8 Freedom and constraint in creativity

Philip N. Johnson-Laird

> One's own free and unfettered volition, one's own caprice, however wild, one's own fancy, inflamed sometimes to the point of madness – that is the one best and greatest good, which is never taken into consideration because it will not fit into any classification, and the omission of which always sends all systems and theories to the devil.
>
> – Dostoyevsky (1864)

Creativity is a mystery, and many people believe that it should remain a mystery. It should not be scrutinized too closely, says the anxious Romantic, because there is a danger in knowing too much about it. If we discover its sources, they may dry up. The cynical Realist asserts a different proposition: Those who cannot create study those who can. Critics and historians have assessed acts of creation by the canons of their day. Artists and scientists have reflected on their own and others' inspirations. And, above all, psychologists have carried out tests to measure creativity, experiments to explore it, exercises to enhance it, and investigations to reveal it in the lives of gifted individuals. Contrary to Romantic and Realist alike, a remarkable amount of imagination has been exercised in studying imagination, and we are none the worse for it. Alas, we are not too much wiser, either.

Few students of creativity have stopped to define what it is that they are studying. On the whole, a priori definitions do not advance science, but impede it. The advance of science, however, enables us to frame superior a posteriori definitions. My goal in this chapter is to arrive at such a definition of creativity. To reach that goal, I shall first analyze the nature of free will – for to be creative is to be free to choose among alternatives. Next, I shall consider constraints on creativity – for what is not constrained is not creative. These considerations will enable me to propose a computational theory of creative processes. The theory postulates a relation between computational power and the temporal demands of creating new ideas. I shall explore this idea in a particularly tractable domain – musical creativity – and establish certain existence proofs in the form of computer programs that generate bass lines and tonal chord sequences. The theory is computational because

I am grateful to my colleague Roy Patterson for encouragement and for developing the software that turns raw numbers into musical sounds.

I want to avoid a defect that has been common to theories of creativity from Wallas (1926) to Bateson (1979): an amalgam of vagueness and incompleteness that takes too much for granted. If quantum electrodynamics and meteorology can be modeled computationally, then it should be possible to do the same for theories of creativity. It seems a sensible strategy to start with the hypothesis that the processes of creation are computable. If they are not, then they may be beyond the scope of science, or else the foundations of computability theory may be wrong.

A working definition

The best recent dictionary of psychology (Reber, 1985) offers the following definition:

> *creativity* A term used in the technical literature in basically the same way as in the popular, namely, to refer to mental processes that lead to solutions, ideas, conceptualizations, artistic forms, theories or products that are unique and novel.

With one caveat – to which I shall return presently – I applaud this definition. But I fear that its emphasis on mental processes may win few adherents in certain quarters, and so let me first defend it.

In a fascinating historiometric project, Simonton (1984) has analyzed many sorts of data and showed that there are systematic relations between such data and the chances of an individual becoming an acknowledged genius. Simonton defines genius in terms of fame and influence and makes no distinction between creativity and leadership. He writes: "when the most famous creators and leaders are under scrutiny the distinction between creativity and leadership vanishes because creativity becomes a variety of leadership." No one can doubt that certain creators may become leaders, or that certain leaders may exercise a high degree of creativity. But not all great creators have their schools of followers or are even judged to be great within their lifetimes. In his day, J. S. Bach was known primarily as an organist, and his compositions were neglected until his genius was recognized by Mendelssohn and other nineteenth-century composers (on whom his techniques had little influence). Conversely, a political or military leader may be a leader by exercise of force, and a spiritual or religious leader may be a leader by virtue of asceticism. Leadership need not depend on imagination, and there is no warrant, other than Simonton's fiat, for identifying it with creativity. The mind of a successful leader may work in quite a different way from the mind of a successful creator, and the danger in equating the two is that the difference between them will be overlooked.

A focus on genius and exceptional acts of creativity also has its dangers. On the one hand, there may be many intangible social factors at work in the creation of major works of art and science. Of course, one should like to understand them, but it is difficult to investigate scientifically what it was about Periclean Athens, Renaissance Florence, Elizabethan London, or fin-de-siècle Vienna that made them such outstanding forcing grounds for the intellect. On the other hand, such a focus leads naturally to the view that there is some sort of discontinuity between prosaic acts of imagination and products of genius. There may be such a discontinuity – indeed, I shall make such a case later in this chapter – but unless we investigate both the mundane and the marvelous we shall never discover its nature.

At this point, let me make my one objection to Reber's definition. It asserts that the products of creation should be "unique and novel." Uniqueness and novelty are, of course, matters that can be determined only by considering what else has been created in other places and in other times. Two astronomers working independently in the middle of the nineteenth century predicted within a year of each other that there was a further planet beyond Uranus. The discovery of Neptune confirmed their prediction. Are we to say that only one of them, the one with priority, exercised a creative process of thought? By no means. I shall assume that the critical factor is that the product of a creative process should be novel *for the creator,* not merely remembered or perceived. If Pierre Menard, the eponymous hero of Borges's (1970) story, rewrites *Don Quixote* word for word, not by copying it but by becoming Cervantes (or, with more difficulty, by remaining himself), then by my criterion he does not fail to be creative. Nor do we fail to be creative, as is Borges's point, when we reread the novel with a deliberately anachronistic attribution to a twentieth-century author.

