Noname manuscript No.
(will be inserted by the editor)

Towards Autonomous Creative Systems

A Computational Approach

Rob Saunders

Received: date / Accepted: date

Abstract This paper reviews the long-standing debate surrounding the nature of machine intelligence, autonomy and creativity and argues for an approach to developing autonomous computational creativity that models personal motivations, social interactions and the evolution of domains. The implications of this argument on the types of cognitive processes that are required for the development of autonomous computational creativity are explored and a possible approach to achieving the goal is described. In particular, this paper describes the development of artificial creative systems composed of intrinsically motivated agents engaging in language games to interact with a shared social and cultural environment. The paper discusses the implications that this type of approach may have for the development of autonomous creative systems.

Keywords autonomy · computational creativity · artificial creative systems · systems theories of creativity · autopoiesis · intrinsic motivation · language games · evolution of language

1 Introduction

Computational Creativity is a synthetic approach to the study of creativity. There are three clear motivations for the study of computational creativity: (1) to provide insights into the nature of creativity; (2) to produce tools to support human creativity; and, (3) to develop autonomous machines capable of undertaking tasks that currently require human creativity (Colton and Pease, 2011).

This paper focusses on the first of these motivations with specific reference to questions of autonomy in computational models of creativity. What would it mean to produce an autonomous creative system? How might we approach this task? And, how would we know if we had succeeded? To examine the implications for the development of autonomous creative systems, these questions will be explored with reference to our current understanding of human creativity.

1.1 The Lovelace Objection

The question of the autonomy of machines displaying apparently intelligent or creative behaviour pre-dates the manufacture of machines on which to test them. Famously, Ada Augusta, Countess of Lovelace, commented upon her translation of Menabrea's "Sketch of the Analytical Engine", declaring that:

> The Analytical Engine has no pretensions to *originate* anything. It can do whatever we *know how to order it* to perform. It can *follow* analysis; but it has no power of *anticipating* any analytical relations or truths. Its province is to assist us in making *available* what we are already acquainted with."
>
> (Menabrea and Augusta 1842)

Consequently, according to Boden (1990), Lady Lovelace would have credited any creative products of the Analytical Engine to the engineer, not to the engine. Turing

---

This research has been partly supported by the Australian Research Council, Discover Grant DP0666584.

Rob Saunders
Faculty of Architecture, Design and Planning
University of Sydney
Australia
Tel.: +61-2-90366379
Fax: +61-2-93513031
E-mail: rob.saunders@sydney.edu.au

(1948) responded to Lovelace's objection in two ways: firstly, he noted that computers often surprised him, due to a faulty understanding on his part and the complex nature of the processes involved, highlighting the subjective nature of any ascription of creativity. Secondly, Turing explored in some detail the notion of a machine that could organise itself through clearly programmed means but more importantly as a result of its "experiences". Turing's stated position already begins to raise questions about the need for autonomy in intelligent machines if they are to be considered creative. Interestingly, Turing's responses do not focus on the *imitation* of creative behaviour but the development of an independent intelligence.

The importance of creativity as a foundation of machine intelligence continued to be recognised as computer science developed increasingly sophisticated techniques. "Randomness and Creativity" was one of only a small number of topics proposed for the Dartmouth Summer Research Conference on Artificial Intelligence as one of the grand challenges facing the nascent field:

> A fairly attractive and yet clearly incomplete conjecture is that the difference between creative thinking and unimaginative competent thinking lies in the injection of a some randomness. The randomness must be guided by intuition to be efficient. In other words, the educated guess or the hunch include controlled randomness in otherwise orderly thinking.
>
> (McCarthy et al 1955)

Some of the early pioneers attending this workshop went on to develop the first examples of computational creativity, developing discovery systems that reproduced the findings of eminent scientists (Langley et al, 1987). Meanwhile, Lenat (1976) explored the discovery of theories in mathematics and Harold Cohen began his epic development effort, resulting in AARON (McCorduck, 1990). Despite some impressive results the early discovery systems came under increasing criticism for their lack of autonomy, either because of the amount and type of information provided at the outset, or because the systems relied on human supervision to determine when a significant result had been achieved (Lenat and Brown, 1984).

2 Creativity

The need to define the nature of creativity has haunted most attempts to develop models and theories of it. The difficulty of this task is clear from the abundance of definitions that can be found in the literature – Taylor (1988) gives some 50 definitions. One definition of creativity upon which there is a general consensus is that *creativity is the ability to produce work that is novel and valuable*. Mayer (1999) refers to this as the 'basic definition of creativity' and much research in creativity has attempted to identify the specific abilities and behaviours of creative individuals or clarify what is meant when something is described as being novel (original, new, unexpected, surprising) and valuable (useful, appropriate, aesthetic, adapted). Beyond the basic definition, however, precise definitions of creativity diverge dramatically and there is a danger for researchers "to view a part of creativity as the whole phenomenon" (Sternberg and Lubart, 1999, p. 4); often leading to hotly contested views.

Creativity is often identified as a social construction and as such has a strong honorific sense that is as much the result of an audience's appreciation of a work as it is the creator's production. Some researchers reserve creativity solely for those who make a significant contribution to a domain of knowledge (Csikszentmihalyi, 1988). Proponents of these definitions contend that creativity cannot occur in a vacuum and must be studied in the context of the socio-cultural environment of the creator (Gruber, 1974; Gardner, 1993; Simonton, 1984; Martindale, 1990). Other popular definitions of creativity maintain the distinction between personal and social creativity. Boden (1990) defines both psychological creativity (P-creativity) and historical creativity (H-creativity), while Gardner (1993) distinguishes between little-c (mundane) and big-C (eminent) creativity. Within these definitions both Boden and Gardner suggest that creativity has two important but distinct meanings for the research community; the first is a label used to identify the processes employed by a creative individual, the second is a title awarded by society to honour the importance of creative works or creative individuals.

