Gödel's metaphor

MICHAEL R. JACKSON

Between the years 1910 and 1913, Alfred North Whitehead and Bertrand Russell published their monumental study, *Principia Mathematica*. In this work, Whitehead and Russell settled a classical philosophical dispute by rigorously demonstrating that all the truths of traditional mathematics are derivable from a few axioms of formal logic. Of more general significance, however, *Principia Mathematica* also systematically exhibited, for the first time, the technical apparatus of modern symbolic logic. The development of this apparatus greatly encouraged certain positivist philosophers who wished to formulate a 'scientific method' for addressing philosophical issues. These philosophers hoped that with the help of symbolic logic all human knowledge might eventually be analyzed and organized into a vast reductive system similar to that of *Principia Mathematica* and consistent with the results of scientific investigation.

In 1931, however, Kurt Gödel published a paper that struck a mortal blow against this program. In essence, Gödel showed that it is possible to construct a true mathematical formula that is not derivable from the axioms of *Principia Mathematica*, nor from those of any such deductive system. Moreover, this formula, which may be characterized as a 'hole' in the tapestry of Whitehead and Russell's work, is just one of a demonstrably infinite class of such formulas distributed throughout vast regions of mathematics in an essentially unpredictable way. Furthermore, the peculiar but inexorable line of reasoning that Gödel employed in the construction and proof of this formula — a line of reasoning that is now usually referred to generically as Gödel's Theorem — is generalizable to all deductive systems that exceed a certain degree of complexity. That is, Gödel's Theorem shows that any deductive system equivalent to or semantically richer than formal mathematics must inevitably generate truths that are not provable within that system.

The consequences of Gödel's Theorem are now widely understood and accepted. Few philosophers, if any, still harbor the vision of a comprehensive reduction of all human knowledge, and this change in consciousness

Semiotica 98–1/2 (1994), 5–48

0037–1998/94/0098–0005

is directly related to the negative implications of Gödel's proof. In this paper, however, I wish to focus upon *positive* implications of the theorem that have received less attention in the literature. Specifically, I will examine certain unusual properties of the prototypic underivable formula that Gödel constructed and what these properties suggest about the structure of ordinary, though semantically complex, types of meaning. I will begin with a discussion of symbolic logic and how Gödel used this tool to construct his prototypic formula. In so doing, I will identify components of the formula that embody, in logical form, two crucial semiotic mechanisms which I will designate coding and encoded self-insertion, respectively. In the remainder of the paper I will argue that these mechanisms underlie the construction of semantically rich texts and account for many of their peculiar properties, including the encoding of one type of meaning within another, the operation of self-reference and self-allusion, and the generation of metaphoric structures. In addition, I will make some comments about what this suggests about the dynamic process by which meaning is constructed.

The language of logic

Before looking at Gödel's formula, it will be useful to say something about the language in which it is written: symbolic logic. Symbolic logic consists of a set of signs, representing fundamental elements and relationships, and a set of rules for manipulating these signs in deductively valid ways. The elements and relationships represented by the signs can be chosen from a wide range of possible content areas, including those of ordinary experience. In Gödel's paper, however, the apparatus of symbolic logic is applied solely to the mathematics of the natural numbers (0, 1, 2, 3, etc.). That is, the elements and relationships designated by this apparatus are simply natural numbers (or variables for natural numbers) and mathematical operations performed upon these numbers — along with a handful of basic logical relations like 'and' (signified by '.'), 'not' (signified by '$\sim$'), 'or' (signified by '$\vee$'), 'if-then' (signified by '$\supset$'), and the two logical quantifiers 'all' and 'some' (signified, respectively, by '$\forall$' and '$\exists$'). Parentheses and commas are also used, for purely syntactical purposes.

Using this apparatus, it is possible to represent and process information about a wide range of mathematical relationships. Thus, the expression

$$\exists x(x=5+3)$$

signifies the assertion 'for some number, x, the equation $x=5+3$ holds' —

or, to put it more colloquially, 'there is a number that is the sum of 5 and 3'. The more complex expression

$$\forall x(\text{Comp}(x) \vee \text{Prime}(x))$$

asserts that every number, x, is either composite or prime.

It should be noted that the predicates 'Comp(x)' and 'Prime(x)' here are actually complex ones that can be defined in terms of simpler relations. For example, 'Comp(x)' ('x is composite') could be replaced by the expression '$\exists y \exists z((y \neq 1).(z \neq 1).(y \times z = x))$', which asserts that there are two numbers, y and z, neither of which is equal to 1 and whose product is x. (Likewise, Prime(x) could be replaced by the same expression preceded by a '$\sim$'.) In a similar manner, other complex mathematical predicates can be reduced to defining expressions composed entirely of simple operations upon natural numbers. The more complex the predicate, the longer its defining expression will be — but every defining expression, no matter how complicated, will always consist, ultimately, of a small number of basic signs representing fundamental mathematical operations upon specific natural numbers.

Moreover, the defining expressions themselves can be further reduced. It is possible to replace not only the fundamental operations of mathematics but also the natural numbers with more basic defining expressions consisting of more basic signs representing purely logical relationships. This is precisely the insight of *Principia Mathematica* — that the entirety of ordinary mathematics is completely derivable from a handful of logical relations. As a matter of fact, only a very small number of such 'primitive' logical relations is required, since many fundamental notions of logic are reducible to each other.

The reducibility of mathematical and logical relationships illustrates an important property of symbolic logic as a system of signs — namely, that it functions not only as a language but also as a *calculus*. What I mean by 'calculus' is that true expressions of symbolic logic can be changed into other true expressions by the mechanical application of certain inferential rules. In fact, we need give no consideration to meaning at all as long as we string the basic signs together according to certain syntactical rules, determine whether these strings are 'true', and, if so, transform them inferentially into more derived expressions. This conception of logic as a kind of mechanical calculus is really nothing more than a formalization of a classical idea that finds its prototype in geometry — the idea of a purely deductive 'formal system' that begins with a set of axioms or postulates and proceeds to the systematic deduction of more and more complicated and derived truths.

Symbolic logic, however, differs from geometry in that it is much more general. It can be used to construct a wide variety of deductive systems (including geometry); and in both *Principia Mathematica* and Gödel's paper it is employed to analyze the formal system that we know as ordinary mathematics or number theory. The fundamental elements of this system are the constellations of form that we call numbers. As we have already seen, symbolic logic is applied to these elements to produce statements of formal mathematics like '$\exists x(x=5+3)$', which we ordinarily abbreviate as arithmetical or algebraic calculations. It should also be noted, however, that such statements can themselves become the elements of higher order or *meta*mathematical statements — for example, the statement '"$\exists x(x=5+3)$" is provable'. Furthermore, these higher order statements can *also* be represented in the notation of symbolic logic and mechanically processed by the rules of its calculus. It is perfectly proper to do this kind of thing, as long as we remain absolutely clear about which signs of the calculus refer to which levels of logical abstraction or 'type levels'. Thus, in Gödel's paper, numbers are regarded as constructions involving the lowest logical type level; these constructions are then incorporated into the higher type level statements of number theory or formal mathematics, which, in turn, become elements that are addressed by the still higher type level statements of metamathematics, and so forth.[1] In this manner, symbolic logic enables us to illuminate some very intricate patterns of embedded form, and it was this kind of power that encouraged those who developed symbolic logic in the belief that it might be extended to a comprehensive systematization of all knowledge.

As already noted, however, Gödel's Theorem destroyed this hope. Ironically, it did so by working within the framework of symbolic logic rather than outside of it and by exploiting the power of the calculus rather than its limitations. Gödel's approach was to develop the calculus in a new way that demonstrated the existence of subtle reverberations among the levels of logical form described above. These reverberations, which Hofstadter (1979) has referred to as 'strange loops', consist of relations of entailment between constructions at various type levels endowing them with multiple but related meanings and disclosing the existence of extraordinary mathematical truths *expressible* in the notation of symbolic logic but not able to be assimilated to the mechanical functioning of its inferential rules.

Gödel numbering

In order to explain how Gödel demonstrated this, it is necessary to describe a major innovation that he introduced: the technique that is now known as *Gödel numbering*. Roughly speaking, Gödel numbering is a form of *coding* in which an expression of one type level is transformed into an expression of another type level. In the prototypic case, an expression of formal mathematics (i.e. a formula) is transformed into a single number.

The rules governing such a transformation are actually chosen somewhat arbitrarily. For the purposes of this paper, I will employ a set of rules adapted from Nagel and Newman (1958) that stays rather close to Gödel's original scheme. This set of rules begins by assigning a number (known as a Gödel number) to each of ten signs of symbolic logic. These ten signs may be regarded as 'primitive', for all the concepts of ordinary mathematics are derivable from them (see Table 1).[2]

In addition to these signs, which can be said to designate *logical constants*, signs designating *logical variables* can also be given Gödel numbers in the following manner: variables for natural numbers (signified by 'x', 'y', 'z', etc.) are assigned specific prime numbers greater than 10 as Gödel numbers (11, 13, 17, etc.); variables for properties of natural numbers (signified by 'X', 'Y', 'Z', etc.) are assigned the squares of prime numbers greater than 10 as Gödel numbers (11^2, 13^2, 17^2, etc.); variables for properties of properties of natural numbers (signified by '*X*', '*Y*', '*Z*', etc.) are assigned the cubes of prime numbers greater than 10 as Gödel numbers (11^3, 13^3, 17^3, etc.), and so forth.

So far we have presented a coding scheme for transforming the basic signs of symbolic logic into numbers. However, it is also necessary to code information about the *order* in which these signs are arranged. To accomplish this, the position of each sign in a given expression is represented by a successive prime number (2, 3, 5, 7, 11, etc.) which is then raised to the power of the Gödel number of the sign that occupies that

Table 1. *Gödel numbers of primitive signs*

Sign (and its meaning)		Gödel number
~	(not)	1
∨	(or)	2
⊃	(if/then)	3
∀	(all or every)	4
=	(equals)	5
0	(zero)	6
s	(is the successor of)	7
(	(punctuation mark)	8
)	(punctuation mark)	9
,	(punctuation mark)	10

position. The product of the entire series can be defined as the Gödel number of the entire expression. To give a simple example, the formula '0=0' is transformed by these rules first into the numbers 6, 5 and 6 (the Gödel numbers of its constituent signs) and then into the single number $2^6 \times 3^5 \times 5^6$ or 243,000,000, which represents the whole expression. The more complex formula '∀x∼(x=sx)' (which asserts that no number is its own successor) becomes first 4, 11, 1, 8, 11, 5, 7, 11, 9, and then $2^4 \times 3^{11} \times 5^1 \times 7^8 \times 11^{11} \times 13^5 \times 17^7 \times 19^{11} \times 23^9$ — producing a Gödel number that is far larger. (We need not actually calculate this number, which is 66 digits long.)

By a similar line of reasoning, we can also determine the Gödel numbers of expressions other than formulas, as long as these expressions can be put in terms of the primitive constants and variables of symbolic logic. For example, the elements around which formulas are constructed — ordinary natural numbers — can themselves be given Gödel numbers, since they can be defined solely in terms of two of the ten logical constants which were given above: zero (0) and the successor function (s). That is, the number 1 can be defined as the successor of zero, signified by 's0', and hence can be given the Gödel number $2^7 \times 3^6$ (=93,312), the number 2 can be defined as the successor of the successor of zero, signified by 'ss0', and hence given the Gödel number $2^7 \times 3^7 \times 5^6$ (=4,374,000,000); and so forth. By adjusting the base numbers of these sequences, we can incorporate them into the Gödel numbers that represent various formulas. For example, the formula 'x=1' (i.e. 'x=s0') becomes $2^{11} \times 3^5 \times 5^7 \times 7^6$.

With one additional convention, we can also define the Gödel number of an *entire series* of formulas, such as a mathematical proof. In such a proof, we have a series of steps, each consisting of a formula that was derived from the formula of a preceding step by the application of an inferential rule. If the Gödel number of the first formula (or step) is 'a', and the Gödel number of the second formula (or step) is 'b', and so on, up to the Gödel number of the conclusion — let us say it is 'e' — then the Gödel number of the entire argument that constitutes the proof can be defined as $2^a \times 3^b \times 5^c \times 7^d \times 11^e$.