Choice and freedom of will

Many mental processes deliver products that are novel to those who entertain them. If I ask you to multiply two numbers together, or to reverse the order of words in a sentence, then the result may be something that you have never experienced before. Yet both of us, I suppose, would be disinclined to judge your performance as creative. We would not withhold the judgment because you merely did as you were told – Bach merely did as he was told in composing the *Musical Offering* on a theme given to him by Frederick the Great. We withhold it because, unlike musical composition, the tasks I gave you can be performed by rote. That is to say, they can be carried out as a result of a calculation, or deterministic procedure, that gives no freedom whatsoever to the imagination. Of course, you could do these tasks imaginatively. The point is that you do not have to; you will succeed perfectly well using procedures that leave no room for choice.

The concept of freedom that I have invoked refers, of course, to freedom of will – a puzzle that philosophers have agonized over for centuries (see Dennett, 1984, for some avuncular comfort on the topic). The problem of free will and the problem of creativity are, in some respects, one and the same. They can both be solved together.

If a task can be carried out by a process that at no stage calls for a choice to be made, then I shall say, in the parlance of computer scientists, that it is "deterministic." Thus, long multiplication is deterministic, because what is done at each stage is, in principle, wholly determined by the numbers to be multiplied and the current state of the calculation. Human beings, of course, have certain deterministic behaviors, such as the protective blink of the eyelid. They can also choose to carry out deterministic procedures – as when they choose to carry out a long multiplication. But human beings are also able to behave in ways that are not deterministic. By this claim, I do not mean that they violate the laws of nature, nor do I mean that their behavior is necessarily governed by indeterminate quantum events in brain cells. What I mean can be illustrated by asking the reader a simple question:

What are you going to do next?

You could choose – rational individual that you are! – to continue reading this chapter (if only to find out my solution to the riddle of free will). But you could equally well decide that you have had enough of creativity for the time being, and go out for a walk. There are many, many other possibilities. (Indeed, the problem of life is solved if you always have an answer to my question.) Normally, we decide what to do next either in response to events in the world (such as ringing telephones) or in response to mental states (such as boredom). And, normally, the decision is made tacitly: We decide what to do without reflecting on *how* the decision should be made. Such tacit decisions are likely to reflect a number of factors, both conscious and unconscious, that folk psychology refers to as "intuitive" or as "gut reactions." But if we do reflect on the matter, we can decide how to make the choice. We can even decide – should we so wish – to make an *arbitrary* choice.

If we are confronted with two equally appealing alternatives, then, rather than deprive ourselves of both as a result of indecision, we can plump for one alternative, not on the basis of any further rational evaluation but as a result of an arbitrary decision. Buridan's ass is said to have starved to death as result of being unable to decide which of two equally attractive bales of hay to eat. This dilemma seems implausible, because the slightest movement toward one bale or the other is likely to break the conflict. Approach–avoidance conflicts, however, are more likely to have a paralyzing effect. Animals that are arrested by them lack freedom of will. Human beings, however, are wont to say to themselves, "This is ridiculous: I'll have to do something." They may then, as a result of this higher-order reflection about the choice, make an arbitrary decision.

When someone makes a seemingly arbitrary decision, and has no idea on what it was based, psychologists often suppose that it was, in fact, determined by some minuscule aspect of the environment. Psychoanalysts emphasize the role of the internal environment; behaviorists emphasize the role of the external environment. If we knew the state of the individual's unconscious mind, or bank balance, they say, then we could account for the decision, which was entirely determined by such factors. It is merely our ignorance that forces us to treat it as nondeterministic. There are some splendid experimental demonstrations of how factors outside the individual's consciousness can influence the outcome of a decision (Nisbett & Wilson, 1977). Nevertheless, there are decisions that are truly nondeterministic. You may resolve – at a meta level – to make a deliberately arbitrary decision and to ensure its arbitrary nature by recourse to external means. You spin a coin, toss a die, or, if you are a psychologist, consult random-number tables.

What gives us freedom of will is the ability to reflect about *how* we shall make a decision, and thus to choose at a meta level a method of choice. We may decide to set out the pros and cons for the alternatives and attempt to make a rational decision, perhaps by seeking to minimize our maximum regret (cf. Pascal's wager on the existence of God, and Darwin's decision to marry); or we may decide to make a tacit intuitive choice on the basis of a "gut reaction" and without further reflection; or we may decide to make a decision based on some factor outside our control (e.g., consulting the Bible, I-Ching, or the Delphic oracle, or following a spouse's advice, whatever it may be); or we may decide to make an arbitrary choice, either plumping at random for one alternative or selecting an external randomizing mechanism. We may even decide not to decide, but to wait to see how the spirit moves us at the last moment.

The meta-level choice of a decision procedure may itself be made in the usual tacit way. But we can confront the question explicitly and make a conscious decision (at the meta-meta level) about how to choose (at the meta level) the method of choice. Should you continue reading or go out for a walk? You decide to spin a coin to decide whether to choose by minimaxing or by tossing a die. And how did you make *that* decision? There is potentially no end to the hierarchy of decisions about decisions about decisions, and so forth. Similar hierarchies form the basis of consciousness, intentional behavior, metacognition, and the development of an understanding of meaning and inference (Johnson-Laird, in press-a). They seem to depend on the ability to embed mental models within themselves recursively (Johnson-Laird, 1983).