2.1 Approaches to Studying Creativity

Many different approaches have been taken to studying human creativity. The study of exceptional individuals, those who make substantial contributions to human culture, was the dominant focus of early studies of human creativity. A focus on creative genius continues in recent research (Sternberg, 1988; Gardner, 1993), however, the study of exceptional individuals has waned as the study of everyday creativity has increased. Bohm (1968) argued that the creativity of eminent artists and scientists does not arise from particular mental talents but from an ability to overcome inhibitions.

2.1.1 The Four Ps

Approaches to studying creativity have typically focussed on one of the Four Ps proposed by Rhodes (1961): *person*, *process*, *product* and *press*.

Person: Guilford (1950) called for the study of the creative person, leading to an abundance of research highlighting specific characteristics of creative individuals including intelligence, attitudes, and behaviours. While no general consensus has been arrived at regarding the cognitive characteristics of creative people, the research in this area has provided interesting and useful results (Tardif and Sternberg, 1988). For example, Amabile (1985) conducted extensive studies of motivation in creative individuals and found that intrinsic motivations play a more important role for creative individuals than extrinsic motivations.

Process: The detailed study of people engaged in creative tasks has provided valuable insights into the nature of everyday creativity as the extraordinary result of ordinary thought processes (Weisberg, 1988); and, with the availability of technology for studying the brain during creative tasks, there is an increasing use of techniques from neuroscience (Takeuchi et al, 2011; Chermahini and Hommel, 2010; Arden et al, 2010).

Product: Many researchers advocate a product-centric view for the evaluation of creativity (Plucker et al, 2004; Ritchie, 2001; Petroski, 1992). Studies of preference judgements have provided insights into the perception of creativity (Berlyne, 1971; Humphrey, 1973; Gaver and Mandler, 1987). Tardif and Sternberg (1988) argued, however, that the production of novel work is insufficient for the attribution of creativity, rather it has to be considered within a domain-specific context. This argument was supported by the studies of Martindale et al (1988), which showed that the perception of creativity changes with exposure to examples of works.

Press (of the environment): The physical, social and cultural environment that a creative individual operates in has a significant influence (Simonton, 1988; Hennessey and Amabile, 2010). In particular, Gardner (1993) examined the environments within which exceptional individuals were raised and worked in later life.

In describing the press of the environment, Rhodes (1961) focused on the influence of the environment on a person during the creative process, rather than how a creative product is judged by others or the influence that the creative person has on the world. Consequently, Simonton (1990) added a fifth P to this list—**Persuasion**: The ability to persuade others of the value of one's work.

2.2 Cognitive Models of Creativity

Models of creative thought processes have significantly influenced the development computational creativity. Early models of creative thinking relied heavily on introspection. Poincaré (1913) first proposed that creativity could be broken down into a period of conscious work followed by a period of unconscious work. Wallas (1926) proposed that creativity proceeds through a series of four phases: *preparation*, *incubation*, *illumination* and *verification*.

A concerted effort to study creative thinking began with Guilford's 1950 address to the American Psychological Association (Guilford, 1950). As a result, principled and objective accounts of the cognitive processes involved started to appear. Koestler (1964) presented an early example of this type of model as *bisociation*, which appears to be both general and mechanistic, but as Boden (1990) notes: "The thought processes [Koestler] described do happen, and they do seem to be involved in creativity. But because how they happen was not detailed, he did not fully explain how creativity is possible." This apparent lack of specificity has not stopped the further development of the idea of bisociation, which in more recent years has developed into the more detailed computational model of *conceptual blending*, having been informed by the development of frame-based knowledge representations by Minsky (1975) as well as the cognitive architectures of Soar (Laird et al, 1987) and ACT-R (Anderson and Lebiere, 1998).

2.2.1 Boden's Model of Creative Thinking

Boden's model of creative thinking has been particularly influential in the development and understanding of computational models of creative processes (Boden, 1990). The model provides an account of the creative process in terms of two primary processes: the exploration of an existing conceptual space; and, the transformation of conceptual spaces. As described by Boden (1990), conceptual spaces cover broad areas of expertise, e.g., jazz music and poetry, although the constitution of the conceptual space is never formally defined. Boden (2004) argued that transformation is the only process that can lead to genuinely creative ideas: "A merely novel idea is one which can be described and/or produced by the same set of generative rules as are other, familiar ideas. A genuinely original, or creative, idea is one which cannot." While it is tempting to equate the transformation of a conceptual state to H-creativity, the history of human creativity does not support this view, which has hailed the creativity of both the "young turks" who have revolutionised domains and "old masters" who have spent lifetimes perfecting their craft (Galenson, 2008).

Boden's model of creativity poses two significant challenges for the development of autonomous creative systems. The first is that, while the model is expressed in computational terms, little detail is given about representation of conceptual spaces or the mechanisms required to produce, maintain, explore and transform them.

Boden (1990) has argued that conceptual spaces are developed and solidified as individuals acquire knowledge in a domain but again no detail is given about the nature of this domain knowledge or the types of learning processes that may be involved. Boden (1995) argues that increased exposure to a domain produces richer, deeper cognitive maps which may act as a guide to search. Meanwhile, researchers have developed more detailed formalisms of these ideas (Wiggins, 2006), including possible mechanisms for the structuring of conceptual spaces (Thornton, forthcoming).

