



UvA-DARE (Digital Academic Repository)

Relevance Theory as model for analysing visual and multimodal communication

Forceville, C.

DOI

[10.1515/9783110255492.51](https://doi.org/10.1515/9783110255492.51)

Publication date

2014

Document Version

Submitted manuscript

Published in

Visual communication

[Link to publication](#)

Citation for published version (APA):

Forceville, C. (2014). Relevance Theory as model for analysing visual and multimodal communication. In D. Machin (Ed.), *Visual communication* (pp. 51-70). (Handbooks of communication science; No. 4). De Gruyter Mouton.
<https://doi.org/10.1515/9783110255492.51>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

The following is a pre-print version of a chapter that has been published as:

Forceville, Charles (2014). Relevance Theory as model for analysing visual and multimodal communication.” In: David Machin (ed.), *Visual Communication*. Berlin: Mouton de Gruyter, pp. 51-70.

Please check the published version if you want to quote from it. ChF

Relevance Theory as model for analyzing visual and multimodal communication (version 26/12/'12) In: David Machin (ed.), *Visual Communication*. Berlin: Mouton de Gruyter

Charles Forceville
Dept. of Media Studies
University of Amsterdam

INTRODUCTION

In this chapter I will promote the usefulness of Relevance Theory as a model for the analysis of visuals and multimodal discourse. The discussion will not be restricted to visuals alone, since most of the time visuals are accompanied by information in one or more other modalities. This means I will first have to say something about multimodality. If the study of multimodal discourse is to develop into a respectable scholarly humanities discipline, each of the modes/modalities partaking in multimodal discourse needs to be theorized on its own. The question is thus first of all what, and how, a mode can communicate on its own. Now before anything else, this means we are faced with the formidable question of how “mode” is to be defined. Various attempts have been made (e.g., Jewitt 2009, Kress 2010, Elleström 2010), but the definitions diverge, and it is too early to accord any of them authoritative status. However, if there is no agreement on what constitutes a “mode,” *any* dimension of discursive meaning could qualify for modal status, and that would make the concept uselessly vacuous. I will not attempt my own definition here but be practical and consider as modes: (1) written language; (2) spoken language; (3) visuals; (4) music; (5) sound; (6) gestures; (7) olfaction; (8) touch. To this list all the boundary problems adhere that are typical of any categorization (why not split up the visual mode into color, size, orientation, and a host of others? where does sound end and music begin? shouldn't gestures be subdivided into facial expressions, arm/hand movements, and bodily postures?). However, these problems are not crucial, as long as it is acknowledged, first, that many categories allow for subdivisions and, second, that they tend to have fuzzy borders. Still, most categories have *prototypical* members, and it is via these that we can distinguish categories from one another (Lakoff 1987).

Elaborating on my earlier work (Forceville 1996: chapter 5, 2005, 2009; see also Yus 2008), I will here sketch how discussions of visual and multimodal discourse can be embedded in a more general theory of communication and cognition: Sperber and Wilson's Relevance Theory/RT (Sperber and Wilson 1986, 1995; Wilson and Sperber 2004, 2012). The focus of attention will be the visual mode, sometimes accompanied by the written verbal mode, but the idea is that the reasoning developed here is generalizable to other (combinations of) modes. Such a project should benefit both multimodality theory, which urgently needs more rigorous analytic models than have hitherto been

proposed for it, and RT, which while claiming to hold for all forms of communication has been mainly applied to its spoken verbal varieties. Let me acknowledge straightaway that while I intend to do justice to RT, its pioneers and advocates may not agree with parts of this project, if only because my use of the theory must be selective. Extending concepts that were developed primarily with the verbal modality in mind to cover a very different modality such as the visual means that some of these concepts need to be relaxed, or may even be found to be inapplicable. So be it. Scholars' models are there to account for reality, and if there is a mismatch between them it is the model which will have to be adapted. That being said, I also do not want to alienate non-linguists by using more of the sometimes daunting RT terminology than is necessary. So while I am happy to stand corrected if I should make demonstrable mistakes, and hope that RT scholars will help further improve my proposals, I take full responsibility for what some might consider a bastard version of RT.

A CRASH COURSE IN RELEVANCE THEORY

Relevance Theory is much indebted to Grice's (e.g., Grice 1975) Cooperative Principles. Grice distinguished four "maxims" that govern communication: the maxim of quantity (make your contribution optimally informative); the maxim of quality (don't lie, don't say things for which you have insufficient evidence); the maxim of relation (be relevant); and the maxim of manner (be perspicuous and thus avoid obscurity, ambiguity, prolixity, and be orderly (see Wilson and Sperber 2012: 3). Sperber and Wilson (henceforth S&W) contend that actually the maxim of relation, or relevance, is the only necessary one, the other Gricean maxims being subservient to it. Moreover, they see the Relevance Principle not as prescriptive, a rule to be obeyed, but as hardwired in humans' brains, something we simply can't help relying on. In this section I will briefly, and with no claim to exhaustiveness, discuss those concepts of RT that I think are essential for theorizing visual and multimodal discourse.

Communicative versus informative intention

Typical communication is ostensive-inferential (S&W 1995: 50-54): it is clear to both sender and addressee that the sender wants the addressee to be aware that she directs a message to him and that he is to infer relevant information from this message. (To avoid ambiguity in the choice of pronoun, I adopt S&W's convention to make the communicator consistently female and the addressee consistently male.) Within ostensive-inferential communication, S&W distinguish between two types of intentions, defined as follows:

Informative Intention: to make manifest or more manifest to the audience a set of assumptions **I** (S&W 1995: 58, emphasis in original).

