# On transformational creativity

**Graeme Ritchie**
Department of Computing Science
University of Aberdeen
Aberdeen AB24 3UE
Scotland
gritchie@csd.abdn.ac.uk

## Abstract

The work of Boden on the nature of creativity has been extremely influential, particularly her promotion of *transformational* creativity. We consider how these ideas could be made more precise, starting from foundational assumptions.

## 1 Introduction

In recent years, there has been a renewed interest in the development of theoretical models or methodological frameworks for studying computational creativity in general [Ritchie, 2001], [Wiggins, 2001; 2003], [Pease *et al.*, 2001]. Our aim here is to contribute to that formal line of research, by examining the highly influential proposal by [Boden, 1992; 1998] that high levels of creativity result from the *transformation of a conceptual space*. We consider how this might make sense within a formal framework and hence how it could be tested empirically. We start by setting out our general premises (Section 2), then provide an informal summary of what we regard as the situation to be modelled (Section 3). In Section 5 we consider some possible formalisations of Boden's "conceptual spaces", and how creativity-related notions might be defined in them. We conclude by setting out what we see as the logical path to verifying hypotheses about the fruitfulness of transformational creativity.

## 2 Basic aims and assumptions

The existence of the words *creative* and *creativity* suggests that in the general society at large there is a general notion of 'being creative'. However, this is no guarantee that there is a single, coherent, consistent, precisely definable notion of 'creative' which is amenable to scientific scrutiny. The use of these words in ordinary discourse may be highly vague, very unsystematic and completely inconsistent [Buchanan, 2001]. It is plain, nevertheless, from the wealth of academic writing on creativity that there is a widespread belief, or perhaps hypothesis, amongst philosophers, psychologists, and others that such a concept can be defined. Moreover, artificial intelligence researchers sometimes claim (or assume) that this definition will be in terms of some formal characteristics of the creative process, in a way which could be applied to computational models. (A exception is [Bundy, 1994]: 'Creativity... does not correspond to some well-defined family of computational processes.') A suitably framed definition of creativity may help to set out explicitly what we regard as the overall arrangement involved in (machine) creativity, thereby allowing disagreements about substantive issues to be brought out into the open, and also clarifying the set of subproblems that remain to be solved. For example, if we adopt a formal model which relies on the notion of 'similarity', then elucidating how exactly this construct should be defined can be seen as a further intellectual task in the quest for a complete theory of creativity.

We base our approach on certain methodological axioms, which are broadly plausible and probably tacitly adopted by most research in this area, as follows.

**Naturalness.** Any technical definition of "creative" (or "creativity") which is to be used in discussing the behaviour of computer programs must capture fairly accurately the original ordinary language use of the term. We cannot be arbitrary or circular. If claims such as 'this program has been creative' are to have a comprehensible and plausible meaning, then our definition of 'creative' has to reflect the meaning of the ordinary language term. If, for example, we were to define 'creative' to mean 'computationally efficient', then verification of machine creativity would be much more straightforward, but we would be open to accusations of not addressing the real question. The precise formal definition for scientific use should have a significant amount in common with the ordinary usage, despite the messiness of the latter.

**Basis in human behaviour.** We should be guided by the way that the word 'creative' is ordinarily used when talking of non-machine (human) creativity, for two reasons: firstly, that is the original, established usage; secondly, to rely on instances of machine creativity (the problem we wish to analyse) would risk circularity in claims about the nature of that process. We shall therefore, when discussing issues relating to computational creativity, allude to instances of human creativity (as do most writers on creativity and AI, including – copiously – Boden).

**Observable factors only.** In human creative activities, there are certain aspects which are knowable, such as the attributes of the artefact created, the other comparable artefacts in existence, possibly the other artefacts the creating individual was

aware of. What we usually do not know is the mental or emotional processes by which the individual produced the artefact (although we may know other aspects of the action, such as the time taken). Hence, it is routine to make judgements of creativity (in humans) on the basis of what is known, often focussing on the attributes of the artefact(s). If our formal definition of creativity, for analysis of computer systems, is to mimic our judgements of humans, then it too should be based only on comparably observable factors, without adding extra information about the internal workings of the computer program. This may be our most contentious working assumption, as some would argue that the inner workings of a computer program are critical in deciding its creativity; in particular, [Boden, 1992, pp.39-40] advocates just such consideration of the underlying process (for both humans and computers). We suggest that this would move away from the way human creativity is normally judged. (It also risks circularity when asking the question 'which computational mechanisms give rise to creativity?' – see next point.) There is a fine but important distinction between the production of the artefact, and the devising of the production-method; if we happen to know the method, we can treat it as an abstract artefact and consider the creativity it manifests, relative to other production-methods.