The second, and perhaps more serious, challenge to the development of autonomous creative systems based on Boden's model of creativity is the absence of any role for motivation or attention in the model. Boden (1994) offers two explanations for not addressing questions of motivation in her model of creativity: computational models of motivation have not been sufficiently well defined; and, motivation is concerned with the "why" and not the "how" of creativity.

It could also be reasonably argued that Boden's model can be extended at a later time to add a model of motivation but motivation does far more than provide an impetus to engage in a creative process. Bohm (1968) argued that understanding why artists and scientists are so deeply interested in their work will provide explanations about how some people are capable of producing creative work while others with similar mental capabilities are not. Amabile (1985) showed the profound impact that intrinsic and extrinsic motivations, both positive and negative, have on creativity.

In addition, Csikszentmihalyi's studies of flow, where creative individuals experience an altered state of consciousness when engaged in highly productive creative activity, suggest that attentional control and selectivity are two of the most significant predictors of creative performance (Csikszentmihalyi, 1997). Researchers cannot afford to ignore these findings if they are to develop autonomous computer programs capable of independently governing their own creative processes.

2.3 The Challenge of Autonomy

The omission of self-motivation or self-direction from Boden's model of creative thinking is also surprising when one considers the long tradition of empirical research that has established autonomy as one of the core characteristics of the creative personality (Sheldon, 1995). But then, as Smithers (1997) argues, the term "autonomous" has long been wasted in AI, robotics and Artificial Life by its ill-disciplined use; variously being used as a synonym for "intelligent", "independent", "self-sufficient", and "self-regulating". Smithers argues that the kinds of artificial systems that have been the subject of study in these fields are better described as either *automatic* (self-acting) or *cybernetic* (self-regulating) and that the term *autonomous* (self-governing) must remain something for researchers of AI, robotics and Artificial Life to aspire to. In particular, Smithers notes that:

> Unlike cybernetic systems, which must have some means for self-regulation, but which have the laws of regulation either built into them or somehow externally introduced, autonomous systems must have some means of forming their own laws of regulation as well as the means to regulate their behavior with respect to them.
>
> (Smithers 1997, p. 94)

Smithers survey of the meaning of "autonomous" across biology, philosophy, ethics and law found that in "all these cases the underlying notion is one of self-law making, or self-governing, and it is closely related to the concepts of self-identity and self-determination." In law and politics, Smithers notes that "autonomy is closely connected to issues of self-identity of ethnic groups, national identity, and self-determination" and that "identity, normally of a group in this context, is something which comes about through interaction with the environment, in this case, with other people or peoples." In philosophy, "a person is morally autonomous if and only if his or her moral principles or guidelines are his or her own." Autonomy plays a similar and important role in ethics but here the situation is often made more complex by the interaction of the autonomy of different kinds of identities. Finally, Smithers sees that the notion of autonomy is essentially the same in biology, especially in Maturana and Varela's development of autopoiesis as a theory of life in single cells (Maturana and Varela, 1980). According to Varela (2000)[1], a system is autopoietic if;

1. it has a semi-permeable boundary,
2. the boundary is produced from within the system, and
3. it encompasses reactions that regenerate the components of the system.

---

[1] As reported in Luisi (2003)

The theory of autopoiesis thus attempts to capture the essential process that generates the identity of living organisms. It is from this continuous process of identity formation that the concept of autonomy is developed by Varela (1979): the identity of an autonomous system is the result of a continuous and ongoing construction process that is neither fixed from the start nor given to it, as is the case for allopoietic systems. Autopoiesis as found in biological systems is thus a mechanism for being autonomous, but not necessary the only one. Varela and others have attempted to apply similar ideas to more abstract versions of autopoiesis, as a way of understanding cognition (McMullin and Varela, 1997; Bourgine and Stewart, 2004) and sociological phenomena (Luhmann, 1984).

3 Systems Theories of Creativity

According to systems theorists, the mechanisms of the mind are insufficient to explain creativity. Vygotsky (1971/1930) first proposed a systems theory of creativity, which highlighted the reciprocal relationship between creative individuals and their socio-cultural environment. In Vygotsky's thoery, creative individuals are both influenced by their personal understanding of their socio-cultural environment and through their actions cause changes in their environment that affect themselves as well as others (Lindqvist, 2003). This recognition that the creativity of an individual is necessarily embedded within a socio-cultural context is the basis of the theories of the creative systems that follow.

3.1 Csikszentmihalyi's Systems View of Creativity

Csikszentmihalyi (1988) presented a systems view of creativity, which he later extended to produce the Domain Individual Field Interaction (DIFI) theory of creativity (Feldman et al, 1994), as a description of the components and interactions involved in a creative system. In Csikszentmihalyi's view, an individual's role is to bring about some transformation of the knowledge held in the domain; the field is a set of social institutions that selects from the variations produced by individuals those that are worth preserving; and, the domain is a repository of knowledge held by the culture that preserves ideas or forms selected by the field.

The interactions between the components of a creative system, illustrated in Figure 1, occur in cycles. In a typical cycle, an individual takes some knowledge provided by the culture and transforms it, if the transformation is deemed valuable by the individual's society, it will be included in the domain of knowledge held by its culture, thus providing a new starting point for the next cycle of transformation and evaluation.

