

Creativity and Artificial Intelligence

Margaret A. Boden

School of Cognitive and Computing Sciences

University of Sussex

Brighton, Sussex BN1 9QH

ENGLAND

Abstract

Creativity is a fundamental feature of human intelligence, and a challenge for AI. AI techniques can be used to create new ideas in three ways: by producing novel combinations of familiar ideas; by exploring the potential of conceptual spaces; and by making transformations that enable the generation of previously impossible ideas. AI will have less difficulty in modelling the generation of new ideas than in automating their evaluation.

1 Why AI Must Try to Model Creativity

Creativity is a fundamental feature of human intelligence, and an inescapable challenge for AI. Even technologically oriented AI cannot ignore it, for creative programs could be very useful in the laboratory or the market-place. And AI-models intended (or considered) as part of cognitive science can help psychologists to understand how it is possible for human minds to be creative.

Creativity is not a special “faculty,” nor a psychological property confined to a tiny elite. Rather, it is a feature of human intelligence in general. It is grounded in everyday capacities such as the association of ideas, reminding, perception, analogical thinking, searching a structured problem-space, and reflective self-criticism. It involves not only a cognitive dimension (the generation of new ideas) but also motivation and emotion, and is closely linked to cultural context and personality factors [Boden, 1994]. Current AI models of creativity focus primarily on the cognitive dimension.

A creative idea is one which is novel, surprising, and valuable (interesting, useful, beautiful...). But “novel” has two importantly different senses here. The idea may be novel with respect only to the mind of the individual (or AI-system) concerned or, so far as we know, to the whole of previous history. The ability to produce novel-

ties of the former kind may be called P-creativity (P for psychological), the latter H-creativity (H for historical). P-creativity is the more fundamental notion, of which H-creativity is a special case.

AI should concentrate primarily on P-creativity. If it manages to model this in a powerful manner, then artificial H-creativity will occur in some cases. (In what follows, I shall not use the letter-prefixes: usually, it is P-creativity which is at issue.)

2 Two Types of Creativity

There are two main types of creativity, involving different ways of generating the novel ideas. Each results in surprises, but only one can (sometimes) lead to the “shock” of surprise that greets an apparently impossible idea [Boden, 1990]. Both types include some H-creative examples, but the creators celebrated in the history books are more often valued for their achievements in respect of the second type of creativity.

The first type involves novel (improbable) combinations of familiar ideas. Let us call this “combinational” creativity. Examples include much poetic imagery, and also analogy – wherein the two newly associated ideas share some inherent conceptual structure. Analogies are sometimes explored and developed at some length, for purposes of rhetoric or problem-solving. But even the mere generation, or appreciation, of an apt analogy involves a (not necessarily conscious) judicious structural mapping, whereby the similarities of structure are not only noticed but are judged in terms of their strength and depth.

The second, and potentially deeper, type is “exploratory-transformational” creativity. Here, the generation of novel ideas involves the exploration, and sometimes also the transformation, of structured conceptual spaces.

Many human beings – including (for example) most professional scientists, artists, and jazz-musicians – make a justly respected living out of exploratory creativity. That is, they inherit an accepted style of thinking from

their culture, and then search it, and perhaps superficially “tweak” it, to explore its contents, boundaries, and potential. But human beings sometimes transform the accepted conceptual space, by altering or removing one (or more) of its dimensions, or by adding a new one. Such transformation enables ideas to be generated which (relative to that conceptual space) were previously impossible.

The more fundamental the transformation, and/or the more fundamental the dimension that is transformed, the more different the newly-possible structures will be. The shock of amazement that attends such (previously impossible) ideas is much greater than the surprise occasioned by mere improbabilities, however unexpected they may be.

If the transformations are too extreme, the relation between the old and new spaces will not be apparent. In such cases, the new structures will be unintelligible, and very likely rejected. Indeed, it may take some time for the relation between the two spaces to be recognized and generally accepted.

3 Computer Models of Creativity

Combinational creativity is studied in AI by research on analogy. Most AI-models of analogy generate and evaluate analogies by taking advantage of the programmer’s careful pre-structuring of the relevant concepts. This guarantees that their similarity is represented, and makes it likely that it will be found by the program [e.g. Forbus *et al.*, 1994].

