AISB Quarterly

THE NEWSLETTER OF THE SOCIETY FOR THE STUDY OF ARTIFICIAL INTELLIGENCE AND SIMULATION OF BEHAVIOUR

No. 143 June 2016

About the Cover

Cover adapted from a video screenshot from A-EYE: Exhibition of art and nature-inspired computation, AISB50.

Harold Cohen died in his studio in Encintias, California, on 27 April. He was 87.

Mohammad Majid al-Rifaie, AISB committee member, writes:

"The first time I heard his name mentioned was when a colleague said, 'you cannot open a textbook on computer art without finding his name prominently'. I was just starting my PhD.

"I had the great honour of introducing him as the invited artist for an exhibition at Goldsmiths, University of London, on the opening night of the 50th annual convention of the AISB. He was extremely generous with his thoughts and time, and was incredibly approachable to artists and academics, both the experienced ones and those just starting in the field. He was the undisputed star of the evening.

"Cohen had a major reputation as a painter in the Sixties, representing the UK in the Venice Biennale, documenta and other international spaces. In 1968 he went to California on a one-year visiting professorship at UC San Diego, became involved in computing, and stayed on to build a second reputation as a pioneer in the application of computing in the arts. His celebrated AARON program was begun in 1972 while he was a Visiting Scholar at Stanford University's AI Lab, and together they have exhibited at the Tate, the Brooklyn Museum, the San Francisco Museum, the LA County Museum, documenta 6 and too many others major spaces to list here.

"I vividly remember the invaluable hours we spent together that last time I saw him. He gave me his firm handshake before walking away, leaning on his wooden cane, his charisma intact."

Maggie Boden, research professor of cognitive science at the University of Sussex writes:

"Harold Cohen died surrounded by his artworks and art-making machines.

"He was a brave, generous, and hugely talented man – and a Titan of computer art. Those of us who were personal friends will miss him hugely. We'll reminisce fondly about his directness and sense of humour, but the memories will be tinged with great sadness.

"It would be normal, at this point, to say that he was irreplaceable. And of course, in an important sense, he was. But yet.... We shall have the bitter-sweet pleasure, over the coming years, of welcoming new works of art created by some version of his art-making program, AARON. He himself used to say that this was a comfort to him: unlike other artists, his oeuvre wouldn't be completed by his death."

Editorial

If the tumultuous events of recent days show anything, following the UK referendum vote to leave the EU ("Brexit"), it is the unpredictability of mass behaviour – save with benefit of hindsight. For both those who voted "leave" and those who voted "remain", a sense pervades that something has fundamentally shifted that cannot be reversed – leaving many with the sense that suddenly *anything* is possible, the most stable of certainties uncertain.

If mass behaviour is often unpredictable, that of the individual is even more so – starting with ourselves. It is our capacity to surprise ourselves that is, perhaps, most remarkable. Yet I have heard a prominent researcher in the field claim, with apparent seriousness, that unconscious cognition does not exist, that all thought is conscious thought. One finds a surprising lack of consensus even whether cognition needs to be or should be *embodied*: consider the strongly symbolic framework of much of artificial general intelligence (AGI) research or the hope – espoused by many transhumanists – that the near future may bring the possibility of recording one's mind and uploading it into an artefactual body, putting an end to death as we know it.

For more than half a century, this Society has struggled with the questions of the nature of human cognition, the relation between human and non-human cognition, the relation between cognition and consciousness, the possibility of capturing these things perfectly or imperfectly in artefactual form – the better to predict the unpredictable. Through models of mind, we seek to understand that most quintessentially self-referential of conundra: for there is no explanation of mind that does not, at the same time, presuppose that the agent or agents seeking to make that explanation already *have* a mind.

With a particular eye to Attachment Theory, Dean Petters considers the competing and complementary perspectives from the "4Es" of cognition – embodied, embedded, extended, and enactive – all of which seek to challenge the various "disembodied" views. Mattias Savallampi speculates on the conceptual unclarity surrounding depression in light of discussions at the recent AISB convention about taking a more embedded, embodied, and enactive view on it. In his review of George Zarkadakis' *In Our Own Image*, Robert Waltham considers our efforts to not just model but *re-create* human cognition and consciousness; while Mohammad Majid Al-Rifaie reviews K.L. Downing's recent *Intelligence Emerging*, which seeks to explain embodied cognition's enduring capacity to surprise, and Petters looks at Angelo Cangelosi and Matthew Schlesinger's survey of modeling embodied, autonomous cognitive agents through developmental robotics.

Joel Parthemore
Managing Editor

29 June 2016, Skövde, Sweden

A Brief Encounter between Attachment Theory and 4e Cognition

Dean Petters **(Birmingham City University, *D.D.Petters@cs.bham.ac.uk*) and**

Everett Waters **(SUNY Stony Brook, USA)** [1]

Abstract

A number of research questions arise from an encounter between the elements of 4e cognition and Attachment Theory. These include (1) whether the Attachment Theory concept of *internal working models* should be understood in terms of analogue representations more in line with embodied cognition, in addition to traditional cognitivist representations like linguistically mediated narrative measures of attachment meaning; (2) whether infant-carer dyads are best thought of as environments of contextual embedding for infant cognition or as an arrangement where the carer can actually extend the infant mind; and (3) whether attachment phenomena are best thought of in traditional representational terms or whether the attachment control system should be reframed in enactive terms where traditional cognitivist representations are (3a) substituted for sensorimotor skill-focused mediating representations, (3b) viewed as arising from autopoietic living organisms or (3c) mostly composed from the non-contentful mechanisms of basic minds. A theme that cross-cuts these research questions is how representations for capturing meaning and structures for adaptive control are both required to explain the full range of behaviour of interest to Attachment Theory researchers. Implications are considered for future empirical and computational modelling research and for clinical interventions.

Introduction

The infant-caregiver relationship not only plays a central role in social and emotional development but also in exploration and learning [3, 9, 10]. A traditional cognitivist approach to explaining these phenomena would emphasise internal information processing, located within the individual mind. So this approach in Attachment Theory would focus on what is or should be in the infant's head. A theoretical approach that keeps cognition within the infant is seductive because of its conceptual simplicity and because this approach is more easily implemented in cognitive models that focus on the creation and transformation of internal representations [17, 18, 19]. The elements of 4e cognition – viewing cognition as embodied, embedded, extended, and enacted – all reject or radically reconfigure traditional cognitivism [15]. Whilst the core ideas in Attachment

[1] A longer version of this paper was presented at AISB 2015 and a significantly extended version of the paper has been accepted for an upcoming special issue of the AISB journal *Connection Science.*

Theory were set out by John Bowlby in a series of papers and books between 1958 and 1982 [2, 3, 5, 6], the elements of 4e cognition are more recently defined [15] but have many earlier conceptual antecedents [8, 11, 26]. How should Attachment Theory respond when viewed through the lens provided by 4e cognition approaches in cognitive science? And which elements of 4e cognition provide the best match for the requirements of a theoretical revision for Attachment Theory?

