Margaret Boden: Can computer models help us to understand human creativity?

(From http://onthehuman.org/2010/05/can-computer-models-help-us-to-understand-human-creativity/)

Creativity and computers: what could these possibly have to do with one another? “Nothing!,” many people would say. The two are simply incompatible.”

Well, I disagree. Computers and creativity make interesting partners with respect to two different projects. One, which interests me the most, is understanding human creativity. The other is trying to produce machine creativity—or anyway, machine “creativity”—in which the computer at least appears to be creative, to some degree.

What is Creativity?

Human creativity is something of a mystery, not to say a paradox. One new idea may be creative, while another is merely new. What’s the difference? And how is creativity possible? Creative ideas are unpredictable. Sometimes, they even seem to be impossible — and yet they happen. How can that be explained?

Before we can hope to explain creativity, we need to know what’s meant by the term. In fact, people use it in rather different ways — so, when discussing it, they can end up talking at cross purposes.

Here, let’s agree that creativity is the ability to come up with ideas or artefacts that are new, surprising, and valuable. “Ideas,” here, includes concepts, poems, musical compositions, scientific theories, cooking recipes, choreography, jokes … and so on. “Artefacts” include paintings, sculpture, steam-engines, vacuum cleaners, pottery, origami, penny-whistles … and more.

Creativity isn’t a special “faculty,” confined to a tiny elite: it’s an aspect of human intelligence in general. Nor is it an all-or-none affair. Rather than asking “Is that idea creative, Yes or No?,” we should ask “Just how creative is it, and in just which way(s)?” Asking that question will help us to see just what sorts of psychological process could have brought the new idea about.

Creative ideas, then, are new. But of course, there’s new—and there’s new. Ask a teacher, for instance. Children can come up with ideas that are new to them, even though they may have been in the textbooks for years. Someone who comes up with a bright idea is not necessarily less creative just because someone else had it before them.

We need to make a distinction between “psychological” creativity and “historical” creativity. (P-creativity and H-creativity, for short.) P-creativity is coming up with a surprising, valuable idea that’s new to the person who comes up with it. It doesn’t matter how many
people have had that idea before. But if a new idea is H-creative, that means that (so far as we
know) no-one else has had it before: it has arisen for the first time in human history.

Clearly, H-creativity is a special case of P-creativity. For historians of art, science,
and technology, H-creativity is what’s important. But for someone who is trying to understand
the psychology of creativity, it’s P-creativity which is crucial. Never mind who thought of the
idea first: how could anyone manage to come up with it, given that they had never thought of
it before?

If “new,” in this context, has two importantly different meanings, “surprising” has three.
First, An idea may be surprising because it’s unfamiliar, or even unlikely—like an outsider
winning the Derby. This sort of surprise goes against statistics.

The second sort of surprise is more interesting. An unexpected idea may “fit” into a style
of thinking that you already had—but you’re surprised because you hadn’t realized that
this particular idea was part of it.

And the third sort of surprise is more interesting still. This is the astonishment you feel
on encountering an apparently impossible idea. It just couldn’t have entered anyone’s head,
you feel—and yet it did. What on earth can be going on?

The Three Roads to Creative Surprise

“What is going on” isn’t magic—and it’s different in each type of case. For creativity can
happen in three main ways, which correspond to the three sorts of surprise.

The first involves making unfamiliar combinations of familiar ideas. Examples include
poetic imagery, collage in painting or textile art, and analogies. Think of a physicist comparing
an atom to the solar system, for instance, or call to mind some examples of creative
associations in poetry or in political cartoons.

In all these cases, making—and appreciating—the novel combination requires a rich store
of knowledge in the person’s mind, and many different ways of moving around within it. If the
new combination is to be valued, it has to have some point. It may or (more usually) may not
have been caused by some random process—like shaking marbles in a bag. But the
ideas/marbles have to have some intelligible conceptual pathway between them for the
combination to “make sense.”

The other two types of creativity are interestingly different from the first. They involve
the exploration, and in the most surprising cases the transformation, of conceptual spaces in
people’s minds.

Conceptual spaces are structured styles of thought. They aren’t originated by one
individual mind, but are picked up from one’s culture, or occasionally borrowed from other
cultures. They include ways of writing prose or poetry; styles of sculpture, painting, or music;
theories in chemistry or biology; fashions of couture or cooking … in short, any disciplined way
of thinking that’s familiar to (and valued by) a certain social group.
Within a given conceptual space, many thoughts are possible, only some of which may have been actually thought. Some spaces have a richer potential than others. Noughts-and-crosses is such a restricted style of game-playing that every possible move has already been made countless times. But that’s not true of chess, where the number of possible moves, though finite, is astronomically large. So is the space of possible sonnets, or screenplays, or fugues.

