What is the difference between real creativity and mere novelty?

devastating argument against those claiming computers can be creative: Computers can't create anything. For creation requires, minimally, *originating* something. But computers originate nothing; they merely do that which *we* order them, via programs, to do. Boden finds this argument "too quick and too simple" (p. 6), and she begins her book by decomposing Lovelace's attack into four questions (p. 7) around which her entire essay then revolves:

Q1 Can computational ideas help us understand how *human* creativity is possible?

Q2 Could computers (now or in the future) ever do things which at least *appear to be* creative?

Q3 Could a computer ever appear to *recognize* creativity?

Q4 Could computers themselves every *really* be creative?

Boden's answers to these questions, and her defense of those answers, constitute her essay. My comments, with defense, constitute my reaction and refutation. In outline, they look like this:

	Boden	**Bringsjord**
Q1	Yes.	No, not really.
Q2	Yes – but a *guarded* yes.	Yes, obviously!
Q3	Yes – but a *guarded* yes.	Yes, obviously!
Q4	No, probably not.	No.

Q4, Boden tells us, is actually beside the point, and she relegates discussion of it to the final chapter:

> For reasons explained in the final chapter, I would probably answer 'No' to the fourth question. Perhaps you would, too. However, this hypothetical moral decision-making about imaginary artificial creatures is irrelevant to our main purpose: understanding human creativity. For even if we answer 'No' to the fourth Lovelace-question, the affirmative answers to the first three questions can stand. (p. 11)

Ah, but when we reach the final chapter we find the book's big surprise (p. 274): John Searle's (1980) Chinese Room (CR) answer to Q4 is a negative one which implies that his answer to Q1 would be "perhaps, but not at a fundamental level." The point is a twofold one: Searle's negative answer to Q4 implies, *contra* Boden (p. 11), an anti-Boden answer to Q1; *and,* this implication is one Boden herself affirms (p. 274). The implication is straightforward: CR supposedly shows that executing a computer program cannot give to that which executes it *bona fide* understanding; such execution can only bestow a mindless ability to move symbols around. An affirmative answer to Q1 presupposes that "computational psychology" (which comprises for Boden the writing of programs Searle parodies) can provide genuine insights into human creativity, so embracing CR means at best a half-hearted "Perhaps" on Q1, or, as I put it on my side of the chart above, "No, not really" – since mindlessly moving squiggle-squoggles around is somewhat unlikely to reveal how *Hamlet* came to be.

Let Qn-_ ⇒ Qn-_, abbreviate the form of the implication we've just isolated. What other relations of this sort are true? Boden herself affirms (p. 10)

$$\text{Q2-Y-guarded} \Rightarrow \text{Q3-Y-guarded}$$

by reasoning which would also sanction

$$\text{Q2-Y-obviously!} \Rightarrow \text{Q3-Y-obviously!}$$

These conditionals, conjoined with the one we unearthed above,

$$\text{Q4-N} \Rightarrow \text{Q1-N, not really,}$$

imply that if the answer to Q4 *is* "No," and if the answer to Q2 *is* "Yes, obviously!," then Boden's entire project (which is in large part an attempt to demonstrate that Q2 and Q3 are to be answered with a *sedulous, reflective* "Yes") is threatened by inferences of a sort noncreative computers of today can effortlessly perform.

Why is the answer to Q2 an emphatic and obvious affirmative? The argument is short and simple: First, note that the "could" in Q2 is for Boden an "in principle" could (e.g., see p. 10). Well, it is surely possible, in principle, that computers of the future will be judged creative on purely behavioral grounds. This can be established by a thought-experiment in which computers ostensibly do all sorts of creative things – an imagined future in which our silicon-based friends generate Balzacian novels, engage in conversations with the literati about Shakespearean sonnets, and generate symphonies that would have Beethoven himself salivating. Remember, the point isn't that such a future will as a matter of fact arrive (on that score Boden is herself prudently agnostic, hence the *guarded* affirmative to Q1 and Q2); the point is that it is in principle possible that our future holds AIs which *appear,* on behavioral grounds, to be creative. "Hold on," you say, in synchrony with Boden, "for a computer to appear to be creative *in the sense intended,* its internal workings would have to be of the right sort." True enough, but the objection is surmounted by adding to our "gedankenexperiment" a pinch more imagination: we have only to watch the innards of our "creative" computers being probed by skeptical cognitive scientists, resulting in the discovery of unbelievably complex systems, whose details are impenetrable, but whose broad strokes suggest that they are *n*th generation descendants of today's AI systems. (Traditional symbolicist systems were in 1995 given the capacity to evolve in wild ways with help from connectionist-based sensors and effectors, and by 2005 our synthetic rival to Beethoven arrives.)