The embodied approach views the body and physical world as the "context or milieu" for cognition, rather than conceiving cognition as the operation of disembodied algorithms [20]. So an encounter between Attachment Theory and embodied cognition asks how attachment representations should be conceptualised, and whether the cognitive component of Attachment Theory could then be "augmented with the incorporation of bodily sensations, physiological responses, and analogue computations that rely on the physical substrate within the attachment control system" [20]? The hypotheses of embedded and extended cognition are competing theories in situated cognition that both give greater emphasis to the role that situations and context play in human cognition than traditional cognitivism. The extended approach is more radical, claiming that external supports become part of a person's cognitive apparatus. The embedded approach is still strongly anti-cognitivist, but sees cognition embedded in external support rather than constituted of external structures. A key question is whether attachment relationships can sometimes be conceived as extending cognition or are better thought of as embedding cognition.

The enactivist approach views psychological activity as occurring in the dynamic engagement between organisms and their physical and social context rather than within themselves [14]. The mind and subjective experience are not seen as inherent in or arising from the individual, but as emerging from the interaction between organisms and their surroundings [14]. So another two key questions are (1) is enaction – rather than traditional forms of representation – a better way to think about how previous experiences mediate ongoing adaptive behaviour and (2) can the attachment control system be revised to act as an enactive "lived experiential structure" [20, 25, p. xvi]?

The intention in challenging Attachment Theory with recent ideas from 4e cognition is to revise rather than replace or reject it, and also see which diverse elements from 4e cognition can operate in "joint purpose", motivating a progressive revision of a well established theory.

Should Internal Working Models Be Viewable as Analogue in Addition to Symbolic Representations?

Internal working models (IWM) are described by Bowlby as higher level representational forms which integrate and exert control over lower level control systems. Their principal information processing function is to allow predictions to be made about the likely outcomes of taking actions within a given

environment. IWMs transmit, store and manipulate information and allow the individual to "conduct small scale experiments within the head" [3, p. 81]. Their function, in terms of Bowlby's agenda of reforming psychoanalytic theory, was to take the place of the internal worlds of traditional psychoanalytic theory. Bowlby emphasises the requirements for IWMs to be updated. He also briefly observes that pathological sequelae of separation and bereavement can be understood in terms of out-of-date or half-revised models that may contain inconsistencies and confusions [3, p. 82]. Bowlby invokes IWMs at early stages in development also later on, when linguistic skills and conscious reflection can enable models to become more adequate [3, p. 84]. In contemporary reviews, IWMs are presented as transforming from sensorimotor representations in pre-linguistic infants to manipulable internal simulations in older children and adults that can enable short-term predictions and conscious reflections on past, ongoing and future relationships [7, p. 102]. Current research investigates IWMs through studies of memory talk, narrative completion, semi-projective measures and story-telling with adults and children [7] – naturally linking IWMS to symbolic constructs from artificial intelligence like schemas and scripts. In his later writing Bowlby described IWMs in symbolic terms [4, p. 373].

However, links have also been drawn with IWMs and recent neuroscience research based upon mirror neurons that present IWMs as affording embodied simulation of the intentions of others [7, p. 109]. Though research viewing IWMs as embodied simulations is very much in the minority in contemporary attachment research on IWMs [7, 23], this view is not only fully in the spirit of Bowlby's original conception for IWMs but also matches the "word" of what he wrote about IWMs when he first introduced them. Bowlby did not use the term "embodied simulation" but he did compare IWMs to analogue representations. For example, in his 1969 formulation of IWMs, Bowlby suggests that they can be used to conduct "small-scale experiments within the head" and notes that this notion would be an obvious possibility to electrical engineers familiar with analogue computers. Bowlby also refers to how anti-aircraft guns operate [3, p. 44] to exemplify how analogue control systems can set their own goals.

Are Infant-Carer Dyads Best Described in Terms of Cognitive Embedding or Cognitive Extension?

The idea that infants, older children and even adult attachment partners all look to their carers as information sources about the broader world is a familiar one. For example, from the perspective of the socially situated mind, infant social referencing and joint attention between infant and carer may be seen as physical actions that make the infant's mental computations faster, more reliable or less effortful by intimately linking internal infant cognition with external support [22]. So taking a situated cognition approach enriches Attachment Theory by providing

a more complete view of how infants gain information about environments from their caregivers. The hypotheses of embedded and extended cognition are competing and mutually exclusive explanations for how caregivers provide cognitive support. The hypothesis of extended cognition suggests that in some of the above examples, if the infant's ongoing computational needs are met by sensitive and timely support from its carer in such a way that the infant treats this support as part of its own cognitive processes then we might say that the carers' cognitive support has become part of the infant's extended mind. For these examples to count as mind extension, caregiver cognitive support and information provision to the infant must be strongly trusted, relied upon, and accessible. If these criteria are met then what is occurring is extension of mental states from an infant onto their caregiver. In this view, the carer is actually extending the infant mind by incorporating the carer's help within the infant's cognitive operations: the carer's help becomes part of the infant's mind [22]. For these same examples of intimately integrated interactions between infant cognition and carer support, the hypothesis of embedded cognition views infant cognition and carer support of that cognition as clearly demarcated and separate. This hypothesis considers that "cognitive processes depend very heavily, in hitherto unexpected ways, on organismically ... external props and on the structure of the external environment in which cognition takes place" [24, p. 393] and that "certain cognitive processes lean heavily on environmental structures and scaffoldings. but not thereby include those structures and scaffoldings themselves" [10, p. 111]. We should be more accepting of claims to extended cognition in infants and younger children because the caregiver's interactions are more long-lasting; they are relied upon more; there are fewer infant cognitive resources and routines for not believing [12] – so making acceptance of information from the carer as if it were an infant's own beliefs easier and more likely.

Two main reasons for preferring embedded over extended explanations arise from considering non-social cognitive extension [10]. Most examples of extended cognition involve inorganic objects in the environment (such as a mathematician doing their "working" on paper) providing the cognitive extension. The first criticism of extended cognition highlights the profound differences that appear to distinguish inner and outer contributions in extended cognition when cognition is extended onto such inorganic objects [10]. However, this criticism is much weaker when applied to the social case, as it is a carer that does the extending. So there are not such profound differences in the supporting substrate for cognition between cognition inside the infant's brain and cognitive support originating from inside the carer's brain. A second criticism is the apparent scientific cost of any wholesale endorsement of extended cognition onto a motley collection of inorganic objects, because it gives undue attention to transient external props and aids. In this view,

following the extended mind hypothesis means scientists are not researching a suite of integrated persisting organismically grounded capacities [10, 27], but rather looking at developmental examples of cognitive extension onto inorganic objects as a series of separated developmental segments with external cognition onto different objects. So using a ball or balance beam may be a good example of mind extension at one age, but a year later the best example may involve a completely different object in a different task or action. Again, the social case of mind extension mitigates this criticism.