**Genuinely empirical questions.** In AI, interest in creativity naturally leads to the general question 'can computer programs be creative?' (Boden's 'Lovelace Question' 2, or perhaps 4), the slightly more particular question 'which computational mechanisms lead to creativity?', and the specific question 'has this computer program been creative (on this occasion)?' These questions are usually treated as empirical questions which can be falsified or corroborated by building programs which generate artefacts (physical or abstract), and then studying these programs. However, this appearance of being empirical is illusory unless we can define what we mean by 'creative/creativity', and define it in terms of factors which are (at least in principle) observable (for example, [Ritchie, 2001] proposes some such factors). Moreover, our definition(s) should describe what behaviour we would regard as creative without building in, prematurely, proposals about *how* that behaviour might be achieved. If we can maintain a separation between our observational vocabulary and our theoretical models of possible mechanisms, then we can, without circularity, treat questions such as 'which computational mechanisms lead to creativity?' as empirical issues. If we incorporate our hunches about the best way to achieve creativity into our definitions of what observable behaviour constitutes creativity, then we have, to a large extent, undermined the empirical nature of the investigation. (This is to some extent a reiteration of the *Observable factors only* point, but with a different motivation.)

## 3 Informal overview

We start by summarising briefly (and informally) what we believe to be a reasonable approximation to the notion of creative activity, simplified in ways which should be conducive to subsequent formalisation. This should give some idea of the level of abstraction that we are adopting. Notice that the next few paragraphs set out a set of *simplifying assumptions*, not empirical claims; the use of bald declarative statements should not imply that these are assertions of fact. These assumptions are generally derived from consideration of what goes on in human creative activities.

In keeping with past work, we shall regard a (potentially) creative action as resulting in a specific item, an *artefact*, although this need not be a concrete object ([Wiggins, 2001] uses the term *concept* in a comparable role). The action takes place within the context of a society, in which there are various individuals. The judgement that an artefact manifests creativity, or that an action constitutes creativity, is made by individual(s) within the society. In particular, each individual may have idiosyncratic opinions or criteria about the type of artefact being produced, so these judgements are always relative to the one making the judgement.

Within the society, there a finite (and small) set of what we will call *medium types* and also a finite and small set of *genres*. An artefact belongs to a medium type; that is, a medium type is a class of artefact, and that class is determined by the basic form of the artefact – whether it is a sequence of words and punctuation, or a two-dimensional array of coloured pixels, or a sequence of musical notes, etc. These medium types embody as few claims or assumptions about the higher-level analyses that might be imposed upon the data – they do not represent the metre of the poem, the possible scenes depicted by a painting, or the key of a melody.

A genre is a culturally-defined type of artefact, either very broad or quite narrow (e.g. an impressionist painting, a symphony, a story). It will have an associated medium type, indicating the basic data which represents the artefact. Artefacts do not simply belong to genres: instead, each individual can make a judgement about the extent to which an artefact conforms to the norms of a genre; for example, not all sequences of words count as poems. We distinguish between the genre, which is shared (in some sense) amongst the members of the society, and an individual's assessment of a particular artefact with respect to that genre. Individuals also make judgements about the *quality* of an artefact, and these judgements are relative to some particular genre (a text may be a poor story but an excellent poem).

The word *creative* is sometimes applied to a person, sometimes to an action, sometimes to an artefact; [Boden, 1995, p. 170] refers to a *thought* as being creative. Here, driven by our desire to consider only observable data, we shall standardise to regarding it as a property of an artefact relative to a set of other artefacts of the same medium type (and usually of the same genre). That is, a judgement would be of the general form 'Artefact $A$ displays creativity relative to artefacts $\{A_1, \ldots, A_n\}$'. The choice of which artefacts are relevant in context for this comparison is not simple. For a human-created artefact, it might be the other exemplars that the creator was already familiar with (although human judges might well make comparisons with exemplars that they themselves are familiar with, overlooking Boden's distinction between P-creative and H-creative). For a computer program, the comparison set might be some corpus of examples available to the program designer [Ritchie, 2001] or some knowledge base used for case-based generation [Ram *et al.*, 1995].

## 4 Spaces and transformation

[Boden, 1992] makes central use of the term *conceptual space* (as the abstract location of the entities produced by creative acts) but does not define it precisely, although (p. 73) she asserts the need for it to be elaborated. Subsequent debate in this area has speculated on what this notion might mean. For example, could it be considered to be the traditional *search space* of AI problem-solving [Perkins, 1995; Wiggins, 2001]? ([Boden, 1992, p. 77] says a search space is 'one example... of a conceptual space'.)