Communicative Intention: to make it mutually manifest to audience and communicator that the communicator has this informative intention (S&W 1995: 61).

Specifically, “the true *communicative intention* is [...] the intention to have one’s informative intention recognised” (1995: 29, emphasis in original). The communicative intention, then, is the sender’s intention to get one or more addressees to accept certain assumptions **I** as (probably) correct, while the informative intention pertains to the contents of these assumptions. In spoken communication, it is usually difficult to distinguish between the two: as soon as Mary starts talking to Peter, he is aware of her communicative intention, and thanks to the contents of her message he is aware of the informative intention. It is nonetheless useful to keep the two apart. If I tell you that I won the lottery, and you understand my utterance, I have succeeded in fulfilling my informative intention; but only if you believe me, I have also succeeded in fulfilling my more encompassing communicative intention. Put differently: if you understand what I say to you, I have informed you; but only if you accept what I say, I have communicated with you. It is to be noted that the informative intention is also fulfilled while the communicative intention is not when, for instance, the addressee eavesdrops on the sender or when a sender quasi-accidentally gives an addressee access to the contents of a message while seemingly addressing somebody else (e.g., via a “mistakenly” sent reply-to-all e-mail). Even though the receiver may obtain valuable, indeed highly relevant information, in neither of these cases does the sender *address* the listener or reader. This means that the assumptions that are to be communicated are not made *mutually manifest* to the sender and the addressee, and so these examples do not exemplify ostensive-inferential communication.

For my purposes it is useful to emphasize one specific dimension of ostensive-inferential communication mentioned in passing by S&W, namely the signal(s) that one person wants to communicate something to another in the first place. In Forceville (1996) I considered this an aspect of the communicative intention. However, since this is incommensurate with S&W’s claim that fulfilment of the communicative intention *presupposes* fulfilment of the informative intention, it seems wiser to refrain from burdening the communicative intention with the signaling of a desire to communicate as such. Let me briefly expand on this matter (see also S&W 1995: 51). It may be the case that Mary makes mutually manifest her desire to communicate, but not (yet) her informative intention: she may clear her throat, whistle at Peter to draw his attention, run toward him frantically waving her arms, or emphatically try to catch his eye. In ordinary circumstances this will be a preamble to fulfilling her informative, and hopefully her communicative intention, but this need not be so. Peter may not hear her in a storm, she may stand behind sound-proof glass, or – having suffered a stroke – she may only produce incomprehensible babble. In these cases Mary’s intention to engage in ostensive communication, but not her informative, let alone her communicative intention may be recognized by Peter.

Effect (benefit) and effort (cost)

RT presupposes that human beings are a goal-directed species *par excellence*. We have micro- and macro-goals and are acutely aware that so do our fellow human beings. Inasmuch as we stand to gain from cooperation, at least from the within-group variety (a tenet held by socio-biologists – see Tomasello 2008, De Waal 2009 – and other scholars taking Darwin’s evolution theory as the best explanation for human behavior; see Boyd *et al.* 2010), it is not surprising that we are in principle prepared to help each other achieve

our goals. Effective communication is crucial in this respect. The central idea of RT is that a communicator cannot help but presume to be optimally relevant to her addressee. For a message to be relevant to a given addressee, that message must have an “effect” on the sum total of knowledge, beliefs, and emotions (called the “cognitive environment,” S&W 1995: 38) of that addressee. The time at which the message is communicated and the place where the addressee processes it is also part and parcel of the addressee’s cognitive environment. Typical examples of relevant messages are telling an addressee something he did not know (“today the shops close at 4 o’clock”), strengthening an assumption he already entertained (“remember you promised to repair my bike?”), or weakening an assumption he held (“I will probably not have time after all to go out for a drink with you tonight”). Importantly, sharing an emotion or evaluation with an addressee also counts as an “effect” (“isn’t this Gorgonzola exquisite?” “this is the saddest story I ever heard”). The greater the effect a message has on the cognitive environment of the addressee, the greater its relevance. Put differently, the greater the real or potential difference the uptake of the message will have on the addressee’s future decisions, actions, and attitudes, the greater the relevance. This means that a message’s relevance can vary from being minute and extremely short-lived to being vast and life-changing. But the benefits of relevance are offset by its costs, called “effort” in RT. The more mental effort an addressee must invest to recover a message’s effect, the less relevant it becomes. Relevance thus arises from a balance between effect and effort.

One of the most interesting and far-reaching claims of RT is that since he trusts his interlocutor to be optimally relevant, an addressee will typically stop interpreting a message as soon as it achieves relevance. The addressee assumes that the sender has selected the best possible or available stimulus, given the circumstances, to get across the message – which means that it is neither too short or simple, nor too long or complex. So as soon as the addressee deems he has found an interpretation of the message that strikes him as relevant, he will stop searching for further relevance. This means that, typically, in face-to-face conversation the first interpretation the addressee hits on is the one the sender intended him to derive. Thanks to this mechanism, communication can be as quick and, usually, efficient as it is (just imagine that an addressee would have to work out the possible meanings of each individual sentence, or indeed each individual word of a message...). In poetry, a text-genre that is known to require rereading, pondering, and reassessing, a reader may be aware that he needs to invest more effort to achieve optimal relevance than stopping at the very first interpretation, and he may indeed settle for an aesthetically pleasing ambiguity that he would find intolerable in most other forms of communication (see Pilkington 2000).