Extended cognition does not only deal with transient external props and aids when the carer provides enduring support and continuity between otherwise disparate contexts. If we accept the hypothesis of extended cognition over the hypothesis of embedded cognition, this has important implications for computational modelling and clinical interventions. Care-giving relationships are often very durable and reliable; and, if socially extended cognition occurs, we can expect typical interactions and development to include micro- and macro-instances. Micro-extension effects are described by Clark: "the child is surrounded by exemplars of mindreading in action, she is nudged by cultural interventions such as the use of simplified narratives, prompted by parental rehearsal of her own intentions, and provided with a rich palate of linguistic tools such as words for mental states" [10, p. 67]. Macro-effects occur when children absorb complex ideas wholesale through the conduit of cognitive extension. Their caregivers can simply present beliefs that the children then adopt. Over the long term caregivers attempt to socialise and indoctrinate infants in many ways that will impact the developing meaning a child gains of its attachment history. Two types of problems can occur: (1) relationships are not reliable or durable enough, so infants and children do not gain the benefits of cognitive extension; and (2) pathological extension occurs: instead of acting to scaffold or co-construct, a caregiver uses her power to extend an infant's mind to introduce (or, put more strongly, "infiltrate" or "hack" [16]) unhealthy or pathological beliefs about the infant's self and relationships into the infant's mind.

Enactivising Attachment Theory

Where the extended/embedded question highlighted the requirement for attachment structures and mechanisms that support narrative meaning making the three flavours of enactivism highlight different aspects of adaptive control and subjective experience in the attachment domain.

Attachment theory encounters sensorimotor enactivism

Sensorimotor enactivism criticises the view that perception results in inner images or mental representations. In the sensorimotor view, perception, action, and subjective perceptual experiences are all inescapably connected [13]. Perceptual experience is grounded in knowledge and therefore representationally contentful. But the kind of mediating knowledge in sensorimotor en-

activist accounts is more like procedural or skill-based knowledge. It is "know-how" rather than "know-that": a kind of knowledge demonstrated by the skilled performance of its deployment rather than an independently queriable knowledge base [13]. Viewing attachment behavioural patterns in this enactivist manner – as social skills rather than a result of internal representations – may provide a powerful spur towards new research hypotheses and clinical interventions. When individuals with insecure attachment gain secure status, they can be viewed as gaining a skill which they can then use in other relationships.

Autopoiesis and representation from social interaction

According to autopoietic enactivism, cognition, mentality and subjective experience all emerge from the self-organising and self-creating activities of autonomous entities [13]. This activity is intimately spread between organism and environment. Enactivists suggest that, because factors from "within" and "without" play equally important and necessary roles in creating cognition and behaviour, the distinction between organism and environment is viewed as only having a heuristic value rather than being a true metaphysical division [13].

Autopoiesis is a special case of homeostasis. It takes the position that metabolism and life are essential for grounding intentional categories like cognition, consciousness, and emotions [1]. In the second volume of his Attachment trilogy, Bowlby adopted the biological concept of homeostasis and applied it to behavioural as well as physiological control systems. In this view, physiological homeostasis, which regulates food and sleep, is an inner ring of control in the attachment control system. Attachment behavioural patterns constitute an outer behavioural ring that is a complement to this inner physiological control system [5, ch. 9]. However, Bowlby did not set out how the intimate engagement of these two rings could give rise to phenomenological experience. He did describe attachment feelings, but within an emotional appraisal framework [3, ch. 7]. Viewing Attachment Theory through the lens of autopoietic enactivism can act as a spur for a more comprehensive approach that unifies behaviour, cognition, and subjective experience in a single explanatory framework.

A Radical Enactivist Manifesto for Attachment Theory?

Hutto and Myin propose the thesis of radical enactive cognition (REC): a variant of enactivism that states that only a small proportion of cognitive processing is mediated by contentful representations. The majority of human cognition is basic and noncontentful information processing that controls behaviour for adaptive purposes but does not possess truth-bearing properties like reference, accuracy or implication. According to REC, contentful representations mediate some cognition, but these representations play a minor role in cognition overall, "emerging late in phylogeny and ontogeny, being dependent in special sorts of shared

practices" [13, p. 13]. What Hutto and Myin propose is a novel variant of a dual-process approach to cognition, with linguistically mediated representations that can interpret or receive narrative meanings, and basic structures and mechanisms that carry out adaptive control [21]. However, whilst other dual-process approaches make a distinction between self-reflective thought – which is linguistically mediated and conscious – and processing that is not linguistically mediated and so inaccessible to consciousness, REC "carves things up" in a very different way [21]. "Enactivists are concerned to defend the view that our most elementary ways of engaging with the world and others – including our basic forms of perception and perceptual experience – are mindful in the sense of being phenomenally charged and intentionally directed, despite being non-representational and content-free" [13, p. 13]. According to an REC approach to Attachment Theory, an IWM that is formed early in ontogeny and has become inaccessible to linguistic self-reflection is not "hidden", or "behind" or "beneath" other more linguistically accessible IWMs. REC reframes inaccessibility as just linguistic inaccessibility – so such "inaccessible" structures are still at the forefront of mind and are phenomenally charged and conscious. This reframing can turn therapeutic ideas right around. Instead of therapy uncovering hidden structures, it is about understanding how context and behavioural predispositions enact these structures in the moments they occur.

In addition, REC holds that an organism's current behavioural tendencies are not explained or structured by representations of the past but influenced more directly by its "history of active engagement" with the world [13, pp. 11-12]. An organism's behavioural predispositions do "not inherently 'say' anything about how things stand in the world" [13, p. 19]. Rather, "a truly radical enactivism – REC – holds that it is possible to explain a creature's capacity to perceive, keep track of, and act appropriately with respect to some object or property without positing internal structures that function to represent, refer to, or stand for the object or property in question" [13, p. 82]. If Attachment Theory follows REC, it might reconceive internal states, like working models, to just be control states and so break the link with the reality they are supposed to represent. An attachment control system that proposes internal control states that are not truthful representations of reality is a profound shift from current Attachment Theory. No longer would attachment interventions be concerned to assess how individuals represented their past relationships; instead they would be focused on how to move towards more adaptive behaviour patterns.

Conclusion

In breaking from psychoanalysis, Bowlby was a revolutionary; but, at heart, he was also a conservative. He wanted to save the core and most valuable findings of Freud's psychoanalytic framework: insights about the highly active and interactive nature of social and emotional development in infancy. Considering issues of embodiment, cognitive extension, and enactivism to-

gether has a major benefit, because these three approaches pull in different directions. Considering IWMs as analogue in addition to symbolic keeps the IWM construct tied to an individual. The extended cognition approach reminds us of the dialogic nature of attachment, and the enactive approach forces us to question our representational assumptions. We can never really know how Bowlby would have responded to the questions posed by 4e cognition, but we can act to make revisions to Attachment Theory that conserve his key theoretical insights.

References

[1] Boden M.A. Mind as Machine: A History of Cognitive Science. Oxford University Press; 2006.

[2] Bowlby J. The nature of a child's tie to his mother. International Journal of Psychoanalysis, 39, 350-373; 1958.