Fortunately, there are at least two constraints likely to curtail the hierarchy of meta-level decisions from towering ever upward. First, the business of life demands that we reach decisions rather than get lost in speculation about how to reach them. The buck must stop somewhere. The decision at the highest level is always necessarily tacit. It can never be made explicitly, for if it were, there would, of course, be a still higher level at which a decision was taken to use the particular explicit technique, and this higher decision must have been tacit (on pain of an infinite regress). Second, there are constraints between the levels. If we decide to choose at random between alternatives at one level, and one of these alternatives is itself a random method of choice, then we might as well go straight to the latter. It can be rational to decide to make a random decision (in certain games, for example), and it can be rational to decide to make a rational decision (in certain other games, for example). But can it be rational to decide (at a meta-meta level) to decide at random (at a meta level) between making a rational or random decision? I think not.

We are free not because we are ignorant of the roots of many of our decisions, which we certainly are, but because we know that we can choose how to choose, and we know that among the range of options are those arbitrary methods that free us from the constraints of any ecological niche or any rational calculation of self-interest. This fact lies behind Dostoyevsky's deepest beliefs, epitomized in the quotation at the beginning of this chapter, and behind the Existentialists' fascination with gratuitous acts. One demonstrates freedom (if not imagination) in acting arbitrarily.

Freedom in creation

Freedom of choice occurs par excellence in acts of creation. When an artist paints a picture, at each point there are several possible brush strokes that could be made. When a musician improvizes a melody, at each point there are several possible notes that could be played. When a scientist imagines how a phenomenon might be explained, at each point there are several lines of thought that could be explored. When a speaker expresses an idea, at each point there are several possible forms of expression. In every case, the set of choices is constrained by largely tacit mental criteria that determine the genre and the individual's style. Sometimes, perhaps, these criteria reduce a particular set of possibilities to just a single item, but in general there is a range of options.

How is the choice among options made? Sometimes it depends on principle. But if the principles by which an individual creates were to completely determine every choice, then apart from the first stroke on the canvas, the first note on the piano, the first line of thought or word in the sentence, everything would be inevitable. There would be only as many works as there were beginnings. It follows that some choices are arbitrary. Among them are those occasions on which the individual explicitly exercises freedom of will and knowingly makes – or attempts to make – an arbitrary choice from among the set of viable possibilities. The mind certainly contains a system for making arbitrary choices. Nondeterminism in a deterministic device, such as a computer, is simulated by borrowing a technique from the casino at Monte Carlo and choosing at random. However, people tend to be rather poor at making genuinely random choices when they are asked to do so in the psychological laboratory. The departures from true randomness do not count against the existence of a mental system for arbitrary decisions, but rather imply that the mechanism lacks access to anything like a random-number generator and that perforce one such decision may influence others. Evidently, if the brain is governed by quantum indeterminacies, it is unable to exploit them.

In short, creativity depends on arbitrary choices and thus on a mental device for producing, albeit imperfectly, nondeterminism. Unlike calculation and other deterministic procedures, which yield the same response in the same situation, a genuine process of imagination could deliver a different response the second time around if the same stage of the process could be reinstated exactly. At each step, there may be more than one possible continuation. What determines the set of possible continuations is a matter that I shall take up next.

Criteria, genres, and the paradox of creativity

Creativity is like murder – both depend on motive, means, and opportunity. Society has, as I have already noted, dramatic effects on the creation of works of the imagination. There are grounds for supposing that these effects can be loosely divided into (a) diffuse global factors that affect motivation and opportunity and (b) specific cultural factors that influence the means of production – genre, paradigm, and style. The former are indeed the province of historiometricians. For instance, Simonton (1984, p. 170) observes that political instability in one generation depresses in the next the likely number of major creators in discursive fields such as science, philosophy, and literature. Specific cultural factors are the province of critics and historians. Cultural practices lead to the crystallization of artistic genres and scientific paradigms. These frameworks are the products of earlier creative processes, which are transmitted, often with significant modifications, from one generation to the next.

As Reber's definition allows, people create solutions, ideas, conceptualizations, artistic forms, theories, and products. I want to distinguish, however, between creation within an existing genre or paradigm and the creation of a new framework itself. Creation within a framework depends on access to its principles – the criteria or constraints of the framework – and ultimately these principles must be embodied within the mind of the creator. Even the invention of a new framework (i.e., new principles) must meet certain other criteria – not everything goes – but, as we shall see, these criteria are unlikely to be embodied in the mind. The creative process therefore depends on criteria. Conversely, any outcome that lies outside all frameworks is likely to be judged as uncategorizable rather than as creative.

When I talk of criteria, it is natural to think of the sorts of principles that are spelled out in theoretical treatises and in works on aesthetics and the philosophy of science. In fact, I have a different and broader notion in mind that I can bring out by reminding the reader of the central paradox of creativity (Perkins, 1981, p. 128). People are better critics that creators. The paradox, of course, is that if they have the knowledge to judge the products of a creative process, then they ought to be able to use it to generate such products in the first place. The resolution of the paradox depends on two factors. First, the explicit knowledge that is consciously accessible to a critic is by no means sufficient for the generation of ideas, which depends on other tacit forms of knowledge. Second, this tacit knowledge is not automatically available to consciousness.

The mind appears to depend on a set of separate processors that communicate data one from another, but that are not privy to each other's internal operations or representations; there are various versions of this hypothesis (Fodor, 1983; Johnson-Laird, 1983; Rumelhart, McClelland, and PDP Research Group, 1986). Hence a particular form of mental representation may be used by one ability, but not by another. Consider, for example, whistling a tune. It depends on a mental representation of a sequence of musical intervals. The ability to transcribe the tune in a musical notation also depends on a mental representation of a sequence of musical intervals. You can whistle a tune, but can you write it down? Most people can whistle, but even among those who can cope with musical notation, only a few can write down a tune as a result of being able to whistle it. The task is difficult because the melodic representations for whistling are not available to the symbolic process of writing.

