

A preliminary framework for description, analysis and comparison of creative systems

Geraint A. Wiggins

Centre for Cognition, Computation and Culture, Department of Computing, Goldsmiths' College, University of London, New Cross, London SE14 6NW, UK

Received 23 January 2006; accepted 15 April 2006

Available online 12 June 2006

Abstract

I summarise and attempt to clarify some concepts presented in and arising from Margaret Boden's (1990) descriptive hierarchy of creativity, by beginning to formalise the ideas she proposes. The aim is to move towards a model which allows detailed comparison, and hence better understanding, of systems which exhibit behaviour which would be called "creative" in humans. The work paves the way for the description of naturalistic, multi-agent creative AI systems, which create in a societal context.

I demonstrate some simple reasoning about creative behaviour based on the new framework, to show how it might be useful for the analysis and study of creative systems. In particular, I identify some crucial properties of creative systems, in terms of the framework components, some of which may usefully be proven *a priori* of a given system.

I suggest that Boden's descriptive framework, once elaborated in detail, is more uniform and more powerful than it first appears.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

One of the few attempts to address the problem of creative behaviour in the early days of AI was that of Margaret Boden, perhaps best summarised in her book, *The Creative Mind* [2]. A common criticism of Boden's approach is that it is rather lacking in detail, and that it is not clear how the various components fit together to give a real account of creative behaviour.

Boden's ideas have been debated at some length [17,16,13,12,9,8,3], but little attempt has been made to give a mechanism through which they can be applied formally (and thence automatically). In the current paper, rather than entering into the debate above (which I leave for future work), I will attempt to make Boden's descriptive hierarchy more precise. In doing so, I will suggest some additions to the theory, which may or may not be implicit in Boden's account and show how some of the distinctions over which she has been challenged may perhaps be supported. The formalisation will also make it possible to

identify desirable and undesirable properties of creative systems in abstract terms.

Next, I will define some crucial properties of (artificial) creative systems in terms of the framework, including some which might be proven *a priori* and some which may be usefully detectable during the activity of the system.

I will conclude by suggesting that, once formalised, the uniformity and power of Boden's proposal becomes rather more clear than before, and that it may not be idly dismissed as vague, as it sometimes has been in the past.

In proceeding with this discourse, there is a hurdle which must be negotiated, which has been dogging the creative systems community for some time: what does "creative" mean, and what is "computational creativity"? First, however, I summarise my starting point in the literature of AI creativity research: the writing of Margaret Boden, so that definitions can be given in that context.

2. Background: Boden's analysis of creative systems

Boden (1990) aims to study the idea of AI-based simulation of creativity from a philosophical viewpoint.

E-mail address: geraint@city.ac.uk

URL: <http://soi.city.ac.uk/~geraint>

She begins by setting out two taxonomies of creative behaviour, in two orthogonal ways.

First, she makes the distinction between H- and P-creativity: creativity which is “historical” or “psychological”, respectively, the latter being interchangeable with “personal”, should that be more natural for the reader. The distinction is between the sense of creating a concept which has never been created before at all, and a concept which has never been created before by a specific creator. This distinction will be only tangentially relevant to my argument here, but before proceeding, I note that this is not a simple binary choice, but rather multi-dimensional, context-based one: it would be possible, for example, for a creative behaviour to be only P-creative in one society, but H-creative in another; from the point of view of the second society, only the H-creativity matters.

Second, in Boden’s work, there is the distinction between *exploratory* and *transformational* creativity, which is directly relevant here, and so needs a little more explanation. Boden conceives the process of creativity as the identification and/or location of new conceptual objects in a *conceptual space*. Subsequent authors have sometimes imagined the conceptual space to be the state space of Good Old Fashioned AI [15], though it is not clear that Boden intends her proposal to be taken so specifically or literally. I discuss this elsewhere [18].

If we do go along with Boden’s conception, for the sake of argument, then the creative act might be said to be exploring a space of partial and complete possibilities, and this is the kind of creativity which Boden calls *exploratory*. The existence of such a conceptual space begs a question (at least to the AI researcher): what rules define the space? If there are rules which define the space, then, presumably, those rules can be changed, producing what might be thought of as a paradigm shift. This kind of change is *transformational creativity* to Boden.

However, it is not clear from Boden’s writings about these ideas how she would define the constraints which give rise to a particular conceptual space, and therefore what is the difference, in terms of the new concepts discovered, between exploring the space and transforming it. I will argue below that there is at least one way we can coherently make such a distinction. First, however, it is necessary to sharpen slightly the philosophical tools that Boden introduced.

3. Terminology: What is “computational creativity”?

Some 40-odd years ago the discipline of Artificial Intelligence was identified and named. One issue that dogged it then, and still does now, is “what – exactly – is the intelligence it claims to create and/or simulate and/or emulate?”. Many attempts at answers have been made, some more successful than others. One definition that seems to have some hope of enduring circumvents the problem of saying what intelligence *is*, by restricting its definition to where it *resides*:

“The performance of tasks, which, if performed by a human, would be deemed to require intelligence”.¹

This definition is in many ways unsatisfactory: for example, it does not include many of the fundamental aspects of robotics and machine vision which we do normally include in the field of AI. However, it will do nicely for the current purpose, because it does capture the parts of AI which are concerned with higher cognitive function, such as mathematical reasoning, construction of language semantics, and artistic pursuits including painting and music.

The zenith of human intelligence is very often portrayed as the ability to create, and to create radically new and/or surprising things. For example, in humans, the acts of elaborating a new and elegant mathematical proof, or designing a subtle new experiment to investigate the existence of a new sub-atomic particle (the prediction of which is itself a creative contribution), or writing a novel, painting a watercolour, or composing a sonata are all deemed the height of creativity, and therefore the height of (one aspect of) intelligence.