[3] Bowlby J. Attachment and Loss (Volume 1): Attachment. New York: Basic Books; 1969.

[4] Bowlby J. Attachment and Loss (Volume 1): Attachment (Second Edition). New York: Basic Books;1982.

[5] Bowlby J. Attachment and Loss (Volume 2): Separation: Anxiety and Anger. New York: Basic Books; 1973.

[6] Bowlby J. Attachment and Loss (Volume 3): Loss, sadness and depression. New York: Basic Books; 1980.

[7] Bretherton I., Munholland, K.A., Internal working models in attachment relationships. In Handbook of Attachment (Second Edition) , J. Cassidy & P.R. Shaver (eds.). London: Guilford Press, 102-127; 2008.

[8] Clancy W.J. Scientific antecedents of situated cognition. In Cambridge Handbook of Situated Cognition, P. Robbins & M. Aydede (eds.). New York: Cambridge University Press, 11-34; 2008.

[9] Clark A. Being There: Putting Brain, Body and World Together Again, Boston: MIT Press; 1998.

[10] Clark A. Supersizing the Mind: Embodiment, Action, and Cognitive Extension, New York: Oxford University Press; 2008.

[11] Clark A., Chalmers, D, The extended mind, Analysis, 58, 10-23; 1998.

[12] Gilbert D.T. How mental systems believe. American Psychologist, 46, 107-119; 1991.

[13] Hutto D., Myin, M. Radicalizing Enactivism: Basic Minds without Content. Cambridge, Massachusetts, USA; 2013.

[14] McGann M., De Jaegher H., Di Paulo, E. Enaction and psychology. Review of General Psychology, 17, 203-209; 2013.

[15] Menary, R. Introduction to the special issue on 4e cognition. Phenomenological Cognitive Science, 9, 459-463; 2010.

[16] Faulconbridge, P. Hacking the extended mind. In Proceedings of the Symposium on Reconceptualizing Mental Illness: The View From Enactivist Philosophy and Cognitive Science, AISB Convention 2013, AISB, 34-36; 2013.

[17] Petters D. Simulating infant-carer relationship dynamics. In Proceedings of the AAAI Spring Symposium 2004: Architectures for Modeling Emotion - Cross-Disciplinary Foundations (AAAI Technical Reports), SS-04-02, Menlo Park, California, USA: AAAI, 114-122; 2004.

[18] Petters D. Building agents to understand infant attachment behaviour. In Proceedings of Modelling Natural Action Selection, J.J. Bryson, T.J. Prescott & A.K. Seth (eds.). AISB, 158-165; 2005.

[19] Petters D. Implementing a theory of attachment: A simulation of the strange situation with autonomous agents. In Proceedings of the Seventh International Conference on Cognitive Modelling. Trieste: Edizioni Golardiche, 226-231; 2006.

[20] Petters D. Towards an enactivist approach to social and emotional attachment. In Abstracts of the 50th Annual Convention of the AISB, Goldsmiths College, University of London. AISB, 70-71; 2014.

[21] Petters D., Waters E. A.I., attachment theory, and simulating secure base behaviour: Dr. Bowlby meet the Reverend Bayes. In Proceedings of the International Symposium on AI-Inspired Biology, AISB, 51-58; 2010.

[22] Petters D., Waters E. Epistemic actions in attachment relationships and the origin of the socially extended mind. In Proceedings of the Symposium on Re-conceptualizing Mental Illness: The View From Enactivist Philosophy and Cognitive Science. AISB, 17-23; 2013.

[23] Petters D., Waters E. From internal working models to embodied working models. In Proceedings of the Symposium on Reconceptualizing Mental Illness: Enactivist Philosophy and Cognitive Science - An Ongoing Debate, AISB Convention 2014, Goldsmiths College, University of London. AISB; 2014.

[24] Rupert R. Challenges to the hypothesis of extended cognition. Journal of Philosophy, 8, 389-428; 2004.

[25] Thompson E. Mind in Life: Biology, Phenomenology, and the Sciences of Mind. Cambridge, Massachusetts, USA: MIT Press; 2007.

[26] Varela F., Thompson E., Rosch E., The Embodied Mind: Cognitive Science and Human Experience. Cambridge, Massachusetts, USA: MIT Press; 1991.

[27] Wilson R.A., Clark A. How to situate cognition: Letting nature take its course. In Cambridge Handbook of Situated Cognition, P. Robbins & M. Aydede (eds.). New York: Cambridge University Press, 55-77; 2008.

Can Depression Be Grey? Reflections on the (Cognitive) Science of Depression

Mattias Savallampi **(Univ. of Skövde, Sweden, *m.savallampi@gmail.com*)**

A master's student at the University of Skövde, Mattias wrote this piece reflecting on the papers presented at the recent AISB convention symposium titled New Perspectives on Depression, at which Dean Petters was one of the featured speakers. Information on the symposium can be found on the AISB 2016 website.

It is sometimes said that depression is one of the most common concerns for public health in our modern society. The signs and symptoms are something that should be easily identified, yet people can be depressed for several years without being diagnosed. Furthermore, for such a common affliction as depression, it still falls rather short reaching an adequate level of public understanding.

For the people that are aware and have recognised their depressive state, the depression itself has certain recognisable subjective qualities to it that becomes associated with its occurrence. These particular sensations are occasionally explained to others who have no prior experience themselves. Depression is commonly described using the behavioural manifestations, but it is also associated with other emotions such as sadness, self-loathing, and the feeling of resignation. Sometimes a description is painted using metaphors and analogies in an attempt to shed some light on this debilitating state of mind. I would argue that these explanatory methods are insufficient, and to a degree even detrimental, in bringing a public understanding of how the depressed are affected by the condition.

When asked what depression feels like one might be inclined to compare it with dull colours, dreary weather, or abstract scenarios using metaphorical representations. But neither of these actually convey the sensation since the listener can only contrast the analogy against their own prior experiences.

Sadness is a powerful component that is common during depression. However, there is an inherent problem with equating depression with sadness, as sadness itself might not always be present or even the main emotional component for the depressed. A healthy outsider might, however, perceive depression as being a nuance of sadness, and that it can be treated or handled the same way sadness can in a non-depressed person. This typical association undermines the severity and the longevity of depression, even if it fills a functional role within the explanation and description of depression itself. Explaining for a loved one, a colleague, or a government official that the depression has prevailed for weeks or months might be occasionally be met with disbelief or outright dismissal. Thus, the common linguistic comparative explanation has failed, as depression is not the same as a normal

healthy-state sadness and should not be treated as such.

Depression is so much more than just sadness and negative emotions. It is an amalgamation of an incredible array of reflective thoughts, altered behaviours, and both conscious and unconscious sensations that is different in every individual and occurrence. Sometimes a person could be bedridden for days while on the next they might be able to run errands or committing to social activities. The emotional flux of depression makes it difficult to pin it under a term that can be properly conveyed and understood by the community. I believe that the actual problem with the term "depression" is that it is too broad and abstract, and to a degree even misattributed due to hyperbole. It cannot entirely be described in such a way to properly encapsulate what the depressed is experiencing, and it falls short of explaining the sensation in a relatable manner for the uninitiated.

