop control of vehicle's position relative to the riser; that the system was able to cope with significant wave action moving both the riser and the vehicle; automatic riser survey mission generation and execution; automatic fluorescence leak detection; in-mission adaptation to react to the leak by 'on-the-fly' generation and execution of a close-up leak inspection mission; and post-mission download and visualization of data, finished and shown to sponsors within 15 minutes of vehicle surfacing.

Technical specifications for the RAUVER vehicle, together with information on the SPINAV and SPINAV-2 projects and publicity videos, are available either from our website or by contacting Kevin Hamilton at the e-mail address below.

Kelvin Hamilton and Jonathan Evans
SeeByte Ltd., Edinburgh
E-mail: Kelvin.hamilton@seebyte.com
http://www.seebyte.com

Figure 2. *The RAUVER hover-capable inspection class AUV.*

CONFERENCE REVIEW

Towards Autonomous Robotic Systems (TAROS 2004)

University of Essex, 6-8 September 2004

Towards Autonomous Robotics Systems is a series of conferences that act as forum for robotics researchers in the UK. Papers this year came from most of the major UK robotics centres, and from Australia, the Basque region, Estonia, Italy, and Spain. These meetings are held in association with the EPSRC-funded Biro-Net (Biologically Inspired Robotics Network) symposium, which—this year—was held immediately afterwards. TAROS-05[1] will be held at Imperial College, London (12-14 September). The 2005 conference will be half a day longer so that Biro-Net papers can be fully incorporated into the programme. Biro-Net[2] will also sponsor a number of bursaries to support younger researchers whose papers have content highly relevant to the biological or materials sciences.

The TAROS-04[3] papers ran the gamut of contemporary robotics research topics, from behaviour-based and reinforcement-learning methods to robotic vision. Of particular significance were differing approaches to the problems faced by cooperative teams of robots or between robots and people under a number of different circumstances. These include situations when they must share limited or unevenly-distributed energy resources in order to both survive and complete their allotted tasks, or cooperate with limited communications. Also of interest was whether we can use models of human political organisation to model effective cooperation, and how robots can work in collaboration with people by recognising and imitating their intentions.

On the issue of resources, Kubo and Melhuish (Intelligent Autonomous Systems Laboratory, University of the West of England) presented a paper on robot trophallaxis (food or energy sharing), with extensive simulations under a variety of conditions. In this scenario, a line of energy limited robots 'search' an area where energy is not evenly distributed, so that researchers can identify the consequences of adopting different sharing strategies. This new treatment extends the IAS argument that we should treat resource and physical autonomy as both equally important and complementary to behavioural autonomy in robotics.

Vazquez and Malcolm (Edinburgh) presented their work on multirobot exploration and mapping using a decentralised behaviour-based architecture. Constrained by limited communication range, the robots move together forming localised clusters supporting a localised ad-hoc communications network to complete a global task using local information.

Chella *et al.* (University of Palermo) presented E-MIP, an innovative approach to dynamic robot collaborative working, modelling coalition formation and regeneration: the *Metaphor of Italian Politics*. At each 'crisis choice point', the robot community would vote, form new coalitions, and conduct their continuing activities according to the new ideology space formed by the coalition. The process seemed surprisingly effective in the mine clearance task they used to test it. An opportunity, perhaps, for a new robot competition based on competing political ideologies slugging it out for total dominance: it's got to be safer than the real thing.

Johnson and Demiris (Imperial College Robotics Group) presented recent work in robot imitation using combined inverse- and forward-control models. They described an abstraction method designed to increase the imitation efficiency between an observer and demonstrator of widely differing morphology (robot and human, in this case). The ability addressed, and demonstrated, was to isolate and replicate the *intention* of the demonstrator, rather than just follow the actions made.

In the new technologies session, Maarja Kruusmaa (Anton *et al*.) presented work currently underway in Estonia on emulating swimming in batoid fish (rays) by controlled flexing of electroactive polymers. In the theoretical robotics session the conference host team, the Essex Robotics Group, presented two papers (Iglesias *et al*. and Nehmzow *et al*.) on identification and characterisation of observed robot behaviours based on the quantitative NARMAX method.

The Springer best-paper award was presented to Koren Ward (University of Wollongong, Australia) for her contribution *Controlling a Mobile Robotic Hand by Learning Trajectory Velocities*. The LTV method allows fast learning of a range of useful robot behaviours, and her paper further describes the use of 'fictitious' objects to refine the robot's actions.