Every well-formed expression of formal mathematics, therefore — be it a component of a formula representing a number, a whole formula representing an assertion about numbers, or a series of formulas representing an entire proof — can be assigned its own unique Gödel number. This Gödel number will encode, within its factorial composition, the exact structure of the expression from which it was derived. Furthermore, the nature of this encoding is such that it involves no loss of information. That is, the expression from which the Gödel number was derived can always be completely recovered by a reciprocal *decoding* process of factoring and restoration of the original signs. For example, 243,000,000 can be broken back down into its (unique) fundamental factorial structure, $2^6 \times 3^5 \times 5^6$, which, in turn, can be translated back into the formula that was originally encoded: '$0=0$'. Likewise, the Gödel number of a series of formulas can also be decoded, but this will require a two-stage procedure — first, to determine the number and order of the formulas in the series by deriving a series of exponents affixed to successive prime numbers (like $2^a \times 3^b \times 5^c \times 7^d \times 11^e$ in the above example), and then to determine what each formula is by decoding each exponent.

Because Gödel numbering involves no loss of information, it can be formally treated as a logical operation. This operation, which may be called the 'Gödel transform', is not a primitive one, since it is a coding function defined in terms of substitutions and rearrangements of primitive signs. However, we can signify it, as we can signify any complex operation, with a simple sign — say the letter 'G'. Using this notation, we can formally represent the encoding process with expressions like

$$G_1(2) = 4{,}374{,}000{,}000$$

and

$$G_2(\forall x \sim (x = sx)) = 2^4 \times 3^{11} \times 5^1 \times 7^8 \times 11^{11} \times 13^5 \times 17^7 \times 19^{11} \times 23^9.$$

I have added the subscripts '1' and '2' to the 'G's' here because the Gödel transforms in these two expressions operate at different logical type levels. That is, the Gödel transform in the first expression operates on *numbers* and is therefore equivalent to an ordinary mathematical operation (regarded here as type 1), whereas the Gödel transform in the second expression operates on *mathematical statements or operations themselves* and is therefore equivalent to a *meta*mathematical operation of a higher logical type (regarded here as type 2).[3]

We can also treat the *decoding* process as a logical operation and represent it with a simple sign. Since the decoding process is the reciprocal of the encoding process, it can be represented by the sign '1/G'. Thus we have

$$1/G_1(93{,}312) = 1$$

and

$$1/G_2(243{,}000{,}000) = (0 = 0).$$

Using this notation, we can summarize certain features of the coding process. For example, as noted above, every natural number can be assigned its own Gödel number. Since Gödel numbers are themselves natural numbers, they too can be assigned Gödel numbers, as can Gödel numbers of Gödel numbers, and so forth. Such successively encoded Gödel numbers quickly become astronomical — but we can easily represent them with expressions like '$G_1(G_1(G_1(2)))$'. By the same token, we can represent the successive *decoding* of such astronomical Gödel numbers by expressions of the form '$1/G_1(1/G_1(1/G_1(\ldots)))$'. We can also represent alternating cycles of encoding and decoding by expressions of the form '$G_1(1/G_1(\ldots))$', '$G_2(1/G_2(\ldots))$', etc. Note that such alternating expressions can be reduced, since 'G' and '1/G' represent reciprocal processes that ultimately cancel each other out. Hence, $1/G_1(G_1(2))=2$, $G_2(1/G_2(93{,}312)=93{,}312$, and so forth.[4]

The Gödel Formula

We can now turn to Gödel's Theorem. This theorem states that within any minimally rich calculus of symbolic logic it is possible to construct true expressions that are not derivable from the operations of the calculus itself. As noted above, Gödel demonstrated this by actually constructing such a formula and showing by a peculiar but compelling line of reasoning that that formula must be both nonderivable and true. In addition to this, I have noted that Gödel's nonderivable formula has a variety of semantically interesting properties that can illuminate the nature and functioning of semantically rich texts. It will be useful, therefore, to present this formula now and to discuss it in some detail. Gödel's construction may be represented by the following expression, which I will call Formula G:

$$\forall x \sim \mathrm{Dem}(x,\mathrm{sub}(n,13,G_1(n)))$$

In considering this formula, it will be helpful to begin by emphasizing that it represents a statement about numbers, no different, in principle, from other statements of number theory — such as '$\exists x(x=5+3)$' or '$\forall x(\mathrm{Comp}(x) \vee \mathrm{Prime}(x))$'. On the other hand, it represents numerical relationships that are very complex. To be more specific, the formula asserts that every number, x, does *not* stand in a particular arithmetical relationship (represented here by 'Dem') to a specific number, represented by the expression '$\mathrm{sub}(n,13,G_1(n))$'. In other words, the complex term '$\mathrm{sub}(n,13,G_1(n))$' designates a single number that is constructed in a particular and rather peculiar way (which will be explained below); and the formula as a whole states that there is no number that has the Dem relationship to this particular, peculiarly constructed number.

In order to explain what the Dem relationship is, it is necessary to bring in some additional information. All numbers that are specified in the formula, including the complex number sub(n,13,G_1(n)), are Gödel numbers. Thus, each one represents, *in code*, some expression of symbolic logic (be it a part of a formula, a whole formula or a series of formulas). To be more specific, the complexly constructed number sub(n,13,G_1(n)) represents, in code, a single complexly constructed *formula* of number theory. Again, the specifics of this will be elaborated below, but it is essential to keep in mind that the number sub(n,13,G_1(n)) represents (in code) a mathematical formula, for it is this fact that has determined the way in which the Dem relation has been constructed. The actual mathematics of the Dem relation are extremely complex, although they can be reduced, in principle, to a large number of elementary mathematical computations. We can see through much of this complexity, however, when we learn that the Dem relationship is precisely the one that holds between the Gödel number of a formal deductive proof of number theory and the Gödel number of its conclusion. (In the example that was given above, these were the numbers $2^a \times 3^b \times 5^c \times 7^d \times 11^e$ and e, respectively.) Once this fact is comprehended, it should become clear that the Dem relationship mathematically encodes the logical structure of formal deductive proof just as Gödel numbers mathematically encode the logical structures of specific formulas. This is why it is named 'Dem' — to remind us that although it is a purely mathematical relationship, it represents *in code* a formal demonstration of deductive logic.

We can now further clarify the meaning of the Gödel construction. It asserts that no number stands in the Dem relation to the number sub(n,13,G_1(n)). Again, what is asserted is a purely mathematical fact; but given what we now know about the Dem relation and its connection with Gödel numbering, we can see that the formula also asserts, *in code*, a *meta*mathematical fact: namely, that the formula that corresponds to the Gödel number sub(n,13,G_1(n)) is not the result of any proof sequence, i.e., that it cannot be formally proven by the mechanical calculus of symbolic logic, even in principle. Furthermore, this mathematical assertion acquires considerable significance when we consider an additional piece of information: not only is the number sub(n,13,G_1(n)) the Gödel number of a particular and peculiarly constructed formula; it is the Gödel number of Formula G itself, the very formula in which it is embedded. To put it another way, *the Gödel construction asserts, in code, its own unprovability*.

In order to show how this works, we must discuss the meaning of the expression 'sub(n,13,G_1(n))', beginning with the term 'sub.' Like the Dem relation, the sub relation, considered in itself, is strictly arithmetical. Unlike the Dem relation, however, the sub relation can be characterized as a *triadic operation*, which means that it operates on three numbers, as input, and produces a fourth number, as output. The three input numbers in the present case are signified by 'n', '13' and '$G_1(n)$' (with the letter 'n' standing for a natural number that has yet to be specified). All three of these are Gödel numbers, and their encoded meanings will be discussed below. For the time being, however, we will consider only their mathematical properties as they are processed by the sub operation in the following manner: the first number, n, is decomposed into its fundamental factorial structure; the second number, 13, is then matched against all exponents of all factors making up that structure; and wherever a match is found, the factor that has the matching exponent is replaced by the fundamental factorial structure of the third number, $G_1(n)$. The bases of the resulting factorial structure are then replaced by consecutive prime numbers to create the fundamental factorial structure of a new number, the output number or 'result' of the operation sub(n,13,G_1(n)).[5]

Just as the Dem relation has a *meta*mathematical meaning, so, too, does the sub relation; and once again, this meaning is determined by the meanings of the numbers upon which it operates. The number 'n' is the Gödel number of a formula that has not yet been identified; the number 13 is the Gödel number of the variable 'y' — a variable which may be found in that formula; and the number $G_1(n)$ is the Gödel number of the Gödel number of that same formula. Therefore, the *meta*mathematical meaning that corresponds to the purely arithmetical meaning of the sub operation given above may be characterized as the insertion of the Gödel number of a formula into the place formerly occupied by a variable ('y') within that formula — or, to put it another way, *encoding and inserting a formula into itself*. This operation of *encoded self-insertion* constitutes the semantic core of the Gödel construction and is the source of many of its peculiar properties. But more than this, I will argue below that it represents a special case of a much more fundamental semantic form, one that lies at the heart of many kinds of text construction and leaves its stamp on a wide range of semiotic phenomena. At present, however, let us continue with the analysis of Formula G.

It is still necessary to determine the number signified by 'n', or — what amounts to the same thing metamathematically — the formula (with the Gödel number 'n') whose encoded self-insertion will produce the Gödel construction. It should be noted that while this formula is not itself Formula G, it will bear a close relationship to Formula G. Specifically, it must be a *preliminary* version of Formula G, a kind of schema from which Formula G can be derived by the operation of encoded self-insertion. The determination of this preliminary formula turns out to be straightforward once its defining properties are fully understood: the formula must be one which contains one or more instances of the variable 'y' and which will become Formula G when each of these instances of 'y' is replaced by the formula's own Gödel number, 'n'. There is only one formula that meets these requirements — namely that which I will call Formula F:

$$\forall x \sim \mathrm{Dem}(x,\mathrm{sub}(y,13,G_1(y))).$$

In constructing this formula we have also determined how to derive the number 'n', for, as noted above, 'n' is the Gödel number of Formula F, by definition. In fact, from a strictly arithmetical point of view, the only reason we had for constructing Formula F was to determine what 'n' would be. Now that we have done so, however, we are in a better position to appreciate the larger meaning of Formula G. Considered in itself, Formula G asserts a purely mathematical fact — specifically, that the number produced by the (arithmetical) operation $\mathrm{sub}(n,13,G_1(n))$ does not bear the (arithmetical) Dem relation to any other number. But as we have now seen, there is also a second meaning to Formula G, one which can be extracted from it by decoding the numbers n, 13 and $G_1(n)$ and inferring the metamathematical counterparts of the Dem and sub relations. This second meaning is the assertion that a certain formula is not demonstrable within the calculus of symbolic logic and that that formula is the one which is obtained by the encoded self-insertion of Formula F — in other words, that that formula is Formula G itself.[6]

In summary, therefore, the Gödel construction asserts its own unprovability. It does so, however, in a complex way in which one meaning encodes another and in which the Gödel formula refers to itself not by a direct process of naming or description but rather by an indirect process of what might be called historical-recursive reconstruction. That is, the formula designates itself, via the expression '$\mathrm{sub}(n,13,G_1(n))$', by describing indirectly and in code the recursive logic underlying its own construction — i.e., the logic of its own systematic constructability from a preliminary general schema (Formula F) into a final expression (Formula G) by the operation of encoded self-insertion. Without this peculiar property of historical-recursive reconstruction, the self-reference of Formula G would fall into the vicious circle of an infinite regress.

Having constructed Formula G, it was relatively easy for Gödel to go on to evaluate its truth status, to show that the formula is impossible to derive from the axioms of number theory, yet is true. The nonderivability of Formula G follows from the fact that if the formula were derivable there would be an inconsistency between the fact of its derivability and both the arithmetical and metamathematical relations asserted by the formula; and it can be shown that this inconsistency would entail, as a logical consequence, the invalidation of the very deductive apparatus upon which this putative derivation would be based. On the other hand, once we conclude that Formula G is not derivable, we must also conclude that it is true, for the nonderivability of the formula is, in effect, the very fact that the formula asserts. As noted above, the combination of these two properties — nonderivability and truth — gives Formula G a peculiar and very important status in number theory, for the existence of even one nonderivable truth demonstrates the incompleteness of symbolic logic as a calculus. Furthermore, it turns out that this incompleteness goes far beyond the single instance embodied in Formula G, for Gödel's method can be indefinitely extended to construct limitless variations on nonderivable truth and, eventually, a large number of entirely new kinds of mathematical problems. The extension of Gödel's method has therefore led to a revolution in mathematics since 1931 that can be characterized as the entry of that discipline into the postmodern era.