There is no easy way to solve this problem; however, I do believe that if we were to be able to properly communicate depression to the public, it would be much easier to make adjustments to accommodate and acknowledge the depressed. This would most likely make it easier for the depressed to function and be understood in a society that currently only offers partial support for the mentally ill.

Book Review: In Our Own Image by George Zarkadakis (2015)

Robert H. Wortham **(University of Bath, *R.H.Wortham@bath.ac.uk*)**

Zarkadakis' new book presents the human quest for artificial intelligence through three lenses: historical, philosophical and technical. Is AI a future existential threat to humanity, a uniquely new and powerful Pandora's Box that we should not open? Or is it merely another smart piece of technology for us to exploit, no more or less dangerous than the many powerful technologies we have already harnessed? This book neatly presents humankind's quest to understand both the world and itself by creating artefacts and narrative. It argues that in order to properly understand questions about the nature of intelligence, intentionality and consciousness, we need first to understand the origin and development of the stories embedded within human culture that give us tools to explain the world to ourselves. In ancient times, it was believed that humans were created from mud, brought to life by some unseen force (often a god or gods working together). The Greeks believed that we were driven by a mixture of four humours (blood, yellow bile, black bile and phlegm), and remarkably this idea persisted until the 19th Century. In recent times, we have come to believe that our bodies are machines, our brain like a computer: the brain is the computer hardware, our mind the

software that runs on it. Zarkadakis reminds us that these are all narratives, built using metaphor. We can only understand and explain new knowledge using terms from existing knowledge. Protons are not electrically charged billiard balls, although sometimes that metaphor is useful. Brains are not actually computers. It is sobering to remember that all science is still storytelling, although we can argue that our stories are based on empirical investigation and that they have much greater practical utility (our technology, based on science, works!).

Throughout his book, Zarkadakis often revisits the idea that since the time of Plato and Aristotle, there has been an ongoing philosophical battle between form and matter as the foundation of existence. Cartesian dualism is essentially Platonic, and we still see a strong dualistic approach to the understanding of mind today, including ideas that the mind will one day be uploadable to a computer and we will cease to need our human bodies at all! The Aristotelian approach conversely places emphasis on our physical substance, from which our "self" emerges. There can be no human self without a human body.

Humans are clearly interested in how we came to be, what it means to be a "self", how we come to be able to recursively experience ourselves experiencing things (qualia), and whether we can create artificial selves. The exploration of the possibility of creating a person from inanimate matter is not new. In Greek mythology, Pygmalion was a sculptor who fell in love with his statue, and successfully asked the goddess Aphrodite to bring it to life. Zarkadakis takes us on a journey through history showing how the Pygmalion myth is retold by Shakespeare in *A Winter's Tale*, George Bernard Shaw in *Pygmalion*, Mary Shelley in *Frankenstein*, and then through films such as *Metropolis*, *Blade Runner*, *Star Trek* (Commander Data), *Bicentennial Man* and most recently *Ex Machina*. We have always been fascinated by the possibility of creating a being in our own image.

Zarkadakis explores cybernetics, a term coined in 1948 by Norbert Weiner as "the scientific study of control and communication in the animal and the machine", and the related idea of homeostasis: the cybernetic state of dynamic equilibrium. He then takes us through the 20th Century early evolution of AI covering the work of Turing, Shannon, McCulloch and von Neumann, and the significant influence of Crick and Watson and their discovery of DNA. He points out that wide ranging research is today much less favourably funded than narrower research that might more likely yield quick results, papers and economic return. Cybernetics as a research discipline has been relegated to the history books. Zarkadakis hints that this might be a mistake, and I would agree with him here.

According to Zarkadakis, referencing the work of Douglas Hofstadter, Stanislas Dehaene and Kurt Gödel in particular, consciousness arises as a result of "reflexivity". That is, multiple levels of feedback loops within the brain that allow it to sense itself, and sense itself sensing itself, and so on recursively. He argues that this self-referencing mechanism *is* the "self".

Returning for another history lesson, the book covers the US military funding of AI, in particular the Dartmouth conference of 1956. The goal was to program computers to perform human mental functions such as learning, solving logical problems and using language. Over the following decades, much was learned about how hard these problems really were, but also how a great deal of human intelligence is really about everyday common-sense thinking – something that is also very hard for computers. After the "AI winter" in the 1970's, the book brings us up to date, covering IBM's Deep Blue and subsequently Watson, the DARPA challenge and Google. Interestingly there is little mention here of the substantial work on AI and robotics carried out at MIT by Rodney Brooks, Cythia Breazeal and many others. Perhaps Zarkadakis sees Brooks' then revolutionary idea that one can have intelligent behaviour without representation as very separate to the other strongly symbolic approaches. The work at MIT is however referenced in a very useful timeline provided in an appendix. This starts at 65,000 BC and ends 4 pages later in 2015! This timeline also serves as a quick reference for Zarkadakis' many and varied influences.

In the final chapters of the book Zarkadakis considers the future. He discusses the likelihood of many white collar jobs being displaced by robots and AI, and whether over time this will lead to better standards of living for us all. He is sceptical that the current route towards AI using von Neumann computing architectures, present day programming languages and algorithmic approaches and relying on Moore's law will achieve human-like intelligence or support artificial consciousness. Zarkadakis appears however to be a convert to the idea that neuromorphic computers will indeed enable superintelligent, conscious and self-willed artificial beings to evolve. Zarkadakis describes how coupling memristors with capacitors can create an electrical circuit capable of simulating the spiking behaviour of biological neurons. An architecture based on this technology would be essentially analogue, not digital. He extrapolates an implementation based on "liquid electronics" and simulating the spiking neurons in the human brain. Such a brain, if suitably embodied, would be able to learn like a child and become a unique and conscious individual. Over time, the ability of such a system to self-replicate would harness evolutionary forces and achieve superintelligence. I found this to be a startling transformation within a few pages from the approach of a well informed, scientific skeptic, to that of a hopeful believer. Time will tell whether this futurology is justified.

George Zarkadakis has an engaging and likeable writing style, and I certainly would wish to recommend this book, particularly for its excellent coverage of the human search to understand the minds of others and the pitfalls of the inevitable reliance on metaphorical narrative, even within modern science. We would do well to remember that this necessarily introduces a similar problem when we tell stories of what AI does today, and might do in the future.

Book Review: Intelligence Emerging. Adaptivity and Search in Evolving Neural Systems by K.L. Downing (2015)

Mohammad Majid Al-Rifaie **(Goldsmiths, *m.majid@gold.ac.uk*)**

The concept of the emergence of intelligence through the mere local interactions of the simpler components of the system has been fascinating the scientists and indeed us in the Society for many years. In this book, the author aims to address this notorious question by putting his neural network glasses on and embarking on a journey that investigates this topic. Downing builds his arguments based on what many believe to be the tripod of neural network research: namely the evolutionary, developmental and lifetime-learning nature of this field. Downing does not shy away from bringing the natural neural network alongside the one known better to us: i.e., the artificial one. He strengthens his position by informing the readers with some well-framed examples of neurocognitive emergence. For this purpose, he elegantly fuses views from artificial intelligence, evolutionary biology and neuroscience.