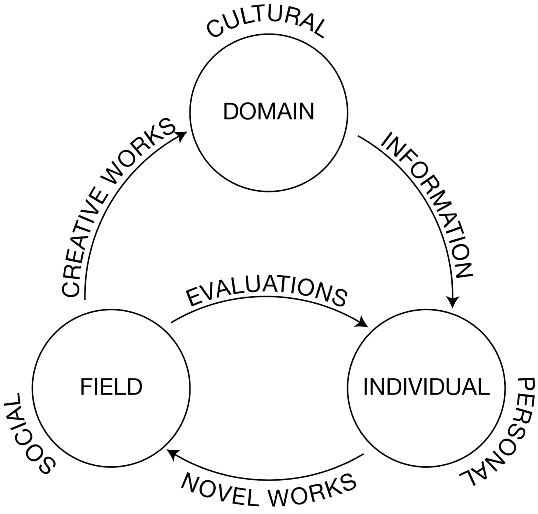

Fig. 1 The interactions between the individual, field and domains in the DIFI model of creativity.

In Csikszentmihalyi's model, creativity is not to be found within any one of the system's components, but in the interactions between them (Csikszentmihalyi, 1999). Clearly, however, the DIFI model reserves the title "creativity" to transformations of knowledge by individuals that have a significant impact on the domain, i.e., H-creativity.

3.2 Iba's Autopoietic Theory of Creative Systems

Iba (2009) introduced a systems theory of creativity, building on the theory of autopoietic social systems proposed by Luhmann (1984). Luhmann's theory of autopoietic social systems considers society and mind as two structurally coupled autopoietic systems, where society is an autopoietic system whose element is communication and mind is an autopoietic system whose element is consciousness. Autopoietic social systems are closed and must constantly reproduce through communication. Conscious minds (or psychic systems) are similarly closed, i.e., they cannot give or receive consciousness. Consequently, minds are mutually inaccessible and communication is only possible through the interaction of three parts; the *information* held in the mind of the speaker, the *utterance* that is shared from speaker to receiver, and the *understanding* that results from the perturbation caused by the utterance in the receiver's mind.

Iba's theory proposes a third autopoietic system, a creative system, where the elemental unit is discovery, such that the *ideas*, *associations* and *consequences* of a discovery are only meaningful within a creative system. Consequently, in Iba's theory, creative systems are closed to the outside with respect to discoveries. Iba's theory attempts to capture the “contingent nature of creativity”, or the fact that creative processes do not build purely on successes but that they come about as a result of chaining together many ideas, both good and bad. As Sawyer (2003) noted, Darwin produced many ideas during the course of his life's work that were wrong in hindsight but proved to have an important role in the development of his transformational theories. Iba's theory also attempts to resolve the problem of the “intrinsic nature of creativity” to produce a theory that accommodates both P-creativity and H-creativity and explain how communication between members of a group can transform the creative process such that it promotes the generation of creative ideas (Sawyer, 2003, 2008).

4 Artificial Creative Systems

Artificial creative systems are agent-based simulations that explicitly model social and cultural aspects of creativity as protocols that govern communication between agents. This section presents three artificial creative systems developed to explore the role of communication in creative systems.

4.1 The Emergence of Shared Values in Creative Fields

The Digital Clockwork Muse (Saunders and Gero, 2001) implements an artificial creative system where each agent is capable of independently generating and evaluating artefacts. Agents communicate artefacts and evaluations with other agents according to a simple protocol, which consists of two rules:

1. If an agent generates an artefact and evaluates it to be suitably novel and valuable to be considered “potentially creative” then it may send the artefact to another agent for evaluation.
2. If an agent receives an artefact and evaluates it to be suitably novel and valuable to be considered “creative” then it will inform the sender of its evaluation and store the artefact in the domain for other agents to access.

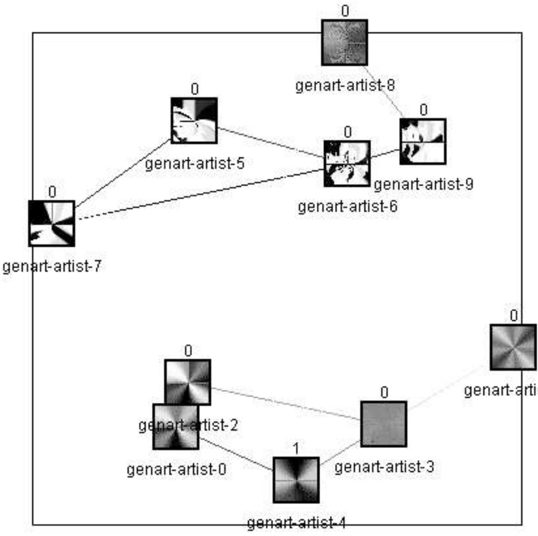

Fig. 2 The formation of two subfields, or ‘cliques’ within an artificial creative system. Each square is an agent, the lines between agents represent rewards for generating ‘creative’ products.

The communication protocol ensures that while judgements of artefacts are arrived at independently by agents, a consensus must be reached between at least two agents before an artefact can be considered “creative” and added to the domain. The artefacts in the Digital Clockwork Muse are small bitmap images. Agents generate the images using an interactive evolutionary algorithm, similar to the one devised by Sims (1991). The images are evaluated using the Wundt Curve (Figure 3), a model of interest proposed by Berlyne (1960), which favours novel artefacts that are similar-but-different from previously learned prototypes. By varying the model of interest, each agent may have a different preferred novelty.

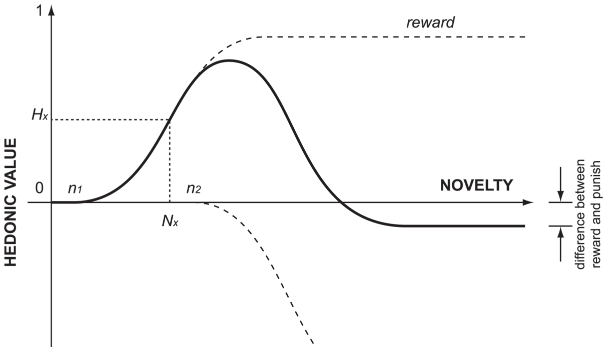

Fig. 3 The Wundt Curve: A model of interest for generating rewards based on novelty.