Throughout human society, creative individuals and groups are valued very highly, and creative behaviour is fundamental in that society. For example, musical behaviour is a uniquely human trait (notwithstanding our anthropomorphic tendencies in terminology such as bird- and whale-“song”, which are in fact much more like language than music); further, it is also ubiquitously human: there is no known human society which does not exhibit musical behaviour in some form.²

It seems, then, that at least commonplace definitions of creativity place it as a primary determiner of human intelligence. As such, we might expect it to be at the forefront of AI research, but it is not – at least, not explicitly. Boden [1] raised the issue for what seems to be the first time in AI literature, but was apparently rebuffed:

“... when (in the early 1970s) I included a chapter on creativity in my book *Artificial Intelligence and Natural Man* [(1), chapter 11], most people – including my AI-friends – asked me in puzzlement ‘Why on earth are you doing that?’”. ([5], p. 11)

Perhaps creativity is to AI researchers what intelligence seems to be to some of those computer scientists and philosophers who continue to argue against AI: that feature which is best left alone, lest we cease to be distinguishable from machines, and belief in whose attainment is viscerally beyond the pale.

¹ This definition is now part of the AI folklore, being attributed to various authors including Minsky and Turing. Its true original source is obscure to me; I apologise, therefore, to the original author for lack of citation.

² For an unscientific thought-demonstration of music’s ubiquity, the reader may wish to try to name a social or ceremonial occasion in which music is not usually involved at some level. The legal courts are the only example this author has found.

Returning, then, to the topic of the present section: how can we define “creativity” and thence “computational creativity”? It turns out that one good way to do so is to adapt the definition of intelligence I gave above. This is a good strategy for two reasons: like “intelligence”, “creativity” is ill-defined, but we do tend to know it when we see it; worse than “intelligence”, “creativity” as a word is overloaded, and is usable in distinctly different and *confusing* ways. First, then, I suggest that a useful working definition of “creativity” is

“The performance of tasks which, if performed by a human, would be deemed creative”.

Note that this definition includes the production of creative artefacts, which may be deemed “more” or “less creative” in some usages of the C-word.

From this, my personal definition of “computational creativity” is

“The study and support, through computational means and methods, of behaviour exhibited by natural and artificial systems, which would be deemed creative if exhibited by humans”.

This “study and support” may, of course, include simulation.

Having now given a definition, however intangible, of the word “creativity”, I will now forswear most of its diverse usage, on the grounds that it is already defined in too many ways to carry yet another meaning, no matter how precise. I shall use the following terminology:

Creative system A collection of processes, natural or automatic, which are capable of achieving or simulating behaviour which in humans would be deemed creative.

Creative behaviour One or more of the behaviours exhibited by a *creative system*.

Novelty The property of an artefact (abstract or concrete) output by a *creative system* which arises from prior non-existence of like or identical artefacts in the context in which the artefact is produced.

Value The property of an artefact (abstract or concrete) output by a *creative system* which renders it desirable in the context in which it is produced.

Armed with these terms, I will attempt to avoid use of the word “creativity” itself altogether, and thus avoid the confusion it brings. The exception to this will be in the use of terminology introduced by other authors.

It is perhaps worth mentioning some particular terminological issues at this point. Some authors ([10], for example) seem to view “surprise” as a property of a creative system. I would argue that “surprise” is a property of the receiver

of a creative output: it is an emotion generated by either the novelty of the output, or (cynically) by the unexpected ability of the creative system to produce something of value. It is also common in everyday language to hear an artefact being called “creative” (for example, “That’s a creative painting!”). What this usually means is that the perceiver finds a blend of novelty and value in the artefact; it needs to be distinguished from the claim that the artefact is itself a creative system, which could be expressed by the same sentence.

Having pinned down the terminology a little, I now proceed to discuss my preliminary formalisation of Boden’s informal framework.

4. A framework for the description of creative systems

4.1. A universe of possibilities

Boden’s combination of the idea of the conceptual space with distinct notions of exploratory and transformational creativity has some consequences which are left implicit in her published work. Most fundamentally, for transformational creativity to have any meaning, there must be a universe of possibilities, which I shall call \mathcal{U} , which is a *non-strict superset* of the conceptual space at any given point in the creative process. To see the reason for this, let us first define \mathcal{U} .

Definition 1 (Universe). The *universe*, \mathcal{U} , is a multidimensional space, whose dimensions are capable of representing anything, and all possible distinct concepts correspond with distinct points in \mathcal{U} .

For parsimony, we could restrict \mathcal{U} to be capable of representing just the things which are relevant to the domain in which we wish to be creative – but this would rule out cross-domain transfer of ideas, by processes such as analogy, which would be undesirable in general. (I return to analogy and other means of guiding exploratory creativity below.) To make the proposal as state-space-like as possible, I allow that \mathcal{U} contains all abstract concepts as well as all concrete ones, and that it is therefore possible to represent both complete and incomplete artefacts. Henceforward I will refer to both incomplete and complete concepts simply as “concepts”, except where the distinction is significant. It follows from the inclusion of incomplete or abstracted concepts that we should also admit the most incomplete concept of all, the empty concept, which I will denote by \top , and that it should be a member of \mathcal{U} .

To summarise, the following points are axiomatic to my formulation. These axioms cannot be stated within the formulation, because they describe its own properties, and not just those of the system it models.

Axiom 1 (Universality). All possible concepts, including the empty concept, \top , are represented in \mathcal{U} ; thus, \mathcal{U} is the type of all possible concepts. $\top \in \mathcal{U}$.

Axiom 2 (Non-identity of concepts). All concepts, c_i , represented in \mathcal{U} are mutually non-identical. $\forall c_1, c_2 \in \mathcal{U}. c_1 \neq c_2$.

We need \mathcal{U} because, if a conceptual space were *equal to* \mathcal{U} (and \mathcal{U} were therefore superfluous), any point in \mathcal{U} could be reached by exploration. Therefore, transformation would be unnecessary. So, for transformational creativity to be meaningful, all conceptual spaces, \mathcal{C} , are required to be non-strict subsets of \mathcal{U} . Therefore, the type of conceptual spaces, Σ , is the set of subsets of \mathcal{U} whose members include \top .

Axiom 3 (Universal Inclusion 1). All conceptual spaces are non-strict subsets of \mathcal{U} . $\forall \mathcal{C} \in \Sigma. \mathcal{C} \subseteq \mathcal{U}$.

Axiom 4 (Universal Inclusion 2). All conceptual spaces include \top . $\forall \mathcal{C} \in \Sigma. \top \in \mathcal{C}$.