Conceptual space is central to Boden's approach to creativity, primarily because she is interested in how *changes* to the space cause, or even constitute, creative acts. ([Boden, 1992]'s first mention of conceptual space is: '... changing the existing rules to create a new conceptual space'(p. 46)). Boden says that, although working within an existing space may produce interesting results (*exploratory creativity*), a higher form of creativity results from making changes to the space: *transformational creativity*. This concept, and the claim that it represents the highest form of creativity, has been very influential.

Our aim is to examine the hypothesis that transformational creativity is/leads to a higher form of creativity. We cannot hope, in the present state of the field, to set out evidence for or against this claim. Instead, we will discuss what might count as relevant evidence.

(It should be noted that it is not entirely clear whether Boden regards her statements about the superiority of transformational creativity as a hypothesis, rather than a *definition* of what counts as creative. This blurring continues in a review of [Boden, 1992] which refers to the idea as a 'thesis' and a 'definition' in adjacent sentences [Turner, 1995, p. 147]. We shall adopt the interpretation that it is a hypothesis.)

The 'transformation of spaces' is very much an *analysis* that is posited by Boden (and others) of the implications of particular artefacts – it is not part of the raw data. An art scholar might characterise an early Cubist painting as "transforming the space", but what the artist has actually done is produce a painting. The spaces are not given to us, nor is the transformation. If we are to have a formal description of space-transformation and hence transformational creativity, we have to show how such analyses are grounded in the actual artefacts. To be even more cautious, we should try to formalise the situations which might provoke these transformational analyses, so that we can then weigh up whether the transformational account is the best explanation, or whether there are alternative ways of describing what is going on.

Given our assumptions, the scenario we are interested in analysing is, informally, "artefact $A$ causes individual $P$ to adopt a different space"; this could be either by creating a new genre, or by restructuring an existing genre. That is, we are trying to elucidate the relationships between several things: an artefact, an individual, and two spaces (before and after); the genre should perhaps be included.

The artefact must be sufficiently similar to previous artefacts for it to be relatively clear which norms or spaces are relevant. It must at least be of a known medium type. However, it may also be that it has to belong at least in some peripheral way to an existing space (genre).

However, this supposes that an individual seeks a revised space. If not, then no transformation occurs. The artefact remains the same, rated as before. The question of whether an artefact "requires" a transformation is not wholly formal (although that is the aspect we are focussing on here). There are issues to be considered within the psychology and sociology of art and science: when does the individual feel the need to transform? These lie beyond the scope of the current paper.

It is not clear whether advocates of transformation have in mind changes to the *available* space (all logically possible artefacts of the genre), or whether the alterations might be to the way that the artefacts are distributed within that space: contrasts and similarities, degrees of variation, etc. A radical rearrangement of artefacts within the logically possible space might reflect the kinds of changes that Boden and others have offered as instances of transformation. Distinguishing between these empirically may be exceedingly difficult.

## 5 What is a space?

Let us consider, more abstractly, the requirements for a "conceptual space" in Boden's sense.

We start with a minor terminological point. In one sense, the space is just a set of artefacts. Although for the purposes of some analyses it may be feasible to consider this simple perspective, we shall more often want to consider how the space is *characterised*; that is, what finite rules or structures indicate how artefacts fall within the space. Only when there is some characterisation independent of the actual artefacts (i.e. something other than a simple extensional list of known artefacts) can we consider issues such as whether a new artefact does or does not fall within the space, or how a space may or may not be altered. A space may be infinite, yet have a finite (and computationally malleable) definition. In what follows, we shall use the terms *artefact-set* for the actual set, and *space-definition* the more compact definition of possible artefacts; where the sense is obvious, we may just use the term 'space' for either of these.

Also, it could be argued that the kind of analysis we are considering here involves *two* spaces: that imposed by judgements of the extent to which an artefact conforms to a genre (following [Ritchie, 2001], we can call this *typicality*), and a further space induced by judgements of the *quality* of the artefact. We shall return to this in Section 7 below.

There are four crucial functions that a space must fulfil if it is to support this analysis of creativity.

**Membership.** It must be possible to determine the membership of an artefact with respect to a definition. This membership may not be a binary decision, but may be graded in some way. Or the membership may be more insightfully viewed as the positioning of the artefact relative to other artefacts in various ways (for example, along multiple dimensions). However, there must be some notion of "membership" or "positioning within the space", and this must be decidable: some parts of the definition may rely on vague or subjective terms, but the computational structure of the definition must not be circular or otherwise flawed.