During simulation runs, sub-fields, or "cliques", emerged with shared values such that agents in a clique would typically favour images produced by other agents in the same clique over those produced by agents outside the clique, as illustrated in Figure 2. Cliques would form when a sub-group of agents shared images frequently as this caused the prototypes stored in the memories of each agent to become more similar. Cliques could become stable if the agents also shared a similar preference for novelty as this would increase the likelihood that novel images produced by an agent would be highly valued by the other members of its clique.

4.1.1 Other Computational Models of Creative Societies

Other multi-agent models of social creativity have examined the relationships between individuals and fields. Gero and Sosa (2002) explored the emergence of "gatekeepers" in creative fields, i.e., individuals with the ability to strongly affect the contents of the domain. Bown (2008) developed multi-agent models to explore cohesion, competition and maladaptation in the evolution of musical behaviour. Colton et al (2000) present a computational model involving multiple agents working together to explore a mathematical domain.

4.2 The Evolution of Language in Creative Domains

Creative domains, as described by Csikszentmihalyi (1988), are dynamically maintained and distributed across fields, with each individual in a field having only a partial view of the whole domain. The knowledge in domains exists in a variety of forms and is often expressed in domain-specific languages that evolve over time.

Steels (1995) developed an agent-based model of the evolution of languages based on the repeated playing of language games, similar to those described by Wittgenstein (1953). In Steels' "guessing game", an agent, the initiator, selects an object from the environment and describes it using a simple utterance to a second agent, the recipient, who attempts to identify the object in the environment. Success or failure to identify the object described by an utterance provides reinforcement signals to inform the learning processes of both the initiator and the recipient. Through the repeated playing of many language games, a group of agents is driven to adopt common meanings for their utterances and, as a consequence, a shared lexicon emerges (Steels, 1996). The languages evolved share some interesting features of natural languages including words that have multiple meanings and multiple words having the same meaning.

Saunders (2011) developed an artificial creativity system that supports the evolution of a domain-specific language for describing coloured shapes of varying type, size and hue. Individuals are modelled as curious design agents: each agent is capable of generating new shapes and assessing their novelty. If a generated shape is appropriately novel, the agent produces an utterance and communicates the shape and its description to another agent by playing the guessing game.

Simulations showed that agents could develop a common lexicon for describing coloured shapes, indicating that the knowledge in the domain could be successfully distributed across a field. The distributed nature of the domain in these simulations opens up new opportunities for modelling social and cultural aspects of creativity, including the transfer of knowledge between domains and the maintenance of domains through education, see Saunders (2011) for details.

4.2.1 Other Computational Models of Creative Domains

Meme and Variations (MAV) is a computational model of how cultural evolution is driven by a society of interacting individuals, based on the premise that useful novel ideas are variations of existing ones (Gabora, 1995). Axelrod's model of the dissemination of culture, while not attempting to model cultural creativity, illustrates the significance that individual acts of communication can have on the formation and multi-cultural societies (Axelrod, 1997). Miranda et al (2003) developed a model of the evolution of simple musical forms using a language game such that compositions are shared through agents performing for each other and the success of a tune is measured by the ability of another agent to reproduce it.

4.2.2 Embodied Artificial Creative Systems

Curious Whispers is a project to develop artificial creative systems using simple robots to explore opportunities afforded by embodiment for computational creativity (Saunders et al, 2010). The current implementation of Curious Whispers (version 2.0) uses a small group of mobile robots, each equipped with speakers, microphones and a movable plastic cover, see Figure 4.

Each robot is capable of generating songs, evaluating the novelty and value of a song, and performing those songs that they determine to be potentially creative to other members of the society. Each robot also listens to the performances of the other robots and if it values a song attempts to compose a variation.

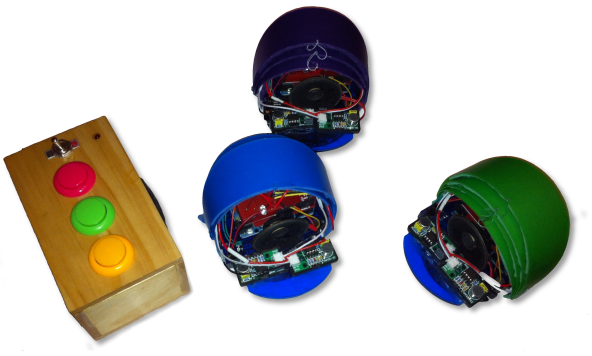

Fig. 4 Curious Whispers 2.0, an embodied artificial creative system consisting of robots, capable of generating and performing simple songs, and a simple three-button synthesiser to allow humans to participate. (Photo: Emma Chee)

The robots have been programmed to take advantage of their embodiment in order to reduce their computing requirements. By closing their plastic covers, the robots can play songs to themselves during composition sessions. This allows them to make use of the same hardware and software that they use to analyse the songs of other robots and removes the need for simulation that would require a more powerful central processing unit.

Unlike the simulations described above, humans are welcome to join this tabletop society by playing songs to the robots using a simple synthesiser. If the robots consider the songs played by humans to be interesting they will adopt them. Using this simple interface, humans are free to introduce domain knowledge from their culture, e.g., fragments of well-known songs, into the collective memory of the robot society.