So far, I have done nothing more precise than Boden's informal characterisation; I have merely pinned the ideas down to a specific formulation and pointed out a logical consequence (the necessity for the existence of \mathcal{U} as distinct from \mathcal{C}). It is in the definition of \mathcal{C} , in terms of its own constraints, rather than its relation to \mathcal{U} , that we first find the necessity to clarify the existing ideas.

4.2. Defining the conceptual space

Boden [1] does not explicitly acknowledge the existence of (an equivalent of) \mathcal{U} . Instead, she loosely defines her conceptual spaces in terms of a set of definitional rules, which we must therefore assume to be generative. However, she blurs the distinction between the rules which determine membership of the space (*i.e.*, which select members of \mathcal{U} to be members of a particular \mathcal{C} , in my terms), and other rules which might allow the construction and/or detection of a concept represented by a point in the space. To remedy this, let us take two distinct rule sets, \mathcal{R} and \mathcal{T} , being rules which constrain the space and rules which allow us to traverse it, respectively. In AI terms, then, \mathcal{T} might be thought of as encoding a search strategy, perhaps including heuristics.

In order to introduce these sets of rules, we need a language in which to express them; it will be convenient to arrange that both sets can be expressed in the same language.³ I will call this language \mathcal{L} , and formalise it as the set of all sequences composed of some alphabet, \mathcal{A} . Therefore, by definition

$$\mathcal{R} \in \mathcal{L}, \quad \mathcal{T} \in \mathcal{L}.$$

Once given the language, and rule sets expressed in it, we need an interpretation function, $\llbracket \cdot \rrbracket$, which is a partial function⁴ from \mathcal{L} to functions yielding real numbers in $[0, 1]$. At this point, we will use a real value greater than or equal to

³ This is always possible: a language can always be the union of two others, given any necessary renaming apart.

⁴ Partial, because I have not required that \mathcal{L} contain only sequences which are well-formed with respect to $\llbracket \cdot \rrbracket$. The reason for this will become clear below.

0.5 to mean Boolean `true` and less than 0.5 to mean Boolean `false`; the need for the real values will become clear below. This Boolean decision procedure will allow us to choose the members of \mathcal{U} we want in \mathcal{C} , assuming that \mathcal{R} is well-formed with respect to $\llbracket \cdot \rrbracket$:

$$\mathcal{C} = \llbracket \mathcal{R} \rrbracket(\mathcal{U}).$$

4.3. Exploring the conceptual space

A similar approach is required for the application of the search strategy encoded in \mathcal{T} , though a little more computational mechanism is required. We need a means not just of *defining* the conceptual space, irrespective of order, but also, at least notionally, of *enumerating* it, in a particular order, under the control of \mathcal{T} – this is crucial to the simulation of a particular creative behaviour by a particular \mathcal{T} . By analogy, again, with the familiar approach to state space search, I introduce a further interpreter $\langle \langle \dots \rangle \rangle$, which, given three well-formed subsets of \mathcal{L} , computes a function which maps c_{in} , a totally ordered subset of \mathcal{U} , to c_{out} , another totally ordered subset of \mathcal{U} . As with \mathcal{R} and $\llbracket \cdot \rrbracket$, I assume that \mathcal{T} does not contain subsequences which have no interpretation under $\langle \langle \dots \rangle \rangle$. The three arguments of $\langle \langle \dots \rangle \rangle$ are, respectively, the rule set defining the conceptual space, \mathcal{R} , the traversal strategy, \mathcal{T} , which the function is intended to apply, and another set, \mathcal{E} , which I will define below; this gives \mathcal{T} the possibility of being informed by \mathcal{R} and \mathcal{E} . The resulting function operates on members of \mathcal{U} and not just on members of \mathcal{C} , as one might expect, because it is necessary to describe and simulate behaviours which are not completely well-behaved. More on this below.

The ordering on c_{in} and c_{out} indicates the order in which the concepts in them are to be next considered for further development under \mathcal{T} , so the input and output of the function are successive states of a kind of agenda:

$$c_{out} = \langle \langle \mathcal{R}, \mathcal{T}, \mathcal{E} \rangle \rangle(c_{in}).$$

However, note that this formulation is more powerful than the standard formulation of AI state space search because the function is allowed to select arbitrarily many members of c_{in} and is not limited to the head(s) of the sequence. This is a key feature, because it admits search strategies which rely on the combination of or comparison between agenda items.

It follows that we would begin some of our creative processes by computing

$$\langle \langle \mathcal{R}, \mathcal{T}, \mathcal{E} \rangle \rangle(\{\top\}).$$

We now have all the mechanism we need to model Boden's exploratory creativity as presented in 1990 [1].

4.4. The value of two rule sets, \mathcal{R} and \mathcal{T}

Importantly, separating the rules out into the sets \mathcal{R} and \mathcal{T} has given us the ability to consider alternative versions of \mathcal{T} with any given \mathcal{R} , and, perhaps less obviously, *vice versa*. We can partition \mathcal{C} into \mathcal{C}_1 , concepts which have

already been discovered, and $\mathcal{C}_?$, concepts which have not. Some versions of \mathcal{T} may be effective in traversing \mathcal{C} and in finding members of $\mathcal{C}_?$; some may be less so; and some may be good at finding members of $\mathcal{C}_?$ in some parts of \mathcal{C} and not in others. Further, some elements in \mathcal{C} may not even be accessible under \mathcal{T} . So now we have, for example, the ability to simulate two composers working in the different ways within the same style, which was less clear in Boden's simpler formulation.

It is worth noting, also, why \mathcal{T} does not supplant \mathcal{R} as a primary component of the framework: this is because, in any real simulation of creativity, there is a distinction (as noted by, for example, [14]) between something being an X (where X is the kind of thing to be created) and it being a *valued* X . \mathcal{E} defines the value; \mathcal{R} defines the nature of the created artefact. In particular, in the societal context, \mathcal{R} represents the agreed nature of what the artefact is, in the abstract; this is distinct from \mathcal{T} , which defines the way a particular agent produces an artefact in practical terms.