**Similarity.** Discussions of creativity, both informal and formal, usually involve (sometimes tacitly) some notion of "similarity" between artefacts. If all the artefacts of a given genre were completely incomparable, every artefact would, by definition, be 100% novel. The very idea of new but unnovel artefacts assumes some form of similarity. It is conceivable that the metric of similarity could be formally unrelated to the space-definition, but this seems unnatural and unlikely. A more elegant approach would be to have a definitional apparatus which directly led to some distribution of the artefacts within the space, with some means of comparison. As with membership, this comparison could result in a score of some kind, or could be a qualititative statement of a set of differences (e.g. along various dimensions). In this way, the attributes of an artefact that affect membership would also affect similarity. Also, any useful notion of similarity must rely on some higher-level representation of the data than the rudimentary data types we used to distinguish medium types. A distance metric which simply compared word-strings, or pixel arrays, would miss the kinds of similarity that are important for questions of creativity. For example, a pixel-by-pixel comparison of a portrait and a landscape by the same painter, or even two landscapes, would probably be classed as very distant (dissimilar), whereas two portraits in different styles might be classed as similar.

**Exploration.** Boden's *exploratory creativity* consists of following an organised path through the artefact-set. Most supposedly creative programs can be viewed as doing this, which is unsurprising, as the standard way (perhaps the only manageable way) to construct any generating program is to have a well-defined set of possibilities and move through them systematically. The formal definition of conceptual space must allow this. (It might be hard to contrive a definition which did not.)

Once again, the level of basic data types (word strings, pixels, etc.) offers an uninteresting way to organise exploration. Instead, exploration should be channelled by whatever space is postulated for the set of artefacts. (Inspection of some discussions in the literature on computational creativity suggests that adopting this more relevant kind of space has been an implicit assumption of most authors, but it is helpful to make the point explicit.)

**Change.** This aspect is the central topic for discussion here: how can a space be altered ("transformed"), particularly in response to a new artefact? More precisely, how can the space-definition be altered in a way that will have suitable changes for the associated notions of membership, similarity, and exploration?

These seem to be the most critical desiderata for a conceptual space. The basic operations that can be performed on a space $S$, corresponding to the four aspects listed above, therefore appear to be:

- **Locate** $A$ **within** $S$ (this could be a numerical rating of membership, or could be a more complex rating where the space-definition ascribes attribute-values, or positions on dimensions, to an artefact).
- **Rate artefacts** $A$**,** $B$ **for similarity (w.r.t.** $S$**)** (where the space-definition leads to degrees of similarity)
- **Given (existing) artefacts** $A_1, \ldots, A_n$**, generate a (new) artefact** $A$ (where $A$ is in some suitable sense "within" the space)
- **From** $S$**, create a revised space-definition** $S'$ **in the light of artefact** $A$.

From this it can be seen that the requirements for a conceptual space are very underdetermined by the writings of Boden and others.

## 6 Some possible formal structures

The world of formal models contains a wide variety of structures that could be used to construct space-definitions. We shall consider a few of these here, and comment upon their suitability for supporting the operations that spaces can undergo (particularly change).

### 6.1 State-Transition/Derivation models

This category covers, abstractly, both formal rewrite grammars and traditional automata ([Hopcroft and Ullman, 1979]), and heuristic search programs [Nilsson, 1971]: various symbolic rules define possible choices, and there is a notion of a *derivation* - a combination of, or sequence through, the rules - which defines the valid items. Whether this is what [Boden, 1992] means by a *generative rule* approach is hard to say.

In such a model, membership would be a binary decision, depending on whether the artefact was the result of a derivation. Similarity is not a natural feature of such a system, but some measure could perhaps be contrived based on the rules used within a derivation. Inasmuch as one can make formal sense of the General Theory of Verbal Humour (GTVH) [Attardo and Raskin, 1991], it uses a derivation model to define the space of jokes, defining similarity by attaching less weight to rule variation at the later stages of the derivation (roughly speaking).

Exhaustive generation of possibilities is straightforward in such a model, and there are natural points where additional domain-specific heuristic information could be injected: in the choice of rule for each step in the derivation.

Alterations to a space-definition could be made in various ways, but generally adding a construct (e.g. a transition or expansion rule) would lead to an expanded artefact-set, and removing a construct would shrink the artefact-set.

### 6.2 Multiple dimensions

Perhaps the most intuitively natural structure for a conceptual space, judging by writings on this topic, is a set of *dimensions*, where an artefact can lie at specific (ordered) points on each dimension.

If each artefact can be allocated values for the dimensions, then membership can be defined by specifying some subspace of the full multi-dimensional space. A further possibility would be to have a *weight* associated with each dimension (intuitively, reflecting the importance of that factor) (cf. [Ritchie, 2001]). This would map each basic $n$-tuple to

another ($n$-dimensional) vector, but in a space with a different distribution of artefacts. In this case, membership could also be computed as a numerical value by aggregating the weighted vector components. (Formally, such arrangements, and variations of them, have been explored in multi-attribute decision theory [Keeney and Raiffa, 1976]).

Of various possible similarity measures, an obvious one would be the Euclidean distance between the two vectors, either in the original $n$-dimensions, or in the weighted space.

