10 Encoding/decoding*

Stuart Hall

Traditionally, mass-communications research has conceptualized the process of communication in terms of a circulation circuit or loop. This model has been criticized for its linearity - sender/message/receiver - for its concentration on the level of message exchange and for the absence of a structured conception of the different moments as a complex structure of relations. But it is also possible (and useful) to think of this process in terms of a structure produced and sustained through the articulation of linked but distinctive moments - production, circulation, distribution/consumption, reproduction. This would be to think of the process as a 'complex structure in dominance', sustained through the articulation of connected practices, each of which, however, retains its distinctiveness and has its own specific modality, its own forms and conditions of existence. This second approach, homologous to that which forms the skeleton of commodity production offered in Marx's *Grundrisse* and in *Capital*, has the added advantage of bringing out more sharply how a continuous circuit - production-distribution-production - can be sustained through a 'passage of forms'.[1] It also highlights the specificity of the forms in which the product of the process 'appears' in each moment, and thus what distinguishes discursive 'production' from other types of production in our society and in modern media systems.

The 'object' of these practices is meanings and messages in the form of sign-vehicles of a specific kind organized, like any form of communication or language, through the operation of codes within the syntagmatic chain of a discourse. The apparatuses, relations and practices of production thus issue, at a certain moment (the moment of 'production/circulation') in the form of symbolic vehicles constituted within the rules of 'language'. It is in this discursive form that the circulation of the 'product' takes place. The process thus requires, at the production end, its material instruments - its 'means' - as well as its own sets of social (production) relations - the organization and combination of practices within media apparatuses. But it is in the *discursive* form that the circulation of the product takes place, as well as its distribution to different audiences. Once accomplished, the discourse must then be translated - transformed, again - into social practices if the circuit is to be both completed and effective. If no 'meaning' is taken, there can be no 'consumption'. If the meaning is not articulated in practice, it has no effect. The value of this

*This article is an edited extract from 'Encoding and Decoding in Television Discourse', CCCS Stencilled Paper no. 7.

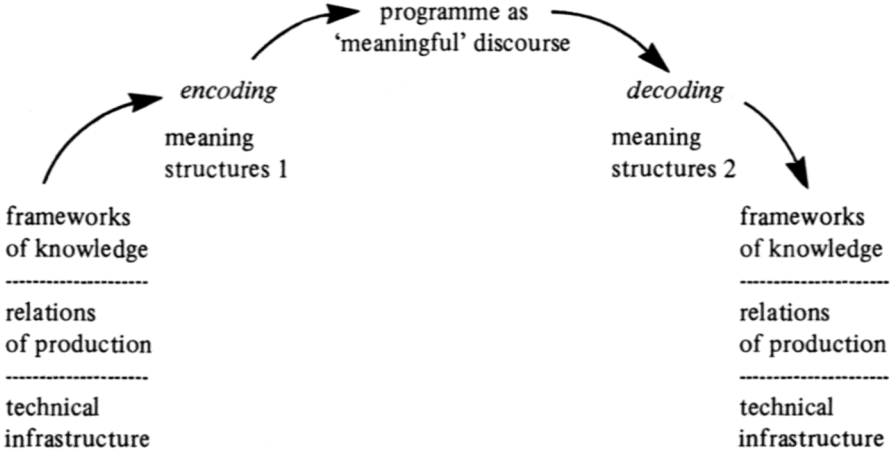
approach is that while each of the moments, in articulation, is necessary to the circuit as a whole, no one moment can fully guarantee the next moment with which it is articulated. Since each has its specific modality and conditions of existence, each can constitute its own break or interruption of the 'passage of forms' on whose continuity the flow of effective production (that is, 'reproduction') depends.

Thus while in no way wanting to limit research to 'following only those leads which emerge from content analysis',[2] we must recognize that the discursive form of the message has a privileged position in the communicative exchange (from the viewpoint of circulation), and that the moments of 'encoding' and 'decoding', though only 'relatively autonomous' in relation to the communicative process as a whole, are *determinate* moments. A 'raw' historical event cannot, *in that form,* be transmitted by, say, a television newscast. Events can only be signified within the aural-visual forms of the televisual discourse. In the moment when a historical event passes under the sign of discourse, it is subject to all the complex formal 'rules' by which language signifies. To put it paradoxically, the event must become a 'story' before it can become a *communicative event.* In that moment the formal sub-rules of discourse are 'in dominance', without, of course, subordinating out of existence the historical event so signified, the social relations in which the rules are set to work or the social and political consequences of the event having been signified in this way. The 'message form' is the necessary 'form of appearance' of the event in its passage from source to receiver. Thus the transposition into and out of the 'message form' (or the mode of symbolic exchange) is not a random 'moment', which we can take up or ignore at our convenience. The 'message form' is a determinate moment; though, at another level, it comprises the surface movements of the communications system only and requires, at another stage, to be integrated into the social relations of the communication process as a whole, of which it forms only a part.