4.5. Evaluating members of the conceptual space

To do full justice to Boden's model as presented in 1998 [4], we need one further set of rules, \mathcal{E} , such that $\mathcal{E} \in \mathcal{L}$. This is the set of rules which allows us to evaluate any concept we find in \mathcal{C} and determine its quality, according to whatever criteria we may consider appropriate – and, of course, it is not hard to imagine that \mathcal{T} might be related to or dependent on \mathcal{E} , which is why the interpretation function includes \mathcal{E} in its parameters. However, I am making no attempt here to discuss or assess the value of any concepts discovered: while this issue is clearly fundamentally important ([4,14,11]), it can safely be left for another time. Suffice it to say here that the existing function $\llbracket \cdot \rrbracket$ will be adequate to select those results of the creative process which are “valued” by \mathcal{E} , thus:

$$\llbracket \mathcal{E} \rrbracket (\langle \langle \mathcal{R}, \mathcal{T}, \mathcal{E} \rangle \rangle^\diamond (\{\top\})),$$

where

$$\mathcal{F}^\diamond(X) = \bigcup_{n=0}^{\infty} \mathcal{F}^n(X),$$

\mathcal{F} being a set-valued function of sets.

4.6. Characterising an exploratory creative system

To summarise, we now have the machinery to describe an exploratory creative system in these terms with the following septuple:

$$\langle \mathcal{U}, \mathcal{L}, \llbracket \cdot \rrbracket, \langle \langle \cdot, \cdot \rangle \rangle, \mathcal{R}, \mathcal{T}, \mathcal{E} \rangle.$$

4.7. Exploring and transforming

Before proceeding to the formality of transformational creativity, there are some more issues to discuss in the exploratory context.

It follows from my characterisation of \mathcal{T} as a search engine that there may be a fitness hypersurface associated with any combination of \mathcal{C} and \mathcal{T} . The “landscape” so defined may be arbitrarily – perhaps extremely – convoluted. This means that it is possible to imagine finding c , a member of $\mathcal{C}_?$ which is, in general, very hard to find, but doing so *without changing* \mathcal{T} . Finding such a concept would presumably mark the creator as successful, especially if other creators' \mathcal{T} s were less fortunate, for the discovery is unlikely. So here is a case where an exploratory creation might well be very significant – perhaps more significant than many transformational creations.

Now, consider the converse situation. Suppose that a concept c is a member of \mathcal{U} , but not a member of \mathcal{C} , and that we *transform* \mathcal{C} into \mathcal{C}' , by transforming \mathcal{R} into \mathcal{R}' – I discuss how this can happen below. Now we have exhibited transformational creativity, which, according to Boden, is more significant than exploratory creativity. But it may be the case merely that

$$\mathcal{C}' = \mathcal{C} \cup \{c\},$$

in which case it is hard to argue that the transformation is any more significant than the exploration in the account immediately above, unless, of course, c is very significant in itself.

Now let us consider a third possibility, one which was not available to Boden because of her conflation of my \mathcal{R} and \mathcal{T} : it is possible in principle for a concept which exists in \mathcal{C} – and so is sanctioned by the constraints in \mathcal{R} – to be unreachable by the rules specified in \mathcal{T} . This is an important point: it distinguishes what is *in principle* possible in a creative domain from what is *actually* possible according to the properties of a given creator. Therefore, another possibility for reaching the elusive discovery, c , above, is that $c \in \mathcal{C}$,

but the rules of \mathcal{T} fail to make it accessible:

$$c \notin \llbracket \langle \langle \mathcal{R}, \mathcal{T}, \mathcal{E} \rangle \rangle^\diamond (\{\top\}) \rrbracket.$$

So we have to introduce a new, different notion of transformational creativity – one which transforms not \mathcal{R} , the rules constraining the conceptual space, but \mathcal{T} . It is not hard to imagine that we can transform \mathcal{T} into some \mathcal{T}' such that c becomes accessible to our search.

From an external viewpoint, these different events are probably often indistinguishable, but the point is that they all fall short of Boden's informal definition of transformational creativity (that is, in the terms used here, changing \mathcal{R}) – which she argues is generally more significant than the exploratory kind.

So by making the argument more precise, we can demonstrate a potential weakness in Boden's characterisation: the boundary between exploratory and transformational creativity is ill-defined.⁵ We are now in a position to argue

⁵ This author is by no means the first to note this point.

that transformational creativity is unnecessary, and to conflate \mathcal{U} and \mathcal{C} , thus producing a simpler characterisation.

However, I will argue that there is indeed a valid distinction between a kind of creative behaviour that might be called “transformational” and a kind of creative behaviour that might be called “exploratory”. Before I can do so, however, we must consider transformational creativity in more detail.

4.8. Transformational creativity

Having gone some way towards formalising Boden’s notion of exploratory creativity, we are now in a better position to say what transformational creativity actually is. It is at this point that we begin to see the benefits of this laborious formalisation. In this section, I discuss transformational creativity informally; I will treat it more formally in a later section.

Boden characterises her “transformational creativity” as the kind of creative behaviour concerned not with finding members of \mathcal{C} , in a given conceptual space \mathcal{C} , but with transforming the rule set defining \mathcal{C} so as to produce a new conceptual space, \mathcal{C}' . In my terms, however, transformation might be achieved in two essential ways: by transforming \mathcal{R} or by transforming \mathcal{T} (recall that, although changing \mathcal{T} does not, by definition, change \mathcal{C} , any given \mathcal{T} cannot guarantee to reach all the elements of \mathcal{C} – so a new \mathcal{T}' may make a different subset of \mathcal{C} available). Transforming \mathcal{R} corresponds with changing the rules of the creative game being played – and, it seems, with what Boden calls “transformational” creativity. The new, second kind of transformation, of \mathcal{T} , more naturally applies to the creative individual’s *modus operandi* only – there seems to be no explicit analogue of this in Boden’s formulation. Of course, it is possible for both kinds of transformation to happen at once. To distinguish between these two different kinds of transformation, I will use the terms *\mathcal{R} -transformational* and *\mathcal{T} -transformational*; however, in using Boden’s terminology, I shall continue to use the unadorned “transformational”.

Now, as mentioned earlier, Boden concludes, from what she portrays as historical precedent, that her transformational creativity (*i.e.* \mathcal{R} -transformational) is somehow more significant than exploratory creativity. This claim deserves some more scrutiny in the light of my division of the creative rules into \mathcal{R} and \mathcal{T} . Let us consider the difference between these two.