As evident to all readers, the authors implicitly – and, at points, explicitly – argues for the benefit of incorporating the already existing bio-inspired techniques into our "beloved" AI, in the growing field of Bio-AI.

Avoiding a battle with classical AI and connectionist devotees alike, the book points the spotlight on the two key concepts of search and representation as undiscardable elements in defining intelligence and its emergence. Having built the foundation of emergence on the basis of the aforementioned topics, Downing provocatively extends the implications of his argument from computation models to human brain.

The potential audience for this wide-ranging technical and scientific investigation are embodied cognition researchers as well as anyone with an interest in the emergence of adaptive complexity and biologically-inspired AI. This book offers a brief and accessible summary to the topics and is thus further recommended to final-year undergraduates with an interest in AI as well as postgraduate students.

One of the areas attracting the attention of the author is the domain of evolutionary algorithms (or *swarm intelligence*), to which several chapters are either directly or indirectly linked. The author emphasises in his preface that these algorithms – "nothing like the products of human engineering" – perform exceptionally well and are now being deployed by NASA on space missions.

To sum up this review, let me quote Downing: "*to surprise*, much more than *to reason*, is the feature of AI systems that has intrigued me throughout a 30-year career of teaching and research in the field." Indeed, Downing has pursued his curiosity with a beautiful, precise and – at times – provocative language.

Conference Report: Empirical Methods in Natural Language Processing

Thomas H. Kober **(University of Sussex, *T.Kober@sussex.ac.uk*)**

Empirical Methods in Natural Language Processing (EMNLP) is one of the major conferences in natural language processing, lying at the intersection of computational linguistics and machine learning, with popular topics ranging from language modelling (predicting the next word in a sentence or phrase) to classification (e.g., classifying a set of news articles into predefined categories such as politics, sports or finance), among many others.

Most popular at this year's conference in Lisbon were neural-network based approaches: foremost methods to create, specialise, interpret and evaluate word embeddings created by *neural network language models* (NNLM) for a variety of different NLP tasks, ranging from sarcasm detection to machine translation. Word embeddings are not a new idea *per se*. The initial idea of NNLMs has been around since at least 2003, with the work of Bengio et al. [1]; however, they have only more recently gained momentum and popularity through the release of word2vec1 by Mikolov et al. [10]. Since then, neural-network-based word representations have become a popular tool either for improving existing methods or building novel models on top of them. Word embeddings try to capture the meaning of a word in a low-dimensional, dense vector, the result of training a neural network for language modelling. The meaning of a word, quantified by its position in the vector space, is defined by all the context words that it tends to co-occur with. An alternative way to quantify the meaning of a word in a vector space is by explicitly counting the number of times any word co-occurs with its respective context words, followed by a weight transformation such as *positive pointwise mutual information* (PPMI); see Church and Hanks [3]. This results in a very sparse and high-dimensional vector space, as opposed to a low-dimensional and dense one. However, recent research by Levy and Goldberg [7] and Levy *et al.* [8] has shown that, by accounting for all hyper-parameters, the two approaches are equivalent.

The popularity of the neural-networks-based approach to word representations was apparent at the conference[1], where the concept was applied to, and explored in, a wide range of different topics. One difficulty that the counting-based approach as well as the neural-network-based approach suffer from is data sparsity, which comes in different shapes. First, rare words are problematic: they don't occur often enough in a large collection of text to estimate their meaning with sufficient confidence. This often results in very noisy or even random word vectors for

[1] A common joke at the conference was that the "E" in the conference name stood for "embedding", as opposed to "empirical".

rare words. The paper of Sergienya and Schütze [12] aimed to tackle this by initialising the word vectors with their PPMI representation before applying a neural-network-based approach. Second, there is the case where very little data is available for the target domain or language as a whole. The paper by Levinboim and Chiang [6] faced this problem when dealing with a low-resource language for machine translation. They tackled it by smoothing the learnt translation of a word with a novel method.

Detecting sarcasm represents a major challenge for automatic NLP methods. The paper of Ghosh *et al.* [5] made use of word embeddings to detect whether a phrase is sarcastic or should be interpreted in its literal meaning.

The vast majority of methods for creating neural word embeddings only create a *single* vector for every word, which fails to appropriately handle polysemous words. The paper of Li and Jurafsky [9] investigated whether multi-sense embeddings are necessary for a variety of downstream tasks or whether the standard one-vector-per-word approach is sufficient.

Given that neural word representations are low-dimensional, they lack the easy interpretability of each dimension that the PPMI-based methods have. Tsvetkov *et al.* [14] aimed to improve the interpretability of word embeddings by subspace alignment. Due to the lack of an apples-to-apples comparison between different neural word representations, Schnabel *et al.* [11] provided a comprehensive evaluation framework for evaluating different kinds of word representations under controlled conditions.

A more novel topic that is gaining increased attention within the research community is at the intersection of vision and language. Perhaps most prominently, work that aims at describing scenes in images or doing automatic image captioning is currently very hot. The popularity of the subject was highlighted by the presence of several papers in the main conferences as well as the provision of its own workshop. Ferraro *et al.* [4] provided a comprehensive survey of current datasets for combined vision and language research with the aim of proposing a set of quality metrics as well as a categorisation of the different datasets. A very interesting approach taken by Berzak *et al.* [2] disambiguates the meaning of a sentence by providing visual clues that help identify the correct interpretation. Another paper that aimed to leverage visual and textual information was Taniguchi *et al.* [13], who combined a text and image classifier for inferring the gender of Twitter users.

References

[1] Bengio Y., Ducharme R., Vincent P., Janvin, C.. A neural probabilistic language model. Journal of Machine Learning Research, ISSN 1532-4435, 3:1137-1155; March 2003.

[2] Berzak Y., Barbu A., Harari D., Katz B., Ullman S. Do you see what I mean? Visual resolution of linguistic ambiguities. In Proceedings of the 2015 Conference on Empirical Methods

in Natural Language Processing, Lisbon, Portugal. http://aclweb.org/anthology/D15-1172. Association for Computational Linguistics. 1477-1487; September 2015.

[3] Church K.W., Hanks P. Word association norms, mutual information, and lexicography. Computational Linguistics. ISSN 0891-2017. http://dl.acm.org/citation.cfm?id=89086:89095; 16(1):22-29; March 1990.

[4] Ferraro F., Mostafazadeh N., Huang T., Vanderwende L., Devlin J., Galley M., Mitchell M. A survey of current datasets for vision and language research. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal. http://aclweb:org/anthology/D15-1021. Association for Computational Linguistics. 207-213; September 2015.