From this general perspective, we may crudely characterize the television communicative process as follows. The institutional structures of broadcasting, with their practices and networks of production, their organized relations and technical infrastructures, are required to produce a programme. Using the analogy of *Capital,* this is the 'labour process' in the discursive mode. Production, here, constructs the message. In one sense, then, the circuit begins here. Of course, the production process is not without its 'discursive' aspect: it, too, is framed throughout by meanings and ideas: knowledge-in-use concerning the routines of production, historically defined technical skills, professional ideologies, institutional knowledge, definitions and assumptions, assumptions about the audience and so on frame the constitution of the programme through this production structure. Further, though the production structures of television originate the television discourse, they do not constitute a closed system. They draw topics, treatments, agendas, events, personnel, images of the audience, 'definitions of the situation' from other sources and other discursive formations within the wider socio-cultural and political structure of which they are a differentiated part. Philip Elliott has expressed this point succinctly, within a more traditional framework, in his discussion of the way in

which the audience is both the 'source' and the 'receiver' of the television message. Thus - to borrow Marx's terms - circulation and reception are, indeed, 'moments' of the production process in television and are reincorporated, via a number of skewed and structured 'feedbacks', into the production process itself. The consumption or reception of the television message is thus also itself a 'moment' of the production process in its larger sense, though the latter is 'predominant' because it is the 'point of departure for the realization' of the message. Production and reception of the television message are not, therefore, identical, but they are related: they are differentiated moments within the totality formed by the social relations of the communicative process as a whole.

At a certain point, however, the broadcasting structures must yield encoded messages in the form of a meaningful discourse. The institution-societal relations of production must pass under the discursive rules of language for its product to be 'realized'. This initiates a further differentiated moment, in which the formal rules of discourse and language are in dominance. Before this message can have an 'effect' (however defined), satisfy a 'need' or be put to a 'use', it must first be appropriated as a meaningful discourse and be meaningfully decoded. It is this set of decoded meanings which 'have an effect', influence, entertain, instruct or persuade, with very complex perceptual, cognitive, emotional, ideological or behavioural consequences. In a 'determinate' moment the structure employs a code and yields a 'message': at another determinate moment the 'message', via its decodings, issues into the structure of social practices. We are now fully aware that this re-entry into the practices of audience reception and 'use' cannot be understood in simple behavioural terms. The typical processes identified in positivistic research on isolated elements - effects, uses, 'gratifications' - are themselves framed by structures of understanding, as well as being produced by social and economic relations, which shape their 'realization' at the reception end of the chain and which permit the meanings signified in the discourse to be transposed into practice or consciousness (to acquire social use value or political effectivity).



Clearly, what we have labelled in the diagram 'meaning structures 1' and 'meaning structures 2' may not be the same. They do not constitute an 'immediate identity'. The codes of encoding and decoding may not be perfectly symmetrical. The degrees of symmetry – that is, the degrees of 'understanding' and 'misunderstanding' in the communicative exchange – depend on the degrees of symmetry/asymmetry (relations of equivalence) established between the positions of the 'personifications', encoder-producer and decoder-receiver. But this in turn depends on the degrees of identity/non-identity between the codes which perfectly or imperfectly transmit, interrupt or systematically distort what has been transmitted. The lack of fit between the codes has a great deal to do with the structural differences of relation and position between broadcasters and audiences, but it also has something to do with the asymmetry between the codes of 'source' and 'receiver' at the moment of transformation into and out of the discursive form. What are called 'distortions' or 'misunderstandings' arise precisely from the *lack of equivalence* between the two sides in the communicative exchange. Once again, this defines the 'relative autonomy', but 'determinateness', of the entry and exit of the message in its discursive moments.