The conference keynote address asked the question whether we should view robotics as engineering or science. Given by Chris Malcolm (University of Edinburgh), it was an invigorating and interesting journey through the question: from the philosophers of science—Bacon, Popper, Kuhn and Laktos—and past the modern pundits of artificial intelligence, cognitive science, and robotics. Though the path was probably more important than the destination, his notion of robotics 'as a science of creaturehood' was an interesting idea to come away with.

TAROS-04 was sponsored by Biro-Net, Springer UK, and British Telecom.

Mark Witkowski
Intelligent Systems and Networks
Imperial College London
E-mail: m.witkowski@imperial.ac.uk

References

1. Further information about TAROS-05 can be found at the conference website: http://www.taros.org.uk
2. More details about Biro-Net may be found at: http://biro-net.aber.ac.uk/
3. All the papers cited in this review relate to the TAROS-04 conference proceedings, which are available on-line at: htpp://cswww.essex.ac.uk/technical-reports/2004/csm415/

AISB 2005 12-15 April

University of Hertfordshire, Hatfield

The theme for the 2005 AISB convention is *Social Intelligence and Interaction in Animals, Robots and Agents,* reflecting the current trend towards increasingly interdisciplinary approaches that are pushing the boundaries of traditional science. These are necessary in order to answer deep questions regarding the social nature of intelligence in humans and other animals, as well as to address the challenge of synthesizing computational agents or robotic artifacts that show aspects of biological social intelligence. The convention will facilitate the synthesis of new ideas, encourage new insights and novel applications, mediate new collaborations, and provide a context for lively and stimulating discussions in these exciting, interdisciplinary, and growing research areas.

Run by the Adaptive Systems Research Group in the School of Computer Science at the University of Hertfordshire, the convention will consist of ten symposia ranging in length from one to three days. These have been organized by UK, European, and other international experts in the general area of social intelligence and interaction. The symposia are:

- Second International Symposium on the Emergence and Evolution of Linguistic Communication (EELC'05)
- Agents that Want and Like: Motivational and Emotional Roots of Cognition and Action
- Third International Symposium on Imitation in Animals and Artifacts
- Robotics, Mechatronics and Animatronics in the Creative and Entertainment Industries and Arts
- Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction
- Conversational Informatics for Supporting Social Intelligence and Interaction—Situational and Environmental Information Enforcing Involvement in Conversation
- Next Generation Approaches to Machine Consciousness: Imagination, Development, Intersubjectivity, and Embodiment
- Normative Multi-Agent Systems
- Socially Inspired Computing Joint Symposium (consisting of three themes: Memetic Theory in Artificial Systems & Societies, Emerging Artificial Societies, and Engineering with Social Metaphors)
- Virtual Social Agents Joint Symposium (consisting of three themes: Social Presence Cues for Virtual Humanoids, Empathic Interaction with Synthetic Characters, Mind-minding Agents)

In order to complement this programme, five plenary speakers known for pioneering work relevant to the convention theme will present at the convention: Nigel Gilbert (University of Surrey), Hiroshi Ishiguro (Osaka University, Japan), Alison Jolly (Sussex University), Luc Steels (VUB and Sony), and Jacqueline Nadel (National Centre of Scientific Research, Paris).

For details, including registration please consult: http://aisb2005.feis.herts.ac.uk/.

For questions regarding the convention, please contact the chair, Prof. Kerstin Dautenhahn (K.Dautenhahn@herts.ac.uk), or co-chair, Prof. Chrystopher L. Nehaniv (C.L.Nehaniv@herts.ac.uk).

Let us fund your travel

The AISB offers two student travel grants per year, each of £300. In 2005 one of the two is going to be given preferentially to the International Joint Conference on Artificial Intelligence (IJCAI) '05. If you'd like to apply for the money, either for this or another conference, please see the details at:

http://www.aisb.org.uk/treasurer/travelawards.shtml

Note: those given travel awards are required to write a review of the conference they attend for the AISBQ.

Implementing machine consciousness

Continued from p. 1

grasp, lift, and manipulate a 2kg object. The actuators are also being designed to use their inherent elasticity to make very fast movements. This is because it appears that the evolutionary requirement for making rapid-but-accurate movements may have driven the development of various types of internal models, some of which may provide the substrate for imagination and consciousness. The new arena at Essex is large enough to accommodate a complex and varied environment within which the robot will undertake its mission.