Doesn't Boden dispose of Searle's CR? Perhaps. But another Q4-N rationale, which she touches upon in her final chapter, may well support the "Q4-N ⇒ Q1-N, not really" conditional. This rationale for Q4-N is a variant on what Boden calls (pp. 278–281) the "consciousness argument" (a variant specified in Bringsjord 1994): (a) Creativity requires inner, first-person-point-of-view, *something-it's-like-to-be* consciousness. (b) No computer could ever be conscious in this sense. Ergo; (c) no computer can ever be creative. Boden would simply pronounce this line of reasoning "iffy" (p. 281) and go her merry way. This would seem to be a dangerous response, however, given that a number of thinkers have argued explicitly for (b), for example Jackson (1982); Kripke (1971); Nagel (1974); Searle (1992); and Bringsjord (1991). But why might (a) be true? Connected to the Boden-covered area of computer generated literature, the supporting argument, encapsulated, is this: in order to produce sophisticated fiction, an author, whether human or artificial, must adopt the points of view of the characters involved, but if X adopts the point of view of Y, then X itself has a point of view. Boden's reaction would doubtless be: "But look at the point-of-view-*free* computational accounts of literary creativity canvassed in my Chapter 7!" This, of course, is nothing but a statement of faith – faith that these techniques can work. Well, here we can trade. I too have faith, faith based on failure: after toiling for three years as a PI in a well-funded group effort (known as *Autopoeisis*) to realize, by following the computational techniques Boden praises in her seventh chapter, the sort of storytelling AI she thinks possible, I have no reason whatever to think (b) anything but depressingly true.

What is the difference between real creativity and mere novelty?

Alan Bundy

Department of Artificial Intelligence, University of Edinburgh, Edinburgh EH1 1HN, Scotland; a.bundy@ed.ac.uk

1. Introduction. In *The creative mind,* Boden advances both a general thesis about creativity and a more specific one. The

general thesis is that creativity can be scientifically understood in computational terms. This general thesis has my enthusiastic and wholehearted support. The specific thesis is that what distinguishes "real creativity" from "mere novelty" is the "mapping, exploration, and transformation of conceptual spaces." "Mere novelty" is just the generation of a new object from an existing conceptual space. It is this specific thesis that I wish to investigate.

2. *Mathematical creativity?* Consider the following AI case study. Larry Wos and his coworkers at Argonne National Laboratory in Illinois have used an automated theorem prover to solve open conjectures in mathematics (Wos 1993). Many of these open conjectures were suggested by human mathematicians. The conjectures came mostly from areas of mathematics in which human mathematicians had developed few intuitions about guiding proofs, for example, ternary boolean algebra. They were fed to a fairly standard but very efficiently encoded resolution theorem prover (initially AURA, more recently OTTER), which used brute force methods guided by a few simple, syntactic heuristics. Sometimes millions of intermediate formulae are generated in the search for a proof. The resulting proofs have sometimes been published in the mathematics literature (Wos et al. 1983).

Is this an example of artificial creativity? According to Boden's specific thesis, the answer is an unequivocal *no;* this is mere novelty. The automated theorem prover is merely generating something new from an existing conceptual space. The conceptual space here is the search space defined by the application of resolution to the axioms of a mathematical theory and the open conjecture.

On the other hand, mathematics is often regarded as one of the pinnacles of human intellectual achievement. The proof of an open conjecture by a human mathematician would usually be regarded as an act of real creativity, not mere novelty, especially if this open conjecture has defeated other mathematicians. However, the proof of open conjectures often appears to involve no more than what Boden characterizes as "mere novelty," the combination of rules of inference in a predefined mathematical theory.