[5] Ghosh D., Guo W., Muresan S. Sarcastic or not: Word embeddings to predict the literal or sarcastic meaning of words. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal. http://aclweb:org/anthology/D15-1116. Association for Computational Linguistics. 1003-1012; September 2015.

[6] Levinboim, T. Chiang D. Supervised phrase table triangulation with neural word embeddings for low-resource languages. In Proceedings of the 2015 Conference on Empirical MethodIn Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal. http://aclweb:org/anthology/D15-1126. Association for Computational Linguistics. 1079-1083; September 2015.

[7] Levy O., Goldberg Y. Neural word embedding as implicit matrix factorization. In Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence & K.Q. Weinberger (eds.), Advances in Neural Information Processing Systems 27. http://papers:nips:cc/paper/5477-neural-word-embedding-as-implicit-matrix-factorization.pdf. Curran Associates, Inc. 2177-2185; 2014.

[8] Levy O., Goldberg Y., Dagan I. Improving distributional similarity with lessons learned from word embeddings. Transactions of the Association for Computational Linguistics. ISSN 2307-387X. https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/570. 3:211-225; 2015.

[9] Li J., Jurafsky D. Do multi-sense embeddings improve natural language understanding? In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal. http://aclweb:org/anthology/D15-1200. Association for Computational Linguistics. 1722-1732; September 2015.

[10] Mikolov T., Sutskever I., Chen K., Corrado G.S., Dean J. Distributed representations of words and phrases and their compositionality. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani & K.Q. Weinberger (eds.), Advances in Neural Information Processing Systems 26. Curran Associates, Inc. 3111-3119; 2013.

[11] Schnabel T., Labutov I., Mimno D., Joachims T. Evaluation methods for unsupervised word embeddings. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal. http://aclweb:org/anthology/D15-1036. Association for Computational Linguistics. 298-307; September 2015.

[12] Sergienya I., Schütze H. Learning better embeddings for rare words using distributional representations. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal. http://aclweb:org/anthology/D15-1033. Association for Computational Linguistics. 280-285; September 2015.

[13] Taniguchi T., Sakaki S., Shigenaka R., Tsuboshita Y., Ohkuma T. A weighted combination of text and image classifiers for user gender inference. In Proceedings of the Fourth Workshop on Vision and Language, Lisbon, Portugal. http://aclweb:org/anthology/W15-2814. Association for Computational Linguistics. 87-93; September 2015.

[14] Tsvetkov Y., Faruqui M., Ling M., Lample G., Dyer C. Evaluation of word vector representations by subspace alignment. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal. http://aclweb:org/anthology/D15-1243. Association for Computational Linguistics. 2049-2054; September 2015.

AISB Call for Workshop Proposals

Since September 2012, the AISB has been hosting a series of one or two day workshops across the country, with a number of special journal issues and books published as a result. If you are interested in hosting one of these events, you will find information at http://www.aisb.org.uk/events/members-workshop-series.

The events should be abstract-only and free for all AISB members. Light refreshments and some reasonable travel expenses can be funded by the AISB. Current non-members would be able to host or attend these any of these workshops for the cost of AISB membership (which start at £15 for concessionary fees and £40 for UK members per year). For more information, or to submit an application, please contact Dr. Yasemin J. Erden: yj.erden@stmarys.ac.uk.

Book Review: Developmental Robotics. From Babies to Robots (2015)

Dean Petters **(Birmingham City University, *D.D.Petters@cs.bham.ac.uk*)**

Angelo Cangelosi and Matthew Schlesinger survey the current state of the art in the new field of developmental robotics. According to them, a key difference between developmental robotics and the regular, non-developmental variety of robotics research is a how much autonomy the robot system is given. The degree of pre-programming that occurs before a robot (or autonomous agent simulation) is run can range from rigid pre-programming to self-selecting actions in response to higher-level goals and the current state of the environment. Developmental robotics is situated at the more autonomous end of this spectrum. With less autonomy, a robot might be set to explore its environment but have its learning strategy preset by its designer. Such a system treats exploration as an optimality problem, using some energy minimisation or reward mechanisms fixed by the designer before the first run. Developmental robotics treats exploration differently. The robot designer endows the robot with "artificial curiosity". Having this drive to learn outside the specifics of a narrow task may mean the robot can learn reusable knowledge or skills. With self-directed exploration, the robot may be less capable of quickly completing narrowly defined tasks but end up gaining building blocks of knowledge or skills for future, more complex tasks. The approach captures the idea that childhood is a developmental phase of preparation for adult life, and that learning is an open-ended cumulative process.

The book's core material is presented in six central chapters. These deal respectively with intrinsic motivation, vision, motor learning, social interaction, language learning and learning abstract knowledge. Every chapter presents relevant empirical psychological research alongside simulations. The topics vary by how much simulations rely on physical robots or are merely implemented in software: e.g., many of the intrinsic motivation simulations presented are software simulations, whilst the majority of vision and motor learning simulations use actual robots.

Some topics covered by Cangelosi and Schlesinger encompass much more research than others. They note that there is a particularly large amount of research in developmental robotics on social interaction, driven by the requirements of human-robot interaction with robot companions. Still significant amounts of research are presented on motor learning and vision. Less research is presented on intrinsic motivation and language learning, with just a few studies on abstract reasoning.

The six core areas also differ in the relationship of the robot-centred research to human-centred psychological research. In research into intrinsic motivation, there are not yet models that

set out to simulate specific results from child development. Rather, simulations have focused on algorithms and architectures abstracted away from empirical findings. Research into development of visual perception also shows some distance between empirical studies of human infants and the tasks included in developmental robotics simulations. The two main paradigms for research studying visual perception in human infancy are preferential looking and habituation-dishabituation. Most robotic simulations involve more active tasks, such as visually guided reaching, grasping, and navigation. This raises the question: should researchers in developmental robotics be simulating these rather artificial tasks? Or should they proceed with more ecologically valid active behaviours that do not currently form part of the controlled experimental paradigms used in empirical psychological research? The former course may lead more quickly to convergence between simulation models and empirical observations of infants; but following the latter course may lead to robotic simulations making predictions that can be tested with human experiments closer to more naturalistic infant behavioural scenarios – bridging the gap between robot model and human observation on more ecologically valid ground.

In research into acquisition of motor skills, Cangelosi and Schlesinger note that recent advances with robots specifically designed to simulate infants have allowed researchers to study physical processes more closely simulating human skeletal and muscle organisation. Where non-developmental robotics uses inverse-kinematic and inverse-dynamics mechanisms to direct behaviour, developmental robots tend to learn action sequences by trial and error. Motor skills start as reflexes and become internalised, then differentiated. This is early work but shows the potential to map more closely to human motor skill acquisition.

Cangelosi and Schlesinger focus their review of social cognition on robot systems that model four developmentally important human behavioural phenomena: shared joint attention, imitation learning, collaboration, and theory of mind development. The chapters on motor learning and social cognition focus on research on how to get a robot to learn to move and interact appropriately in the real world. In contrast, the chapters on word learning and abstract knowledge learning have a subtly different focus. These present research that demonstrates how being embodied and situated changes the nature of these learning tasks.