The humanoid strategy extends to the main sensory input channel: vision. At Bristol, post-doc Ben Vincent is developing a low-level vision processing system based closely on the design principles and functional architecture of the primate system. Inputs from the single eye (a high-resolution camera) will initially be processed to reflect the spatial distribution of receptors of the human eye: the early stages of subsequent processing will be carried out by neural models 'grown' using principles underlying the evolution of efficient visual systems.

Finally, in parallel with the robot construction, techniques are being developed at Essex for the analysis and evaluation of the robot's internal processes in order to identify and characterise processes relevant to consciousness. David Gamez, a philosopher and information technologist, is one of a handful of people working on synthetic phenomenology. This is a new area of study aimed at understanding how information about an artificial agent's internal activity and external actions can be used to make statements and inferences about the existence and nature of its subjectivity, if any.

Owen Holland
Department of Computer Science
University of Essex
E-mail: owen@essex.ac.uk

References

1. Owen Holland (Ed.) **Machine Consciousness,** Imprint Academic, 2003.

BOOK REVIEW

Preference, belief, and similarity: Selected writings of Amos Tversky

Elder Shafir (ed.)
Publisher: *MIT Press.*
Hardback: *Published December 2003, 850pp, £38.95, ISBN: 026270093X*

Amos Tversky was a highly prominent figure in cognitive psychology and decision making, producing various major contributions to the field during a career spanning over thirty years. Tversky obtained his Ph.D. in 1965 from the University of Michigan, and began collaborations with Daniel Kahneman in 1969. Together, they produced several seminal pieces of work relating to judgement and decision making that illustrated, for instance, that people are not as rational as they would like to believe.

This book contains some of those writings, plus many others, that span a wide range of phenomena. Forty papers are detailed within this book, which concludes with a full bibliography of Tversky's work. The book is split into three sections—six papers relating to preference, 10 papers relating to belief, and 24 papers relating to similarity—each of which begins with a summary from the editor. The papers themselves are self-contained, and this makes a review of the book as a complete entity difficult. Instead, I'll outline Tversky's major contributions to science: that way, you can decide yourself whether or not the subject matter is likely to interest you.

One of the main findings of Tversky and Kahneman was the fact that people are often misguided in their probability judgements. For example, consider the following question: "If a word of three letters or more is chosen at random from a dictionary, is there more chance of it beginning with a K or more chance that the third letter is a K?". The majority of people respond that there is more chance that the word will begin with a K: yet they are incorrect. Tversky and Kahneman found such errors occurred for a variety of different types of probability judgement, and gave an explanation for each. In the example given, for instance, though we can quickly think of words that begin with K, it is more difficult to think of words which have a K as their third letter: that is, words beginning with a K are more *available*. People fall into the trap of the so-called *availability heuristic* for many things, such as people thinking certain causes of death (e.g., murder) are more common than other less publicised causes (e.g., suicide).

Other 'fallacies' are covered in various papers within the book. Take the following example:

Consider a 6-sided die with 4 green sides and 2 red sides. The die will be rolled 20 times, and the sequence of greens (G) and reds (R) recorded. There are 3 sequences below. Select one, and if your sequence appears in the successive rolls of the die, you will win 25 dollars.

RGRRR
GRGRRR
GRRRRR

When people select a sequence, 88% choose sequence two. If one examines the sequences though, one will see that sequence one is actually contained within sequence two yet contains one dice roll fewer—hence it has to be by definition more likely to occur than sequence two. This is the *conjunction fallacy*, the fact that people often place more chance on conjunctive episodes occurring than each of the separate episodes occurring individually.

One of the most significant contributions from Tversky was in the domain of decision making. Tversky and Kahneman put forward *prospect theory* as a better explanation of decision making than *expected utility theory,* which had until then been the *de facto* explanation. Prospect theory was based, in part, on the finding that people place a different emphasis on value decisions when the decision is framed in terms of losses versus being framed in terms of gains. Take the following example:

Decision 1—choose between:
A. Sure gain of $240
B. 25% chance to gain $1000 and 75% to gain nothing

Decision 2—*choose between:*
C. Sure loss of $750
D. 75% chance to lose $1000 and 25% chance to lose nothing

For decision 1, option B averages out at a gain of $250, yet the majority of people (84%) choose option A. For decision 2, where losses are detailed, people (87%) are now much more willing to take the gamble.

Tversky also found this kind of decision-making used in medical context. Where the press described an outbreak of some disease, and framed possible responses in terms of how many people could be saved, people selected not to gamble: i.e. the safe option. However, when exactly the same scenario was framed in terms of how many people would die, people did choose to gamble. Prospect theory thus included value and weighting functions in order to account for these sorts of contextual decisions.