But some – broadly connectionist – AI-models allow for a more flexible representation of concepts. One example is the Copycat program, which looks for analogies between alphabetic letter-strings [Hofstadter, 1995]. Copycat’s concepts evolve as processing proceeds, and the model allows a wide range of (more or less daring) analogies to be generated, and evaluated. The degree to which the analogies are obvious or far-fetched can be altered by means of one of the system-parameters.

The exploratory-transformational type of creativity can also be modelled by AI-systems. For conceptual spaces, and ways of exploring and modifying them, can be described by computational concepts. Usually, this involves considerable domain-expertise – in music, for example [e.g. Cope, 1991], or graphic art [Cohen, 1995], or architecture [e.g. Koning and Eizenberg, 1981]. Occasionally, a “creative” program is said to apply to a wide range of domains – as EURISKO, for instance, does [Lenat, 1983]. But to make a generalist program useful in a particular area, such as genetic engineering or VLSI-design, considerable specialist knowledge has to be provided if it is not to generate hosts of nonsensical (as opposed to merely boring) ideas.

Most current AI-programs focussed on creativity are concerned only with defining and exploring conceptual spaces. They may allow for highly constrained tweaking, but no fundamental novelties or truly shocking surprises are possible. This is true, for instance, of the BACON suite, whose heuristics for scientific discovery are carefully pre-programmed, and whose data are deliberately prestructured so as to suit the heuristics provided [Langley *et al.*, 1987].

However, a few AI-systems attempt also to transform their conceptual space, sometimes in relatively unconstrained ways. Transformational systems include AM and EURISKO [Lenat, 1983], and certain programs based on genetic algorithms – some of which have produced valued structures that the human experts say they could never have produced unaided.

Most GA-programs only explore a pre-given space, seeking the “optimal” location within it. But some also transform their generative mechanism in a more or less fundamental way. For example, GA-work in graphics may enable superficial tweaking of the conceptual space, resulting in images which, although novel, clearly belong to the same family as those which went before [Todd and Latham, 1992]. Or it may allow the core of the image-generating code to be lengthened and complexified, so that the novel images may bear no family-resemblance even to their parents, still less to their more remote ancestors [Sims, 1991]. Similarly, some work in evolutionary robotics has generated novel sensory-motor anatomies and control systems as a result of GAs that allow the length of the “genome” to be altered [Cliff *et al.*, 1993].

4 The Evaluation of New Ideas

A main reason why most current AI-models of creativity attempt only exploration, not transformation, is that if the space is transformed then the resulting structures may not have any interest or value. Such ideas are novel, certainly, but not creative. (We saw in Section 1 that “creativity” implies positive evaluation.)

This would not matter if the AI-system were able to realize the poor quality of the new constructions, and drop (or amend) the transformation accordingly. A truly automatic AI-creator would have evaluative mechanisms sufficiently powerful to do this. At present, this is very rarely so (an exception is artificial co-evolution in which the fitness function evolves alongside the several species involved [Ray, 1992]). Notoriously, AM produced many more useless items than powerful mathematical ideas, and although it did have heuristics of “interestingness” built into it, its evaluations were often mistaken by human standards. And some “adventurously” transformational programs embody no evaluative criteria at all, the evaluation being done interactively by human beings

[Sims, 1991].

There is no reason in principle why future AI-models should not embody evaluative criteria powerful enough to allow them to transform their conceptual spaces in fruitfully creative (including H-creative) ways. But for such computerized self-criticism to be possible, the programmers must be able to express the values concerned sufficiently clearly for them to be implemented. Even if the values are not predetermined, being represented instead as an evolving fitness function, the relevant features must be implemented in and recognized by the (GA) system.

To some extent, this can be achieved implicitly, by defining a culturally accepted conceptual space so successfully that any structure that can be generated by the program will be accepted by humans as valuable [Cohen, 1995; Koning and Eizenberg, 1981]. But the structures generated within newly transformed spaces will need types of evaluation different (at least in part) from those implicit within the original space, or previously provided in explicit form.

It is even more difficult to express (verbally or computationally) just what it is that we like about a Bach fugue, or an impressionist painting, than it is to recognize something as an acceptable member of one of those categories. And to say what it is that we like (or even dislike) about a new, or previously unfamiliar, form of music or painting is even more challenging.