Suppose, as Boden supposes, that \mathcal{R} defines a set of concepts which is largely agreed among all creative agents interested in the area defined by \mathcal{R} . Then, almost by definition, any change in \mathcal{R} has the force of a paradigm shift (even if only a little one), if it is valued highly enough by the *existing* evaluation rule set, \mathcal{E} , because it changes the *agreed* rules of the game. To ground this in an example, consider Kekulé’s discovery of benzene rings, cited repeatedly by Boden [1] as an example of transformational creativity. The idea was new because it allowed *rings* of carbon

atoms, and not just chains. But the evaluation system was independent of the shape used: it was a meta-theoretic evaluation of whether the theory explained the chemical data. Thus, Kekulé’s new rule set was valued more highly under the *existing* evaluation rules than the pre-existing solutions.

On the other hand, \mathcal{T} , as I have proposed it, is not global or agreed: it is the “technique” of the individual creator. Therefore, a change in \mathcal{T} is on a different scale from a change in \mathcal{R} : it may perhaps accelerate the agent’s progress towards a good solution; it may even make accessible concepts which were not previously available to this particular agent, but it will not change the nature of space of possibilities, and therefore will not constitute a paradigm shift. An archetype here would be the comparison between an expert organist, capable of convincingly harmonising a chorale melody, at first sight, in the style of J.S. Bach, and a beginning harmony student, struggling to do so for the first time. The rules of Bach Chorale harmony (\mathcal{R}) may be common to both, but the techniques (\mathcal{T}) of the two are not.

For completeness, it is necessary to consider the case where an \mathcal{R} -transformation is not necessarily adopted by all the creative agents working on \mathcal{R} . This case has, of course, been seen many times in history.⁶ It can arise reasonably only where different creative agents working in an initially common \mathcal{R} have different evaluation rule sets, \mathcal{E}_i – the alternative case, where there are initially differing \mathcal{R}_i s, does not correctly describe the example situation. This raises an interesting question of how discovery of new ideas can lead to changes in the evaluation rule set itself; this will be a focus of future work.

4.9. Creative behaviour and the meta-level

An aspect of this discussion which Boden (1990) leaves implicit is the formal relationship between exploratory and transformational creativity – one would need a formalisation of the kind presented here to do so. I now extend that formalisation to cover transformational creativity, and will show that, as informally conjectured by Bundy [7], we can view transformational creativity as exploratory creativity at the meta-level, and thus keep the framework simple.⁷

The idea at the root of Boden’s (\mathcal{R} -)transformational creativity is that of changing the rules which define her conceptual space. In my formulation, there are two such rule sets, \mathcal{R} and \mathcal{T} . So, in my terms, transformational creativity consists in changing either \mathcal{R} or \mathcal{T} or both. The two sets are expressed in the language \mathcal{L} , which means that the result of the transformation(s) must also be in \mathcal{L} . We can place a useful restriction on the results of these transformations: that they be well-formed in terms of whatever interpreter will interpret them. So we need a syntax checker

⁶ The continuing prevalence of tonal harmony in music long after the arrival of modernism is one example.

⁷ Buchanan [6] argues that creative behaviour arises only at the meta-level.

which will select the elements of \mathcal{L} which are well-formed in that sense.

Transformation of either kind means constructing new subsets of \mathcal{L} from old ones. Starting from the empty sequence, or from anywhere else in the space of possibilities, we can do this by application of a search algorithm. If we allow ourselves access to a meta-language, $\mathcal{L}_\mathcal{L}$, for \mathcal{L} , which can describe the construction of new members of \mathcal{L} from old ones, we can pair it with an appropriate interpreter, to allow us to search the space of possibilities. Since the syntax checking task mentioned above is a meta-level structural one (with respect to \mathcal{L}), we can use $\mathcal{L}_\mathcal{L}$ to describe this task too, again given an appropriate interpreter. Finally, we need to be able to evaluate the quality of the transformational creativity, with some function Ω .

By now, the reader will see where this argument is going. We can specify interpreters, $\llbracket \cdot \rrbracket$ and $\langle \langle \dots \rangle \rangle$, which will interpret a rule set $\mathcal{T}_\mathcal{L}$ applied to an agenda of potential sequences in \mathcal{L} ; if we are careful, we can specify such an interpreter which will work for both \mathcal{L} and $\mathcal{L}_\mathcal{L}$. Finally, we can express our evaluation function, Ω , as a set of sequences, $\mathcal{E}_\mathcal{L}$, in $\mathcal{L}_\mathcal{L}$, and use $\llbracket \cdot \rrbracket$, to execute it. This time, we can allow the real-valued output of the interpreter to be used either for comparison or (as before) selection, depending on context. Our transformational creativity system can now be expressed as

$$\langle \mathcal{L}, \mathcal{L}_\mathcal{L}, \llbracket \cdot \rrbracket, \langle \langle \dots \rangle \rangle, \mathcal{R}_\mathcal{L}, \mathcal{T}_\mathcal{L}, \mathcal{E}_\mathcal{L} \rangle,$$

or, in other words, as an exploratorily creative system working at the meta-level of representation.

The only connection we have now not considered is that (if any) between \mathcal{E} and $\mathcal{E}_\mathcal{L}$. I suggested above that, for a transformationally creative act to be valued, it would normally need to be valued under the criteria that governed the original search space. This begs the question of how to relate \mathcal{E} , which is defined over the exploratory/object-level universe \mathcal{U} , to the transformational/meta-level universe, \mathcal{L} .

Informally, and minimally, the transformation is valued if it admits a new concept which is valued to the available object-level conceptual space. We can express this, in terms of the exploratory creative system described above, by saying that $\mathcal{E}_\mathcal{L}$ is the rule set which selects pairs of $\mathcal{R}_\mathcal{L}$ and $\mathcal{T}_\mathcal{L}$, such that

$$\{c|r \in \langle \langle \mathcal{R}_\mathcal{L}, \mathcal{T}_\mathcal{L}, \mathcal{E}_\mathcal{L} \rangle \rangle^\diamond(\{\mathcal{R}\}) \wedge t \in \langle \langle \mathcal{R}_\mathcal{L}, \mathcal{T}_\mathcal{L}, \mathcal{E}_\mathcal{L} \rangle \rangle^\diamond(\{\mathcal{T}\}) \\ \wedge c \in \llbracket \mathcal{E} \rrbracket \langle \langle r, t, \mathcal{E} \rangle \rangle^\diamond(\{\perp\}) \neq \emptyset$$

where \diamond is as before. In other words, $\mathcal{E}_\mathcal{L}$ is the rule set which selects pairs of $\mathcal{R}_\mathcal{L}$ and $\mathcal{T}_\mathcal{L}$ such that new concepts are added to the conceptual space under consideration, and that those new concepts are valued by \mathcal{E} .