In the same way, the tacit criteria for generating ideas are not available to conscious critical processes. Because critical criteria are easy to communicate to other people but insufficient for creation, whereas generative abilities are unconscious and ineffable, critical judgment tends to be considerably in advance of the ability to create works of the imagination. The paradox of creativity therefore leads ineluctably to the view that there are many criteria on which the creator must rely and that by no means all of them are available to overt inspection. Some of these criteria are common to many practitioners; they constitute the genre or paradigm. Other criteria are unique to individuals; they constitute an individual style of thought within the more general framework.

The principles that I have described amount to a theory of creativity at the computational level (Marr, 1982) – a theory of *what* has to be computed, namely, nondeterministic choices among the options characterized by a set of criteria. A theory at the algorithmic level must specify *how* the choices are made. In fact, I shall argue that, depending on the demands of the creative task, there are three possible classes of procedures. The crucial determinants of which sort of procedure is used are whether creation occurs within a framework or is intended to produce a new framework, and whether or not there is any opportunity to revise the creative product. I shall consider the resulting types of creativity in the next three sections of this chapter.

Creation within a genre in real time

If it is necessary to work rapidly within a framework – typically an artistic genre – with no opportunity for revision, then a sensible procedure is for the principles governing the genre and the individual's style to be used at each point in the process of *generating* ideas. They will constrain the set of options as tightly as possible. If they still leave open several options at any point, then a rapid arbitrary choice can be made from among these alternatives.

This procedure is likely to be used whenever there is time pressure on the creator. Thus, I hypothesize that it is used in all extemporaneous performances, including the making of artifacts in media that allow no second chances, the spontaneous use of natural language in discourse, and the improvisation of music, dance, and other art forms. As a test case, I shall consider musical improvisation because it can be treated largely as the syntactic organization of sounds into patterns, without having to worry about what, if anything, those patterns might represent.

Musical improvisation is governed by principles that must be in the musician's mind and that must suffice for the generation of music in real time. There is no opportunity to go back and revise an improvisation, and the musician cannot afford to make mistakes (i.e., to choose notes that do not make up a satisfactory melody of the appropriate variety). Many great composers – Bach, Mozart, and Liszt, for example – were consummate improvisers. Beethoven is another particularly interesting case, because he improvised with such fluency and brilliancy that his extemporaneous works were considered by some of his contemporaries to be better than his compositions (Sonneck, 1967). Yet his notebooks show that he composed with the greatest of difficulty. The two skills evidently depend in part on different underlying processes, as is borne out by the existence of composers who cannot improvise and improvisers who cannot compose.

What is common to most forms of improvisation is a reliance on two quite separate mental components: first, a long-term memory for a set of basic structures, such as the chord sequences of modern jazz or the ragas (scalic patterns) of Indian music; second, a set of tacit principles that underlie the improvisatory skill. We know that these two components exist because the basic structures are accessible to consciousness, and musicians can talk about them, write them down in a suitable notation, and teach them to neophytes. But, this explicit knowledge is not sufficient to enable a musician to improvise. Hence, there is a second component, which is relatively inaccessible to consciousness. Some musicians are aware of a few of its principles, but no one has complete access to them. Musicians learn to improvise by imitating other virtuosos and by experimenting with various possibilities. They learn to improvise by improvising, and they thereby develop their own particular styles within a genre.

A jazz musician can make up melodies that fit a large variety of different chord sequences. These chord sequences are known by heart, and the same basic sequence is used over and over throughout a piece. The computational problem in improvisation is therefore to produce in real time an acceptable melody that fits the chord sequence, and the tempi of modern jazz may call for melodies to be extemporized at an extremely rapid rate (e.g., 10 to 12 notes per second). A plausible conjecture about the solution to this problem can be based on the differences between the basic structures and the tacit principles.

The chord sequences are not made up during performance, and they may be the work of several musicians over a long period of time. Hence, it is expedient to do as much work as possible in the construction of the chord sequences so that they provide a rich structure that is latent with possibilities for the improviser. I shall presently outline a theory of this process.

The tacit skills have to run efficiently in real time. They govern the choice of notes to fit the harmonic implications of the chord structure and to make a good melody. I conjecture that these principles embody as little *computational power* as possible. What this conjecture means, in practice, is that the principles should place a minimal load on memory for the results of intermediate computations. In other words, the principles should take as input the basic chord sequence and deliver an

Figure 8.1. A fragment from an improvised bass line.

improvised melody as directly as possible without any internal representation of an intermediate form.

As an initial test of this conjecture, I have established an existence proof of the feasibility of such a computational architecture, because I have shown that it is possible to produce passable improvised bass lines without using intermediate representations.

The double bass player in modern jazz improvises a bass line to fit a given tonal chord sequence. Figure 8.1 presents a fragment from a typical bass line, together with the chords on which it is based. The bass line is rhythmically simple, just a steady four beats to the bar, though other styles are more complex. The actual timing of the notes depends on an exquisite sense of the metrical pulse of jazz. Nevertheless, the bass line allows us to approach the improvisation of a melody without the complications of rhythm; for a discussion of rhythm, see Johnson-Laird (1987).