The application of this rudimentary paradigm has already begun to transform our understanding of the older term, television 'content'. We are just beginning to see how it might also transform our understanding of audience reception, 'reading' and response as well. Beginnings and endings have been announced in communications research before, so we must be cautious. But there seems some ground for thinking that a new and exciting phase in so-called audience research, of a quite new kind, may be opening up. At either end of the communicative chain the use of the semiotic paradigm promises to dispel the lingering behaviourism which has dogged mass-media research for so long, especially in its approach to content. Though we know the television programme is not a behavioural input, like a tap on the knee cap, it seems to have been almost impossible for traditional researchers to conceptualize the communicative process without lapsing into one or other variant of low-flying behaviourism. We know, as Gerbner has remarked, that representations of violence on the TV screen 'are not violence but messages about violence':[3] but we have continued to research the question of violence, for example, as if we were unable to comprehend this epistemological distinction.

The televisual sign is a complex one. It is itself constituted by the combination of two types of discourse, visual and aural. Moreover, it is an iconic sign, in Peirce's terminology, because 'it possesses some of the properties of the thing represented'.[4] This is a point which has led to a great deal of confusion and has provided the site of intense controversy in the study of visual language. Since the visual discourse translates a three-dimensional world into two-dimensional planes, it cannot, of course, *be* the referent or concept it signifies. The dog in the film can bark but it cannot bite! Reality exists outside language, but it is constantly mediated by and through language: and what we can know and say has to be produced in and through discourse. Discursive 'knowledge' is the product not of the transparent representation of the 'real' in language but of the articulation of language on real relations and conditions. Thus there is no intelligible discourse without the operation of a code. Iconic

signs are therefore coded signs too - even if the codes here work differently from those of other signs. There is no degree zero in language. Naturalism and 'realism' - the apparent fidelity of the representation to the thing or concept represented - is the result, the effect, of a certain specific articulation of language on the 'real'. It is the result of a discursive practice.

Certain codes may, of course, be so widely distributed in a specific language community or culture, and be learned at so early an age, that they appear not to be constructed - the effect of an articulation between sign and referent - but to be 'naturally' given. Simple visual signs appear to have achieved a 'near-universality' in this sense: though evidence remains that even apparently 'natural' visual codes are culture-specific. However, this does not mean that no codes have intervened; rather, that the codes have been profoundly *naturalized.* The operation of naturalized codes reveals not the transparency and 'naturalness' of language but the depth, the habituation and the near-universality of the codes in use. They produce apparently 'natural' recognitions. This has the (ideological) effect of concealing the practices of coding which are present. But we must not be fooled by appearances. Actually, what naturalized codes demonstrate is the degree of habituation produced when there is a fundamental alignment and reciprocity - an achieved equivalence - between the encoding and decoding sides of an exchange of meanings. The functioning of the codes on the decoding side will frequently assume the status of naturalized perceptions. This leads us to think that the visual sign for 'cow' actually *is* (rather than *represents*) the animal, cow. But if we think of the visual representation of a cow in a manual on animal husbandry - and, even more, of the linguistic sign 'cow' - we can see that both, in different degrees, are *arbitrary* with respect to the concept of the animal they represent. The articulation of an arbitrary sign - whether visual or verbal - with the concept of a referent is the product not of nature but of convention, and the conventionalism of discourses requires the intervention, the support, of codes. Thus Eco has argued that iconic signs 'look like objects in the real world because they reproduce the conditions (that is, the codes) of perception in the viewer'.[5] These 'conditions of perception' are, however, the result of a highly coded, even if virtually unconscious, set of operations - decodings. This is as true of the photographic or televisual image as it is of any other sign. Iconic signs are, however, particularly vulnerable to being 'read' as natural because visual codes of perception are very widely distributed and because this type of sign is less arbitrary than a linguistic sign: the linguistic sign, 'cow' possesses *none* of the properties of the thing represented, whereas the visual sign appears to possess *some* of those properties.

This may help us to clarify a confusion in current linguistic theory and to define precisely how some key terms are being used in this article. Linguistic theory frequently employs the distinction 'denotation' and 'connotation'. The term 'denotation' is widely equated with the literal meaning of a sign: because this literal meaning is almost universally recognized, especially when visual discourse is being employed, 'denotation' has often been confused with a literal transcription of 'reality' in language - and thus with a 'natural sign', one produced without the

intervention of a code. 'Connotation', on the other hand, is employed simply to refer to less fixed and therefore more conventionalized and changeable, associative meanings, which clearly vary from instance to instance and therefore must depend on the intervention of codes.