More relevant to the current paper, however, are the underlying reasons for the productivity of Gödel's method, and, in particular, what might be described as *the importation into mathematics of textual semiosis* in the form of mathematically defined operations of self-reference, multivocality and metalinguistic embedding through a process of encoded self-insertion. In the remainder of this paper I will argue that the above components of Gödel's construction — and particularly the operation of encoded self-insertion represented by the expression '$sub(n,13,G_1(n))$' — are merely what might be called the local, mathematical variants of a deep semantic form that underlies an extremely wide range of textual phenomena, including some of those which seem — superficially, at least — to be among the furthest removed from mathematics. In order to show this, it will be necessary to consider first the representation of ordinary meanings within symbolic logic.

Gödel and ordinary semiosis

In discussing the Gödel construction, we have so far limited our focus to purely mathematical issues. As noted at the beginning of this paper, however, the apparatus of symbolic logic has an extremely wide scope of applicability, one that potentially includes not only the formalisms of mathematics but more practical, substantive phenomena as well. In that more substantive use, the constructions of symbolic logic represent the meanings of statements about the empirical world by analyzing them into universal and particular components. The universal components are predicate terms (represented in symbolic logic by capital letters) that stand for simple qualities or complex relations, while the particular components are individual terms (represented in symbolic logic by small letters) that stand for specific objects and events.

Keeping in mind this more substantive use of symbolic logic, let us now examine whether the mathematical operations that Gödel discovered have counterparts in the ordinary semantic realm, or, more correctly, whether both kinds of operations — mathematical and ordinary semantic — may be rooted in more fundamental logical and semiotic relationships. In particular, it is appropriate to ask whether there are ordinary semantic prototypes corresponding to the substitution and coding functions represented above as 'sub' and 'G'. I will argue here that there are such prototypes, that these prototypes are evident in such fundamental notions as sense, reference, framing, multivocality and metalanguage, and that the recursive application of these prototypes leads us into the sphere of rich textual semiosis. I will begin by considering the operation of coding.

I have argued elsewhere that coding is a fundamental semiotic process and that a wide variety of textual phenomena can be analyzed in terms of the coding concept (Jackson 1991). Let us therefore ask whether it is meaningful to talk about a coding function, C, that encodes higher semantic levels into lower ones in a manner analogous to that in which Gödel's function G transforms mathematical formulas into Gödel numbers. In the semantic sphere, this would mean, in the first instance, that formulas of symbolic logic standing for concepts and propositions would be encoded, in some manner, into individual terms standing for objects and events. That is, meanings considered abstractly would be transformed, via this coding function, into particular terms representing those meanings as concrete instances.

In fact, this transformation is already implicitly recognized in the philosophical literature, although its status as a type of coding has, to my knowledge, never been discussed as such. I am referring here to the relationship between *sense* and *reference*. In natural language, this is roughly equivalent to the grammatical transformation by which predicates that are used to classify objects and events (i.e., to endow them with sense) are converted into descriptive phrases or names uniquely specifying such objects and events (i.e., identifying them as particular referents). To give an example, the predicate or concept of being a tree is represented in symbolic logic by the expression 'Tx' (which should be read 'x is a tree'). Technically, this is a propositional function that can be regarded as signifying the category 'tree' in purely abstract terms — i.e., as a unit of sense. It is also possible, however, to designate a specific tree by an individual term such as 't' (which may be glossed here as 'the tree' or 'this tree'), thus identifying the tree as a particular instance — i.e., a particular referent — of the category 'tree'. As indicated above, this relationship between the sense and reference of the word 'tree' can be conceptualized as a form of coding. We can express this fact in the formula

$$C_2(Tx) = t.$$

In this expression, 'C_2' is analogous to the Gödel numbering function 'G_2', 'Tx' is analogous to the mathematical formula that is encoded, and 't' is analogous to the Gödel number that results.[7] We can likewise express the decoding of the referential tree back into its sense by the formula

$$1/C_2(t) = Tx.$$

Moreover, we can extend this concept of coding to a variety of expressions of the calculus. For example, the proposition 'John plants the tree' — expressed in the calculus as 'jPt' — immediately signifies the sense or *idea* of a person named 'John' planting a particular tree. This sentence can be transformed, however, via an application of the coding function, $C_2(jPt)$, into the individual term 'e', which represents the specific *event* to which the sentence refers — that is, the event of John's planting the tree. The same can be said for other expressions of the calculus which signify sense in the form of categories or propositions and which can be encoded into individual terms signifying specific referents.[8]

Let us consider next a coding function that we may designate 'C_1' and that corresponds to G_1. It will be recalled that G_1 encodes one number into another number, and we therefore specify that C_1 will encode one individual term into another such term. Using the tree example, we can specify this transformation as

$$C_1(t) = t'.$$

We are no longer talking about the encoding of sense into reference, for we are *beginning*, here, with a particular thing — the tree — and encoding it into some other particular thing, designated by 't′'. But what might this latter thing be? Since it is a particular object or event that encodes the tree, we may interpret it as a *signifier* of the tree. That is, 't′' may refer to the words 'the tree' written on a piece of paper, or the speech event corresponding to these words, or even to the letter 't' as used above. It should be kept in mind that while 't′' designates a signifier, it designates it as a material object or event, and it should also be kept in mind that the C_1 function can be recursively applied in expressions such as

$$C_1(C_1(t)) = t''$$

and

$$C_1(C_1(C_1(t))) = t'''$$

to designate a signifier of a signifier, a signifier of a signifier of a signifier, etc.[9]

Encoded self-insertion

We are now in a position to consider the notion of encoded self-insertion. In order to discuss this mechanism, we can no longer use a simple example like that of the tree, for the operation of encoded self-insertion applies *only to texts and text-creating events.* As an illustration, we may imagine an artist painting a picture and including within it a picture of the picture that he or she is painting. As a matter of fact, certain avant-garde artists like Magritte have used this form to raise questions about the nature of textual semiosis. The completed painting can be represented by the expression

$$\text{sub}(p,s,C_1(p)).$$

It should be noted that in this case C_1 designates visual coding, which is no different, in principle, from linguistic coding as described above (see Jackson 1991: 42–43); 'p' represents the painting in general terms, and '$C_1(p)$' represents the picture of the painting within the painting. The letter 's' designates the place or *slot* within the painting where the picture of the painting occurs, and the term 'sub' expresses the operation in which the slot is filled in.[10] What this example illustrates in visual coding, we can also illustrate in linguistic coding, and in slightly more complex form in the following sentence:

This sentence has five words.

The meaning of the above sentence can be approximated by the expression

$$F\text{sub}(t,s,C_1(t)),$$

where 'F' represents the property of having five component parts, 't' represents the sentence (without specifying its grammatical subject), 's' represents the syntactical slot of the grammatical subject, '$C_1(t)$' represents, roughly, the phrase 'this sentence' and '$\text{sub}(t,s,C_1(t))$' represents the whole sentence.[11]

It is not difficult to find examples of encoded self-insertion in art and literature, especially in textual genres that are primarily comic or surreal. An example is Mel Brooks's movie *Spaceballs*, in which some of the characters locate a videotape of the very movie they are in and fastforward through it to find out what they must do next. This example is a particularly clear one, for Brooks plays on the culture and technology of the videotape in order to frame the insertion as explicit and complete. But most instances of encoded self-insertion are semantically much more complicated, involving, for example, insertions of partial texts, parallel texts and embodiments of the framing function (such as the author), often featuring some sort of interaction across the textual frame. Cartoons have liberally drawn on these devices since their very earliest days, as illustrated by the classic technique of beginning the story by having the characters jump out of the artist's ink bottle. A more recent example is an episode of *Rocky and Bullwinkle* in which a villain becomes so incensed by the condemnation of the narrator that he climbs out of the frame of the cartoon, grabs the narrator, and drags him back into the cartoon (in animated form), where he ties him up and gags him.

In order to discuss the coding of such texts, it is necessary to consider the logical structure of a certain class of actions. These are actions that involve some sort of *text production*, such as painting, writing, stating or performing. Logically, every such action consists of a complex relationship among things of two kinds. The things of the first kind are people and instruments that are involved in producing the text and those of the second kind are the signifiers of the text itself. Although both kinds of things are material in nature, there is a fundamental difference between the two, for the things of the second kind — the signifiers of the text — are objects or events that involve at least *two* successive layers of coding. The first such layer is the encoding of sense into reference by the C_2 function, and the second is the encoding of reference into signification by the C_1 function. This was illustrated in the example of the tree, above, first by the transformation of the predicate 'Tx' into the individual term 't', and then by the transformation of 't' into 't″', which signifies the name or descriptive phrase that refers to the actual tree. In the following pages, I will often 'unpack' such formal signification by using the expression 'C_1C_2' to show the double coding that goes into it. In this case, the signifier that names the tree — represented here by 't″' — is equivalent to '$C_1C_2(Tx)$'.

Keeping this in mind, we can return to the more complicated examples of textual self-insertion mentioned above. The semantics of these texts resembles that of the simpler encoded self-insertions illustrated by the painting, the sentence and the movie, but there are two additional features: (1) only a part or component of the first term occurs in the third (encoded) term; and (2) this part or component represents some aspect of actual text-production that is not normally encoded as such. For example, the classic cartoon device of Koko the clown jumping out of the artist's ink bottle can be analyzed as

$$sub_2(aiPx,x,C_1C_2(kJi))$$

where 'aiPx' stands for the artist (a) using the ink bottle (i) to produce (P) an unspecified text (x), and 'kJi' stands for Koko (k) jumping (J) out of the ink bottle (i). The self-insertion consists in the occurence of the same element, 'i', in both the first and the third terms; and the disruption of the textual frame resides in the fact that this element, in its uncoded form, is one of those involved in the production of the text itself. Note that the substitution function here differs from those above in that it is of a higher logical type (as indicated by the subscript '2'); this is because it operates not upon an object but rather upon a *concept* — that is, it operates upon the logical structure of the text-productive situation in order to show the transference of content across the textual frame. The *Rocky and Bullwinkle* text can be similarly analyzed as the encoded self-insertion of a text-productive term that stands for the narrator.

We can also use this notation to show another complication of textual semiosis: the creation of multiple layers of texts within texts. This kind of layering may involve repeated self-insertions, or the injection of new content, or various patterns in which the two are interspersed and/or interwoven. Obviously, this can lead to considerable complexity, but the basic technique can be illustrated by a relatively straightforward example from the television sitcom *Murphy Brown*. In this episode, Murphy Brown, a television journalist, is invited to make a guest appearance on a sitcom about a television journalist called *Kelly Green*. After much hesitation, she agrees to do so, moved primarily by an ignoble desire to outshine her journalistic competitors, particularly Connie Chung. The day after making the appearance, however, Brown is chagrined when Chung, played by herself, unexpectedly makes an appearance and chastises Brown for her lack of professionalism in failing to turn down the invitation as she, Chung, had done.

As is so often the case, the humor here is semantically complex, but one aspect of it resides in the playful semiosis of texts within texts — that is, in the fact that Chung plays herself saying of herself that she does not play herself. This situation can be understood as the product of three successive operations of encoded self-insertion. The first of these involves the two innermost textual layers and consists of the noninvolvement of the *character* Chung (c) in the production (P) of the text in which she would meet Kelly Green (cMg):

$$\text{sub}_2(\sim\text{cPx,x,}C_1C_2(\text{cMg})).$$

This formula becomes '$\sim cP(C_1C_2(cMg))$' when the insertion is actually carried out; and we can use this simplified version to represent the first insertion when it becomes the encoded term in the next operation of encoded self-insertion. This next self-insertion represents the character Chung's saying (S) to Brown that she has refused to make the guest appearance:

$$\text{sub}_2(\text{cSy,y,}C_1C_2(\sim cP(C_1C_2(cMg)))).$$

This expression becomes '$cS(C_1C_2(\sim cP(C_1C_2(cMg))))$', which, in turn, serves as the encoded element in the final operation of encoded self-insertion that represents the *real* Chung's guest performance (P) on the *Murphy Brown* show:

$$\text{sub}_2(\text{cPz,z,}C_1C_2(cS(C_1C_2(\sim cP(C_1C_2(cMg)))))).$$

When carried out, this operation reduces to

$$cP(C_1C_2(cS(C_1C_2(\sim cP(C_1C_2(cMg)))))),$$

which shows the entire multilayered situation of Chung playing herself saying of herself that she does not play herself. The three textual frames are represented here by the 'C_1C_2' terms.