(En/de)coding versus inferencing

Each verbal message contains “objective” information: knowledge of the denotation of the words in it, and of the grammatical rules that determine the relation in which these words stand. Both the sender and the addressee of the message need to be in possession of the linguistic code (knowledge of vocabulary and grammar) to encode, respectively decode, this information. But this linguistic information, which is called the “logical form” of a sentence (“it is in virtue of its logical form that a conceptual representation is involved in logical processes and enters into relations such as contradiction or implication with other conceptual representations,” S&W 1995: 72), needs to be further developed

before it achieves the status of “explicit information” or, in S&W’s terminology, “explicatures.” Explicatures are propositions that are capable of being evaluated as true or false. Inferring explicatures comprises reference-attribution, disambiguation, and enrichment. If Mary tells Peter, “John will come soon,” Peter must instantly decide which “John” Mary means (answer: the first one he thinks of, John Baker, since otherwise, he trusts, Mary would have said “John Carpenter” to keep him from thinking she means John Baker [= reference attribution]). He also needs to decide whether, in the given situation (which is part of both Mary’s and Peter’s mutually manifest cognitive environment) “soon” means, say, “within minutes,” “within hours” or “within days” (= enrichment”). And in this same situation, say, it is also clear that Mary refers to John arriving, not coming in a sexual sense (= disambiguation).

But Peter knows that Mary has uttered “John will come soon” with a reason: she expects him to derive certain relevant effects from this information. That is, after having derived the explicatures of the message, he now needs to infer the relevant information it contains. He will do this by combining her message with information that is in his cognitive environment. Let us imagine that John will come to dinner. This is mutually manifest to Mary and Peter. In that light, it may be that Mary intends Peter to derive the implicated message: “I need to stop reading and start laying the table,” or “I am to go and run to get a bottle of wine from the liquor store,” or “Mary suggests we now stop our marital row in order be in a decent mood when our guest arrives,” or “Mary wants to reassure me that though John is late, he will turn up,” or all of them, or any of a range of others. These inferences, which are based on combining explicatures and assumptions that are mutually manifest to the interlocutors are not of the explicit but of the implicit kind, and are called “implicatures.” Sperber and Wilson distinguish between strong implicatures, which *must* be derived for the message to make sense, and weak implicatures, which the addressee derives at his own discretion.

Mary’s utterance presupposes a number of things – actually very many things. First of all, of course, Mary presupposes that Peter understands English. If Mary knew that this is not the case, she would not have chosen the phrase “John will come soon” in the first place (she might have tried another language, or attempted to mime the information). More generally, she presupposes that Peter in one way or another has an interest in John’s coming or not coming, and in the implications of this event. Even more generally, Mary takes numerous facts and evaluations pertaining to Peter, herself, and John, as well as their relationships to be mutually manifest to Peter and herself.

The more the uptake of a message depends on the addressee’s derivation of explicatures and self-evident, hard-to-miss-or-disagree-about implicatures, the more the message counts as “strong” communication. By contrast, the more its uptake relies on implicatures whose recruiting is not very self-evident, the more the message verges toward “weak” communication. One of the most insightful aspects of the RT model is in my view the following: the closer the message is to the “strong” end of the continuum, the more the responsibility for the derivation of implicatures resides with the sender of the message, while conversely the closer it is to the “weak” end of the continuum, the more the responsibility for deriving it rests with the addressee. If I shout “Fire!” at you, I take strong responsibility for conveying the message that there is a fire, which will probably lead you to get out of the building as soon as possible. If there then should turn out to be no fire at all, you would be fully justified to be indignant with me, saying I lied.

If, by contrast, I say, “I seem to perceive a somewhat smoky smell” you take far more responsibility yourself for deriving the message “there is a fire,” and you shouldn’t complain if this was not the case after all.

Relevance is always relevance to an individual

Most of the examples S&W (1995) use pertain to verbal exchanges between two individuals, Mary and Peter, who know each other very well (they may be married). This information is essential, since it means that Mary and Peter have a lot of mutually manifest information and history in their cognitive environments due to their shared past experiences. They thus presumably can often be very concise in their exchanges, since many things need not be spelled out between them. In the scenario of John about to arrive, if instead of her husband Peter Mary had addressed a neighbor who happened to have dropped in, she might have said, “John Baker, a friend of ours, will arrive in the next fifteen minutes or so for dinner,” hoping that the neighbor would derive the implicature he should go home. If Mary uttered this sentence to both her husband Peter and the neighbor, she would expect each of them to derive different implicatures.

Face-to-face verbal communication is the prototype form of communication in S&W’s model. In this form of communication, implicatures are not just derived by the addressee from combining explicatures, cognitive environment, and awareness of the sender’s goals, but can also be triggered by the sender’s intonation, facial expression (see de Brabanter 2010), and body posture, as well as the spatio-temporal circumstances in which the exchange takes place. When sender and addressee are no longer in the same space (as in telephone conversations) or time-frame (as in voice-mail messages), this inevitably affects what is, and is not, possible in the communication between them. In the two last examples, facial expressions and body postures cannot play a role, while in the last example there is not (as in a telephone conversation) an opportunity for immediate feedback or the correction of errors.