There are several theories of how a bass player decides what note to play next. The player might merely choose any note in the range of the instrument that is among those making up the current chord. But this procedure would leap wildly around from a low note to a high note in a most unmelodic way, and it would fail to use ''passing'' notes, that is, notes that are not in the current chord but that pass from one such note to another (e.g., the second bar of Figure 8.1, which contains two passing notes to the chord of F7: the chromatic F♯ and the more consonant G that leads back to a major note of the chord, A).

A second method was suggested by Ulrich (1977). He argued that the melodic improvisations of jazz musicians are made up from existing motifs – fragments of melody – that are woven together to form a new melody. The musician does not make up new melodies, but rather modifies existing motifs to fit the current harmonic situation. This idea was implemented in a program devised by Ulrich and in a more sophisticated program developed by Levitt (1981). Levitt's program takes as input both a chord sequence and an existing melody. It divides the melody into units of two bars, which are then reused in a different order and in variant forms that fit the current harmony. The program is deterministic; that is, given the same input melody and chord sequence, it produces the same improvisation. Jazz musicians, however, do not perform in this way. They use motifs some of the time, but no one, apart perhaps from a complete beginner, uses them all of the time. It is easier, in the long run, as any competent performer will attest, to make up new melodies than to remember a vast array of motifs and to modify them to fit the chord sequence; see, for example, the reminiscences of David Sudnow, 1978, a well-known sociolinguist

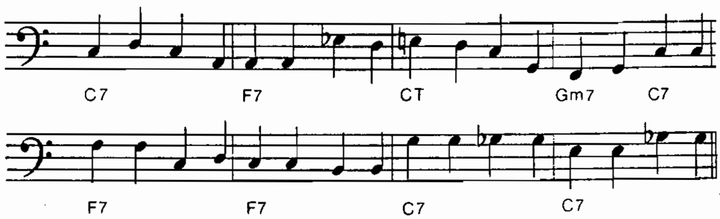

Figure 8.2. A fragment from the output of the program generating bass lines.

who learned to play jazz piano. There is no more evidence for creation solely by motif than there is for conversation solely by cliché.

Perhaps the most plausible hypothesis is that bass players choose their notes in order to meet two sets of criteria. The first set reflects the player's tacit knowledge of harmony, including both the notes that are concordant with chords and those that may be used in passing. The second set reflects a tacit knowledge of the appropriate "contour" underlying successful melodies. Here, the general idea is likely to be that after a series of small steps in scale, a step of a rather larger interval, and vice versa, makes for a pleasing melody. These principles are embodied in the two sets of criteria used by my computer program.

The program takes a chord sequence as input and delivers a viable bass line as its output. To produce each note, it first generates a small step, a large step, or the same note again, according to the rules of a grammar for contours derived from a corpus of improvisations. The grammar is "regular"; that is, it can be used with the minimum of computational power, which does not need any memory for the results of intermediate computations. The program then selects a pitch that meets the constraint of the step size and a set of harmonic principles. Where more than one possible note meets the various constraints, a random choice is made between them. Figure 8.2 presents a fragment from an entirely typical output of the program. The output, which also contains a rudimentary accompaniment based on the chord sequence, is played by a further program, devised by Roy Patterson and Rob Milroy, that synthesizes the sound of the double bass and the accompaniment.

The program is quite competent, but it lacks two abilities of the jazz player. It makes no specific use of chromatic runs of several passing notes (see the second bar of Figure 8.1), and it makes no use of motifs, which occasionally are featured in bass performances. Likewise, the program commits a minor solecism that revealed the existence of a special category of passing notes of which the author was previously unaware. The modifications to rectify these shortcomings do not require a larger memory for intermediate computations, but merely a slightly larger buffer for what has just been played.

The reader may be worried about the use of a grammar. It is often claimed that a creator "breaks the rules" in order to produce a more original work of art. Likewise, although a grammar may capture a genre, individuals have their own unique styles. Both these objections are instructive, but not decisive. If a creative process breaks the rules, then either it must make a choice at random regardless of the consequences or it must be governed by yet further criteria. These criteria can in turn be captured in a grammar. Hence, the breaking of a rule can be described by yet another rule (or else it is merely an arbitrary infraction). If an individual has a unique style, then it must depend on idiosyncratic biases in choosing alternatives. A grammar can likewise be framed to capture this style.

The procedure used by the bass program exemplifies a computational architecture in which criteria are used directly in the generation of a creative product. Because the criteria suffice to define a genre, the output is guaranteed to be at least viable, and the procedure can also be of weak computational power, which requires a minimal memory for the results of intermediate computations. Hence, the generative stage yields only a relatively small number of possible options, all of which meet the desired characteristics. Where there is more than one possible continuation, an arbitrary choice can be rapidly made. The choice has to be arbitrary because all the criteria are used in the generative process. The procedure can also rely on a long-term memory for structures produced using a greater degree of computational power. They enable the finished product to have more intrinsic interest than could be generated solely by an "on-line" procedure. The procedure as a whole is highly efficient, but it is feasible only if there is some way for previous experience to yield a mental representation of the criteria that govern the generative stage. I have elsewhere likened this computational architecture to Lamarck's theory of evolution (Johnson-Laird, 1987). He proposed that what an organism acquires by adapting to its environment can be conveyed to its progeny, and thus acquired constraints can guide the process that generates species.

Creation within a framework in stages

The generation of ideas by tacit principles does not always yield a product that meets the creator's critical criteria. In the case of improvisation, there is nothing to be done about such shortcomings. But in many genres there is no pressure to produce a performance in real time, and consequently there is an opportunity to revise unsatisfactory works. This possiblity is indeed the norm for most forms of creativity.