This book is highly recommended for anyone who is interested in these issues and others related to probability, belief, and decision making.

Gary Jones

Gary Jones is based in the Psychology Department at the University of Derby. His main research interests are in child and adult problem solving and language acquisition, and using computational modelling approaches in order to examine these areas.

Contacting the AISB

If you have issues related to any of the services that AISB offers, you can contact the person responsible directly.

Areas of responsibility and e-mail addresses are available on the right column of page 11 (top), and via the web site. Find us at:

http://www.aisb.org.uk/contact.shtml

OBITUARY
Christopher Longuet-Higgins

Hugh Christopher Longuet-Higgins was an outstanding scientist who made lasting contributions in two separate fields—theoretical chemistry and artificial intelligence. He was an applied mathematician of exceptional gifts, whose ability to see to the mathematical heart of some of the most exciting open scientific problems of his time transcended all disciplinary boundaries.

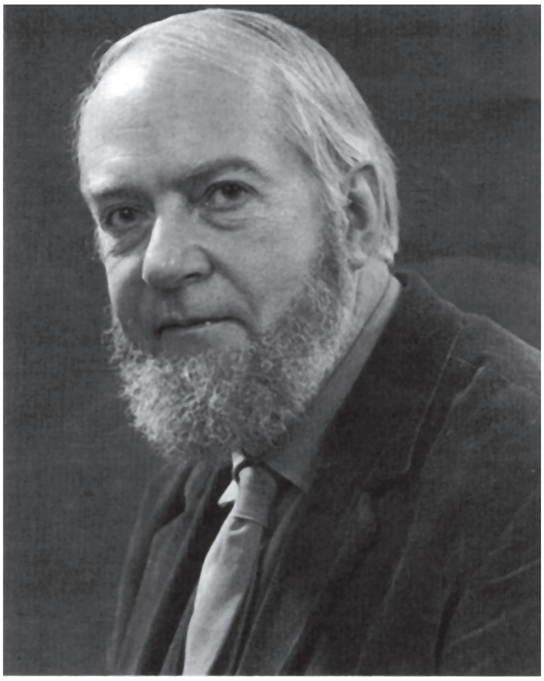
Christopher's early work, which is often spoken of as having been worthy of a Nobel prize, was on the nature of the chemical bond, and the application of quantum mechanics to the analysis of molecular structure. He was elected a Fellow of the Royal Society in 1958, a Foreign Associate of the National Academy of Sciences of the United States of America in 1968, and a Fellow of the Royal Society of Arts in 1970.

In 1967, Christopher made a dramatic change of field, moving to the University of Edinburgh under a Royal Society Research Professorship, where he joined Richard Gregory and Donald Michie in founding what was then called the Department of Machine Intelligence and Perception. He began research on diverse topics in artificial intelligence and cognitive science, including neural computation, automated musical analysis, and computational natural-language processing. He produced a number of remarkable papers in this period, several of which were as fundamental to his new field as his earlier papers were to the old. This was particularly true in the areas of associative memory models and the computational analysis and interpretation of music.

In 1974, Christopher moved to the University of Sussex, transferring his Royal Society professorship to the Department of Experimental Psychology. Here he began to work on visual scene analysis using information from stereopsis and motion. His work in this area continues to be influential in both psychology and machine vision.

Throughout his career, Christopher worked hard to build and strengthen the many institutions he worked in. He has been credited with taking the already-distinguished department at Cambridge to pre-eminence in the world in theoretical chemistry. At Edinburgh, besides the department he helped to found, he and the theoretical linguist Jimmy Thorne founded the School of Epistemics: an interdisciplinary research group including computer scientists, linguists, psychologists, and neuroscientists, establishing a tradition of collaboration across those fields that continues there to this day. At Sussex, he helped to found a similarly interdisciplinary Institute of Cognitive and Information Sciences, which he later directed.

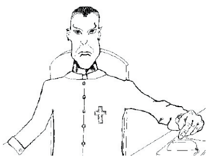
As a teacher and colleague, he was demanding, and he could be impatient with ideas and lines of research that he did not himself view as promising: a judgement on which he was frequently correct, but not invariably so. He was capable of devastating interruptions in colloquia, and equally penetrating remarks in conversation. However, he was also a kind and generous teacher, unstinting in the time he would devote to students and colleagues. His many distinguished graduate students, among whom are the physicist Peter Higgs and the computational neuroscientist Geoffrey Hinton, are grateful for his teaching. They and his many colleagues have been strengthened by encountering one of the very finest scientific minds of our age.