We do *not* use the distinction - denotation/connotation - in this way. From our point of view, the distinction is an *analytic* one only. It is useful, in analysis, to be able to apply a rough rule of thumb which distinguishes those aspects of a sign which appear to be taken, in any language community at any point in time, as its 'literal' meaning (denotation) from the more associative meanings for the sign which it is possible to generate (connotation). But analytic distinctions must not be confused with distinctions in the real world. There will be very few instances in which signs organized in a discourse signify *only* their 'literal' (that is, near-universally consensualized) meaning. In actual discourse most signs will combine both the denotative and the connotative *aspects* (as redefined above). It may, then, be asked why we retain the distinction at all. It is largely a matter of analytic value. It is because signs appear to acquire their full ideological value - appear to be open to articulation with wider ideological discourses and meanings - at the level of their 'associative' meanings (that is, at the connotative level) - for here 'meanings' are *not* apparently fixed in natural perception (that is, they are not fully naturalized), and their fluidity of meaning and association can be more fully exploited and transformed.[6] So it is at the connotative *level* of the sign that situational ideologies alter and transform signification. At this level we can see more clearly the active intervention of ideologies in and on discourse: here, the sign is open to new accentuations and, in Vološinov's terms, enters fully into the struggle over meanings - the class struggle in language.[7] This does not mean that the denotative or 'literal' meaning is outside ideology. Indeed, we could say that its ideological value is strongly *fixed* - because it has become so fully universal and 'natural'. The terms 'denotation' and 'connotation', then, are merely useful analytic tools for distinguishing, in particular contexts, between not the presence/absence of ideology in language but the different levels at which ideologies and discourses intersect.[8]

The level of connotation of the visual sign, of its contextual reference and positioning in different discursive fields of meaning and association, is the point where *already coded* signs intersect with the deep semantic codes of a culture and take on additional, more active ideological dimensions. We might take an example from advertising discourse. Here, too, there is no 'purely denotative', and certainly no 'natural', representation. Every visual sign in advertising connotes a quality, situation, value or inference, which is present as an implication or implied meaning, depending on the connotational positioning. In Barthes's example, the sweater always signifies a 'warm garment' (denotation) and thus the activity/value of 'keeping warm'. But it is also possible, at its more connotative levels, to signify 'the coming of winter' or 'a cold day'. And, in the specialized sub-codes of fashion, sweater may also connote a fashionable style of *haute couture* or, alternatively, an informal style of dress. But set against the right visual background and positioned by the romantic sub-code, it may connote 'long autumn walk in the woods'.[9] Codes of this order

clearly contract relations for the sign with the wider universe of ideologies in a society. These codes are the means by which power and ideology are made to signify in particular discourses. They refer signs to the 'maps of meaning' into which any culture is classified; and those 'maps of social reality' have the whole range of social meanings, practices, and usages, power and interest 'written in' to them. The connotative levels of signifiers, Barthes remarked, 'have a close communication with culture, knowledge, history, and it is through them, so to speak, that the environmental world invades the linguistic and semantic system. They are, if you like, the fragments of ideology'.[10]

The so-called denotative *level* of the televisual sign is fixed by certain, very complex (but limited or 'closed') codes. But its connotative *level*, though also bounded, is more open, subject to more active *transformations*, which exploit its polysemic values. Any such already constituted sign is potentially transformable into more than one connotative configuration. Polysemy must not, however, be confused with pluralism. Connotative codes are *not* equal among themselves. Any society/culture tends, with varying degrees of closure, to impose its classifications of the social and cultural and political world. These constitute a *dominant cultural order*, though it is neither univocal nor uncontested. This question of the 'structure of discourses in dominance' is a crucial point. The different areas of social life appear to be mapped out into discursive domains, hierarchically organized into *dominant or preferred meanings*. New, problematic or troubling events, which breach our expectancies and run counter to our 'common-sense constructs', to our 'taken-for-granted' knowledge of social structures, must be assigned to their discursive domains before they can be said to 'make sense'. The most common way of 'mapping' them is to assign the new to some domain or other of the existing 'maps of problematic social reality'. We say *dominant*, not 'determined', because it is always possible to order, classify, assign and decode an event within more than one 'mapping'. But we say 'dominant' because there exists a pattern of 'preferred readings'; and these both have the institutional/political/ideological order imprinted in them and have themselves become institutionalized.[11] The domains of 'preferred meanings' have the whole social order embedded in them as a set of meanings, practices and beliefs: the everyday knowledge of social structures, of 'how things work for all practical purposes in this culture', the rank order of power and interest and the structure of legitimations, limits and sanctions. Thus to clarify a 'misunderstanding' at the connotative level, we must refer, *through* the codes, to the orders of social life, of economic and political power and of ideology. Further, since these mappings are 'structured in dominance' but not closed, the communicative process consists not in the unproblematic assignment of every visual item to its given position within a set of prearranged codes, but of *performative rules* – rules of competence and use, of logics-in-use – which seek actively to *enforce* or *pre-fer* one semantic domain over another and rule items into and out of their appropriate meaning-sets. Formal semiology has too often neglected this practice of *interpretative work*, though this constitutes, in fact, the real relations of broadcast practices in television.