5 Towards Autonomous Creative Systems

The previous section has presented three implementations of artificial creative systems, each of which explores different aspects of the role of communication in the generation, expression, evaluation, and description of creative artefacts. Understanding that communication is an essential component of any creative system, whether human or artificial, provides a different perspective on the nature of creativity. Returning to the questions from the beginning of this paper, we can explore the implications of this perspective on the relationship between creativity and autonomy.

5.1 What would it mean to produce an autonomous creative system?

An autonomous creative system, in the sense that Smithers (1997) demands, will have to be self-governing but will also need to be open to communication. Free from interference, an autonomous creative system will develop its own values, meanings, languages and practices. Interacting with an autonomous creative system will require a period of mutual learning and negotiation through repeated communications, such as it occurs in language games. Sustained communication building on a common ground, no matter how ambiguous, will provide opportunities to "educate" the autonomous creative system.

Building artificial creative systems that support communication between all agents, whether human or artificial, provides an opportunity to study the processes of mutual learning, negotiation and education in simple systems, e.g., the interactions between humans and the tabletop society of singing robots, as outlined above.

5.2 How might we begin to develop an autonomous creative system?

The approach proposed here is to bootstrap the development of autonomous creative systems by starting with very simple artificial creative systems and iteratively reducing the dependence on fixed rules that govern the processes of generation, evaluation and communication.

The development of more sophisticated models of communication, such as the ones being developed by Oudeyer and Kaplan (2006) for developmental robotics, may be of particular importance in the bootstrapping process. As Wittgenstein (1953) noted, there are potentially countless types of language games that arise from the activities that people engage in:

> 1. Giving orders, and obeying them
> 2. Describing the appearance of an object, or giving its measurements
> 3. Constructing an object from a description (a drawing)
> 4. Reporting an event
> 5. Speculating about an event
>
> (Wittgenstein 1953)

It is not necessarily the goal here to only replicate natural languages—all forms of creative expression are communicated in languages, whether they are written and oral, visual, musical, gestural, etc. Focussing on the communicative abilities of computational systems to learn from and experiment with languages, in whatever mediums they arise, will provide agents with a solid basis for engaging in creative acts within a dynamic social and cultural context.

5.3 How would we know if we have succeeded?

The recent interest in developing methods of evaluating computational creativity has resulted in a number of tools for evaluating creative systems based on product and/or process (Ritchie, 2001; Colton and Pease, 2011). Jordanous (2011) has conducted the most comprehensive and rigorous review to date on the evaluation of computational creativity, and has developed a Standardised Procedure for Evaluating Creative Systems (SPECS).

Any autonomous creative system will have the features of one or more coupled autopoietic systems, in particular, the creative process will be confined to its boundary. This raises significant challenges for the evaluation of an autonomous system's creativity. Clearly, ad hoc evaluations based on versions of the Turing Test will be of little or no use in determining the creativity of such systems. Even if, as Colton and Pease (2011) suggest, these tests include an interactive component, the products of the system and the responses of its members cannot be assessed by humans uncoupled from the system. Thus, evaluating the creativity of autopoietic systems will require us to look at the coupling of systems, whether artificial or human, as a means of establishing a common reference frame. Consequently, the evaluation of autonomous creative systems may have to take on practices more akin to anthropology (Saunders and Gero, 2002).

6 Conclusion

Human creativity is a complex phenomenon that emerges from the interactions between people, processes, products, and environments. While systems theories of creativity have made strong cases for the importance of interaction between individuals and their social and cultural environments, computational creativity has been dominated by a focus on the simulation of cognitive processes and the replication of extraordinary talents. The central argument of this paper is that no model of creativity can be complete without an account of the interactions between individuals and their social and cultural environments.

This paper has presented an alternative approach for the development of computational creativity, informed by systems theories of human creativity. Artificial creative systems are computational models of creativity built around communication between self-motivated agents capable of independent generation and evaluation of novel and valuable products. In the Digital Clockwork Muse, interactions between curious agents resulted in the emergence of (sub)fields of individuals with shared values that change over time as the field collectively explores a space of possibilities. The introduction of language games to an artificial creative system enabled agents to not only exchange artefacts and evaluations but to negotiate descriptions and meanings for their products. Exploiting the embodiment of an artificial creative system implemented as a group of mobile robots opened the system up to the injection of domain knowledge from human cultures.

As Smithers (1997) argued, autonomy (self-government) of computational systems must remain something that we aspire to. A significant leap of faith is required to imagine that the artificial creative systems described in this paper could ever become autonomous in this sense. Accepting the essential nature of communication, however, provides a different perspective on the development of autonomous creative systems. In aspiring to creative autonomy, it seems critical that we find ways to take advantage of the rich physical, social and cultural environments as valuable resources, rather than looking at them as yet another problem to solve (Clark, 1998).

The artificial creative systems approach introduced here has explored the nature of interaction and has identified communication between individuals using grounded languages as an effective way to exploit social and cultural environments in the service of creativity. Ultimately, the ability to learn grounded languages with which to express and describe concepts may prove to be the central challenge facing the developers of autonomous creative systems.

Acknowledgements This research has been partly supported by the Australian Research Council, Discover Grant DP0666584.