The space would set bounds on exploration, but would not offer any particular structure for carrying it out, beyond exhaustively trying every valid value on every dimension.

Change could consist of revising one or more of the criteria for allocating values (on dimensions) to artefacts. This could be thought of as altering the assessment (perhaps unconscious) of the various attributes of the artefact. (Whether this describes a perceptual or a conceptual change is unclear.) For example, it could be argued that the rise of impressionist painting involved a change in the public's notion of what counted as a "realistic" painting, or that public ideas of "obscenity" have varied over the centuries. Such a change, of whatever size, leaves the dimensionality, and arguably the dimensions, of the space intact, but would change where some artefacts lie within the space, so that some previously highly typical items might become peripheral, or vice versa.

If weights are associated with dimensions, then these could change, altering the importance assigned to each property. This is a plausible match to the notion of changes in "taste" or "fashion". For example, acceptance of abstract paintings does not necessarily involve a change in assessment of what counts as "realistic", but it does demand a change in the importance attached to this property. Again, changes could be small or great, and would not alter the dimensions of the possible space, but would in general change the distribution of artefacts through the space, perhaps quite dramatically, and might or might not count as "transformation".

A less straightforward change might be to add dimensions to the space. (Although removing a dimension is also a logical possibility, it is hard to see how this can be distinguished from assigning zero weight to the dimension.) While this superficially sounds more radical than the previous two types of change, whether it makes a real difference depends on how the assessment function(s) associated with the new dimensions rate the relevant artefacts, and how much weight is attached to these factors. It is conceivable that new dimensions could be added without having much impact on the overall landscape of the genre (in terms of the abstractions of Section 5 above), if they are allocated very low weights.

### 6.3 Constraints

Boden refers to changes in constraints as a kind of space-transformation, so we should consider a constraint-based specification, as in certain problem-solving representations [Mackworth, 1977]. However, although constraint-solving is an elegant and efficient mechanism for determining the values for a related set of variables, it still leaves open the structures which the variable values characterise. A constraint-solver could be imposed upon a basic representation which is declaratively defined in some other way (e.g. by generative rules). In a multi-dimensional formalisation (see above), the subspaces of interest could be stated by imposing constraints upon values of coefficients. Indeed, if the variable domains are ordered sets (more especially, if they are numerical), then the constraint-based model is inherently multi-dimensional. Only if the variable domains do not lend themselves to interpretation as dimensions is it formally distinct.

Conventionally, any assignment of values to the variables satisfies all the constraints counts as a solution. Here, we could regard the constraints as specifying the artefact-set. To have a looser notion of the artefact-set, we could allow value-assignments which merely satisfy *most* of the constraints. As in the multi-dimensional case, weights could be attached to constraints, allowing a combined rating of how well a value-assignment meets the constraints.

There is no obvious, natural definition of similarity. In the version where not all the constraints have to be satisfied, then perhaps a measure could be based on a comparison of the set of constraints which the two artefacts satisfy.

The constraints may license many combinations of values for variables, so exploration would consist of enumerating these in some order. Some constraint models also contain *preferences* (or *soft constraints*), which do not make rigid stipulations (which would eliminate some possible values) but instead indicate an ordering on possible values, making some more "preferred". These could be used to impose order upon the enumeration of value combinations.

Change could occur by the addition or removal of constraints. Unlike the state-transition/derivation models, where addition of components expands the artefact-set (or leaves it untouched), here addition shrinks the artefact-set. Conversely, removing constraints can only widen (or leave unchanged) the artefact-set. Boden's informal discussions give the impression that constraint removal is the route to creativity. However, some forms of artistic innovation can be seen more naturally as the imposition of further constraints: the *dogme* film movement, or pointillist painting, or (according to [Buchanan, 2001]) haiku.

### 6.4 Connectionist networks

There is not univeral agreement on the usefulness of connectionist models in creativity:

> To peg the definition [of creativity] on a nonconnectionist model is to hitch it to a fading star. [O'Rourke, 1994, p.547]

> ...it is somewhat futile to look to connectionism for useful insights about creative insight. [Schank and Foster, 1995, p. 137]

Nevertheless, connectionist models typically afford very natural notions of the facilities needed.

A network can very easily represent degrees of membership and can show degrees of similarity between inputs. Exploration is perhaps less obvious, but could be organised in a generate-and-test manner.

It is not clear what the natural notion of change would be for a network representation. Various kinds of alterations to the internal topology are possible, although their effects

on the space (as embodied in the membership and similarity judgements) would be very indirect and probably hard to predict.

## 7 The role of quality

As mentioned in Section 3, an important part of judgements about creativity is some assessment of the "quality" of the artefacts. In a formal description, there are various possible ways in which quality might be structured, and in which typicality and quality might be interlinked. Whereas the discussion of typicality (Section 5) considered binary membership (yes-no), graded degrees of membership (a fuzzy set) and some form of structured allocation of a "position" within the space, in the case of quality only the latter two would be plausible – a yes-no decision on quality is excessively unrealistic, even for this simplified discussion.