Reflection

So far, I have discussed some of the relatively straightforward and explicit ways in which texts can refer to themselves. But actual textual self-reference is much more likely to occur in a variety of covert and subtle ways, requiring the interpreter to take an active role in decoding the mechanisms by which it is effected. For reasons that will be explained below, these more covert forms of self-reference seem to be most readily detectable in texts whose content centers around the intentionality and reflectivity of the human subject. Before considering such texts in detail, therefore, it will be helpful to take a closer look at human self-reflection and some ways in which this phenomenon might be represented and analyzed with the help of the concepts of coding and encoded self-insertion.

Self-reflection occurs when a person constructs some kind of representation of self. This representation can take a number of forms, but we may regard it, roughly, as an element of thought or imagination that in some way encodes an action or trait of the person in whom it occurs. For example, a person might reflect upon himself or herself while he or she is engaged in the act of riding a bicycle. If we represent the state of riding a bicycle by the predicate term 'R' and the person who is engaged in this activity by the individual term 'r', then a person reflecting upon himself or herself while in the act of riding a bicycle can be represented by the encoded self-insertion '$\text{sub}(r,s,C_1(r))$'. In this expression, '$C_1(r)$' represents the rider's self-image or self-concept and 's' represents the functional or mental 'space', as it were, where the self-image or self-concept operates. Incidently, the full predicate term encoded by 'r' may be defined here as '$\text{Rsub}(x,s,C_1(x))$', rather than 'Rx', thus representing the predicate as the state of *reflectively* riding a bicycle. So far, the semantics of this situation resembles that of simple textual self-insertion.

In real life, of course, we can assume that the self-reflection of the human subject occurs in a variety of more complex ways as well. For example, it might involve the encoding of fragmentary, transient or functionally associated elements of the self, presumably shifting among these forms and generating multiple layers of coding. I have already discussed, above, how some of these complications can be analyzed as partial encoded self-insertions and multiple textual frames. In addition, there is another class of complications which I have not yet discussed. These are instances in which reflectivity is repeatedly encompassed *wholesale* within itself, a situation which must therefore be analyzed in terms of *total*, rather than partial, encoded self-insertions. An example of this type of situation would be a person riding a bicycle and reflecting upon the fact that he or she is reflecting upon his or her act of riding a bicycle. Another, somewhat different, example would be one person reflecting upon the fact that another person is reflecting upon his or her own act of, say, looking at the first person. These examples both illustrate the phenomenon of reflection upon reflection itself and, as such, must be represented by encoded self-insertions that are in some sense totalizing. More specifically, it can be shown that the first example can be represented by the formula

$$\text{Rsub}(\text{sub}(a,s,C_1(a)),t,C_1(b))$$

and the second example by the formula

$$\text{sub}(a,s,C_1(a))\text{Lsub}(b,t,C_1(b)),$$

when 'a', 'b', 's' and 't' are suitably defined in each case.[12] The first formula can be said to represent an example of *successive* reflection, and the second an example of *interactive* reflection. It can also be shown that aspects of the above formulas can be combined into still more complex formulas representing still more convoluted forms of reflection with both successive and interactive elements. Thus, the formula

$$sub(sub(a,s,C_1(a)),t,C_1(b))Lsub(c,u,C_1(c))$$

represents one person reflecting upon the fact that another person is reflecting upon the fact that he or she is reflecting upon his or her own action of looking at the first person (again, with suitable definitions for the individual terms).[13]

I will not attempt to enter any further into these complexities, which go well beyond the scope of this paper and would take us, in any case, into the realm of cognitive psychology. I will note, however, that such convoluted forms of reflection do not present problems, in principle, to an analysis in terms of encoded self-insertion. In the pages that follow, therefore, when I discuss reflective subjectivity I will generally use simple schemas of the form $sub(x,s,C_1(x))$, but the reader should understand that the analyses that will be presented also apply, in most cases, to the more complex forms of reflection which I have just illustrated.

Covert self-reference

Let us return now to the question of texts that are *implicitly* or *covertly* self-referential. These texts are difficult to describe in the abstract, but they form a rather clearly recognizable group once they have been identified. They can be characterized, roughly, as texts whose content is covertly embodied or reflected in its own production, and whose production, therefore, points implicitly back to that content. To give an example, several years ago I conducted an interview study in which I asked people to talk about things that had brought them a sense of self-esteem (Jackson 1984). One of the most striking things about the discourse of my subjects was the way in which its form often seemed to reflect and refer back to its own content. For example, one subject, who described self-esteem related to his success as a salesman, seemed to be 'selling' his point of view, while another subject, who described self-esteem related to her experience as a mother, seemed to take a 'mothering' role toward me in the interview — and so forth.

This phenomenon of form reflecting content has been recognized in the psychoanalytic concept of *transference*, in which the patient's behavior toward the analyst comes to resemble the events he or she is describing. However, the phenomenon seems to be more than a mere clinical aberration, for it can be observed in a wide range of other social, political and cultural situations as well. We are all familiar with the entertainer who sings about love seductively, the reporter whose story becomes part of the news event, the performer whose life resembles one of his or her roles, and the author who has a personal relationship with his or her characters. In each case, internal textual content seems to be reflected, somehow, outside the textual frame in the very act by which that content is constructed. Given the widespread occurrence of such texts, we might suspect that their structure results, in part, from some fundamental characteristic of semiosis itself. In fact, a close analysis will show that this is the case. Furthermore, it will show that the logic that underlies this kind of situation is very closely related to that which underlies Gödel's Theorem and the construction of Formula G.

In order to explain how this is so, it will be useful to examine a particular text in some detail. It has often been noted that certain Negro spirituals like *Go Down Moses* have had a double significance in Afro-American culture. Overtly, they celebrated the triumph of the children of Israel over the power of the pharaoh; covertly, however, their performance seemed to pose an implicit challenge to the rule of the white power structure. We might say that the very singing of these ostensibly conservative songs expressed, in code, revolutionary sentiments immanent in their own content.

How might we analyze this kind of situation? Let us begin with a consideration of the overt meaning of such a text — that of the children of Israel rising up against Pharaoh. This overt meaning can be summarized as

$$\mathrm{sub}(c,s,C_1(c))Rp,$$

where '$\mathrm{sub}(c,s,C_1(c))$' stands for the children of Israel, 'p' for Pharaoh, and 'R' for the predicated action of rebelling. Note that I have represented the children of Israel here as a (collectively) *self-reflective* protagonist. By analogy to the definition of the bicycle rider (r) in the previous section, 'c' can be defined here as the encoding of '$\mathrm{sub}(x,s,C_1(x))Rp$'. I will refer to this latter expression as 'Formula F'. The use of this term is not coincidental, for it plays exactly the same role here as Formula F plays in Gödel's Theorem.[14]

Next, the covert meaning of the text can be determined by applying the decoding function $1/C_3$ to the above expression, which produces:

$$\mathrm{sub}_2(F,x,c)R_2(Py).$$

This expression requires a careful analysis. Each term in it is the decoded counterpart of a corresponding term in the first expression, and the entire expression represents a *meta*-idea corresponding to and encoded within the overt meaning of the text represented by the first expression. While that overt meaning concerns a relationship of rebellion (R) between two material subjects — the children of Israel and the pharaoh — the covert meaning represented by the second expression concerns a relationship of *meta*-rebellion (R_2) in which one *idea* challenges another idea. The concept of meta-rebellion may seem rather peculiar at this point, but I will clarify its meaning in the course of the following analysis.

The two ideas that are involved in this relationship can be characterized as the abstract conceptions which the children of Israel and the pharaoh respectively encode — but with some important additional features. The first idea, signified by '$sub_2(F,x,c)$', embodies not only an abstract decoding of the children of Israel (as a self-reflective protagonist) but also a meaning-creating sub_2 function. This function operates upon Formula F by replacing all of the x's in this formula with c's. When this replacement occurs, the result is none other than the original expression representing the overt idea of the text. Thus, the first idea represented in the *covert* meaning of the text is the *overt* meaning of the selfsame text — or more correctly, it is the text-productive ideational activity by which that overt meaning was constructed.[15]

The second idea represented in the covert meaning of the text is the one that is said to be challenged (or more correctly *meta*-challenged) by the first idea. This second idea, signified by 'Py', is the abstract sense of the predicate 'pharaoh' — which we may characterize here as that of a powerful despot who conquers and enslaves other races. In the context of the Negro spiritual this predicate clearly has reference to American white culture (as well as that of ancient Egypt). We can therefore characterize the covert meaning of the text as asserting that the overt meaning — or more correctly, the construction of this overt meaning in the production of the text — poses a challenge to the concept of absolute domination of one race by another, and more specifically to the domination of American blacks by American whites. Thus, the text of the spiritual is self-referential by way of its covertly encoded meaning, which describes both the construction of the text itself and the challenge that the text poses to the dominant culture.

Similar analyses can be applied to all the other covertly self-referential texts that were mentioned above. In each case, the overt meaning or content of the text can be decoded to reveal a second meaning that is concealed, so to speak, in the form of the creation or performance of the text itself and the relationship (or rather meta-relationship) of that form to the semantic environment out of which it emerges. Thus, the singing of the love song bears a meta-seductive relationship to the persona of the listener, which is temporarily substituted for that of the singer's imagined partner; the reporting of the news event has meta-political implications for the roles of those whose lives are defined by publicity and power relations; and so forth. More specifically, each of these texts is covertly self-referential because it fulfills two necessary conditions: first, it portrays at least one character who can be regarded as self-reflective, self-conscious or intentional and who can therefore encode the reflective act by which the text itself is created, performed or interpreted; and second, each text is produced in a material environment that is capable of fitting itself to the semantic forms represented in the overt content of the text and that is, in fact, subtly altered to do so by the activities of those involved in the text production.

How does this occur? Evidently, those who are involved in the production of such texts — be they authors, performers or members of the audience — sense the possibility of fulfilling the above conditions and unconsciously conspire to carry them to fruition. Thus, the performer sings in a sultry manner and the listener is drawn into rapt attention; the journalist takes on a crusading tone and the politician becomes defensive; and so on. On the other hand, the process can also be resisted, as when the listener or the politician attempts to minimize, deny or ignore the force of the communicative situation. Such resistance is difficult, however, for it, too, can become part of the covert meaning of the text, as the listener's resistance is assimilated into the very drama that it opposes. It is the nature of this kind of communication to draw its participants into subtle collusion, for it defines them, willingly or unwillingly, as *signifiers* of the type of issue overtly addressed by the text itself; and when this occurs, the communicative situation becomes the covert metalinguistic exemplar of a meta-form that is, in all its most crucial respects, perfectly isomorphic to the overt content expressed within the frame of the text. This is why I have referred to the structures of the communicative situations cited above as meta-rebellion, meta-seduction, meta-politics, and so on.[16]

Productions like the ones I have been discussing occur in a wide variety of situations because they exploit a fundamental fact of the semiotic situation — namely, that the realities on both sides of the textual frame are, in some sense, ultimately the same reality. Just as the world referred to within the text can be fashioned to mirror the one beyond it, so too the material world of the signifying situation can be systematically rearranged to reflect what is in the text. The self-reflective activity of the human subject is the dynamic principle that mediates the two, and, as such, it tends inexorably to align them with each other. This same dynamic principle — the reflectivity of the human subject — is the source of many of the other peculiar properties of texts, an idea which I will develop further below.

Allegory and metaphor

So far, I have indicated how sense and reference, frames and layers, self-reflection, and various forms of self-allusion can be analyzed in terms of coding and encoded self-insertion. In the following pages I will suggest some ways in which aesthetic devices like allegory, metaphor and connotation might be similarly analyzed. I will begin with the simplest of these devices, allegory.

Although allegory can take a number of forms, it will be useful to distinguish two major types. The first of these is the classic form of allegory, in which abstract concepts, traits or ideals are represented by specific persons, objects and actions in order to make a moral or philosophical point. An example of this kind of allegory is John Bunyan's seventeenth-century work, *Pilgrim's Progress* (Bunyan 1678). In *Pilgrim's Progress*, the journey of a man named Christian is used to represent the spiritual journey of all Christians. Each person who helps or hinders Christian along his way is named for some common human attitude that he or she exhibits — such as Piety, Ignorance or Ill-Will — and each character displays, in his or her actions, a typical human situation and how this situation impacts upon the Christian way of life. This type of allegory seems to involve an encoding of the C_2 type — that is, of sense into reference — in something like pure form, since abstract cultural categories are encoded into specific individuals and actions that exhibit them prototypically.