Someone who comes up with a new idea within a particular thinking-style is being creative in the second, exploratory, sense. If the new idea leads on to others (still within the same space) whose possibility was previously unsuspected, so much the better. Exploratory creativity is valuable partly because it can enable someone to see possibilities they hadn’t glimpsed before. They may even start to ask just what limits, and just what potential, this style of thinking has.

All professional artists and scientists do this sort of thing. Even the most mundane street artists produce new portraits every day. They are exploring their space, though not necessarily in an adventurous way. Occasionally, they may realize that their sketching-style enables them to do something (convey the set of the head, or the hint of a smile) better than they’d been doing before. They add a new trick to their repertoire, but in a real sense it’s something that “fits” their established style: the potential was always there.

What the street-artist—or Picasso, in a similar position—may also do is realize the limitations of their style. Then, they have an opportunity to change it.

The limits of the thinking-style, or of some particular aspect of it, may be slightly pushed, slightly altered, gently tweaked. They may even be changed so decisively that ideas which previously were unthinkable now become possible. The deepest cases of creativity involve someone’s thinking something which, with respect to the conceptual spaces in their minds, they couldn’t have thought before. The supposedly impossible idea can come about only if the creator transforms the pre-existing style in some radical way.

But how can that possibly happen? And how could computers help us to find the answer?

How Can Computers Throw Light on Creativity?

To understand how exploratory or transformational creativity can happen, we must know what conceptual spaces are, and what sorts of mental processes could explore and modify them.

Styles of thinking are studied by literary critics, musicologists, and historians of art, fashion, and science. And they are appreciated by us all. But intuitive appreciation, and even lifelong scholarship, may not make their structure clear. (An architectural historian, for instance, said of Frank Lloyd Wright’s Prairie Houses that their “principle of unity” is “occult”.)
This is the first point where computers are relevant. Conceptual spaces, and ways of exploring and transforming them, can be described by concepts drawn from artificial intelligence (AI).

AI tries to get computers to do the many different sorts of things that minds can do. Indeed, AI-concepts enable us to do psychology in a new way, by allowing us to construct (and test) hypotheses about the structures and processes that may be involved in thought. For instance, the structure of tonal harmony, or the “grammar” of Prairie Houses (no longer “occult”), can be clearly expressed, and specific ways of exploring the space can be tried out. Methods for navigating, and changing, highly-structured spaces can be compared.

Of course, there is always the additional question of whether the suggested structures and processes are actually implemented in people’s heads. And that question isn’t always easy to answer. But the point, here, is that a computational approach gives us a way of formulating clear scientific hypotheses about the rich subtleties of the human mind.

With respect to understanding creativity, computer models can help us because they can be creative. Or rather, they can at least appear to be creative.

Many people would argue that no computer could possibly be genuinely creative, no matter what its performance was like. It might produce theories as ground-breaking as Einstein’s, or music as highly valued as Beethoven’s … but still, for these people, it wouldn’t really be creative.

Several different arguments are commonly given. For instance: it’s the programmer’s creativity that’s at work here, not the machine’s. The machine isn’t conscious, and has no desires or values—so it can’t appreciate or judge what it’s doing. A work of art is an expression/communication of human experience, so machines simply don’t count. And all ideas have meaning, which is lacking in computers.

Perhaps you accept at least one of those reasons for denying creativity to computers? Very well, I won’t argue with you here. Let’s assume, for the purpose of this discussion, that computers can’t really be creative. That doesn’t mean, however, that there’s nothing more of interest to say.

All the objections just listed accept, for the sake of argument, that the imaginary computer’s performance is indeed very like that of human beings. What we need to focus on here is whether it’s true that computers could, in fact, come up with ideas that at least appear to be creative—and, if so, how?

**Computer Models of Creativity**

Let’s consider combinational creativity first. In one sense, this is easy to model on a computer. For nothing is simpler than picking out two ideas (two data-structures) and putting them alongside each other. A computer could merrily produce novel combinations till Kingdom come.
But would they be of any interest? We saw, above, that combining ideas creatively isn’t like shaking marbles in a bag. There must be some intelligible, though previously unnoticed, link between them that we value because it is interesting in some way. We saw also that combinational creativity typically requires a very rich store of knowledge, and the ability to form links of many different types.