Matters become even more complicated in mass-communication. Let us consider mass-communication to start, for theoretical purposes, when the audience consists of more than one person simultaneously. This makes the example of Mary addressing both Peter and the neighbor a specimen of mass-communication, albeit an untypical one. I proposed that her utterance presumably achieved a different effect in both of them – but Mary probably had a fairly good idea of how it did. By contrast, when the US president delivers her inaugural speech, televised live around the world, her audience consists of many millions of addressees who will, in terms of command of English, gender, class, material welfare, geographical location and a host of other circumstances each have each their own, very different cognitive environments. The presidential televised message (delivered multimodally, drawing minimally on spoken language, visuals and gestures) will thus be processed marginally or vastly differently by each of the millions of addressees, because each of these latter has his own unique cognitive environment and interests. A consequence of this is that even in mass-communication, *relevance is always relevance to an individual*. This crucial point was made by S&W (S&W 1995: 142), but in my view they insufficiently emphasized it, probably because their focus on simultaneous, primarily verbal communication between two individuals did not foreground the necessity to investigate the implications of their insights when considering mass-communication (see also Forceville 1996: chapter 5).

Some points for consideration

Although the basic tenets of RT are fairly simple, its ramifications are by no means so, and there has been fierce debate about its merits and problems (see the discussions in *Behavioral and Brain Sciences* 10(4), 1987, and *Language and Literature* 6(2), 1997). In this section I will restrict myself to a few brief comments on issues that directly bear on the use I make of RT in the analysis of visual and multimodal discourse.

Babbling and lying. RT describes communication; it does not prescribe rules for its correct use. In this sense RT differs from Grice's cooperative principles, which are formulated in a "thou shalt (not) ..." manner. Critics might object that a madwoman's incomprehensible babble cannot be accounted for by RT. A strategy to counter this criticism could be to argue that the madwoman's babble still comes with the *presumption* of relevance (she believes herself to be relevant); or alternatively, we could maintain that the babble does not constitute ostensive-inferential communication – and thus does not count as communication. Notice, though, that "lying" is unproblematically accommodated in RT: liars hope that their addressees are unaware of the misleading nature of the explications and implicatures of their messages, and that these addressees therefore presume the senders to be optimally relevant. It is thus noteworthy that RT "has no independent maxim or convention of truthfulness" (Wilson and Matsui 2012: 209).

The best possible stimulus. According to RT the sender chooses the best possible stimulus that is commensurate with her preferences and/or her abilities. There can be numerous reasons why a sender's stimulus is not ideal. A speaker may choose a stimulus requiring more effort on the part of the addressee than a more straightforward formulation would have had, for instance for reasons of politeness. Or perhaps she has a far from perfect command of the language she uses, and therefore may have to select her words from the limited stock of words she masters. Or she may be gagged and handcuffed, and only hope to warn her prospective savior that the bad guy is standing behind the door with a hatchet, ready to kill him, by uttering muffled sounds at the top of her voice. Or she may subtly want to inform one person in a group at a party of something without alerting the others, making her favor "weak" communication which she hopes will be picked up by her intended addressee and ignored by the others. But given the circumstances it is nonetheless the best stimulus she can, or is willing to, provide.

Emotional effects. Within cognition studies, there is now increasing consensus that emotions are part and parcel of cognition, and RT is no exception. It may well be that the main, or even only, effect of a stimulus is the communication of an emotional response ("This is beautiful/terrifying/disgusting ..."). Even if the intended effect is primarily the activation of certain assumptions in the addressee, these often have affective dimensions as well; arguably, the reverse is no less true. I note in passing that Michael Tomasello emphasizes that successful communication does not only presuppose the transmission of information, but no less the mutually manifest awareness of the communicator's *attitude* to this transmitted information (Tomasello 2008: chapter 3.)

"Symptomatic" communication? What could be an RT approach to relevant information communicated without its sender being consciously aware of conveying it? Undoubtedly, communicators may unwittingly transmit crucially important information. Bordwell and Thompson (2008) coined the label "symptomatic meaning" for this: a communicator may not be aware of saying something important or disgusting or shocking

that happens to be highly relevant to the addressee. But this is not ostensive-inferential communication. The communicator did not *intend* to convey this information – just as certain clouds do not intend to tell those knowledgeable about clouds that it is going to rain, even though this conclusion may well be both important and correct. For this reason I propose to exclude “symptoms” from the realm of communication: they are by-effects of communication, not parts of it. Tanaka (1994) discusses another borderline case, namely that of “covert” communication, arguing it is typical of much advertising. Consider the following situation: we are both at a cocktail party and stand talking in different groups. I know that you are within hearing distance, and I say loudly about you, but not to you, “s/he is such an intelligent person!” This, again, does not count as ostensive-inferential information. After all, even though I may deliberately speak loudly in the hope that you will hear me, I do not make mutually manifest that I want to communicate something to you (i.e., the communicative intention is not fulfilled). That being said, it is sometimes difficult or even impossible to distinguish between weak communication and symptomatic meaning.

RT AND VISUAL & MULTIMODAL COMMUNICATION

In this paragraph, I will revisit the RT concepts discussed in the previous section with an eye to their applicability to visual and multimodal discourse. My central point is that RT can perfectly model visual communication – whether or not accompanied by language or other modes – but that this requires taking into account the affordances and constraints of the visual mode as well as of the fact that most visual and multimodal communication is of the mass-medial kind.