I will not examine this first type of allegory in much detail, since the coding it involves is relatively simple. However, it will be worthwhile to consider a fragment of such a text to illustrate how it is coded allegorically. In one part of *Pilgrim's Progress* (pp. 43–44), Christian attempts to persuade a character named Sloth to accompany him on his journey; but Sloth, who has been sleeping, wishes to continue doing so and resists Christian's exhortations. As is true throughout the book, the names of the characters invite a straightforward decoding that makes the allegorical meaning of the interaction explicit. Thus, we apply the decoding function $(1/C_3)$ to the expression that represents Sloth's resistance of Christian ('sRc'), obtaining:

$$(Sx)R_2(Cy).$$

This formula expresses the allegorical meaning of the interaction — namely, that the *state* of sloth presents a resistance to the *state* of being Christian. As in the previous analyses, the resistance here is more correctly designated *meta*-resistance; but in this case the higher order designation indicates not only a potential alignment of the communicative situation to the content of the text, but also a kind of Platonism inherent in the concept of allegory itself. That is, the very use of this kind of allegory presumes, in some sense, an order of forms that is both immanent in and isomorphic to the world of ordinary appearances.[17]

The second kind of allegory is equally simple, but rather than involving a movement from abstract to concrete, it seems to involve a displacement from one concrete realm to another. The coding here is not that of concept into object, as was the case in the first kind of allegory, but rather that of object into object and event into event, so that events in one material arena are represented by those associated with a second material arena. I am referring here, particularly, to the kind of allegory involved in political and social satire. An example of this kind of allegory is George Orwell's *Animal Farm*, in which the events of early Soviet history are represented by fictitious events on a farm (Orwell 1954). In this story, specific persons and groups associated with the Bolshevik revolution are represented by specific human and animal characters, whose actions satirically mimic the events of that revolution. Thus, Stalin is represented by a pig named Napoleon, the White Russians by a mare named Mollie, the Moscow trials of 1936–37 by an inquisition among the animals, and so on. As with the first kind of allegory, the coding here seems to be of a particularly simple and straightforward type. However, a closer consideration will show that this type of coding is fundamentally different from all the kinds of coding we have examined so far.

All of the coding functions we have so far considered can be ordered hierarchically, for each of them produces a certain kind of output that serves as input for the next coding function 'down'. Thus, the C_3 function operates on a category of meta-sense and produces a category of sense as output; the C_2 function operates on a category of sense and produces a particular referential object as output; the C_1 function operates on a referential object and produces a formal signifier as output; and so on. This hierarchy can be extended infinitely in both directions and the coding functions that it organizes can be shown to represent certain important operations in formal logic and linguistics.[18] For this reason, I will designate the codes that I have discussed so far — that is, the codes that are organized by this hierarchy — as *formal codes*. In contrast to these, I now want to consider another group of codes, which I will call *transformal codes*.

Transformal codes can be seen in many kinds of texts, but perhaps the simplest case is that of uncomplicated metaphor. In metaphor of this kind, the conceptual contents of one semantic field are systematically converted into those of another. For example, in the metaphor of death as a river, the following set of transformations is effected: living→ traveling; facing death→encountering a river; dying→entering into the river; coldness of death→coldness of the water; unconsciousness→ darkness of the water; and so forth. As Lakoff and Johnson (1980) have pointed out, such sets of transformations — and the correspondences they define — can be elaborated more or less indefinitely.

To put it another way, every transformal code is defined by the juxtaposition of two semantic fields that are ordinarily considered to be independent of each other. In the above example, the relevant fields are those of death and rivers, respectively. The first of these fields can be said to concern the *metaphorical object* and the second the *metaphorical vehicle* through which the first is represented. It will be useful to indicate this relationship in subscripts appended to a coding function to designate a particular transformal code. Thus, the transformal code that represents death as a river can be designated by the coding function 'C_{DR}' and the transformations listed in the previous paragraph can be represented as 'C_{DR}(living) = traveling', 'C_{DR}(facing death) = encountering a river', and so forth. More generally, C_{DR}(death) = river. It should be noted that when two metaphors are logically independent, so are the coding functions that define them. For example, the metaphor that represents death as winter is logically independent of the one that represents death as a river, and its defining function, C_{DW}, is therefore independent of the function C_{DR}. That is, C_{DW}(death) ≠ river; rather, C_{DW}(death) = winter.

It should also be emphasized that transformal codes are independent of the formal codes which I have already discussed. Thus, while C_2 and C_1 operate vertically, so to speak, converting sense into referent and referent into signifier, transformal codes can be said to operate *horizontally*, connecting two areas of semantic content across a given level of abstraction. When I discussed the metaphorical coding of death as a river, for example, I indicated that the abstract concept of death is transformed into the abstract concept of a river. This formulation, however, was actually somewhat arbitrary, and I could equally have said that a *particular* death is transformed into a *particular* river. This kind of transformation can be seen in applied metaphor — in this case, for example, in certain baptismal ceremonies, where the initiate's death is represented by the river into which he or she is immersed. In the notation of symbolic logic, the first formulation (abstract metaphor) can be represented by '$C_{DR}(Dx) = Rx$', and the second formulation (applied metaphor) by '$C_{DR}(d) = r$'. Since transformal coding can occur at any level of abstraction, both formulations are equally valid. As a matter of fact, I have already used the second type of formulation in discussing Orwell's *Animal Farm*, although I did not identify it as such at the time. Given what has been said about transformal coding, however, we can now clarify the nature of the kind of satirical allegory represented by Orwell's story: it is the detailed elaboration of a single metaphor governed by a single transformal code that relates persons and events of the real world to those of a hypothetical social arena. In *Animal Farm*, this transformal code (which can be designated 'C_{SF}') creates the extended metaphor of the plot by systematically converting the developments of Soviet history into those occuring on the fictitious farm.

We can also apply the concept of transformal coding to the production of ordinary metaphor. As an example, we may take T. S. Eliot's (1925) metaphorical designation of the members of his generation as 'hollow men'. In this designation, the semantic medium of hollowness is used to represent the semantic object of valuelessness, and we can therefore assume the operation of a transformal code spanning these two semantic fields and represented by the function 'C_{VH}'. The production of Eliot's metaphor can then be represented by the following partial encoded self-insertion, in which 'V' stands for valuelessness, 'H' for hollowness, 'm' for the men of Eliot's generation, 'e' for Eliot himself, and 'W' for the text-productive act of writing:

$$\mathrm{sub}_2(\mathrm{Vm.eWx,x,}C_1C_2C_{VH}(\mathrm{Vm})).$$

This becomes

$$\mathrm{Vm.eW}(C_1C_2C_{VH}(\mathrm{Vm})),$$

which is equivalent to

$$\mathrm{Vm.eW}(C_1C_2(\mathrm{Hm})),$$

a formula that states that the men of Eliot's generation are valueless and that Eliot writes that they are 'hollow men'.

It should be noted that the term that is self-inserted in the above expression ('Vm') is different in two ways from other self-inserted terms we have considered so far: first, it designates a full-fledged proposition that asserts a fact (which is why it is connected by a dot — signifying the logical 'and' — to the next term, which also designates a proposition asserting a fact); and second, unlike the textualized terms we have considered so far, 'Vm' is not merely double encoded (by C_1C_2), but *triple* encoded (by $C_1C_2C_{VH}$). Both of these characteristics bear on the question of truth. For whenever a proposition is encoded and self-inserted, the expression that results not only describes the creation of a text, but also portrays that text as asserting a truth.[19] And the coding functions that define the textual production indicate the type of truth that the text is said to assert: a double encoding (of C_1C_2) indicates that the truth is a literal one, while a triple encoding (such as $C_1C_2C_{VH}$, above) indicates that the truth is a metaphorical one. In other words, the above expressions assert, in effect, not only that Eliot called the men of his generation 'hollow men', but also that Eliot's appellation was true, at least metaphorically.

In summary, it would appear that both the semantic and the epistemological properties of metaphor can be analyzed as resulting from an encoded self-insertion involving a transformal code. And yet, if we consider the full evocative depth of metaphor, it is difficult to avoid the impression that the above analysis is incomplete — perhaps even facile. Analysts of metaphor have frequently suggested that metaphoric power resides not in the essential properties of the metaphoric concept, abstractly considered, but rather in the accidental properties of the far-flung particulars to which that concept is ordinarily applied — what Levinson has called the 'connotational penumbra' of the metaphoric expression (1983: 150). Thus, the force of Eliot's application of the word 'hollow' to the members of his generation would derive not so much from the denoted meaning of the word 'hollow' as from the full range of the connoted properties of the decaying, broken and depleted things to which that word may previously have been applied in the interpreter's experience.

To some extent, the above analysis does recognize this connotational meaning, for I have left the definition of transformal coding somewhat vague, indicating that the range of features that are metaphorically transformed can be extended indefinitely across any semantic field. By itself, however, this conception is not sufficient, for it leaves many important questions unanswered — particularly the question of what makes one metaphor better than another. For if any transformation can be indefinitely extended, then any metaphor is as good as any other. As Ricoeur says, anything resembles anything else, except for a certain difference. In the following pages, therefore, I will explore some of the problems and characteristics of *connotation* through an extension of the above analysis of the self-insertion of transformally encoded terms.

Connotation

In considering the question of connotation, it is useful to draw upon Wittgenstein's (1953) later investigations of language. Much of this work stems from Wittgenstein's analysis of a specific semantic problem, the question of how we apply a single word to all of its multifarious instances.

As an example, Wittgenstein considers the word 'game'. We instinctively apply this word to a wide range of instances — board games, ball games, party games, and so forth — but if someone were to ask what all these activities have in common, it would be exceedingly difficult to give a clear answer. To any formal definition that might be proposed, it would seem to be possible to offer specific exceptions. And yet, all of these activities are, in some sense, related, for each has resemblances to certain of the others, which, in turn, resemble still others in different respects, and so on, across the entire array of games, to the most distant and apparently unrelated members. Wittgenstein characterizes a system of this kind as a 'family resemblance' — that is, 'a complicated network of similarities overlapping and criss-crossing' (1953: I.66) — and he emphasizes that such a system is never reducible to a single exact definition or determination.

This kind of family resemblance system, with its complex and interrelated similarities, also defines what I have called above a semantic field — in this case, the semantic field of games. Wittgenstein's theory, however, gives an additional insight into the nature and structure of such a semantic field. For at the heart of Wittgenstein's analysis is the recognition that a concept like 'game' does not simply emerge out of the objective features of the world; rather, it is *subjectively* constructed, in part, by the members of the culture, who continuously negotiate and renegotiate the linguistic conventions by which the word 'game' is concretely applied. It is in the sum total of these concrete applications that the word 'game' accumulates its particular connotations; and the dynamic process by which meaning is negotiated ensures that the organizing principles behind these connotations can never be exactly specified.

It may be possible, however, to give some indication of the way in which connotations are encoded into a particular use of a particular word. For example, let us imagine that a person uses the word 'game' to identify a specific soccer match. This situation may be represented by the expression

$$pS(C_1C_2(Gs)),$$

where 'p' stands for the person speaking, 'S' for the text-productive act of saying, 'G' for the predicate 'game', and 's' for the soccer match. Since we are assuming here that the speaker is referring to a real game of soccer, we can expand the above to

$$Gs.pS(C_1C_2(Gs)).$$

This expression can be resolved into the (partial) encoded self-insertion

$$sub_2(Gs.pSx,x,C_1C_2(Gs)),$$

which exhibits the fact that the speaker's statement is literally true. So far, this analysis can be described as a decoding of the word 'game' as it is literally applied to the soccer match. We might also describe it as a *denotative* decoding of the word.