This meta/object level distinction raises some interesting questions. The most obvious is: if this relation holds between the object level of our creative domain and the meta-level of transformational creativity, what would it mean to take the same relation and apply it to the

transformational level? However, I will leave these issues for future work.

To conclude the current section, let us return to the issue of the relative values of exploratory and transformational creativity, as introduced by Boden [1]. I have argued above that Boden's suggestion that transformational creativity is innately superior to exploratory creativity is not well founded in terms purely of the creative product. However, the meta-level notion of transformational creativity which I introduce above gives us another means of looking at the question, a means which Boden does not (at least, explicitly) use.

I suggest that, for true transformational creativity to take place, as described in my framework, above, the creator needs to be in some sense *aware* of the rules he/she/it is applying. This follows from the need to explore the space of possible rule sets defining the conceptual space. One might argue that serendipity – a happy accident – might account for creative behaviour, and this can certainly be the case, but that would be a new category, of “serendipitous” creativity, and not transformational creativity, under either of my definitions. I make this point because it fits in very clearly with philosophical notions of art. Self-awareness is generally cited as the property which distinguishes the artist from the craftsman.⁸ That self-awareness, I suggest, is what makes a creator able to formalise his/her/its own \mathcal{R} and \mathcal{T} in terms of the meta-language $\mathcal{L}_\mathcal{L}$. So without that self-awareness, a creator *cannot* exhibit transformational creativity, though, conversely, of course, a creator *with* self-awareness may choose not to exercise it.

5. Applying the framework

5.1. Introduction

A theoretical framework, such as that presented here, may be useful in teasing out philosophical issues, but it may also be useful in giving generalised descriptions of behaviours which might be observed in creative agents. In this section, we use the tools provided in earlier sections to identify certain potentially important behaviours of creative systems, and suggest ways in which they might be addressed. It is to be emphasised that the aim here is not to ground this particular framework as an implementation, but to show how it might be used to describe implementations and some of their properties in useful ways.

5.2. Useful properties of creative agents

The apparent supposition in Boden's work is that creative agents will be well-behaved, in the sense that they will either stick within their conceptual space, or alter it politely and deliberately by transformation. It can be argued,

⁸ Whether this is a valid distinction is an orthogonal issue, not discussed here.

however, that this is not adequate to describe the behaviour of real creative systems, natural or artificial, either in isolation or in societal context. This section identifies some situations not covered by the assumption of good behaviour, and gives names to them. The important point is that some of these situations may appropriately trigger particular events, such as a step of transformational creativity, so it is useful to be able to identify them in the abstract. This leaves us with several general classes of small-scale conditions which might be observed in AI systems, of which we can then assess the creative potential.

5.2.1. Uninspiration

There are various ways that a creative system can fail to be creative in a valued way. These ways can be characterised through the rule set \mathcal{E} .

Hopeless uninspiration is the simplest case, where there are no valued concepts in the universe:

$$\llbracket \mathcal{E} \rrbracket(\mathcal{U}) = \emptyset.$$

This system is incapable, by definition, of creating valued concepts, and as such might be termed ill-formed (if such creative behaviour is the intention).

Conceptual uninspiration arises when there are no valued concepts in the conceptual space:

$$\llbracket \mathcal{E} \rrbracket(\llbracket \mathcal{R} \rrbracket(\mathcal{U})) = \emptyset.$$

I label this form of uninspiration “conceptual” because it entails a mismatch between \mathcal{R} (which defines the conceptual space) and \mathcal{E} (which evaluates concepts within it, and, more broadly, within \mathcal{U}). This condition is contradictory to the purpose of the two rule sets: if \mathcal{R} is supposed to constrain the domain of a creative process, then it is inappropriate for \mathcal{E} not to select some of the elements it admits. As such, like the hopeless case, conceptual uninspiration indicates ill-formation of the intended-creative system.

Conceptual uninspiration can only be remedied by transforming \mathcal{R} , by modifying \mathcal{E} or by aberration (see below), which in itself requires transformation. How \mathcal{R} and/or \mathcal{E} should be modified is an open question.

Generative uninspiration occurs when the technique of the creative agent does not allow it to find valued concepts within the space constrained by \mathcal{R} :

$$\llbracket \mathcal{E} \rrbracket(\llbracket \mathcal{R}, \mathcal{T}, \mathcal{E} \rrbracket^{\diamond}(\{\top\})) = \emptyset.$$

This kind of uninspiration is less serious than the other two, and does not necessarily indicate an ill-formed creative system: it merely indicates that a creative agent is looking in the wrong place. This raises the question of *why* there is such a mismatch. Boden’s underlying assumption seems to be that the conceptual space is in some sense definitive, and, certainly, in a multi-agent environment, it is the only place in the formalism where the consensus about a creative domain can logically be represented. Therefore, I propose that the usual solution to generative uninspiration would be transformation of \mathcal{T} , for the agent concerned, but that the transformation of \mathcal{R} (instead, or as well)

may also be a valid response, noting that such transformation may be non-trivial in a multi-agent environment.

5.2.2. Aberration

Now, consider the following more interesting scenario, which also concerns the relationship between \mathcal{R} and \mathcal{T} . A creative agent, \mathbf{G} , is traversing its conceptual space. From any (partial) concept(s) in the conceptual space, \mathbf{G} ’s technique will enable it to create another (some more) concept(s). Suppose now that the new concept does not conform to the constraints required for membership of the existing conceptual space (note that there is no guarantee that it should do so – there is only an assumption in Boden’s work), and is therefore not selected by $\llbracket \mathcal{R} \rrbracket(\cdot)$. In this case, the set \mathcal{B} given by

$$\mathcal{B} = \llbracket \mathcal{T}, \mathcal{R}, \mathcal{E} \rrbracket^{\diamond}(\{\top\}) \setminus \llbracket \mathcal{R} \rrbracket(\mathcal{U})$$

is non-empty. I term this *aberration*, since it is a deviation from the notional norm as expressed by \mathcal{R} . The choice of this rather negative terminology is deliberate, reflecting the hostility with which changes to accepted styles are often met in the artistic world.