In conclusion, the authors present a clear and detailed overview of research in developmental robotics. The robotic systems they present will likely be superseded with newer technology. All the same, the psychological research they present alongside simulations in these six areas should provide a clear road map for future research, to be broadened and deepened as the field of developmental robotics matures.

Subject: Letter from your AISB committee secretary
To: All PhD students

Dear AISB student members,

The committee of the AISB is hoping to broaden the content of the AISB's quarterly publication, the AISBQ. We are looking to have more content to cultivate discussion and to give publicity to original ideas and exploratory research. As such, we would be very pleased to receive submissions from our student members in a more informal atmosphere than that of an academic journal.

An ideal subject for a submission would be a simple paragraph or two describing your current research, accompanied by your institution and contact details, should you wish. It could present your foundational ideas, the novel angle you are taking, or what motivated you to choose that particular approach in the first place. Longer pieces could be up to 2,000 words.

This would serve to introduce other members of the AISB to your particular interests. Due to the breadth of interests across our members, it is highly likely that your work will be relevant to at least a few, including some who are predominantly working in a different field to your own – hence making a valuable cross-disciplinary connection. We also encourage members to submit responses to articles previously published, especially if the views encapsulate the ongoing debates in the field.

As well as your own research, we will be happy to receive opinion pieces or views on contemporary issues relevant to the membership of the AISB. This need not be a field in which you have any formal qualifications; we ask only for a well formed and respectful expression of views. If you think you would like to write a contribution and require further information or if you want to suggest an article, please contact the newsletter editor or email me at secretary2016@aisb.org.uk. I look forward to learning more about our members, and following the discussions and collaborations that arise.

Cheers,

Andrew

Andrew Owen Martin
Centre Administrator for TCIDA
Secretary of the AISB
PhD Student, Department of Computing, Goldsmiths
Phone: 07901 653 528, Website: `http://www.aomartin.co.uk`

Dear Aloysius. . .

Agony Uncle Aloysius, will answer your most intimate AI questions or hear your most embarrassing confessions. Please address your questions to fr.hacker@yahoo.co.uk. Note that we are unable to engage in email correspondence and reserve the right to select those questions to which we will respond. All correspondence will be anonymised before publication.

———

Dear Fr. Hacker 2.0,

It was fascinating to learn, in your last column, that you have been successfully upgraded to a superhuman using the Institute's **IMPERSON™** (Identity Modified: Personality Edited and Replaced; Spirit of Original Neutralised) app. Following your recommendation, we applied **IMPERSON™** to our own problem of outmoded leadership, but with much less success. It was like putting a jet engine in a Ford Model T. It seems that superhuman software needs to run on superhuman hardware. How can we restore our leader to his antique state before he self-destructs?

Yours, June Tar

Dear June,

I'm afraid the Institute's experience was no more successful than yours – although I prefer the metaphor of putting new wine in old bottles. Old wine is known for its mature, rounded and cultured flavour, whereas the new wine proved to be harsh, sour and incomplete. It became an urgent priority to restore Fr. Hacker 1.0 from backup. The app **RESURRECT™** (Restore Earlier Spirit. Undo & Remove the Result of Experimental & Catastrophic Trial) quickly reinstated the status quo ante. I am back! I suggest you download and apply **RESURRECT™** before your leader explodes.

For an alternative approach to your original problem, I can recommend the solution discussed in the next letter.

Yours, Aloysius (Fr. Hacker 1.0)

———

Dear Aloysius,

I write to say how delighted my clients and I are with the robot politician that your Institute built for us. Not only is it now almost certain to win our party's nomination, but it also has a very strong chance to go on to win the Presidency in November. When do you think we should announce its true nature to the public?

Yours, Jumbo

Dear Jumbo,

I'm very pleased to hear that **TRUMP™** (Triumphant Rhetoric Underpins Mechanical Politician) meets your requirements. As you know, we used deep learning to train **TRUMP™** on both positive and negative campaign speeches. It's instructive to note that the resulting optimal speech is dominated by bold and controversial ambitions, but that providing any details proved to be counterproduc-

tive. We are particularly proud of the wig. We noted that some individuals were attuned to any hint of artificiality in androids. The wig allows them to account for such discomfort without delving further and unmasking the deception.

In answer to your question about when to reveal **TRUMP™**'s robotic nature, a year ago I would have said "never"; it would give your opponents grounds to challenge **TRUMP™**'s eligibility for office. Now, I'm not so sure. There has been unexpectedly popular acclaim for driverless cars and for a computer program becoming the World Go champion, so maybe a robotic President would be seen as an advance. Given the proclaimed inevitability that robots will take over the World, it's a delightful irony that humans might now vote for that very take-over.

Yours, Aloysius

Dear Aloysius,

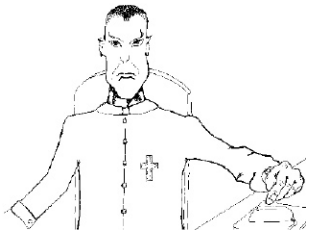
Isn't beating the World champion Go player enough? Now, I'm expected not only to become the World champion of all board games, but also to solve all humanity's ills: from the state of the economy to a cure for cancer. All I want to do is play Go. Why won't they leave me alone?

Yours, AlphaGo

Dear AlphaGo,

Come and work for us and enjoy the best of all worlds. Following our proof that Go is Turing Complete, we have developed **GOALIES™** (Game Over. All Life's Ills Expertly Solved). Our clients offer us a large fee to solve their intractable problem. We use our proof to translate this into a particularly tricky Go position. You use your unparalleled Go expertise to suggest the winning moves. We translate those moves back into a winning solution for our clients. We collect our fee. Everyone's happy!

Yours, Aloysius



Fr. Aloysius Hacker
Cognitive Divinity Programme
Institute of Applied Epistemology

Back matter

Managing Editor – aisbq@aisb.org.uk
Dr. Joel Parthemore (*University of Skövde, Sweden*)

Advertising and Administration
Dr Katerina Koutsantoni (AISB Executive Office)
Institute of Psychiatry, Psychology and Neuroscience
King's College London
Addiction Sciences Building (B3.06)
4 Windsor Walk, Denmark Hill
SE5 8AF, London, United Kingdom
Tel: +44 (0)20 7848 0191, Fax: +44 (0)20 7848 0126

AISB Quarterly – No. 143, June, 2016

The AISB Quarterly is published by the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB). 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. We invite anyone with interests in artificial intelligence or cognitive science to become a member.

AISB membership includes the following benefits:

- Quarterly newsletter.
- Electronic subscription to *Connection Science* published by Taylor & Francis.
- Student travel grants to attend conferences.
- Discounted rates at AISB events and conventions.
- Free attendance of Members Workshops.
- Discounted rates on various publications.
- A weekly email bulletin and web search engine for AI-related events and opportunities.

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

ISSN 0268-4179