Now, how might we expand this analysis to include a *connotative* decoding of the word 'game'? To begin with, we need to consider other instances in which the word 'game' has been applied (or could have been applied) — and more specifically, those instances of games which the current use of the word brings to mind. For the sake of simplicity, let us imagine that we ask the speaker, p, to identify such an instance and that he or she identifies a recent football game, f. We can represent this situation with the expression

$$\text{Gf.sub}_2(\text{Gs.pSx,x,}C_1C_2(\text{Gs})).$$

We can now begin the connotative unpacking of the word 'game' by constructing a transformal code, C_{FS}, which relates the football game and the soccer game. Drawing upon this code, we can resolve the above expression into the following *meta*-encoded-self-insertion:

$$\text{sub}_3(\text{Gf.sub}_2(\text{Gs.pSx,x,}C_1C_2(\text{Gy})),\text{y,}C_{FS}(\text{f})).$$

Next, we push the analysis back one step further by asking p to name another game — this time, one that is brought to mind not by the soccer game, but, rather, by the football game. Let us say that p recalls a particular game of hockey, h, that in some way resembled the football game. We now have

$$\text{Gh.sub}_3(\text{Gf.sub}_2(\text{Gs.pSx,x,}C_1C_2(\text{Gy})),\text{y,}C_{FS}(\text{f})),$$

which decodes into the *meta*-meta-encoded-self-insertion

$$\text{sub}_4(\text{Gh.sub}_3(\text{Gf.sub}_2(\text{Gs.pSx,x,}C_1C_2(\text{Gy})),\text{y,}C_{FS}(\text{z})),\text{z,}C_{HF}(\text{h})).$$

Using this procedure, we can push the analysis back still further to a *meta*-meta-meta-encoded-self-insertion, and so on, indefinitely.

In theory, such an expansion need never terminate, since the set of events which any person might classify as a game has no exact defining constraint. Even if such an expansion could be entirely carried out, however, we would still not have completed the full connotative decoding of the word 'game' (as applied to the soccer match). For the above analysis is a simplification of a semantically much more complex situation in which each game actually serves as a node or collection point for *many* transformations based on a variety of objective similarities to a variety of closely related games. This corresponds to asking p to name *all* the games that each new game in the series brings to mind — the result of which would have to be represented by the conjunction of innumerable expansions like the one above, each signifying a single pathway along the branching structure of the full connotative system, a unique train of associations evoked by the original soccer game and continuing outward more or less indefinitely. Rather than attempting to represent this entire system in all its particulars, we can indicate its general form in the following manner. Any specific expansion representing a specific pathway or train of associations — like the one we began unpacking above — can be summarized by the expression:

$$\overset{x^*}{\underset{x=s}{\mathrm{sub}}}{}^{-1}\,(Gx)$$

which signifies a successive unpacking ('sub^{-1}') of (transformally) encoded self-insertions with respect to the category G, starting with the game s and ending with an arbitrary game, x*. But we are interested in the larger system of *all* such expansions of all lengths and permutations. This totality can be represented by the expression

$$\prod \overset{x}{\underset{x=s}{\mathrm{sub}}}{}^{-1}\,(Gx)$$

in which the large pi indicates the logical product (or conjunction) of all such expansions.[20]

The above formula represents the resemblance family of games insofar as that family is organized by p with respect to the soccer match. To put it another way, the formula delineates the connotations of the word 'game' in a specific instance of its usage by a specific interpreting subject. It should be noted, however, that we are still talking here about connotation in a limited sense. For, as indicated above, connotation includes not only the particulars that are *categorically* designated by a given word, but also those that are merely *accidentally* related to that word by virtue of empirical and psychological association. It will be recalled that our earlier discussion of metaphor, in particular, seemed to lack depth because it did not account for the power of precisely such connotations. In order to further explore this issue, therefore, let us return once again to the question of metaphor and to Eliot's designation of his peers as 'hollow men'.

It will be useful if we begin a reconsideration of Eliot's metaphor by applying the method developed above to unpack the (categorical) connotations of its vehicle, 'hollow'. Such a procedure will produce expansions similar to the ones which were derived for the word 'game'. Specifically, we will obtain expressions of the form

$$\mathrm{sub}_4(\mathrm{Hb.sub}_3(\mathrm{Ha.sub}_2(\mathrm{Vm.eWx,x,}C_1C_2(\mathrm{Hy})),\mathrm{y},C_{AM}(\mathrm{z})),\mathrm{z},C_{BA}(\mathrm{b})),$$

$$\underset{x=m}{\overset{x^*}{\mathrm{sub}}}{}^{-1}(\mathrm{H'x})$$

and

$$\prod \underset{x=m}{\overset{x}{\mathrm{sub}}}{}^{-1}(\mathrm{H'x})$$

where 'V', 'H', 'm', 'e' and 'W' are defined as they were before, and individual terms like 'a' and 'b' stand for the multitude of objects to which the term 'hollow' might be literally applied. It will be noted, however, that these expressions differ in some important ways from the ones that were derived for the word 'game'. First, the innermost formal encoding in the first expression — the 'C_1C_2' that represents the actual production of Eliot's text — while involving the encoded insertion of a proposition ('Hm'), does not involve the encoded *self*-insertion of that (whole) proposition. The reason for this is straightforward: Eliot's characterization of his peers as 'hollow men' is not *literally* true. Insofar as Eliot's text is *metaphorically* true, we can no longer represent this fact in any simple way, for the current analysis is based on the assumption that metaphoric power resides in the connotations of the vehicle and that the meaning of such a metaphor cannot be translated into any clearcut literal equivalent. Thus, the term 'Vm', which represented the translation of Eliot's metaphor in the first analysis, can no longer be regarded as adequate for that purpose, although this term may, in itself, continue to embody a literal truth about the metaphoric object. Second, the transformal encodings that lead from the metaphoric object into the connotative system (represented in the first expression by 'C_{AM}') are not based on similarities that can be described as 'objective' in the usual sense of the term. This reflects the fact that metaphorical resemblance is not so much discovered as it is created — an issue to which I will return, below. Finally, the full system of connotations represented by the last formula is not the resemblance family of Eliot's generation, nor does it embody the semantic field of the word 'hollow' in any ordinary sense, for it represents a metaphoric extension of that word to an object that is usually not considered to lie within its domain. I have therefore attached a prime to the 'H' in the second and third formulas to indicate explicitly this metaphoric extension.[21]

Even with these qualifications, however, the above formulas do not encompass the entire connotative system underlying the metaphor, for they involve only categorical connotations — and, as we have noted above, each of the many (literal) referents of the word 'hollow' evoked by the metaphorical use of that word will, in turn, evoke not only other exemplars of the word 'hollow', but also, potentially, many other things that are only contingently or accidentally related to hollowness. Furthermore, each contingent item will evoke other such contingent items, which, in turn, will evoke still others, and so forth. To give an example, if one branch of associations leads to the recollection of a hollow bone, the next association may evoke not another hollow item but some other particular *bone* — say a fragment of a fossil — and this, in turn, may evoke some other object that is *fragmented* — say a shard of a dish — which, in turn, may evoke some other thing associated with the dish's *breaking*, such a sound, and so forth. Unlike a train of categorical associations, therefore, a train of accidental or contingent associations eventually enters into content that is not necessarily confined to the scope of the original word that set that train in motion. We can assume, however, that every such train must *begin* with one or more categorical associations, for it is only by virtue of some word and some linguistic category that metaphoric connotation is evoked in the first place. This means that an analysis of this kind of connotation must necessarily proceed through two stages. The first stage will involve — for each connotative pathway — a categorical decoding like those we have already carried out and will therefore be represented by an expression like

$$\underset{x=m}{\overset{x^*}{\mathrm{sub}^{-1}}}(H'x)$$

where 'x*' represents the last item for which all previous decodings are categorical (e.g., the hollow bone, above). After this, the connotative decoding will continue, but no longer necessarily constrained by any particular category. Thus, in the above example, if 'Bf' stands for the fossil being a bone, 'Fs' for the shard being a fragment, and 'nRb' for the noise resulting from a breaking, the continued connotative unpacking will be represented as

$$\mathrm{sub}\,(\mathrm{nRb.sub}\,(\mathrm{Fs.sub}\,(\mathrm{Bf.}\underset{x=m}{\overset{x^*}{\mathrm{sub}^{-1}}}(H'x), x^*,$$

$$C_{FB}(y)), y, C_{SF}(z)), z, C_{NS}(n)).$$

The transformal codes C_{FB}, C_{SF}, C_{NS}, etc., are defined by the fact that each particular in the series ('f', 's', 'n', etc.) still shares its predicate ('B', 'F', 'Rb', etc.) with the particular that immediately precedes it, although the shared predicate is not explicitly represented in the preceding term.

The entire expansion can be summarized as

$$\underset{y=x^*}{\overset{y^*}{\mathrm{sub}}}{}^{-1}\left(\underset{x=m}{\overset{x^*}{\mathrm{sub}}}{}^{-1}\,(\mathrm{H'x})\right),$$

and the totality of all such expansions will be represented as

$$\prod \underset{y=x^*}{\overset{y}{\mathrm{sub}}}{}^{-1}\left(\underset{x=m}{\overset{x^*}{\mathrm{sub}}}{}^{-1}\,(\mathrm{H'x})\right).$$

With this expression, we have defined the extended connotative system underlying Eliot's metaphor. As I have already indicated in discussing categorical connotation, above, an extended system of this kind is inexhaustible in principle. The nature of the associative process guarantees that the connotative lines will branch outward indefinitely. Yet there must be constraints upon such a system, for if there were not we would be faced, once again, with the same problem that was raised in connection with metaphor earlier — namely, that any metaphor would be as good as any other. What kinds of constraining principles, then, might be operating here?

It seems to me that there must be two kinds of constraining principles, associated with the categorical and accidental phases of connotation, respectively. In the categorical phase, connotation is constrained by the literal meaning of the metaphoric vehicle — that is, the objective similarities among the particular things to which that metaphoric vehicle ordinarily refers. In the present case, this would confine categorical connotation to the ordinary scope of the word 'hollow'. It was emphasized above that the literal meaning of such a word is never reducible to an exact definition or rule, but seems to reside, rather, in a complicated network of similarities, overlapping and criss-crossing. It was also implied, however, that the members of this network exist in systematic *gradations* of similarity — for they are differentiated out of a common material order governed by underlying objective forces. Thus, Wittgenstein's characterization may be less haphazard than it initially seems. For if categorical connotation is ultimately anchored in *objective* relations of similarity — mediated in linguistic categories and subcategories and grounded in the regular features of the material world — then each member of the resemblance family associated with the word 'hollow' would define a transformally encoded hierarchy of diminishing similarities to itself, down through the whole spectrum of possible hollow things, in progressively more specific variations on indefinitely proliferating new themes. I am envisioning a structure here something like those of the fractals recently discovered by

Mandlebrot and other pioneers of transfinite geometry (see Peitgen and Richter 1986).[22]

When we move into accidental connotation, however, a second kind of constraining principle takes over. In the accidental order, family resemblance no longer functions as a semantic determinant, for similarities here are created in the mind of the subject rather than independently discovered in the world of objects. The only constraint that can function in such a context is the constraint of creative or aesthetic efficacy, the constraint imposed by intuition in choosing a word with just the *right* connotations. Thus, Eliot chooses the word 'hollow' because it entails, ultimately, accidental or contingent connotations that present apt analogies to exactly the traits of character that he perceives in the members of his generation.

This point can be clarified by considering a somewhat more radical metaphor. When Allen Ginsberg describes a cold journey through 'grandfather night' (1956: 11), an image is created through the deep connotations of the word 'grandfather' and the numberless transformal codes that relate them, somehow, to the system of meanings underlying the word 'night'. It is difficult to imagine any translation that could give a literal equivalent to the strange sense of this metaphor, partly because the transformal connections upon which it is based are so intuitive, idiosyncratic and unsystematic. The metaphor does not exploit an analogy, in any 'objective' sense of the term, between the metaphoric vehicle and the metaphoric object. Instead, it seems to involve the selection of a vehicle that is superficially irrelevant, but whose connotative structure branches out in all directions to distant particulars with startling and peculiarly revealing correspondences to previously hidden aspects of the object.

Furthermore, the connotative branches of such a metaphor will join with those of the other metaphors of the text, reinforcing their meanings and that of the text as a whole, while simultaneously being reinforced by them. Consider, for example, the following passage from the same poem (p. 9), in which Ginsberg describes the members of *his* generation:

> angelheaded hipsters burning for the ancient heavenly connection to the starry dynamo in the machinery of night.