There is a corollary; if there is time to revise or to reject the products of a generative process, then the ultimate results are likely to rely on a high degree of computational power. That is, they can be produced making considerable use of a memory for intermediate results. It does not follow that the creators must make greater use of their working memories; if it is possible to leave a permanent record of the product in some external medium, then that record is itself a form of memory for an intermediate result. In certain artistic forms, such as painting and sculpture, the incomplete work is itself such a record. In science and other forms of art, it is possible to record incomplete or tentative ideas in some notation. Writing enhances

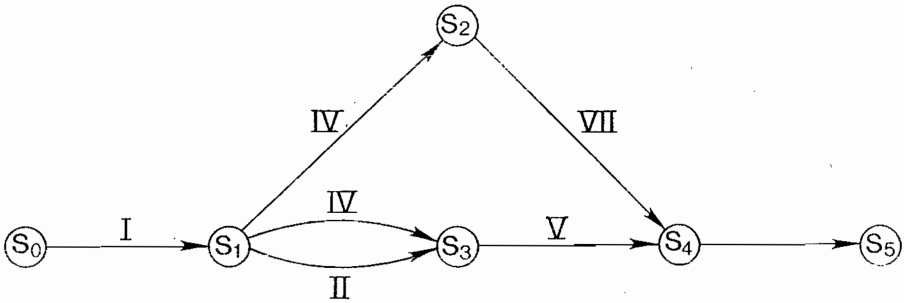

Figure 8.3. A device for generating chord sequences.

the computational power available to a creator by providing an external memory for intermediate results.

We saw in the previous section that jazz improvisations are based on chord sequences that are developed "off line" (i.e., without the need to make them up in real time). Hence, the theory predicts that the development of such chord sequences will demand rather more computational power than the spontaneous improvisations based on them. In order to test this prediction, I have examined a corpus of such chord sequences and have written a computer program to model – in rather abstract terms – the processes that might be responsible for their generation.

There is a large body of musical theory concerning the structure of tonal chord sequences (of the sort that includes those used in jazz). Most of these theories are too vague and incomplete to be directly modeled in computer programs, but they are sufficiently clear to establish their degree of computational power. Thus, theorists from Rameau (1722/1971) to Forte (1979) spell out systems that do not need intermediate representations. Ironically, these theories are equivalent in computational power to "regular" grammars of the sort that I used in accounting for improvisations in real time. A typical fragment from such a theory of chord sequences is illustrated in Figure 8.3. The Roman numerals denote the root of the chord: I = the tonic, V = the dominant, and so forth. (Nonmusicians need not worry about the interpretation of these symbols, but should treat them as strings in an abstract symbolic language.) A chord sequence is generated by starting in the initial state, S_0, and making transitions from one state to another. As each transition along an arrow is made, the symbol above the arrow is generated. Thus, the device produces, for example, the following sequence of symbols: I II V I, which corresponds to the standard tonal sequence of tonic, supertonic, dominant, tonic. A more realistic device would specify the type of chord on each of these roots – say a major triad on I, a minor triad on II, a seventh on V, and a major triad on I. The device shown in the figure is nondeterministic in that in state S_1 there are three different choices available. Some modern theorists have proposed adding probabilities to the different choices in order to model their frequencies of occurrence (Eigen & Winkler, 1983).

Devices that make no use of memory for intermediate results are not powerful enough to generate the tonal chord sequences of modern jazz. They are not powerful enough to generate the chord sequences of classical music, but the case is easier to make for jazz. As any jazz accompanist will attest, a piece has an underlying chord sequence that can be realized in many different superficial variations. Thus, the traditional chord sequence known as the "twelve bar blues" has many alternative forms. Steedman (1982) has outlined a set of rules that generates these alternatives from the underlying basic sequence. His grammar takes for granted the existence of an intermediate representation, namely, for the underlying basic sequence, and it also employs context-sensitive rules that call for intermediate representations in their own right. It might be argued that if the grammar could also generate the underlying sequence, then it would be able to produce variations on it in one pass (i.e., the underlying form would not need to be stored in memory). I have devised a program that produces tonal chord sequences from scratch in order to test this possibility. In fact, the program vindicates Steedman's analysis.

The program takes three stages to generate a chord sequence. Stage 1 employs a grammar containing such rules as

TWO-BARS → | I | Vd |

to generate an *underlying* chord sequence. Many of the rules contain more than one possible expansion, and the program makes a random choice in such cases.

There are many possible variations on the underlying sequence, I Vd, including the following three sequences:

Imj7	VIm7		IIm7		V7		
Imj7	bIII7		bVImj7		bII7		
Imj7	IVm7	bVII7	bIIIm7	bVI7	IIm7	V7	

Stage 2 of the program uses context-sensitive rules similar to those of Steedman to interpolate chords into the underlying sequence according to the "cycle of fifths," one of the major dimensions of tonal space (Longuet-Higgins, 1979). Hence, given as input the sequence

| I | Vd |

the program, on detecting a chord containing the symbol d, can transform it into a seventh (symbolized as "7") and insert a previous chord that is related to it by the cycle of fifths:

| I | II7 V7 |

A further step of the same sort can be used to insert a chord in front of II7:

| I | VIm7 II7 V7 |

By working backward in this way – a procedure that requires a considerable memory for intermediate results – the final result of stage 2, depending on the choices of expansion, might be

| Imj VIIm7 III7 | VIm7 II7 V7 |

Stage 3 employs a further set of context-sensitive rules in order to substitute one sort of chord for another and to make another sort of interpolation. Given the previous string as input, and depending on its choices, it can produce

| Imj7 IVm7 bVII7 | bIIIm7 bVI7 IIm7 V7 |

which is a sequence used by the late Thelonious Monk for one of his compositions.