The simplest approach might be to assume that quality is stated in terms of whatever components make up the underlying conceptual space defining the genre. This would mean that an attribute of an artefact is potentially relevant to determining the quality of that artefact only if it is relevant to deciding membership/location within a genre – no other attributes can be considered. This does not mean that quality values cannot be assigned to items which are within the formal (available) space, but outside the subspace which count as typical of the genre. Discussions of creativity, both informal (Boden) and formal (Wiggins, Ritchie, Pease et al.) assume that highly atypical artefacts can be assigned quality ratings, even high ratings. In fact, the general message from these authors is that the combination of low typicality and high quality is a sign of real creativity, possibly meriting a space-transformation.

Using the same attributes for both typicality and quality is unnecessarily restrictive from the point of view of imitating human judgements in artistic domains, where there may be certain minimum standards for membership of the genre, but other, more subtle, criteria for separating good from bad instances. In certain forms of poetry (e.g. in British nineteenth century culture) a text is a poem if it meets certain basic standards of syntactic and semantic coherence, and conforms to some clear metric structure; rhyming is a further desirable feature. However, the quality of a poem may depend on further factors, such as its emotional effect, or the profundity of its subject matter. Similar remarks apply to jokes, where the attributes that ensure that a text is a joke may not be sufficient to determine whether or not it is a funny joke.

So, for generality, quality has to be based on a different space from typicality. (The formal models discussed earlier would not all be equally plausible for such a space: the state-transition model seems less likely than the multi-dimensional model, for example.) The intuitive interpretation of this would be that judgements of typicality (the extent to which an artefact belongs to a given genre) are based on different factors from judgements about quality.

This means that we have to consider not one but possibly two space transformations. When an artefact stimulates transformation, it could be that change occurs in the typicality space, or the quality space, or both. If both, the magnitude of the two changes need not be the same.

## 8 Discussion

### 8.1 How many spaces?

So far we have focussed mainly on the typicality space, which structures the possible artefacts and their available properties. As Section 7 indicates, there is also some sort of quality space, with its own options for change. Even in the area of typicality, there is the ambiguity between a space which shows all logically possible combinations of values and a narrower space which sets out the currently acceptable combinations. Change is possible in both cases, although easier to formulate for the latter. In Section 6.2, we commented that if weights were attached to the dimensions of a multi-dimensional space, this would map items into another space (of the same number of dimensions), the layout of which would change with changes to the weights.

As mentioned in Section 4, the rules governing a system's search for artefacts affect what might be produced. These rules are yet another area where change might occur [Perkins, 1995],[Wiggins, 2001].

Hence, transformation could occur to any of these *four* "spaces" (or five in the case of the weighted dimensions). These degrees of freedom are not a help in trying to pin down exactly what space-changes give rise to creativity.

For simplicity, we shall continue to talk as if there was only one space involved, since some of the methodological issues are the same for all of them, but the more complicated arrangement should be borne in mind.

### 8.2 Determining the space

Although we attempted to pin down the notion of "(conceptual) space" by listing 4 operations it had to support (locating, comparing, exploring, changing), there is, methodologically, a further process to be considered: determining the space on the basis of a set of artefacts. As noted earlier, all that is available to the theoretician or analyst (of the creative genre under consideration) is a set of artefacts. The space is not given, but must be induced on the basis of the evidence (the artefacts). This is a significant task, even if the analyst narrows the search by opting for one particular type of formal model (such as one of those listed above).

For the types of formal model discussed above, there are learning procedures which will induce definitions given a set of examples, assuming the set of relevant properties is known. This need for pre-selected relevant properties might seem to defer the solution still further, but in the situation under consideration here – some potentially creative activity such as writing, painting, composing – there will be an established culture which will make available a basic set of attributes for artefacts. If an artefact manifests an entirely new property that had not even been thought of as a possible attribute (i.e. was outside the logically possible space), then the process we are sketching may have difficulty even representing the nature of that artefact, which is indeed a problem. However, the situation we are analysing (following the cases that Boden and others talk about) is where an artefact can be seen

to be significantly different from other examples. If the difference is detectable, the properties causing it must be within the system.

In discussions of the creativity of computer programs, the step of figuring out what the relevant space is for a given output set is rarely if ever considered. The 'space' is either a loose metaphorical construct, or is taken to be whatever the computer program uses to structure its computations. The problem with using the latter (leaving aside concerns expressed earlier about inspecting the workings of a program) is that all outputs would, by definition, be within the program's space. This would make it logically impossible for a program to produce output which transcended its own space; programs could never carry out anything beyond 'exploratory' creativity.