The presumption of relevance

The maker of a picture tries to be optimally relevant to her envisaged audience. Whether the picture is an illustration in a book, a political cartoon, an advertisement, or a hand-drawn map, its maker wants to attract the audience’s attention (ostensive communication), convey information and/or attitudes (informative intention) and thus to have an effect on this audience at no unnecessary effort (communicative intention). Usually, this effect is a quite precise one. The political cartoon criticizes a specific person or state of affairs at a certain time and place in history; the illustrations in the manual of your new printer indicate which cord goes where; the advertisement promotes a specific brand or service by focusing on some of its supposed qualities. These are mass-communicative pictures, whose genre their envisaged addressees usually recognize unproblematically. The pictorial message comes with the presumption of relevance. All signal: “hey, look at me, I’m worth your attention.”

Explicatures and implicatures

In order to make RT work for pictures, we need to adopt the distinction between the processes of en/decoding, pragmatically inferring explicatures, and pragmatically inferring implicatures. I think this is feasible. We first need to recognize the elements in the picture that have been depicted. In the *Tintin* panel reproduced in figure 1, we must recognize the patches of light green as grass, the darker green as trees in a wood or

jungle, the grey-white entity in the middle as a simple hut with a thatched roof and a brown door. Furthermore, we presumably recognize the two creatures minimally as a young man and a dog, respectively. We can equate this to the decoding stage. Everybody in possession of the required visual literacy “code” is able to do this. Viewers whose cognitive environment comprises knowledge of *Tintin* derive more information: even when they see the panel in decontextualized form, as in this case, such persons will recognize the redhead as Tintin and the dog as his loyal companion Snowy. We further assume, even though we only see part of Tintin’s left arm, that the hero actually has a complete arm, and that Tintin is walking (not running or flying). That is, we routinely perform reference assignment, disambiguation, and enrichment. Does this amount to claiming that we derive explicatures from this picture? In S&W’s theory explicatures are propositions, and thus can be evaluated as true or false. Critics might insist that only propositional *sentences* are capable of being true or false, which would then lead to the conclusion that pictures never have explicatures, because they do not have a logical form from which propositions can be generated. But I will suggest that we *can* say that the panel invites explicatures such as “this is a young man and a dog” or “these are Tintin and Snowy,” or even “Tintin and Snowy walk toward a hut in a forest.” So we would have to say either that this panel conveys multiple propositions, or that the only intended explicature is “Tintin and Snowy walk towards a hut in a forest,” the other propositions then being implicated premises (see Forceville 2005 for more examples of implicated premises).

The next step is inferring implicatures. Strong implicatures (implicatures for whose derivation the sender takes responsibility) include typically that the young man/Tintin and the dog/Snowy will enter the hut (rather than walk past the hut, or transform into dinosaurs in front of it). Further pragmatic inferences of the implicature variety will be fed by hypotheses about Tintin’s goals in this particular situation, which build both on information from all the previous panels in the album, and on our knowledge of the protagonist’s character and of typical story-plots.



Figure 1. Panel from Hergé, *Tintin et les Picaros* (1976), original in color. © Hergé/ Moulinsart 2013



Figure 2. Cartoon of Barack Obama, original in color. Accessed at 24 December 2012 at <http://noisyroom.net/blog/2012/06/30/obama-lies-taxes-rise/>

In figure 2, the decoding of the picture involves the assessment that a man with a shirt, tie, and jacket with an exceptionally long nose is depicted. Reference attribution, enrichment, and disambiguation lead to the conclusion that this is Barack Obama, at the moment of writing president of the United States. This assessment does not only hinge on resemblance between the depicted person and Obama; after all, the red-and-blue color scheme evokes the original poster in his 2008 election campaign and thus also cues “Obama.” The long nose intertextually refers to Pinocchio. Pinocchio is well-known for the fact that he pathologically lies and that by way of punishment his nose grows a little longer with every lie. I propose that all this amounts to the explicature, “Obama is a liar.” The textual “lies” anchors (in the sense of Barthes 1986; see also Unsworth and Cléirigh 2009) this interpretation, but for many viewers this text is redundant. The next stage is the inference of implicatures, such as that the communicator of this picture counsels viewers to mistrust Obama, or not to vote for him again.

Note that the envisaged addressee needs, again, to be able to recruit a lot of knowledge about the world from his cognitive environment. Somebody who does not recognize Obama, and has no idea that he is/was the president of the United States, and/or is not familiar with the Pinocchio story, will not derive the intended inferences. But it may also be the case that the viewer full well knows about Obama, but nonetheless fails to recognize him in this picture, and is not familiar with the color scheme and the style of the campaign poster. This would perhaps be similar to the failure of communication due to mis-hearing, or not knowing, a crucial word in an interlocutor’s utterance. It is also important that the viewer recruits information that pertains to the moment and place of the discourse under consideration (see Forceville 2005). The meaning of political cartoons, specifically, is very much tied to the situation of the moment of publication, drawing on knowledge that is briefly very salient and thus easily recruited, but quickly forgotten as time passes. This, incidentally, is why cartoons later re-published in book collections or on websites invariably require verbal elucidation to remind the audience of the precise socio-historical context in which they originally appeared.

Finally, it might be objected that one would have to get used to a certain style of drawing a house, a man, a beautiful girl, Barack Obama, before being able to derive pertinent explicatures from pictures. To the extent that this is true, one could counter that, in language, a listener unfamiliar with regional language variations has a less-than-perfect command of the code, and therefore is unable to derive certain explicatures in such dialectical utterances. Once that person has mastered certain words or expressions, he has solved this problem. In my view, this is not really different from having, in some cases, to “learn” the code of a cartoon style (say, Japanese Manga). Here is another example: an architect’s blueprint is a type of drawing that is supposed to be very precise indeed – unlike, for instance, the child’s stylized house with a few token windows and a door. But one has to know the code to derive the explicatures, and it is explicatures, explicit content, that count here.