One of the main features of this program, like the program for base lines, is that even though its principles are well understood, it is impossible to predict its output on any particular occasion. The program is not intended to model directly the mental processes of musicians devising chord sequences. It shows that whatever these processes may be, they call for a considerable degree of computational power, certainly more than is available to the rules implicit in music theory. It would be impossible to capture the superficial variations on underlying forms without exploiting some such degree of power. Likewise, the interpolations must be made one at a time and require a record of the previous state of the chord sequence. Musicians who devise new chord sequences must have access to principles that resemble those embodied in the program. However, they do not have to make up chord sequences in real time. They can write them down and work on them in the same way as composers. There is thus no reason for them to use their memories in the same way as the program; the written chord sequence is available for consultation. Musical literacy has ensured that computational power exerts no psychological price.

The procedure exemplifies a creative computational architecture that depends on several stages. These stages divide the criteria of the genre into those that guide an initial generative process and those that are used to modify its results – to select them, or not, as calling for further work. The generative stage is clearly nondeterministic (i.e., arbitrary choices must be made from among the options defined by the grammar), and the subsequent stages may also be nondeterministic.

The creation of novels, paintings, and other works of art typically is carried out within the conventions of an existing genre. Likewise, the creative processes of a scientist normally occur, as Kuhn (1970) has argued, within the constraints of an existing paradigm. These types of creativity invariably depend on a multistage procedure, though probably of a more complex form than the one I have exemplified. The creator generates ideas making use of some initial constraints, but other constraints may be spread over many stages, or unsatisfactory products may be fed back from an evaluative stage to a generative stage for modification. The reason for this division of labor, as I argued earlier, would be that the generative process has no access to the evaluative criteria. In the case of scientific hypotheses, a major evaluative constraint is a set of empirical observations.

Creation of new frameworks

The invention of a new genre or paradigm is ranked above all other forms of creation, but such events are rare. There appear to be no common principles that account for such transitions within a field – from one genre of art to another, or from one scientific paradigm to another. Moreover, the ultimate success of a major innovation depends on events of which the individual creators (and everyone else) is entirely ignorant. For artistic revolutions, these criteria include both socioeconomic factors and other contemporary developments in the world of the arts; for scientific revolutions, they include the subsequent availability of resources to explore the innovation and, of course, its success in making sense of empirical results – a factor that cannot be foreseen at the time of the innovation. It follows that there are no general criteria or principles that underlie all and only the successful major transitions in a particular domain of art or science.

If there are no principles that govern all major innovations in a domain, then there can be no neo-Lamarckian procedure for this type of creativity. Suppose that there were such a procedure for painting. Given as input the principles of early art, it would have to produce a set of viable alternative developments including the principles of perspective; and given the principles of mid-nineteenth-century painting, it would have to produce those of Cubism among a set of other viable alternatives. Likewise, a neo-Lamarckian procedure for physics would have to produce Newtonian physics from its precursors, and then follow that up with special relativity. Such tasks are obviously impossible if, as I claim, these various revolutionary transitions cannot be accounted for by any common set of principles for the relevant domain.

Because the creation of new genres and paradigms is so difficult, it might depend on an essentially arbitrary or random generative process. New species evolve as a result of the random shuffling of genes, followed by the constraints of natural selection, which eliminate organisms that are not viable. A possible architecture for creativity has the same "neo-Darwinian" design. Its first stage consists of a procedure that combines elements at random to generate a potentially vast number of putative products, and its second stage uses a set of constraints to filter out the products that are not viable. There is a long tradition of such mechanisms for creativity. Some proposals have been satirical, such as Mozart's scheme for musical composition by the shake of a die (O'Beirne, 1971), Swift's machine in the academy of Lagado in *Gulliver's Travels,* and Orwell's machine for writing cheap novels in *1984*. Other proposals are serious, and indeed the generation of ideas at random has been proposed by several authors as the only possible creative process (Bateson, 1979; Campbell, 1960; Skinner, 1953). Yet, shooting first (at random) and asking questions later (in terms of criteria) has a patent and inescapable difficulty. It is grossly inefficient. Most of the products of an arbitrary assembly of elements will not be viable. This point was discovered the hard way in the mid-1960s when computer scientists attempted to build intelligent programs by assembling them at random from simple components (Fogel, Owens, & Walsh, 1966).

The evolution of species is slow, even though there are millions of organisms engaged in the random shuffling of genes. Moreover, species do not evolve in a single step. It was not necessary, for instance, that the complete genetic specification of *Homo sapiens* be assembled as the result of a single random shuffling from an unorganized set of genes – an evolutionary step that is singularly unlikely to have occurred by chance. New species derive from existing species, and complex organisms have gradually evolved by common descent from less complex organisms. Hence, the nature of the elements that are randomly combined can improve the efficiency of the procedure. They should not be simple conceptual atoms, but rather existing sets of interrelated ideas.