### 8.3 Transformation versus tweaking

Boden is adamant that 'transformational creativity' involves radical changes to the space, not minor adjustments. In the types of formal model reviewed above, there is not an obvious distinction between minor and major changes. Any of the adjustments suggested could vary in their extent. What a highly novel artefact exemplifies is a high degree of difference from previous artefacts, but this could take place within the same formal space, for example by a change in the weights attached to the components of an multi-dimensional space.

### 8.4 The metalevel

It has been suggested that genuine creativity involves processing at a metalevel, that transformational creativity consists of metalevel computation and even that it may consist of an exploration at the metalevel comparable to that which goes on at the object level (the main space) [Bundy, 1994; Buchanan, 2001; Wiggins, 2001]. There seems no doubt that if we take the artefacts and the conceptual space, however formalised, to constitute the object level, then any change to this which involves non-trivial computation (e.g. to select a suitable change to the space) necessarily requires metalevel processing. Whether this is in any interesting and substantive way similar (but at a different level) to what goes on in creative exploration without space-change is an open question. [Wiggins, 2001] has shown that it is possible to state the processing at the two levels in a similar formal way, but his account is so general and abstract (search within an unstructured space with an evaluation function) that it does not prove very much. Wiggins' perspective considers the actions of the creating agent. We have taken the perspective of the individual assessing the artefact (with the most minimal assumptions about actual creation), and argued that *if a new space must be found, then this demands some form of metalevel (outside the space) processing*. Thus, whether one attributes the transformation to the creator (Wiggins, Boden, Bundy) or to the individual appreciating the artefact (as here), computing a new structure for the object level necessitates metalevel work.

### 8.5 Empirical questions

If the claim that transformational creativity leads to artefacts which are deemed "more creative" is to be empirical, then a great deal needs to be done. To study the hypothesis as a claim about human creativity, we must, across a range of genres:

(a) for at least one type of formal model, devise criteria by which we can justify one space analysis of a set of artefacts over some other analysis;

(b) define, for our chosen type(s) of formal model(s), what constitutes a "transformation" of a space;

(c) on the basis of the above steps, set out criteria for when an object manifests (or demands) a transformation of typicality space, quality space, or both;

(d) using studies with human subjects, find human-created artefacts which are regarded as examples of high creativity, and (preferably) comparable examples of lesser degrees of creativity;

(e) for the items used in the studies, apply the findings of the first three steps to create an analysis in terms of a suitable formal model;

(f) see what relationship (if any) there is between ratings of creativity and space-transformation.

If the case is to made for the transformational conjecture across a range of medium types and genres, the definitions and criteria listed above will have to be extremely general and abstract, so there will also have to be information about how these abstract principles are made concrete in particular genres (and possibly in different formal models).

In the case of computer creativity, we have to make a decision about methodological strategy: on what do we base our judgement about whether transformation has been involved:

(i) an analysis (like that listed immediately above) of what provides the most suitable description of the output of the program?

(ii) an examination of the processing implemented in the program?

If we opt for (i), then the situation is exactly parallel to the human case, above, differing only in the source of the artefacts for step (d). In case (ii), if we decide that the program's computations do count as transformation, then we then have the further question to consider: is this transformational processing central to the program's processing, or could an elegant and adequate non-transformational account could be given of the computation? This is important, because it might be that the computation could have been organised in a number of ways, some of which are "more transformational" than others. It would not be desirable to allow an implementor to improve the creativity score of a program by reorganising its architecture, while still producing the same output.

This last point is perhaps controversial. Boden does make judgements on the creativity of programs on the basis of their inner workings as do (to a lesser extent) [Pease *et al.*, 2001]. While it is quite legitimate to examine the internal actions of a program in order to further various sorts of analysis (e.g. [Colton *et al.*, 2001]), we have to be careful, as observed in Section 2, about avoiding circularity.

If we (briefly) return to the idea that Boden was not making an empirical claim about creativity, but *defining* creativity,

then the relevant programme is rather different. We should carry out steps (a) – (c) of our list above, and then apply these findings directly to the output of generating programs, to answer the question: 'has this computer program been creative (on this occasion)?' Given our basic assumptions (Section 2), it is not clear where this line of analysis would take us if some program were to be labelled "creative" by this definition, but intuitively was judged not to be at all creative,

## 9 Conclusion

None of the above is to deny, or play down, the importance of other factors in creativity. For example, [Buchanan, 2001] mentions the relevance of the creator's background knowledge and skills, and the effects of experience. The very limited brief we have set ourselves here is to consider how the question of space-transformation might be formalised, in order that various claims about the necessity or effectiveness of transformation might be made empirical.