Relevance to the individual in the mass-audience

Most communication involving pictures is mass-communication. As we saw, mass-communication is more complex than face-to-face communication in the sense that there is no opportunity for adapting the message in the case of misunderstanding or

incomprehension. Experienced mass-communicators of course anticipate this, so they carefully consider the form and contents of their message before releasing it, and may already know fairly precisely who their target audience is, thereby minimizing the risk of miscommunication. But mass-communication remains hazardous, simply because mass-communicators usually have to go for a “one size fits all” message, which “lands” differently in the numerous different cognitive environments of its numerous different addressees (exceptions include customized ads aimed at individual computer users, based on their Google search history). Specifically, mass-communicators have little control over the weak implicatures their mass-audience of individuals will derive.

As all communicators, a picture maker (whether an advertiser, a newspaper cartoonist, a graphic designer, or the passer-by who draws a little map upon your asking for directions) aims for a more or less specific effect in her audience. The friendly passer-by has a relatively easy job: she wants to help an audience of one (you) to reach your destination with the aid of her drawn-on-the-spot map. Since this is Mary-and-Peter communication, she can fine-tune her visual or multimodal (visuals + phrases such as street names) message in interaction with you. But in multimodal communication involving pictures this is an exceptional situation, since usually the audience is a mass-audience. When the communicator is a lecturer standing in front of a class of two hundred students, her verbal communication, gestures, and the diagrams and pictures on her PowerPoint constitute mass-communication in which there is at least a shared time-and-space frame. In most mass-communication, by contrast, the audience does not share spatio-temporality with the maker. Nonetheless the picture maker – just as the USA president delivering her televised speech – still has a fairly specific idea about her audience. The advertiser aims at selling goods or services to, or to increase brand awareness in, a target audience; the cartoonist wants the reader/viewer of the paper/magazine/site where it appears to critically ponder, and momentarily see the humor of, a certain political state of affairs; the graphic designer of a book desires to achieve clarity and perhaps aesthetic appreciation of its layout in its (prospective) reader; the architect hopes to achieve both admiration of the beauty of her building and awareness of its clever, functional design.

Clearly, each of these messages is, as always, highly contextualized. The picture appears at a certain moment in time and place, for a more or less specific audience: the target audience of a specific product flipping through a specific magazine; the reader of a left-wing journal in a certain country, after a certain newsworthy event has happened; the buyer of a Russian classic novel in hardback or of an American management book in paperback; you reading this very chapter sitting in your armchair or behind your laptop having downloaded an illegal copy.

The context within which a mass-audience accesses a message, that is, is to a considerable extent anticipated by its maker. First of all, the message belongs usually to a recognizable genre. A genre is for a discourse what an activity-type, as formulated by Goffman (1974), is for an action: as soon as we recognize that a given discourse belongs to a genre with which we are familiar, we know what conventions to recruit in interpreting it, and what responses are appropriate to it (Fokkema and Ibsch 2000). When we decide that something is an ad, we know the sender is trying to sell us something (see Pateman 1983, Forceville 1996); if we see somebody fall in a melodrama we feel sorry, if we see somebody fall in a comedy we laugh. *Genre is an element of context whose*

importance cannot be overestimated. Genre-attribution moreover occurs mostly subconsciously and in milliseconds, and is in my view the single most important element in the addressee's "cognitive environment" steering his strategy of interpretation of any pictorial or multimodal message (Forceville 2006). Whereas context is endless, and ever-changing, genre-attribution is quite stable and reliable. Indeed, I submit that "genre" more than any other contextual factor helps constrain what the relevance theorist Robyn Carston calls "free pragmatic processes: [...] pragmatic processes that contribute to what a speaker [or picture-maker, ChF] is taken to have explicitly communicated but which are not triggered or required by any linguistic [visual, ChF] property or feature of the utterance [picture, ChF]" (Carston 2010: 265).

This genre-attribution is first of all cued by text-internal semantic and "syntactic" information (I use quotation marks because the notion of grammatical rules applies at best metaphorically to non-verbal modalities: visuals do not have a grammar in the sense that language has, *pace* Kress and Van Leeuwen 1996, 2006; see Forceville 1999). But in addition, the spatio-temporal circumstances in which we encounter a discourse enormously facilitate correct genre-attribution. We usually know what genre to expect before we read/see/hear/feel the discourse itself, due to where and when we encounter it. The multiplex screening of the blockbuster Hollywood film, the JC Decaux billboard showcase on the street or the pop-up banner on internet containing an ad, the bottom-right corner on p. 3 of our newspaper featuring a political cartoon, the manual accompanying your newly-bought machine, the menu at the entrance of a restaurant presenting information about its dishes, prices, and atmosphere ... we already know a lot before actually accessing this information – although of course those out to trick us make deceptive use of these generic cues. In turn, where and when we are likely to encounter a discourse is largely determined by the agency or institution controlling such discourses (see Altman 1999 for a lucid account of how this works in film). The attribution of meaning to a message thus is guided by the chain "text-genre-medium-institution," of course always embedded in the presumption of relevance for the addressee.