Despite its inefficiency, a neo-Darwinian procedure is the only mechanism available if there is no way in which the generative process can be guided by selective constraints. This condition is the basis of modern evolutionary theory, but there is no reason to suppose that it applies to the production of revolutionary ideas. It is more likely that this generative process is guided by some criteria in a multistage procedure. The productive use of knowledge is a central part of genius. Although there are numerous potential heuristics, such as the search for revealing analogies, there are unlikely to be sufficient criteria to yield a tractable computational algorithm for producing successful innovations (Johnson-Laird, in press-b).

Conclusions

If creative processes are computable – and there are, as yet, no grounds for abandoning this hypothesis – then creativity can be defined in the light of the theory that has been advanced in this chapter. Creation yields products with three characteristic properties:

1. They are novel for the individual who creates them.
2. They reflect the individual's freedom of choice and accordingly are not constructed by rote or calculation, but by a nondeterministic process.
3. The choice is made from among options that are specified by criteria.

Although I have not labored the point, there are only three general classes of procedures that meet this definition: a neo-Lamarckian process in which the criteria are used to generate possible products, and an arbitrary choice is made from among them; a multistage process in which criteria are used to generate a work and to modify it, with the possibility of arbitrary choices in either stage; and a neo-Darwinian process in which a wholly arbitrary generation of ideas is followed by selection in terms of criteria. As a source of innovation, however, this last procedure is more likely to be used by nature than by human beings.

References

Bateson, G. (1979). *Mind and nature*. London: Wildwood House.

Borges, J. L. (1970). Pierre Menard, author of the *Quixote*. In D. A. Yates and J. E. Irby (Eds.), *Labyrinths: Selected stories and other writings*. Harmondsworth, Middlesex: Penguin.

Campbell, D. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review, 67,* 380–400.

Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting.* Cambridge, MA: MIT Press.

Dostoyevsky, F. (1972). *Notes from underground.* Harmondsworth, Middlesex: Penguin. (Original work published 1864.)

Eigen, M. & Winkler, R. (1983). *Laws of the game: How the principles of nature govern chance.* Harmondsworth, Middlesex: Penguin.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology.* Cambridge, MA: MIT Press.

Fogel, L., Owens, A., & Walsh, M. (1966). *Artificial intelligence through simulated evolution.* New York: Wiley.

Forte, A. (1979). *Tonal harmony, in concept and practice* (3rd ed.). New York: Holt, Rinehart, & Winston.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge University Press.

Johnson-Laird, P. N. (1987). Reasoning, imagining, and creating. *Bulletin of the British Psychological Society, 40,* 121–129.

Johnson-Laird, P. N. (in press-a). The development of reasoning ability. In G. Butterworth and P. E. Bryant (Eds.), *Proceedings of Stirling Conference on Human Development.* Cambridge University Press.

Johnson-Laird, P. N. (in press-b). Analogies and the exercise of creativity. In A. Ortony (Ed.), *Proceedings of the Illinois Workshop on Analogies.*

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). University of Chicago Press.

Levitt, D. A. (1981). *A melody description system for jazz improvization.* Unpublished MSc thesis, Department of Electrical Engineering and Computer Science, MIT. Cambridge, MA.

Longuet-Higgins, H. C. (1979). The perception of music. *Proceedings of the Royal Society, Series B, 205,* 307–322.

Marr, D. (1982). *Vision: A computational investigation in the human representation of visual information.* San Francisco: Freeman.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

O'Beirne, T. H. (1971). From Mozart to the bagpipe, with a small computer. *Institute of Mathematics and Its Applications, 7,* 11–16.

Perkins, D. N. (1981). *The mind's best work.* Cambridge, MA: Harvard University Press.

Rameau, J. P. (1971). *Treatise on harmony.* New York: Dover. (Original work published 1722.)

Reber, A. S. (1985). *The Penguin dictionary of psychology.* Harmondsworth, Middlesex: Penguin.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations.* Cambridge, MA: MIT Press.

Simonton, D. K. (1984). *Genius, creativity, and leadership: Historiometric inquiries.* Cambridge, MA: Harvard University Press.

Skinner, B. F. (1953). *Science and human behavior.* New York: Macmillan.

Sonneck, O. G. (Ed.). (1967). *Beethoven: Impressions by his contemporaries.* New York: Dover.

Steedman, M. J. (1982). A generative grammar for jazz chord sequences. *Music Perception, 2,* 52–77.

Sudnow, D. (1978). *Ways of the hand.* London: Routledge & Kegan Paul.

Ulrich, J. W. (1977). The analysis and synthesis of jazz by computer. In *Fifth International Joint Conference on Artificial Intelligence* (pp. 865–872).

Wallas, G. (1926). *The art of thought.* London: Cape.

The nature of creativity

Contemporary psychological perspectives

Edited by
ROBERT J. STERNBERG
Department of Psychology
Yale University

The right of the University of Cambridge to print and sell all manner of books was granted by Henry VIII in 1534. The University has printed and published continuously since 1584.

CAMBRIDGE UNIVERSITY PRESS
Cambridge
New York New Rochelle Melbourne Sydney

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1988

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

The nature of creativity.

Includes indexes.

1. Creative ability. I. Sternberg, Robert J.
[DNLM: 1. Creativeness. 2. Psychology. BF 408 N285]
BF408.N354 1988 153.3′5 87–20928

British Library Cataloguing in Publication Data

The Nature of creativity: contemporary
psychological perspectives.

1. Creative ability – Psychological aspects
I. Sternberg, Robert J.
153.3′5 BF410

ISBN 0 521 33036 X hard covers
ISBN 0 521 33892 1 paperback