References

Amabile TM (1985) Motivation and creativity: Effects of motivational orientation on creative writing. Journal of Personality and Social Psychology 48:393–399

Anderson JR, Lebiere C (1998) The atomic components of thought. Erlbaum, Mahwah, NJ

Arden R, Chavez RS, Grazioplene R, Jung RE (2010) Neuroimaging creativity: A psychometric view. Behavioural Brain Research 214:143–156

Axelrod R (1997) The dissemination of culture: A model with local convergence and global polarization. The Journal of Conflict Resolution 41(2):203–226, DOI 10.2307/174371, URL http://dx.doi.org/10.2307/174371

Berlyne DE (1960) Conflict, Arousal and Curiosity. McGraw-Hill, New York, NY

Berlyne DE (1971) Aesthetics and Psychobiology. Appleton-Century-Crofts, New York, NY

Boden MA (1990) The Creative Mind: Myths and Mechanisms. Cardinal, London

Boden MA (ed) (1994) Dimensions of creativity. MIT Press, Cambridge, MA

Boden MA (1995) Creativity and unpredictability. Stanford Education and Humanities Review 4(2)

Boden MA (2004) The creative mind: Myths and mechanisms, (2nd edition) edn. Routledge, London, UK.

Bohm D (1968) On creativity. Leonardo 1(2):137–149

Bourgine P, Stewart J (2004) Autopoiesis and cognition. Artificial Life 10:327–345

Bown O (2008) Theoretical and computational models of cohesion, competition and maladaptation in the evolution of human musical behaviour. PhD thesis, University of London (Goldsmiths), London, UK

Chermahini SA, Hommel B (2010) The (b)link between creativity and dopamine: Spontaneous eye blink rates predict and dissociate divergent and convergent thinking. Cognition 115:458–465

Clark A (1998) Being There: Putting Brain, Body, and World Together Again. A Bradford Book, Cambridge, MA

Colton JC Simon, Pease A (2011) Computational creativity theory: The face and idea models. In: Proceedings of the International Conference on Computational Creativity 2011

Colton S, Bundy A, Walsh T (2000) Agent based cooperative theory formation in pure mathematics. In: Proceedings of the AISB'00 Symposium on Creative and Cultural Aspects and Applications of AI and Cognitive Science

Csikszentmihalyi M (1988) The nature of creativity. In: Sternberg RJ (ed) Society, culture, and person: a systems view of creativity, Cambridge University Press, Cambridge, UK, pp 325–339

Csikszentmihalyi M (1997) Creativity: Flow and the psychology of discovery and invention. Harper Collins, New York, NY

Csikszentmihalyi M (1999) Implications of a systems perspective for the study of creativity. In: Sternberg RJ (ed) Handbook of Creativity, Cambridge University Press, Cambridge, UK, pp 313–335

Feldman DH, Csikszentmihalyi M, Gardner H (1994) Changing the World: A Framework for the Study of Creativity. Praeger Publishers, Westport, CT

Gabora L (1995) Meme and variations: A computer model of cultural evolution. In: Nadel L, Stein DL (eds) 1993 Lectures in Complex Systems, Addison Wesley, pp 471–486, URL http://www.vub.ac.be/CLEA/liane/papers/mav.htm

Galenson DW (2008) Old Masters and Young Geniuses: The Two Life Cycles of Artistic Creativity. Princeton University Press, NJ

Gardner H (1993) Creating Minds. Basic Books, New York, NY.

Gaver WW, Mandler G (1987) Play it again, sam: On liking music. Cognition and Emotion 1:259–282

Gero JS, Sosa R (2002) Creative design situations: Artificial creativity in communities of design agents. In: CAADRIA

Gruber HE (1974) Darwin on Man. University of Chicago Press, Chicago

Guilford JP (1950) Creativity. American Psychologist 5(9):444–454, address of the President of the American Psychological Association at Pennsylvania State College, September 5, 19SO.

Hennessey BA, Amabile TM (2010) Creativity. Annual Review of Psychology 61:569—598

Humphrey NK (1973) The illusion of beauty. Perception 2:429–439

Iba T (2009) An autopoietic systems theory for creativity. In: 1st Conference on Collaborative Innovation Networks

Jordanous AK (2011) Evaluating computational creativity: A standardised procedure for evaluating creative systems and its application. PhD thesis, Department of Informatics, University of Sussex

Koestler A (1964) The Act of Creation. Pan Books, London

Laird J, Rosenbloom P, Newell A (1987) Soar: An architecture for general intelligence. Artificial Intelligence 33:1–64

Langley P, Simon HA, Bradshaw GL, Zytkow JM (1987) Scientific discovery: Computational explorations of the creative processes. MIT Press, Cambridge, MA

Lenat DB (1976) Am: An artificial intelligence approach to discovery in mathematics as heuristic search. Ph.d. thesis, Stanford University, AI Lab., Stanford, CA.

Lenat DB, Brown JS (1984) Why am and eurisko appear to work. Artificial Intelligence 23(3):269–294

Lindqvist G (2003) Vygotsky's theory of creativity. Creativity Research Journal 15(2-3):245–251

Luhmann N (1984) Soziale Systeme: Grundrieiner allgemeinen Theorie. Suhrkamp, Frankfurt

Luisi PL (2003) Autopoiesis: A review and a reappraisal. Naturwissenschaften 90:49–59

Martindale C (1990) The Clockwork Muse. Basic Books, New York, NY.