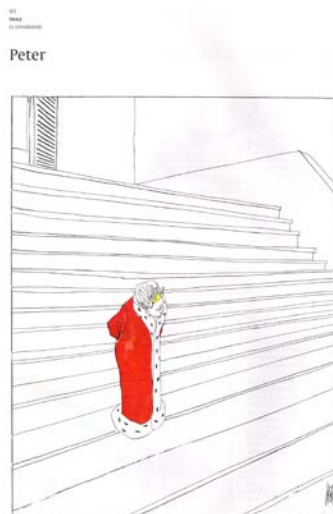


Figure 3. Political cartoon by Peter van Straaten, *Vrij Nederland* 27/1/'07, p. 80, original in color.

To elucidate this point, consider figure 3. Decoding this picture involves assessing that there is a single human figure standing on a big flight of steps, hands behind the back. Most viewers, moreover, would realize that the figure is a royal person (thanks to the crown, which is yellow in the original, and the ermine cloak, which is orange). For viewers knowledgeable about Dutch culture, even more specific information can be decoded: This is the Dutch queen, Beatrix. These enrichments enable certain explicatures, such as: “Beatrix stands on a flight of steps,” “Beatrix is alone,” “Beatrix looks toward the left.” Such viewers also have the following knowledge in their cognitive environment: as soon as, after elections, a new coalition cabinet has been formed, the monarch, surrounded by the ministers, poses on the steps of one of the royal palaces, “Paleis Huis ten Bosch,” for the official press photograph.

But yet other assumptions, retrieved from his cognitive environment, are recruited by the envisaged viewer: this is a cartoon on the “political cartoon” page of the left-wing weekly magazine *Vrij Nederland*. It is made by Peter van Straaten, which the *cognoscenti* know because of one or more of the following pieces of information: there is both the Barthesian anchoring “Peter” above the picture and his signature in the bottom right corner; Peter van Straaten *always* draws the cartoon in *Vrij Nederland*; and his cartoons have a recognizable style. The cartoon would moreover be typically accessed in the week of 27 January 2007, in *Vrij Nederland*. All these fragments of knowledge thus belong to the cognitive environment of the envisaged viewer: the reader of this *Vrij Nederland* issue, who is fully aware, at this moment in time, that it is taking ages for the new coalition government to be formed. All this information leads viewers to infer the following explicature, the one intended by Van Straaten: the Dutch Queen Beatrix impatiently waits for her new government to finally turn up for the official press photograph on the stairs of Paleis Huis ten Bosch, the photograph symbolizing that a new government has been formed.

Communicative versus informative intention

Communication without temporal synchronicity and spatial co-presence may be aimed specifically at an individual (think of a letter, e-mail, SMS, voice-mail), but more often it is not. The presidential speech, the Hollywood film, the advertising billboard, and the tweet are aimed at whoever perceives them as potentially relevant. Clearly any mass-communicator wants to reach as many members of her target audience as possible. Thus we can understand any attention-grabbing device of such a message as signaling its intention to engage in ostensive-inferential communication. Radio jingles, loud noises, scantily-clad ladies or ditto muscle men, big sizes, primary colors, unusual formats ...; the list of devices used to lure people into paying attention is long. Each of them can be considered mass-communicative equivalents of clearing one’s throat, gesturing, or starting to speak in face-to-face communication. In contrast to face-to-face communication, however, it is often quite easy to decline the invitation to heed the message; if so, the informative intention is not fulfilled. If it is, whether the communicative intention is subsequently fulfilled depends, as always, on whether the assumptions conveyed in the message are fruitfully processed in the cognitive environment of the addressee. Misinterpretation may be due to a faulty derivation of explicatures (due, for instance, to inadequate knowledge of the code, or incorrect procedures of disambiguation), or to the derivation of implicatures not intended by the

sender (due to associations evoked by the stimulus in the addressee's cognitive environment that were not foreseen by the sender).

Effect (benefit) and effort (cost)

The mechanisms of effect and effort apply in much the same way in visual and multimodal mass-communication as they do in face-to-face verbal communication. Just as a narcissist at a party may bore us to death with her presumption of the relevance of her self-aggrandizing chatter, so we may find a TV programme completely irrelevant. In the latter case, however, we have the opportunity to zap away after a few seconds, while we may not so conveniently be able to flee the party-bore. (If the bore is a lecturer lecturing, we may, due to considerations of politeness and institutional rules, not so easily run off, even though this is a form of mass-communication.) But equally, just as we are likely to pay close attention to the story of our beloved ones about something important that has befallen them because *they* are superbly relevant to us, so we may be prepared to invest a lot of effort grasping the text and diagrams of a text book whose contents we will be examined about, or the text and pictures in the manual of the DVD recorder we hope to get working. That is, there may be socio-cultural or practical reasons that encourage us to invest an inordinate amount of effort in interpreting a mass-media message.

Whys and wherefores

In this chapter I have fleshed out Relevance Theory's claim that it offers a model for visual and multimodal communication. But since the medium is the message, its central tenets must be tested, and where necessary expanded or adapted, for discourses in genres and media that deviate from the prototypical form of oral verbal communication in Sperber and Wilson's work.