The evaluation of this set of concepts is actually slightly more complicated than the single-concept motivating case outlined above. The aberrant but valued subset, which I call \mathcal{V} here, is calculated thus:

$$\mathcal{V} = \llbracket \mathcal{E} \rrbracket(\mathcal{B}).$$

Because we are working in the extensional limit case, with all the created concepts notionally elaborated, we have to consider the possibility that all aberrant concepts, some aberrant concepts or no aberrant concepts may be valued. I term these *perfect* ($\mathcal{V} = \mathcal{B}$), *productive* ($\mathcal{V} \subset \mathcal{B}$) and *pointless* ($\mathcal{V} = \emptyset$) aberration, respectively.

5.3. Using the properties of creative agents

The characterisations in Section 5.2 are only descriptively useful unless appropriate responses, categorised by condition, can be specified. This section does so. I assume some appropriate learning mechanism(s) which can adapt the rules (expressed in language \mathcal{L} and categorised into \mathcal{R} , \mathcal{T} and \mathcal{E}), from positive and/or negative training sets.

5.3.1. Uninspiration

Hopeless uninspiration has no solution within the specified universe; there is no capacity within the system to solve the problem. Therefore, it is up to the system designer to remedy the problem, like a *deus ex machina*.

Conceptual uninspiration can only be addressed, within the system, by the transformation of \mathcal{R} . In Boden’s terms, this would probably not be appropriate, since the \mathcal{R} set is rather more definitive than in my terms. However, in the general (multi-agent) case, where \mathcal{R} must really reflect some kind of consensus among agents about a particular domain, it would clearly be appropriate to modify \mathcal{R} in some way. Because of the multi-agent aspect, which has

only been mentioned in passing here, I leave the nature of such a modification for a future discourse.

Generative uninspiration, however, can be remedied within the framework. Transformational creativity is required. To transform the set \mathcal{T} in a useful way, we need to identify one or more valued concept(s), in the conceptual space constrained by \mathcal{R} (otherwise, we may have aberration), and to use it (them) to guide the transformation. However, there is a methodological problem here: there is no clear way to pick the concept(s) automatically, except at random or by use of an oracle. The “oracle” might in fact be systematic search of \mathcal{R} (assuming this is possible in finite and feasible time), or, again, the *deus ex machina* of user intervention.

There are some interesting issues to be considered here about the dynamics of this aspect of a creative system. There are obvious possibilities in analogy with the development of creative thinking through education. These, however, are outside the scope of the current paper.

5.3.2. Aberration

In the case of aberration, there is a choice as to whether to view the result as acceptable or not, and therefore we have the three categories, perfect, productive, and pointless. Acceptability is determined in terms of evaluation by whatever audience the agent, \mathbf{G} , is playing to. If a new concept is accepted, then a sensible solution might be to revise the notion of what the correct domain (as constrained by \mathcal{R}) is, so as to include the new concept. This, of course, might have consequences: other new concepts might be included and/or existing ones might be excluded along the way. If the new concept is not accepted under evaluation, then a reasonable recourse would be to adapt \mathbf{G} 's technique, \mathcal{T} . This may have similar consequences with respect to added and existing concepts available to \mathbf{G} : valued concepts may be lost, and new aberrant behaviour may be made possible.

One approach is to use the sets \mathcal{A} and \mathcal{V} to generate training examples to modify \mathcal{R} and \mathcal{T} , using our learning mechanism(s), as follows. Note that there are open questions here about some of the training sets required, since that choice is a major factor in the behaviour of the system. The main issue here is a standard one for AI: how much of what an AI program does is simply programming a computer directly to do something, and how much is emergent behaviour which was not directly programmed? In particular, if we simply train \mathcal{T} to match \mathcal{R} , first, we might be “coaching” our creative agent too directly, instead of allowing it to develop, and, second, in doing so we might be restricting its creative capability.

Perfect aberration yields new concepts, all of which are valued, and so should be added to \mathcal{R} . \mathcal{T} has enlightened us as to new possibilities. We therefore attempt to revise \mathcal{R} , by whatever learning methods are available, in such a way that all the concepts in \mathcal{A} (and \mathcal{V}) are included, so \mathcal{V} is a positive training set, and the negative training set is either \emptyset or $\mathcal{U} \setminus \llbracket \mathcal{R} \rrbracket (\mathcal{U}) \setminus \mathcal{A}$ or some subset of the latter,

depending on the effect desired. This, of course, is subject to the same caution as conceptual uninspiration above: if \mathcal{R} is a representation of an agreed domain between multiple agents, then we need agreement on changing it; the same issue arises in the definition of (any concrete) \mathcal{E} . Again, however, these issues are beyond the scope of the current paper.

Productive aberration means that we need to transform both \mathcal{R} and \mathcal{T} , because we wish valued concepts to become accepted, and unvalued ones not to be generated. \mathcal{V} and $\mathcal{A} \setminus \mathcal{V}$ constitute positive and negative training sets for \mathcal{R} , since \mathcal{R} needs to expand just enough to include only the valued concepts in \mathcal{A} . \mathcal{T} , on the other hand, needs to be transformed to restrict its coverage: $\mathcal{A} \setminus \mathcal{V}$ is a negative training set for \mathcal{T} , while, again, a positive training set might be $\llbracket \mathcal{R} \rrbracket (\mathcal{U})$, or simply \emptyset .

Pointless aberration suggests the need to transform \mathcal{T} only, so as to prevent the unvalued aberrant concepts from being generated. There is a negative training set: \mathcal{A} . Again, the nature of the positive training set is an open question.

5.4. Discussion

These labels and definitions allow us to characterise the behaviour of a given creative system and to identify broad classes of response. This, in turn, will allow comparison of behaviours both between and within the classes defined above, and thus allow better understanding of the field.