In speaking of *dominant meanings*, then, we are not talking about a one-sided

process which governs how all events will be signified. It consists of the 'work' required to enforce, win plausibility for and command as legitimate a *decoding* of the event within the limit of dominant definitions in which it has been connotatively signified. Terni has remarked:

> By the word *reading* we mean not only the capacity to identify and decode a certain number of signs, but also the subjective capacity to put them into a creative relation between themselves and with other signs: a capacity which is, by itself, the condition for a complete awareness of one's total environment.[12]

Our quarrel here is with the notion of 'subjective capacity', as if the referent of a televisional discourse were an objective fact but the interpretative level were an individualized and private matter. Quite the opposite seems to be the case. The televisual practice takes 'objective' (that is, systemic) responsibility precisely for the relations which disparate signs contract with one another in any discursive instance, and thus continually rearranges, delimits and prescribes into what 'awareness of one's total environment' these items are arranged.

This brings us to the question of misunderstandings. Television producers who find their message 'failing to get across' are frequently concerned to straighten out the kinks in the communication chain, thus facilitating the 'effectiveness' of their communication. Much research which claims the objectivity of 'policy-oriented analysis' reproduces this administrative goal by attempting to discover how much of a message the audience recalls and to improve the extent of understanding. No doubt misunderstandings of a literal kind do exist. The viewer does not know the terms employed, cannot follow the complex logic of argument or exposition, is unfamiliar with the language, finds the concepts too alien or difficult or is foxed by the expository narrative. But more often broadcasters are concerned that the audience has failed to take the meaning as they - the broadcasters - intended. What they really mean to say is that viewers are not operating within the 'dominant' or 'preferred' code. Their ideal is 'perfectly transparent communication'. Instead, what they have to confront is 'systematically distorted communication'.[13]

In recent years discrepancies of this kind have usually been explained by reference to 'selective perception'. This is the door via which a residual pluralism evades the compulsions of a highly structured, asymmetrical and non-equivalent process. Of course, there will always be private, individual, variant readings. But 'selective perception' is almost never as selective, random or privatized as the concept suggests. The patterns exhibit, across individual variants, significant clusterings. Any new approach to audience studies will therefore have to begin with a critique of 'selective perception' theory.

It was argued earlier that since there is no necessary correspondence between encoding and decoding, the former can attempt to 'pre-fer' but cannot prescribe or guarantee the latter, which has its own conditions of existence. Unless they are wildly aberrant, encoding will have the effect of constructing some of the limits and parameters within which decodings will operate. If there were no limits, audiences could simply read whatever they liked into any message. No doubt some total

misunderstandings of this kind do exist. But the vast range must contain *some* degree of reciprocity between encoding and decoding moments, otherwise we could not speak of an effective communicative exchange at all. Nevertheless, this 'correspondence' is not given but constructed. It is not 'natural' but the product of an articulation between two distinct moments. And the former cannot determine or guarantee, in a simple sense, which decoding codes will be employed. Otherwise communication would be a perfectly equivalent circuit, and every message would be an instance of 'perfectly transparent communication'. We must think, then, of the variant articulations in which encoding/decoding can be combined. To elaborate on this, we offer a hypothetical analysis of some possible decoding positions, in order to reinforce the point of 'no necessary correspondence'.[14]

We identify *three* hypothetical positions from which decodings of a televisual discourse may be constructed. These need to be empirically tested and refined. But the argument that decodings do not follow inevitably from encodings, that they are not identical, reinforces the argument of 'no necessary correspondence'. It also helps to deconstruct the common-sense meaning of 'misunderstanding' in terms of a theory of 'systematically distorted communication'.