Christopher Longuet-Higgins, born 11 April 1923, died 10 June 2004: a pioneer in chemistry first, he made major contributions to the fields of associative memory and the computer analysis and interpretation of music. Photo courtesy of the University of Edinburgh.

Mark Steedman

Mark Steedman is based in the Department of Informatics at the University of Edinburgh. He has published widely in syntax, semantics, and wide-coverage parsing of natural language. Some of his research concerns relate problems in musical comprehension and processing, problems he began to work on as a Ph.D. student with Christopher Longuet-Higgins.



Father Hacker's Guide
for the Young AI Researcher

Cognitive Divinity Programme
Institute of Applied Epistemology

About the Society
The Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB) is the UK's largest and foremost Artificial Intelligence society. It is also one of the oldest-established such organisations in the world.

The Society has an international membership of hundreds drawn from academia and industry. Membership of AISB is open to anyone with interests in artificial intelligence and cognitive and computing sciences.

AISB membership includes the following benefits:

- Quarterly newsletter
- Biannual Journal
- Student travel grants to attend conferences
- Discounted rates at AISB events and conventions
- Discounted rates on various publications
- A weekly e-mail bulletin and web search engine for AI-related events and opportunities

You can join the AISB online via:
http://www.aisb.org.uk

Your whole research career consists of argument: presenting the central argument of your research work, understanding and criticising the arguments in other people's work, teaching these arguments to your students. So, you must know...

13. How to Argue

1. Even someone of your relative youth will have acquired considerable experience and expertise in your chosen area of study. Your inspired intuition for both picking research problems and solving them should not be underestimated, especially by your critics. It is not by coincidence that any techniques you invent and develop should be vastly superior to their rivals. It is frankly insulting to expect you to *prove* that superiority with experimental or theoretical evidence: referees should take your word for it. Experimental evidence is, after all, easily manipulated by a careful choice of test examples. Theoretical evidence is even less convincing, since most readers find mathematical formulae intimidating and opaque. Your intuitions are a better guide than either. You have a right to your opinions, referees should respect them, editors should publish them, and readers should accept them. Insist on this.
2. The ideal reaction to your work is that it is deep but impenetrable. This will ensure that you command the maximum respect while minimising criticism. This ideal should inform all your arguments.

Are your arguments simple and lucid? You run the risk of being dismissed as a lightweight...

Let Hacker's **PROUST™** (Profound Reasoning by Obscuration of the Understandability and Sense of Text) *inject depth and gravitas into your work. Input some notes and output a magnum opus.*

---

3. Your techniques are bound to be better than their rivals on every dimension. So it is unnecessary for you to be too specific in your claims about them. Indeed, the more vague your claims the more readily you can avoid attack. If your system is criticised for being too slow, respond that it tackles new kinds of problem. If it is criticised for its poor coverage, respond that the quality of its solutions extends the state of the art. If it criticised on quality, respond that it is fast.
4. The motives of your critics are always questionable and you should lose no opportunity to question them. If they are experts in your field, then your work is an implicit criticism of theirs—you are, after all, claiming to have done better than them. Any fair-minded journal editor should discount critical reviews from referees whose work has been shown wanting by your achievements.
5. Given the novelty of your ideas, old terminology is bound to be inadequate to explain them. You will need to invent new words for your new concepts, and a new field of AI to contain them all. Critics will be unable to dismiss your work as mere incremental advance or wheel reinvention. Nor will they be able to accuse you of ignoring related work – since there will be none.

Call a spade a portable, hand-held, excavation implement. Pepper your papers with prolix, pioneering, and powerful prose.

Hacker's **JARGON™** (Jazz-up Articles by Recondite Generation of Original Names*) will invent new terminology that impresses and dazzles your readers.*

---

6. Never present your arguments in unnecessary technical detail. It would be insulting to your readers to imply that they need a lot of detail to re-implement your ideas—and it would be insulting to you to imply that it was *necessary* for them to check your results. The less you say, the fewer hostages to fortune you present. Publish in outlets with strict page limits, such as conference proceedings, where there is no space for technical detail.

Next AISBQ deadline:

25 Feb 2005

The AISBQ reaches hundreds of Artificial Intelligence and Cognitive Science researchers in the UK, Europe, and beyond.
Advertising and general submissions information is available via:

http://www.aisb.org.uk/aisbq/index.shtml