3. *A complexity requirement.* One way to resolve this conflict would be to add an element of complexity into the requirements for creativity, that is, to regard generation from an existing conceptual space as creative rather than merely novel, if the conceptual space is large and complex, and if the generation is from some little explored part of that space. Boden also seems to recognise this possibility. She quotes Dickens's "a squeezing, wrenching, grasping, scraping, clutching, covetous old sinner" as an example of creativity, while admitting (p. 49) that under her definition it is only "mere novelty." She excuses this deviation on the grounds that, although the grammar of English admits "sevenfold strings of adjectives" this part of the conceptual space was previously unexplored. Similar remarks apply to the poetry of Coleridge and the music of Mozart. However, although she describes these counterexamples to her specific thesis, Boden does not draw any general conclusions about the need to modify that thesis nor about the kind of modification required. My proposal of a complexity requirement aims to plug that gap.

The complexity requirement cuts both ways. There are also examples of the mapping, exploration, and transformation of conceptual spaces which we would not want to count as real creativity. Consider mathematics again. Under the specific thesis, real creativity is only involved when the mathematical theory is changed. But changes to a mathematical theory are easily made. Take the axioms of group theory and delete one – or generate a new formula at random and add it as a new axiom. Of course, most such changes will fail Boden's criterion of value. There will be no applications for the new theory or few interesting theorems provable. However, if the value criterion is factored out, the complexity criterion is still required. A simple change to an existing theory counts for less than a radical change – compare the original axiomatisation of group theory to the later modifications to semigroups or rings.

4. *Transformation as generation.* It is no coincidence that similar criteria should apply when assessing creativity both in the generation and in the transformation of conceptual spaces. Transformation is a kind of generation – but at the metalevel. A conceptual space is defined by some kind of grammar. Generation is the application of rules of that grammar. If that grammar is regarded as a data structure then transformation can be effected by a metagrammar, whose rules modify the original grammar. For example, the metagrammar for generating new mathematical theories might have rules for deleting old axioms and adding new ones. Thus, the "real creativity" of transforming conceptual spaces is just the "mere novelty" of generating from an existing conceptual space, but at the metalevel.

This observation further undermines Boden's specific thesis. It also explains the enthusiasm, among some AI researchers, for symbolic representations, especially declarative ones. A declarative grammar is readily viewed both as a procedure for generating novel objects and as a data structure which can be transformed into a new grammar. There is no need for representational redescription; the mapping and exploration phases of creativity are trivial and only transformation using a metagrammar is required. This advantage is lost to some extent by procedural representations and it is totally lost by subsymbolic representations.

These observations might lead us to locate "real" creativity in the representational redescription of nondeclarative representations or in the invention of transformational metagrammars. It is usually a mistake, however, to regard one aspect of an intellectual process as the key with the others playing only a supporting role. "Real" intelligence appears to arise from the interplay of a number of relatively mundane processes.

5. *A self-reflection requirement.* Complexity seems a rather crude criterion for the assessment of creativity. Returning to the automated proof of open conjectures, we can see another objection to be the brute force nature of the search for a proof. Real mathematicians surely reflect more on what they are doing. They bring known methods of proof to direct their search and they sometimes invent new methods of proof during the search. Indeed, the invention of a new proof *method* is often more highly regarded by fellow mathematicians than the new proof itself.

These notions of proof method and self-reflection can also be represented computationally. For example, I have tried to do this in my own work on *proof plans* (Bundy 1991). Proof planning is represented as the exploration of a metalevel conceptual space composed of proof methods. Generation in this metalevel space consists of reasoning about the problems to be solved and the methods available for their solution. This suggests an alternative modification to Boden's specific thesis. "Mere novelty" may arise as the unreflective generation of new objects from an existing conceptual space. "Real creativity" may arise when that generation involves some aspect of self-reflection, that is, the simultaneous reasoning about the generation process at the metalevel.

6. *Conclusion.* I suspect, however, that all attempts to characterise creativity simply in computational terms are doomed to failure. Creativity can be explained computationally, but it is not a *natural computational kind,* that is, it does not correspond to some well-defined family of computational processes. Rather, what we call creativity in folk psychology corresponds to particular aspects of many different kinds of computational processes.

ACKNOWLEDGMENT

Thanks to David Basin, Sean Matthews, Toby Walsh, Geraint Wiggins, and Larry Wos for comments on an earlier draft.