Identifying the criteria we use in our evaluations is hard enough. Justifying, or even (causally) explaining, our reliance on those criteria is more difficult still. For example, just why we like or dislike something will often have a lot to do with motivational and emotional factors – considerations about which current AI has almost nothing to say.

To make matters worse, human values – and therefore the novelties which we are prepared to approve as “creative” – change from culture to culture, and from time to time. In some cases, they do so in unpredictable and irrational ways: think of the fashion-industry, for example, or of rogue memes like the back-to-front baseball-cap. Nor are value-shifts confined to trivial cases such as these: even Bach, Mozart, and Donne were ignored and/or criticized in certain periods.

The scientific criteria of theoretical elegance and coherence, and of experimental verification, are less variable than artistic values. But that's not to say they are easy to define, or to implement.

Moreover, science too has its equivalent of fad and fashion. Even the discovery of dinosaurs was not a cut-and-dried event, but the culmination of a process of scientific – and political-nationalistic – negotiation lasting for several years [Schaffer, 1994]. The important point is that what scientists count as “creative”, and what

they call a “discovery”, depends largely on unarticulated values, including social considerations of various kinds. These social evaluations are often invisible to scientists. For sure, they are not represented in AI-models.

5 Conclusion

Some H-creative ideas have already been generated by AI-programs, though usually by merely exploratory (or combinational) procedures. Transformational AI-originality is only just beginning.

Valuation, thus far, is mostly implicit in the generative procedures used by the program, or interactively imposed by a human being. Only a few AI-models can critically judge their own original ideas. And hardly any can combine evaluation with transformation.

The ultimate vindication of AI-creativity would be a program that generated novel ideas that initially perplexed or even repelled us, but which was able to persuade us that they were indeed valuable. We are a very long way from that.

References

- [Boden, 1990] Margaret A. Boden. *The Creative Mind: Myths and Mechanisms*. Basic Books, New York, 1990.
- [Boden, 1994] Margaret A. Boden, editor. *Dimensions of Creativity*. MIT Press, Cambridge, Massachusetts, 1994.
- [Cliff *et al.*, 1993] David Cliff, Inman Harvey, and Philip Husbands. Explorations in Evolutionary Robotics. *Adaptive Behavior*, 2: 71-108, 1993
- [Cohen, 1995] Harold Cohen. The Further Exploits of AARON, Painter. In Stefan Franchi and Guven Guzeldere, editors, *Constructions of the Mind: Artificial Intelligence and the Humanities*. Special edition of *Stanford Humanities Review*, 4 (2): 141-160, 1995
- [Cope, 1991] David Cope. *Computers and Musical Style*. Oxford University Press, Oxford, 1991.
- [Forbus *et al.*, 1994] Kenneth D. Forbus, Dedre Gentner, and Keith Law. MAC/FAC: A Model of Similarity-Based Retrieval. *Cognitive Science*, 119:141-205, 1994
- [Hofstadter, 1995] Douglas R. Hofstadter and FARG (The Fluid Analogies Research Group). *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. Basic Books, New York, 1995.
- [Koning and Eizenberg, 1981] H. Koning and J. Eizenberg. The Language of the Prairie: Frank Lloyd Wright's Prairie Houses. *Environment and Planning B*, 8:295-323, 1981.

- [Langley *et al.*, 1987] Patrick Langley, Herbert A. Stone, Gary L. Bradshaw, and Jan M. Zytkow. *Scientific Discovery: Computational Explorations of the Creative Process*. MIT Press, Cambridge, Massachusetts, 1987.
- [Lenat, 1983] Douglas B. Lenat. The Role of Heuristics in Learning by Discovery: Three Case Studies. In Richard S. Michalski, James G. Carbonell, and Thomas M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach*, 243-306. Tioga, Palo Alto, California, 1983.
- [Ray, 1992] Thomas S. Ray. An Approach to the Synthesis of Life. In Christopher G. Langton, Charles Taylor, J. Doyne Farmer, and Steen Rasmussen, editors, *Artificial Life II*, 371-408. Addison-Wesley, Redwood City, California, 1992. (Reprinted in Margaret A. Boden, editor, *The Philosophy of Artificial Life*, 111-145. Oxford University Press, Oxford, 1996.)
- [Schaffer, 1994] Simon Schaffer. Making Up Discovery. In [Boden, 1994], 13-51.