The approach outlined in this paper may not be the only route by which claims about transformational creativity can be made concrete and testable, but it is at least, in sketch form, one possibility. Any advocate of transformational creativity as a superior form of creativity who does not offer some comparable route to falsification/corroboration is in a weak position from an empirical point of view.

However, it may well be that published discussions of transformational creativity do not intend to set out an empirical scientific hypothesis. If so, we can lay aside the concerns discussed in this paper. In particular, devising computational architectures which might be particularly useful in building creative programs (in specific genres) is a *different* problem, and one that could be addressed without bothering with any of the issues we have discussed here. Those who wish to build generators of music, poetry, art, jokes, concepts, etc. can move ahead regardless of the status of claims about transformations. Perhaps the conjecture that "transformational is better" is better left as a loose slogan to inspire program designers, rather than viewed as a strict hypothesis.

## References


[Attardo and Raskin, 1991] Salvatore Attardo and Victor Raskin. Script theory revis(it)ed: joke similarity and joke representation model. *Humor: International Journal of Humor Research*, 4(3):293–347, 1991.

[Boden, 1992] Margaret A. Boden. *The Creative Mind*. Abacus, London, 1992. First published 1990.

[Boden, 1995] Margaret Boden. Modelling creativity: reply to reviewers. *Artificial Intelligence*, 79:161–182, 1995.

[Boden, 1998] Margaret A. Boden. Creativity and Artificial Intelligence. *Artificial Intelligence*, 103:347–356, 1998.

[Buchanan, 2001] Bruce Buchanan. Creativity at the metalevel. *AI Magazine*, (Fall 2001):13–28, 2001. AAAI-2000 Presidential Address.

[Bundy, 1994] Alan Bundy. What is the difference between real creativity and mere novelty? *Behavioral and Brain Sciences*, 17(3):533–534, 1994. Open Peer Commentary on [Boden, 1992].

[Colton *et al.*, 2001] Simon Colton, Alison Pease, and Graeme Ritchie. The effect of input knowledge on creativity. In R. Weber and C. G. von Wangenheim, editors, *Case-Based Reasoning: Papers from the Workshop Programme at ICCBR 01*, Vancouver, 2001.

[Hopcroft and Ullman, 1979] J. Hopcroft and J. Ullman. *Introduction to automata theory, languages, and computation*. Addison-Wesley, Reading, Mass, 1979.

[Keeney and Raiffa, 1976] Ralph L. Keeney and Howard Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley and Sons, 1976.

[Mackworth, 1977] Alan K. Mackworth. Consistency in networks of relations. *Artificial Intelligence*, 8:99–118, 1977.

[Nilsson, 1971] Nils J. Nilsson. *Problem-solving methods in artificial intelligence*. McGraw-Hill, New York, 1971.

[O'Rourke, 1994] Joseph O'Rourke. The generative-rules definition of creativity. *Behavioral and Brain Sciences*, 17(3):547, 1994. Open Peer Commentary on [Boden, 1992].

[Pease *et al.*, 2001] Alison Pease, Daniel Winterstein, and Simon Colton. Evaluating machine creativity. In R. Weber and C. G. von Wangenheim, editors, *Case-Based Reasoning: Papers from the Workshop Programme at ICCBR 01*, pages 129–137, Vancouver, 2001.

[Perkins, 1995] David Perkins. An unfair review of Margaret Boden's *The Creative Mind* from the perspective of creative systems. *Artificial Intelligence*, 79:97–109, 1995.

[Ram *et al.*, 1995] Ashwin Ram, Linda Wills, Eric Domeshek, Nancy Neressian, and Janet Kolodner. Understanding the creative mind: a review of Margaret Boden's *Creative Mind*. *Artificial Intelligence*, 79:111–128, 1995.

[Ritchie, 2001] Graeme Ritchie. Assessing creativity. In *Proceedings of the AISB Symposium on Artificial Intelligence and Creativity in Arts and Science*, pages 3–11, York, England, 2001.

[Schank and Foster, 1995] Roger C. Schank and David A. Foster. The engineering of creativity: a review of Boden's *Creative Mind*. *Artificial Intelligence*, 79:129–143 Mind, 1995.

[Turner, 1995] Scott Turner. Margaret Boden, *The Creative Mind*. *Artificial Intelligence*, 79:145–159, 1995.

[Wiggins, 2001] Geraint Wiggins. Towards a more precise characterisation of creativity in AI. In R. Weber and C. G. von Wangenheim, editors, *Case-Based Reasoning: Papers from the Workshop Programme at ICCBR 01*, Vancouver, 2001.

[Wiggins, 2003] Geraint Wiggins. Categorising creative systems. In *Proceedings of Third (IJCAI) Workshop on Creative Systems: Approaches to Creativity in Artificial Intelligence and Cognitive Science*, 2003.