In this passage, each metaphoric vehicle is enriched by the trains of connotations that define its semantic field and the ways in which these connotative trains relentlessly invade the semantic fields underlying the other signifiers of the text. For each signifier, the connotative lines seem to diverge radically in all directions; yet an overarching sense emerges at the level of the whole passage, and one that is all the more unitary and coherent, for the poet has chosen words whose connotations join with and reinforce each other, setting up a kind of harmony that reverberates throughout the text. The ability to use language in this manner, coordinating multitudes of associations among numerous signifiers into a single larger connotative system that also, somehow, intensifies the power of each individual signifier, must be the result of a dynamic and inspired imagination that operates at all levels of connotation simultaneously and that is not reducible to coding as I have used the term in this paper.[23]

Conclusion

We have come a long way, it seems, from the formalisms of mathematics to the fathomless reverberations of poetic metaphor. Throughout this journey, however, I have attempted to show a thread of continuity in the operation of certain underlying semiotic mechanisms — specifically, coding and recursive encoded self-insertion. I have attempted to show this with respect to two types of codes, which I have designated formal and transformal codes, respectively. Formal codes operate on constellations of meanings insofar as they are defined at a given level of abstraction, converting concepts, progressively, into systems of objects in the material world and then into the subsystems of objects that we call signs or texts. Transformal codes, on the other hand, are not defined with regard to any particular level of abstraction, but convert the contents of specific semantic fields into the contents of other specific semantic fields. While these two types of coding are logically independent of each other and rather specialized in their semiotic functioning, we must not fall into the erroneous assumption that they are somehow carried out independently of each other. For if we keep in mind Wittgenstein's analysis of the construction and application of concepts, it becomes clear that the contents of formal and transformal codes are deeply and inextricably interrelated. As a matter of fact, the constructs of formal and transformal coding are merely theoretical ones that have been superimposed upon what is ultimately an integral process — that of the reflective human subject constituting meaning in a dynamic social and material field.

Thus, in language, the abstract sense that is formally encoded in the literal use of a given word is defined by the set of transformally encoded referents to which that word is conventionally applied. This same reciprocity of formal and transformal coding holds for all other kinds of signification, even for the constitution of simple material objects that are not normally regarded as signifiers per se — for every recognizable object always formally encodes a concept which, in turn, is implicitly defined by the set of transformally encoded other things that are conventionally treated as the same kind of object. In the complex interplay of formal and transformal coding, therefore, the reflective function of the human subject constructs the text at all levels of action and conception so that it reverberates with meaning in, so to speak, all directions at once. This reverberation actually occurs within the human subject, of course, as that same reflective function unpacks the text and exhibits the manifold variations of form that were placed there by repeated cycles of formally and transformally encoded self-insertion.

In appropriating the Ginsberg poem, for example, the interpreting subject formally decodes the signifiers of the text, first into their reference, then into their sense, then into a meta-sense that covertly implies that, like the members of the generation it describes so starkly, the poem itself violently searches for meaning; and at the same time, the signifiers are transformally decoded, not only into branching systems of categorical connotations defining literal semantic fields, but also into much broader systems of poetically constrained accidental associations that give specific signifiers metaphoric depth and penetrate into the semantic fields of other signifiers, creating reverberance among the metaphors of the text. I have indicated above that the creative process behind such a text cannot be reductively analyzed, for it seems to involve an intuitive activity that operates at all levels of coding simultaneously. It does seem possible, however, to analyze certain aspects of the semantic relations that result from such a creation; and when such an analysis is carried out we find certain patterns of recursion, encoding and self-insertion manifesting themselves again and again — in the playfulness of direct self-representation, in the intricacies of frames and layers, in the convoluted patterns of reflectivity, in the subtleties of covert self-allusion, and, finally, in transformally encoded metaphoric structures that form the basis for deep textual reverberance.

It is remarkable that these fundamental patterns of coding occur in such a wide variety of texts, from the simplest instances of denotation to the most complex creations of poetic imagination. And it is perhaps more remarkable that a prototype of these patterns would emerge in so unlikely a context as when Kurt Gödel looked into the foundations of mathematics and discovered there traces of the human subject.

Notes

1. Formal mathematics can also be regarded as an artificial or formalized *language*, and metamathematics therefore becomes an artificial or formalized *metalanguage*. The rela-

tionships between such linguistic hierarchies and the type level hierarchies of formal logic are complex but can be understood roughly as follows: a language is defined by a constellation of terms relevant to a particular subject matter and operating at one or more levels of abstraction (or type levels); a metalanguage is then defined by a larger constellation which includes within it terms for everything signified by the terms of the language plus an additional set of terms of at least one level of abstraction (or type level) higher than any term of the language and signifying the functioning of the terms of that language (see Tarski 1935: 230, 235).

2. I have modified Nagel and Newman's rules slightly for purposes of overall clarity. Formulas G and F below are also adapted (with modifications) from Nagel and Newman, although the analysis of Gödel numbering in terms of coding and Gödel transforms ('G functions') is my own. The reader may find it instructive to look at two very different systems of Gödel numbering which are developed, respectively, in Hofstadter (1979: 268–69) and Rucker (1982: 303–304).
3. Taken together, G_1 and G_2 cover all the Gödel numbering rules given above (and in Gödel's paper). I am simplifying the analysis here by making the subscripts '1' and '2' consecutive, which amounts to treating each Gödel transform as though it operated on expressions of a single logical type level (whereas G_2 actually operates on expressions of multiple type levels, as indicated, above, in the rules for assigning Gödel numbers to variables). This simplification is merely a notational convenience and does not affect the following argument, which assumes only that each Gödel transform is of a higher type level than that of any of the signs upon which it operates.

   In addition to the above, it is possible to define higher order Gödel transforms. For example, metamathematical formulas can be converted to mathematical formulas by a set of coding rules which may be collectively designated as 'G_3' — itself a *meta*-metamathematical operation (see note 6). Likewise, meta-metamathematical expressions can be converted into metamathematical ones by a function designated 'G_4' — and so on. It should be noted that the logic that generates this hierarchy entails that lower order Gödel transforms must necessarily be included within the structures of higher order Gödel transforms. That is, G_1 is implicitly entailed within the definition of G_2, which is entailed within the definition of G_3, etc.
4. Another kind of reduction can be effected when Gödel transforms are applied directly to each other. Specifically, any G-function can be used to encode a lower order G-function (or 1/G-function) into one that is still lower, as long as it encodes the argument of that function too. Thus, the operation $G_3(G_2(x))$ can be simplified by applying the G_3 first to the G_2 (producing G_1) and then to the x (producing 11), with a final result of $G_1(11)$. Likewise, $G_4(1/G_3(x)) = 1/G_2(11)$. Similarly, any 1/G function can be used to *decode* a lower order G-function (or 1/G-function) into one that is *higher*, as long as it starts off at least two levels higher than the function it decodes and as long as it decodes the argument of that function too. Thus, $1/G_3(G_1(11)) = G_2(x)$ and $1/G_4(1/G_2(11)) = 1/G_3(x)$. On the other hand, all Gödel transforms are *undefined* when operating on Gödel transforms that are of equal or higher type levels (and all 1/G transforms are additionally undefined when operating on Gödel transforms of only one type level lower). Hence, expressions of the form $G_1(G_1(G_1(\ldots)))$ and $1/G_1(1/G_1(1/G_1(\ldots)))$ can only operate from the inside outward.

   A close consideration of the above will show that it is consistent with the comments of the previous note about lower order Gödel transforms being components of higher order Gödel transforms. (This can be confirmed by carrying out the inner transformations in the examples first.) Moreover, it will show that the subscripts of Gödel transforms can be meaningfully extended to include zero and negative numbers. For example, $G_3(G_2(G_1(0)))=G_1(G_0(6))=G_{-1}(2^7\times 3^7\times 5^7\times 7^7\times 11^7\times 13^7\times 17^6)$. Such nonpositive subscripts indicate the (minimum) number of times that the argument of a Gödel transform has already been encoded from a Gödel number. More significantly, this entire line of reasoning indicates that the hierarchy of logical types can be extended infinitely in both directions and that there is a general meaningful interpretation of *negative type levels*. As will be indicated below, the present application constitutes merely a special case of these larger principles.

5. To further illustrate the operation of the sub function, we may consider a purely numerical example in which that function operates on the three input numbers of 15,000, 1 and 36. The first of these numbers is decomposed into its fundamental factorial structure, $2^3\times 3^1\times 5^4$. The second number, 1, is then found to match the exponent of the middle factor, 3^1, which is therefore replaced by the fundamental factorial structure of the third number, $2^2\times 3^2$, producing $2^3\times 2^2\times 3^2\times 5^4$. The base numbers of this expression are then made consecutive, producing $2^3\times 3^2\times 5^2\times 7^4$ or 4,321,800, the output number.
6. This entire analysis can be represented, in its essentials, by an application of the notational principles that have already been discussed. That is, Formula G can be decoded into its second meaning by the operation of the decoding function defined as $1/G_3$ (see note 3). When this decoding function is applied to Formula G it produces the metamathematical formula

$$\forall X \sim Dem_2(X,sub_2(F,y,n)),$$

each term of which represents the decoded counterpart of a specific component of Formula G. In particular, the sub_2 function operates upon (Formula) F by replacing all of its y's with n's (producing Formula G). The remainder of the formula states that the result of this substitution (Formula G) is not demonstrable within the calculus of symbolic logic. Thus, while Formula G represents Gödel's Theorem in code, the above metamathematical formula states it directly.

7. It should be noted that the analogy between Gödel numbering and the encoding of sense into reference is not a perfect one. For example, while the above analysis indicates that the concept 'tree' must encode into a specific instance of that which it categorizes, the same is not true of the concepts encoded by Gödel numbering. Thus, the Gödel number of the predicate 'Prime(x)' is not itself a prime number (as can be shown from the rules of Gödel numbering). However, this discrepancy is more apparent than real. For just as the Gödel number of 'Prime(x)' is not itself a prime number, so too the referent of the category 'tree' is not a 'real' tree in the absolute sense of a Kantian 'thing-in-itself'; rather, it is a construction of a tree which is very much shaped by the conceptual category upon which it is based. This point can be clarified by considering other predicates that might also characterize the same tree ('green', 'maple', 'swaying in the wind', etc.). Each of these encodes into its own particular referent ('the green thing', 'the maple', 'that which sways in the wind', etc.), all of which are identical empirically but not logically. Thus, the apparent discrepancy between Gödel numbering and the encoding of sense into reference merely represents the functional differences between mathematical statements and ordinary discourse.
8. This raises a number of technical questions which quickly become philosophical. One such question concerns the kinds of referents into which various kinds of sense are encoded. For example, what is the referent of a statement about a plurality of things or of a purely analytic statement or of a false statement? To some extent, this depends upon the theory of reference to which one subscribes (see Lyons 1977: 177–197). In this paper I will employ a very broad conception of reference, assuming that every

meaningful statement has some kind of referent — singular or plural, definite or indefinite, real or fictitious — which encodes the sense of the locution, as elaborated in the text. In adopting this notion I have found it helpful to draw on Reichenbach's notion of event-splitting (1947: 268–269), which demystifies many of the problems of reference by clarifying the logical status of events.

Another question concerns the proper use of symbolic logic to represent sense and reference in the abstract. I will adhere to the following conventions. As indicated in the text, predicate terms of symbolic logic (e.g., 'Tx') will be regarded as representing units of pure sense, devoid of any reference. Individual variable terms (e.g., 'x', 'y', etc.) have neither sense nor reference since they actually serve only a syntactical function. For individual constant terms (e.g., 't') the situation is more complicated. As long as they stand alone, these terms can be regarded as having reference only and can therefore be used to signify the purely referential component of meaning. However, when such terms are embedded in formulas representing propositions (such as 'jPt') they can be regarded as losing their referential function and acquiring instead a sense-signifying function, thus representing only components of the (pure) senses of the propositions in which they are embedded. To put it another way, individual terms in propositions are merely a notational convenience and should theoretically be replaced by quantified expressions. In the present case 'jPt' would become '$\exists x \exists y(Jx.Ty.xPy.\forall w(Jw \supset (w=x)).\forall z(Tz \supset (z=y)))$' where 'J' signifies 'is named John', 'T' signifies 'is a tree', and the components involving the variables 'w' and 'z' assert the oneness of John and of the tree, respectively. I have avoided using such quantified expressions since they would require some technical adjustments and some rather lengthy explanations and would greatly complicate certain of the analyses below without materially affecting the general points I am making. However, the use of quantifiers clarifies a number of subtleties and helps to answer the questions raised in the previous paragraph about specific referents corresponding to specific senses. For example, existentially quantified propositions encode into individual terms that name real referents, singular or multiple; the negations of such propositions encode into individual terms that name fictitious referents; and so forth.

9. Using Peirce's type/token distinction, 't'' is a token of a token, 't''' a token of a token of a token, etc. That is, successive applications of the C_1 function lead into the hierarchy of metalanguages, but only insofar as that hierarchy involves the *concrete material signifiers* by which languages refer to each other. While this conception leaves ambiguities as to which particular tokens are involved, such ambiguities can be resolved by a more complete treatment involving quantifiers (see note 8).