The emphasis in this work is on the further definition and understanding of the three sets, \mathcal{R} , \mathcal{T} and \mathcal{E} , and their relationships to each other, to the creative domain and to the activity they are intended to describe. In any case, what does become clear when one looks in detail at these proposals is that Boden's originals were (intended to be) rather broad-brush, and that when one focuses in, the relationships between the conceptual space, evaluation, and the universe (albeit only implicit in Boden's work) become less, not more, simple.

Three clarifications do seem to emerge naturally from this discussion. First, to be interesting, \mathcal{R} must define a set which is in some sense external to a given creative agent (and, I have supposed, agreed within an agent society – for whence else would it come?); second, \mathcal{T} is the primary characterisation of the agent itself, and in this context, \mathcal{R} is secondary (as in aberration, above); and, third, \mathcal{E} needs to be independent of \mathcal{R} . This last needs a little elucidation, since, at first sight, it looks like a contradiction. The point is that, for transformational creativity to occur, there needs to be aberrant behaviour (unless we allow arbitrary spontaneous behaviour from our agents, which seems inappropriate). Otherwise, unless $\llbracket \mathcal{R} \rrbracket (\mathcal{U})$ is infinite, the creative behaviour will stagnate, and the system will develop no further. While this is, of course, likely to be true of AI creative systems in the foreseeable future, it would be unfortunate if they were condemned to be so for all eternity. We can explain the apparent contradiction as follows: the set $\llbracket \mathcal{R} \rrbracket (\mathcal{U})$ is specific to the domain, and effectively defines

it. But the set constrained by \mathcal{E} need be only the extension in \mathcal{U} of those properties of $[\mathcal{R}](\mathcal{U})$ which are valued. Thus, $[\mathcal{E}](\mathcal{U})$ could be very large, but only a small part of it might be explored, due to the restrictions in \mathcal{R} and \mathcal{T} .

The issue of multi-agent creative systems is becoming increasingly important in the current line of reasoning. The aim of Boden's and my frameworks is to describe the behaviour of creative systems, but no natural creative systems exist in isolation (and, indeed, one might argue that neither do artificial ones). Therefore, the generalisation of these ideas, which has been informally mentioned above on several occasions, to multi-agent systems seems crucial and urgent. Only in this context will the distinctions highlighted above become really clear, as the shared and individual content in the system will need to be made explicit.

6. Summary and conclusion

This paper has presented a small step on the road to a more precise understanding of creative systems, both artificial and natural. I have presented a framework for characterising creative systems and shown that Boden's transformational creativity is actually exploratory creativity at the meta-level. I have given six categorisations of creative behaviour, which can be identified directly from the behaviour of creative systems as described using the formalism, and suggested how the needs of each category of system can be met, from within or from outside the system itself. This raises many questions, not least the issue of interaction between multiple creative agents. These questions will be addressed in future work.

Acknowledgements

I am very grateful to my colleagues in the Intelligent Sound and Music Systems group at Goldsmiths' College, University of London, for on-going discussion about this and other work, and to the Computational Creativity community (especially Simon Colton and Penousal Machado), which has given much useful feedback in its various workshops and symposia. Graeme Ritchie supplied some

particularly helpful comments on earlier related work, during his sabbatical at City University, London in 2002. Several anonymous reviewers gave some very helpful feedback.

References

- [1] M. Boden, *Artificial Intelligence and Natural Man*, Harvester Press, 1977.
- [2] M. Boden, *The creative mind*, Abacus (1990).
- [3] M. Boden, *Modelling creativity: reply to reviewers*, *Journal of Artificial Intelligence* 79 (1995) 161–182.
- [4] M. Boden, *Creativity and artificial intelligence*, *Journal of Artificial Intelligence* (1998).
- [5] M. Boden, *Preface to special issue on creativity in the arts and sciences*, *AISB Quarterly* 102 (1999).
- [6] Buchanan, B., 2000. *Creativity at the meta-level*. *AI Magazine Reprint of keynote lecture to AAAI-2000*.
- [7] A. Bundy, *What is the difference between real creativity and mere novelty?* *Behavioural and Brain Sciences* (1994).
- [8] K. Haase, *Too many ideas, just one word: a review of Margaret Boden's The Creative Mind: Myths and Mechanisms*, *Artificial Intelligence Journal* 79 (1995) 69–82.
- [9] R. Lustig, *Margaret Boden, The Creative Mind: Myths and Mechanisms*, *Artificial Intelligence Journal* 79 (1995) 83–96.
- [10] Macedo, L., Cardoso, A., 2003. *A model for generating expectations: the bridge between memory and surprise*, in: *Proceedings of the Third Workshop on Creative Systems: Approaches to Creativity in Artificial Intelligence and Cognitive Science*. Acapulco, Mexico.
- [11] Pearce, M.T., Wiggins, G., 2001. *Towards a framework for the evaluation of machine compositions*, in: *Proceedings of the AISB'01 Symposium on AI and Creativity in Arts and Science*. AISB.
- [12] D. Perkins, *An unfair review of Margaret Boden's The Creative Mind from the perspective of creative systems*, *Artificial Intelligence Journal* 79 (1995) 97–109.
- [13] A. Ram, L. Wills, E. Domeshek, N. Nersessian, J. Kolodner, *Understanding the creative mind: a review of Margaret Boden's Creative Mind*, *Artificial Intelligence Journal* 79 (1995) 111–128.
- [14] Ritchie, G.D., 2001. *Assessing creativity*, in: *Wiggins, G.A. (Ed.), Proceedings of the AISB'01 Symposium on AI and Creativity in Arts and Science*. AISB.
- [15] S. Russell, P. Norvig, *Artificial Intelligence – A Modern Approach*, Prentice Hall, New Jersey, 1995.
- [16] R. Schank, D. Foster, *The engineering of creativity: a review of Boden's The Creative Mind*, *Artificial Intelligence Journal* 79 (1995) 129–143.
- [17] S. Turner, *Margaret Boden, The Creative Mind*, *Artificial Intelligence Journal* 79 (1995) 145–159.
- [18] G.A. Wiggins, *Searching for computational creativity*, *New Generation Computing* 24 (3) (2006) 209–222.