Martindale C, Moore K, West A (1988) Relationship of preference judgements to typicality, novelty, and mere exposure. Empirical Studies of the Arts 6(1)

Maturana HR, Varela FJ (1980) Autopoiesis and Cognition: The Realization of the Living. Reidel Publishing Company, Dordrecht

Mayer RE (1999) Fifty years of creativity research. In: Sternberg RJ (ed) Handbook of Creativity, Cambridge University Press, Cambridge, UK., chap 22, pp 449–460

McCarthy J, Minsky M, Rochester N, Shannon C (1955) A proposal for the dartmouth summer research project on artificial intelligence. URL http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

McMullin B, Varela F (1997) Rediscovering computational autopoiesis. In: Husbands P, Harley J (eds) Proceedings of the fourth ECAL, MIT Press, Cambridge, MA

Menabrea LF, Augusta A (1842) Sketch of the analytical engine. Bibliothèque Universelle de Genève (82)

Minsky M (1975) A framework for representing knowledge. In: Winston P (ed) The Psychology of Computer Vision, McGraw-Hill

Miranda ER, Kirby S, Todd PM (2003) On computational models of the evolution of music: From the origins of musical taste to the emergence of grammars. Contemporary Music Review 22(3):91–111

Oudeyer PY, Kaplan F (2006) Discovering communication. Connection Science 18(2):189–206

Petroski H (1992) The Evolution of Useful Things. Vintage, New York, NY

Plucker JA, Beghetto RA, Dow GT (2004) Why isn't creativity more important to educational psychologists? potentials, pitfalls, and future directions in creativity research. Educational Psychologist 39(2):83–96

Poincaré H (1913) Mathematical creation. In: The Foundations of Science: Science and Hypothesis, The Value of Science, The Science Press, New York, pp 383—394, translated by G. Bruce Halsted

Rhodes M (1961) An analysis of creativity. Phi Delta Kappan 42(7):305—310

Ritchie G (2001) Assessing creativity. In: Proceedings of the AISB Symposium on AI and Creativity in Arts and Science, York, UK, pp 3–11

Saunders R (2011) Artificial creative systems and the evolution of language. In: Ventura D, Gervás P, Harrell DF, Maher ML, Pease A, Wiggins G (eds) Proceedings of the Second International Conference on Computational Creativity, México City, México, pp 36–41

Saunders R, Gero JS (2001) Artificial creativity: A synthetic approach to the study of creative behaviour. In: Gero JS (ed) Computational and Cognitive Models of Creative Design V, University of Sydney, Sydney, Australia, pp 113–139

Saunders R, Gero JS (2002) How to study artificial creativity. In: Proceedings of Creativity and Cognition 4

Saunders R, Gemeinboeck P, Lombard A, Bourke D, Kocabali B (2010) Curious whispers: An embodied artificial creative system. In: International Conference on Computational Creativity 2010, 7–9 January 2010, Lisbon, Portugal.

Sawyer R (2003) Group Creativity: Music, Theater, Collaboration. Lawrence Erlbaum Associates

Sawyer RK (2008) Group Genius: The Creative Power of Collaboration. Basic Books

Sheldon KM (1995) Creativity and self-determination in personality. Creativity Research Journal 8(1):25—36

Simonton DK (1984) Genius, Creativity, and Leadership: Historiometric Inquiries. Harvard University Press, Cambridge, MA.

Simonton DK (1988) Creativity, leadership, and chance. In: Sternberg RJ (ed) The Nature of Creativity, Cambridge University Press, Cambridge, UK, chap 16, pp 386—426

Simonton DK (1990) Psychology, science and history: A introduction to historiometry. Yale University Press., New Haven

Sims K (1991) Artificial evolution for computer graphics. Computer Graphics 25(4):319–328

Smithers T (1997) Autonomy in robots and other agents. Brain and Cognition 34:88–106

Steels L (1995) A self-organizing spatial vocabulary. Artificial Life 2(3):319–332

Steels L (1996) Perceptually grounded meaning creation. In: Tokoro M (ed) ICMAS96, AAAI Press

Sternberg RJ (ed) (1988) The Nature of Creativity: Contemporary Psychological Perspectives. Cambridge University Press, Cambridge, UK

Sternberg RJ, Lubart TI (1999) The concept of creativity: Prospects and paradigms. In: Sternberg RJ (ed) The Handbook of Creativity, Cambridge University Press, Cambridge, UK, pp 3–15

Takeuchi H, Taki Y, Hashizume H, Sassa Y, Nagase T, Nouchi R, Kawashima R (2011) Failing to deactivate: The association between brain activity during a working memory task and creativity. NeuroImage 55:681–687

Tardif TZ, Sternberg RJ (1988) What do we know about creativity? In: Sternberg RJ (ed) The Nature of Creativity, Cambridge University Press, Cambridge, UK., chap 17, pp 429—440

Taylor CW (1988) Various approaches to the definition of creativity. In: Sternberg RJ (ed) The Nature of Creativity: Contemporary Psychological Perspectives, Cambridge University Press, Cambridge, UK, pp 99–124

Thornton C (forthcoming) A quantitative reconstruction of boden's creativity theory. Artificial Intelligence

Turing A (1948) Machine intelligence. In: Copeland BJ (ed) The Essential Turing: The ideas that gave birth to the computer age, Oxford University Press, Oxford

Varela F (2000) El fenmeno de la vida. Ensayo, Santiago

Varela FJ (1979) Principles of biological autonomy, The North-Holland Series in General Systems Research, vol 2. Elsevier North-Holland, Inc., New York, NY

Vygotsky LS (1971/1930) The psychology of art. MIT Press, Cambridge, MA, original work published 1930.

Wallas G (1926) Art of Thought. Harcourt Brace, New York, NY

Weisberg RW (1988) Problem solving and creativity. In: Sternberg RJ (ed) The Nature of Creativity, Cambridge University Press, Cambridge, UK.

Wiggins GA (2006) A preliminary framework for description, analysis and comparison of creative systems. Journal of Knowledge Based Systems 19(7):449—458

Wittgenstein L (1953) Philosophical Investigations. Blackwell Publishing, London