The first hypothetical position is that of the *dominant-hegemonic position.* When the viewer takes the connoted meaning from, say, a television newscast or current affairs programme full and straight, and decodes the message in terms of the reference code in which it has been encoded, we might say that the viewer *is operating inside the dominant code.* This is the ideal-typical case of 'perfectly transparent communication' - or as close as we are likely to come to it 'for all practical purposes'. Within this we can distinguish the positions produced by the *professional code.* This is the position (produced by what we perhaps ought to identify as the operation of a 'metacode') which the professional broadcasters assume when encoding a message which has *already* been signified in a hegemonic manner. The professional code is 'relatively independent' of the dominant code, in that it applies criteria and transformational operations of its own, especially those of a technico-practical nature. The professional code, however, operates *within* the 'hegemony' of the dominant code. Indeed, it serves to reproduce the dominant definitions precisely by bracketing their hegemonic quality and operating instead with displaced professional codings which foreground such apparently neutral-technical questions as visual quality, news and presentational values, televisual quality, 'professionalism' and so on. The hegemonic interpretations of, say, the politics of Northern Ireland, or the Chilean *coup* or the Industrial Relations Bill are principally generated by political and military elites: the particular choice of presentational occasions and formats, the selection of personnel, the choice of images, the staging of debates are selected and combined through the operation of the professional code. How the broadcasting professionals are able *both* to operate with 'relatively autonomous' codes of their own *and* to act in such a way as to reproduce (not without contradiction) the hegemonic signification of events is a complex matter which cannot be further spelled out here. It must suffice to say that the professionals are linked with the defining elites not only by the institutional position of broadcasting itself as an

'ideological apparatus',[15] but also by the structure of *access* (that is, the systematic 'over-accessing' of selective elite personnel and their 'definition of the situation' in television). It may even be said that the professional codes serve to reproduce hegemonic definitions specifically by *not overtly* biasing their operations in a dominant direction: ideological reproduction therefore takes place here inadvertently, unconsciously, 'behind men's backs'.[16] Of course, conflicts, contradictions and even misunderstandings regularly arise between the dominant and the professional significations and their signifying agencies.

The second position we would identify is that of the *negotiated code* or position. Majority audiences probably understand quite adequately what has been dominantly defined and professionally signified. The dominant definitions, however, are hegemonic precisely because they represent definitions of situations and events which are 'in dominance', (*global*). Dominant definitions connect events, implicitly or explicitly, to grand totalizations, to the great syntagmatic views-of-the-world: they take 'large views' of issues: they relate events to the 'national interest' or to the level of geo-politics, even if they make these connections in truncated, inverted or mystified ways. The definition of a hegemonic viewpoint is (a) that it defines within its terms the mental horizon, the universe, of possible meanings, of a whole sector of relations in a society or culture; and (b) that it carries with it the stamp of legitimacy – it appears coterminous with what is 'natural', 'inevitable', 'taken for granted' about the social order. Decoding within the *negotiated version* contains a mixture of adaptive and oppositional elements: it acknowledges the legitimacy of the hegemonic definitions to make the grand significations (abstract), while, at a more restricted, situational (situated) level, it makes its own ground rules – it operates with exceptions to the rule. It accords the privileged position to the dominant definitions of events while reserving the right to make a more negotiated application to 'local conditions', to its own more *corporate* positions. This negotiated version of the dominant ideology is thus shot through with contradictions, though these are only on certain occasions brought to full visibility. Negotiated codes operate through what we might call particular or situated logics: and these logics are sustained by their differential and unequal relation to the discourses and logics of power. The simplest example of a negotiated code is that which governs the response of a worker to the notion of an Industrial Relations Bill limiting the right to strike or to arguments for a wages freeze. At the level of the 'national interest' economic debate the decoder may adopt the hegemonic definition, agreeing that 'we must all pay ourselves less in order to combat inflation'. This, however, may have little or no relation to his/her willingness to go on strike for better pay and conditions or to oppose the Industrial Relations Bill at the level of shop-floor or union organization. We suspect that the great majority of so-called 'misunderstandings' arise from the contradictions and disjunctures between hegemonic-dominant encodings and negotiated-corporate decodings. It is just these mismatches in the levels which most provoke defining elites and professionals to identify a 'failure in communications'.

Finally, it is possible for a viewer perfectly to understand both the literal and the connotative inflection given by a discourse but to decode the message in a *globally*

contrary way. He/she detotalizes the message in the preferred code in order to retotalize the message within some alternative framework of reference. This is the case of the viewer who listens to a debate on the need to limit wages but 'reads' every mention of the 'national interest' as 'class interest'. He/she is operating with what we must call an *oppositional code.* One of the most significant political moments (they also coincide with crisis points within the broadcasting organizations themselves, for obvious reasons) is the point when events which are normally signified and decoded in a negotiated way begin to be given an oppositional reading. Here the 'politics of signification' - the struggle in discourse - is joined.