The reader may find it useful at this point to review note 4, for all of the rules governing the use of G functions also hold for C functions — only here we are talking not about numbers, mathematics and metamathematics, but rather about objects, language and metalanguage (or more correctly, particulars, predicates and metapredicates). Once again, it is possible to define negative type levels by reducing a formulation like '$C_3(C_2(C_1(t)))$' to '$C_1(C_0(t'))$' and then to '$C_{-1}(t'')$'. In the realm of ordinary semiosis, a nonpositive subscript of a C function is an index of the minimum number of times that its argument has already been encoded from a (token) signifier. The argument itself can be given a type level one lower than that of the C function which operates upon it. Thus, objects are of type zero, (token) signifiers are of type -1, (token) signifiers of (token) signifiers are of type -2, and so forth.

10. I have attempted to keep this explanation relatively clear by treating the painting as though it were a simple physical object. In reality, however, the painting is a signifier of the same kind as those discussed in the previous paragraph. Thus, if the actual

room containing the painting (and pictured within it) is represented by 'r', we may say that $p=C_1(r)=r'$ and that a more complete analysis of the painting is represented by 'sub(r',s',C_0(r'))'. This basic semantic form will occur even in the extreme and limiting case in which the artist excludes the room and paints only the painting. In this limiting situation we will have $p=C_1(m)=m'$ where 'm' represents the *uncoded* material medium — i.e., the blobs of paint — and 'sub(m',s',C_0(m'))' will thus represent the full painting. This analysis can be pushed back still further by decoding the referential object into its sense. In the current situation, 'r' may decode not into 'Rx' but into 'Rsub(x,s,C_1(x))' (and likewise for 'm'), which means, among other things, that there is no picture within the picture within the picture — only a blank spot surrounded by some kind of frame (in the broad sense of Bateson 1954 and Goffman 1974), represented by 'sub(x,s,C_1(x)).'

11. All of the comments of the previous note apply to the analysis of this text as well, except that here we *must* use 'm', where 'm' represents the physical ink on the page. Even with this elaboration, however, the analysis is still not adequate from a purely linguistic point of view, since it concerns a deictic text. In effect, the formula 'Fsub(t,s,C_1(t))' means only that a self-referring sentence has five words (or more correctly, that a self-referring sentence that refers to a self-referring sentence having five words has five words). While this does address the semantics of the text, it does not necessarily spell out the deictic mechanisms by which the phrase 'this sentence' singles out just that particular sentence in which the phrase resides, a task which must be undertaken within the domain of *pragmatics*. Because of this problem — and in order to reduce the complexity of the analysis — I will make no special attempt to analyze deictic words below. For example, I will treat the word 'I' like any other term that refers to a person. For similar reasons I have not attempted to analyze the illocutionary aspects of various speech acts.
12. In the first formula, 'a' is defined as the encoding of a preliminary version of the formula that has the variables 'x' and 'y' in place of 'a' and 'b', respectively, while 'b' is defined as the encoding of a preliminary version of the formula with 'y' in place of 'b' (only), and 's' and 't' are defined as the mental spaces in which reflection and reflection upon reflection respectively occur; in the second formula 'a' and 'b' are given definitions analogous to those in the first formula and 's' and 't' are defined as the mental spaces within the second and first individuals, respectively, as described in the text.
13. In this case, 'a' encodes the whole formula, but with 'x', 'y' and 'z' substituted for 'a', 'b' and 'c', respectively, while 'b' encodes the formula with only the 'y' and 'z' substitutions and 'c' encodes the formula with only the 'z' substitution. The definitions of 's', 't' and 'u' are analogous to those of 's' and 't' in the previous examples. It should be noted that when these definitions of 'a', 'b' and 'c' are altered or exchanged different kinds of convoluted reflection may be represented by the same formula.
14. This is a good place to recall, once again, that the definitions of referents are not strictly determined (see note 7, above). There is nothing requiring that we define the children of Israel as reflective. That is, instead of 'sub(c,s,C_1(c))', the term representing them could be, simply, 'c'. This is true whether or not the children of Israel are known to be reflective, for we are not obligated to encode any particular piece of information about them. Likewise, there is nothing requiring that the term 'c' encode the content of the text in which it occurs. Thus, the term 'c' (in either its reflective or its simple form, above) can be defined as encoding 'sub(x,s,C_1(x))Rp' or 'xRp' or, for that matter, 'xDa' ('x is descended from Abraham') or any other predicate that accurately characterizes the children of Israel, whether or not that predicate is part of the overt

text content. By the same token, the term representing the pharaoh, rather than the one representing the children of Israel, could be cast in reflective form as 'sub(p,s,C_1(p))' and could be defined as encoding a different variant of Formula F — namely, 'cRsub(x,s,C_1(x))'. This course of action would lead to a different analysis of covert self-reference, emphasizing the role of the text in *suppressing* (or rather in attempting unsuccessfully to suppress) the aspirations of Afro-Americans — an analysis that is not altogether implausible. Following this same general line of reasoning, it is possible to analyze some texts as having multiple versions of Formula F encoded within them (and hence multiple self-allusions). In general, the encoded occurrence or nonoccurrence of Formula F (in any of its variants) within the overt meaning of a text is determined not by logical necessity but rather by the semiotics of the situation, as will be described below, and by a logical structure that permits (but does not require) such an analysis.

15. The logic of this situation can be clarified by considering the structure of the semantics involved. If the children of Israel truly recognized themselves as performing the action described in the text of the song (i.e., rising up against Pharaoh), then the semantic form which that reflective action *encoded* is logically indistinguishable from the reflectivity implicit in the subsequent construction (or performance) of the song *about* the children of Israel engaged in that reflective action. That is, the expression 'sub_2(F,x,c)' has a double meaning: (1) the abstract form which is embodied in the concrete (reflective) action of the children of Israel; and (2) the reflectivity that must occur in the creation or appropriation of a text about that concrete (reflective) action. Both of these meanings are exploited in covert self-reference — the first when the overt content of the text is *decoded* (as explained above) and the second when the communicative situation comes to be viewed as a set of metalinguistic signifiers operating *with reference to* the text (as will be explained below).

16. Admittedly, this leads to some rather strange possibilities. Can a performer who sings about a kiss do so in a way that *meta-kisses* the audience? I would not wish to reject such a possibility absolutely, but the relevant question is this: can the participants in the communicative situation take on the forms represented within the text without completely disrupting the text-productive process — and, if so, do they actually do so? The answer will depend partly on the abstractness of the overt meaning of the text and partly on the behavior of those involved in its production and communication.

17. Moreover, it can be argued that a subtler version of this Platonist presumption is inherent in all literary — and, indeed, all semantically rich — language use. This is evident, for example, in the way in which devices like metaphor set up reverberations among levels of abstraction. Thus, when Melville refers to the sea as the 'howling infinite' he conveys a sense that this natural object has both primitive and transcendental characteristics; but decoded, this construction would seem to assert the more abstract notion that the very concept or form of the sea is articulated in a system of both primitive and transcendental relations of meaning. A technical presentation of this argument would require an exhaustive analysis of the allegorical, metaphorical and connotative aspects of Melville's metaphor in terms of what I will define below as formal and transformal codes. It also seems likely to me that formulas representing reflection like 'sub(x,s,C_1(x))' can be understood to apply not only to subjects but also to inanimate objects (such as the sea) if such objects are understood to be encoded and reflectively inserted into a presumed subject or subjects whose representational activity must be considered an essential and defining component of the object itself. This permits an extension of the above analysis of covert self-reference to all kinds of texts, and thus to all situations in which concrete signs can refer to themselves through

the forms that they encode. I will not attempt in this paper to elaborate these ideas in all of their many technical details. In general, however, it is interesting to note the close association between the world view constructed by the Platonist perspective and the semantic reverberance of rich texts.

18. See note 9. Briefly, C functions with positively numbered subscripts greater than 1 represent transformations among isomorphic structures at different logical type levels; those with zero (or negatively numbered) subscripts represent transformations among token signifiers at different levels of metalanguage; and C_1 represents the transition from signified object to signifying token.
19. This is a formalization in the terminology of the current analysis of the principle embodied in Tarski's famous epigram '"snow is white" is true if and only if snow is white' (1944: 343).
20. It should be kept in mind that this formula is not a determinant one like a mathematical formula. If it were, it would constitute the kind of exact definition which was rejected above as a possible organizing principle in the realm of ordinary semantics. Instead, the formula presupposes the interpretive activity of one or more actual human subjects (represented, in this case, by p).
21. The addition of a prime to a predicate term like 'H' is culturally and not logically determined, for the definition of a word as metaphorical rather than literal is ultimately a function of cultural coding in both the linguistic and objective spheres. I have explained this in more detail in Jackson (1991) (see especially pp. 52–55).
22. The analogy between fractals and the structures of such connotative systems is intended to be a loose one. However, there are some striking similarities between the formulas that generate the two. Specifically, a fractal pattern is produced when a certain number (analogous here to the initial application of a word) is recursively iterated by some mathematical function (analogous to recursive iteration by the sub^{-1} function) until a specified output limit is reached (analogous to the final connoted instance, x*) across a continuum of parameters that represents a region of space (analogous to the conjunction of all such connotative expansions defining a region of semantic space). The relationship between the fractal and the structure of the larger system of accidental connotations would appear to be a considerably more complex problem.
23. It may be possible, however, to expand radically the notion of coding to encompass such intuitive processes. See Jackson (1991), where I have distinguished between syntaxic codes, on the one hand, and parataxic and prototaxic codes, on the other. The former are logical, univocal and systematically analyzable, whereas the latter are not. All of the codes which I have discussed in this paper — both formal and transformal — are syntaxic codes, whereas the organizing principles behind connotation and creative text-production would seem to reside in the functioning of parataxic and prototaxic codes.

References

Bateson, Gregory (1954). A theory of play and fantasy. In *Steps to an Ecology of Mind*, 177–193. New York: Balatine Books, 1972.

Bunyan, John (1678) [1909]. *The Pilgrim's Progress From This World to That Which is to Come*. London: J. M. Dent and Sons.

Eliot, Thomas Stearns (1925) [1963]. The hollow men. In *Collected Poems, 1909–1962*, 77–82. New York: Harcourt, Brace and World.

Ginsberg, Allen (1956). Howl. In *Howl and Other Poems*, 9–20. San Francisco: City Lights Books.
Gödel, Kurt (1931) [1962]. *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*, trans. by B. Meltzer. New York: Basic Books.
Goffman, Erving (1974). *Frame Analysis: An Essay on the Organization of Experience*. New York: Harper and Row.
Hofstadter, Douglas R. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books.
Jackson, Michael R. (1984). *Self-Esteem and Meaning: A Life-Historical Investigation.* Albany: State University of New York Press.
—(1991). Categories of codes. *Semiotica* 85 (1/2), 41–72.
Lakoff, George and Johnson, Mark (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
Levinson, Stephen C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
Lyons, John (1977). *Semantics*, Vol. I. Cambridge: Cambridge University Press.
Nagel, Ernest and Newman, James R. (1958). *Gödel's Proof.* New York: New York University Press.
Orwell, George (1954). *Animal Farm*. New York: Harcourt, Brace and World.
Peitgen, Heinz-Otto and Richter, Peter H. (1986). *The Beauty of Fractals: Images of Complex Dynamical Systems*. Berlin: Springer-Verlag.
Reichenbach, Hans (1947). *Elements of Symbolic Logic*. New York: Macmillan Co.
Rucker, Rudy (1982). *Infinity and the Mind.* New York: Bantam Books.
Tarski, Alfred (1935) [1956]. The concept of truth in formalized languages. In *Logic, Semantics, Metamathematics: Papers from 1923–1938*, 152–278, trans. by J. H. Woodger. Oxford: Clarendon Press.
—(1944). The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research* 4, 341–376.
Whitehead, Alfred North and Russell, Bertrand (1910–1913). *Principia Mathematica*, Vols. I–III. Cambridge: Cambridge University Press.
Wittgenstein, Ludwig (1953). *Philosophical Investigations*. Oxford: Basil Blackwell.

Michael R. Jackson (b. 1946) is Staff Psychologist at Wyandotte Hospital and Medical Center in Wyandotte, Michigan. His principal research interests include social science methodology, semiotic approaches to therapeutic change, and textual aspects of dreaming, storytelling, recollection, discourse, and self-reflection. Among his publications are *Self-Esteem and Meaning: A Life-Historical Investigation* (1984) and 'Categories of codes' (1991).