For a computer to make a subtle combinational joke, for example, would require (1) a database with a richness comparable to ours, and (2) methods of link-making (and link-evaluating) comparable in subtlety with ours. In principle, this isn’t impossible. After all, the human mind/brain doesn’t do it by magic. But don’t hold your breath!

The best example of computer-based combinational creativity so far is a program called JAPE, which makes punning jokes of nine general types that are familiar to every ten-year-old. For example: What do you call a depressed train?—A low-comotive, and What’s the difference between leaves and a car?—One you brush and rake, the other you rush and brake. To be able to do this, the program needs a set of templates defining the ‘skeleton’ of each type of joke (e.g. What’s the difference between an x and a y?, What kind of x can y?, and What do you get when you cross an x with a y?), plus rules for finding words to fit the chosen template. Those rules, in turn, need access to a large semantic network (of over 30,000 items), with links representing not only meaning, hierarchy, and synonymy but also phonology, spelling, syllabic structure, and grammatical class.

Filling-in a familiar joke-schema is difficult enough. (Try to work out just what was needed to generate the joke about the depressed train.) But making a one-off jest is usually more demanding. Ask yourself, for instance, what Jane Austen had to know in order to write the opening sentence of Pride and Prejudice: “It is a truth universally acknowledged that a single man in possession of a good fortune must be in want of a wife.” (And why, exactly, is it funny?) To put the relevant knowledge into a computer, alongside (so as not to ‘cheat’) the many other things that Austen happened to know, would take forever. And to enable the program to originate the countless one-off jokes in the book (in Mr. Collins’ preposterous proposal to Elizabeth, for instance) would in practice be impossible.

In short, computer models of combinational creativity can help us to understand, in general terms, how our own combinations can come about—but they will generate valuable new combinations only rarely.

Exploratory creativity is more promising. Indeed, several programs already exist which can explore a given space in acceptable ways.

One example is Harold Cohen’s AARON, a drawing-program that can generate thousands of line-drawings or coloured images in a certain style. These are pleasing enough to be exhibited in galleries around the world. Another is David Cope’s “Emmy”, which composes music in many different styles (based on human composers such as Bach, Mozart, Stravinsky, and Joplin). Still others include architectural programs that design Palladian villas or Prairie
Houses, and programs that can analyse experimental data and find new ways of expressing scientific laws.

A few AI-programs can even transform their conceptual space, by altering their own rules. "Evolutionary" programs, for instance, can make random changes in their current rules so that new forms of structure result. At each generation, the “best” structures are selected, and used to breed the next generation. Several examples evolve coloured images which, like AARON’s, are exhibited in galleries world-wide. These images often cause the third, deepest, form of surprise. In such cases, one can’t see the relation between the daughter-image and its parent. The one appears to be a radical transformation of the other, or even something entirely different.

*Values and Creativity*

There’s no major difficulty in getting an (evolutionary) art program to make transformations: that’s relatively easy. What’s difficult is to state our aesthetic values clearly enough to enable the program itself to make the evaluation at each generation. At present, the “natural selection” is done by a human being. (In scientific domains, the value-criteria can often be stated clearly enough to allow the program to apply them automatically. So these techniques are used, for instance, to help biochemists to design new molecules for pharmaceutics.)

One huge problem here has no special relevance to computers, but bedevils discussion of human creativity too. I said earlier that “new” has two meanings, and that “surprising” has three. I didn’t say how many meanings “valuable” has—and nobody could. Our aesthetic values are difficult to recognize, more difficult to put into words, and even more difficult to state really clearly. (For a computer model, of course, they have to be stated really, really clearly.)

Moreover, they change. They vary across cultures. They are often disputed: different subcultures or peer groups value different types of dress, jewellery, or music. And where transformational creativity is concerned, the shock of the new may be so great that even fellow artists or fellow-scientists find it difficult to see value in the novel idea.

Because creativity by definition involves not only novelty but value, and because values are highly variable, it follows that many arguments about creativity are rooted in disagreements about value. This applies to human activities no less than to computer performance. So even if we could identify and program our aesthetic values, so as to enable the computer to inform and monitor its own activities accordingly, there would still be disagreement about whether the computer even appeared to be creative.

The answer to our opening question, then, is that there are many intriguing relations between creativity and computers. Computers can come up with new ideas, and help people to do so. And computer models of creativity, both in their failures and in their successes, help us think more clearly about our own creative powers.