Why is it useful to adopt RT as the overall model for communication? RT combines the attractiveness of high plausibility and precision in explaining how its central concepts apply to real-life communication. This precision allows for the systematic analysis of all forms of communication in all (combinations of) modes in all media. More specifically, it enables attesting what elements in a message contribute, or fail to contribute, which information towards achieving relevance for one or more addressees. Since communication is geared towards the optimization of relevance, all factors in the discourse that relate to this must be addressed. An asset of the model is that it permits comparisons between modes and media. I have argued that, like language, pictures can convey explicit content – and propose to theorize this explicit content in terms of “explicatures.” This proposal deserves extended discussion: if it is found to be untenable, we would have to conclude that after addressees have decoded visuals (the first stage of interpretation), they skip the derivation of explicatures and immediately move on to the derivation of implicatures. This is a crucial issue, for it would mean that pictures unaccompanied by information in other modalities, specifically the verbal one, could never serve as independent “proof” for anything, since we can construe no propositions from them that are evaluable as true or false. The pertinence for this with regard to visual evidence in documentary films as well as in court cases should be clear.

I have only begun the work to be done. Let me sketch some contours of the broader research program by asking further pertinent questions:

Communicators and addressees. Who are they? Is the communicator a single agent? If not, who/what is formally responsible for the message? Often this is impossible to determine. Someone is formally the maker (the journalist for her newspaper article, the author for her novel, the director for her film), but in mass-communication the communicator is usually part of an institution, in which sometimes many individuals have co-shaped a message. And who is/are the communicator(s) in forwarded mail? A re-tweet? A re-released film? Who are the interlocutors in a TV show or radio show (see Scannell 1996)? Are they the TV show's host and his/her guests? The studio audience? The TV audience? Discussing such questions with reference to the RT model can alert us to essential dimensions of mass-communication, but also suggest refinements of that model.

Institutions and media. Mass-communication takes place in an institutional context. Who have access to the media to send their messages? In what way is the best possible stimulus a communicator can or is willing to provide to achieve relevance guided or constrained by the technical dimensions of the medium? budgets? formats? scheduling slots? censorship? – often in ways that the audience has no way of knowing about? How is the situation for websites? blogs? social media?

Modalities. What are the modalities a given discourse can draw on? Is a discourse monomodal (such as lyric-less music and pantomime) or multimodal (drawing on two or more modalities, as in film, see Forceville 2006). What does each modality contribute to the optimization of relevance? To what extent does a modality allow for a distinction between the levels of en/decoding, and the derivation of explicatures, and (strong/weak) implicatures? While arguably it makes sense to talk about en/decoding non/verbal sound (in a film, or a radio play, for example), could we say the same about music? Do “textures” (pertinent to the tactile modality) require decoding, and do they have explicatures and implicatures? What modalities can be combined, and what does combining modalities yield in terms of meaning potential?

Genre and intertextuality. I have proposed that genre is the single most important element in the pragmatics of context-activation for discourse interpretation. Indeed, the very first thing a would-be addressee of a mass-communicative message does is assess the genre to which the message belongs. He then activates, from his cognitive environment, knowledge of the genre conventions that allow him to derive pertinent explicatures and implicatures. Something similar holds for having (or not having) knowledge of a pertinent “intertext” in one's cognitive environment. Given that individuals possess varying degrees of knowledge about a genre, or an intertext, they are bound to come up with slightly or vastly differing interpretations of a mass-communicative message.

Decoding visual signals and deriving their explicatures and implicatures. It could be argued that the only explicatures in figures 1-3 are “Tintin and Snowy walk toward a hut in a forest,” “Obama is a liar,” and “Queen Beatrix is impatiently waiting for the new Dutch government to be formed,” respectively, and that all other information is part of the implicated premises. Alternatively, we could say that these three pictures are all “multipositional.” Clearly the distinctions between decoding and inferring require further fine-tuning.

All of these questions can be addressed within RT, although (1) the model needs to be expanded; and (2) expert knowledge from many other disciplines needs to be

invoked. Systematic research within the RT model will also help define categories and clarify the always interesting fuzzy border areas between them: when does communication stop being communication? Where does music bleed into sound? Which elements in the visual mode are potentially meaningful – and under what conditions? Which elements in a picture are universally decodable (at least by sane adults) and which require more or less knowledge of (sub)cultural conventions?

Acknowledgments. I thank Belén Soria (University of Granada) for generously commenting on earlier drafts of this paper. Moreover I have benefited from discussing specific sections with Billy Clark and Francisco Yus. But it remains the privilege of the author to be solely responsible for all mistakes and controversial opinions.

Bibliography

- Altman, Rick (1999). *Film/Genre*. London: British Film Institute.
- Barthes, Roland (1986 [1964]). “Rhetoric of the image.” Trans. by Richard Howard. In: *The Responsibility of Forms* (21-40). Oxford: Blackwell.
- Bordwell, David, and Kirstin Thompson (2008). *Film Art* (8th ed.). Boston: McGraw-Hill.
- Boyd, Brian, Joseph Carroll, and Jonathan Gottschall, eds (2010). *Evolution, Literature & Film: A Reader*. New York: Columbia University Press.
- Brabanter, Philippe de (2010) “Uttering sentences made up of words and gestures.” In: Belén Soria and Esther Romero (eds), *Explicit Communication: Robyn Carston’s Pragmatics* (199-216). Basingstoke: Palgrave MacMillan.
- Carston, Robyn (2010). “Explicit communication and ‘free’ pragmatic enrichment.” In: Belén Soria and Ester Romero (eds), *Explicit Communication: Robyn Carston’s Pragmatics* (217-285). Basingstoke: Palgrave MacMillan.
- Elleström, Lars, ed. (2010). *Media Borders, Multimodality and Intermediality*. Basingstoke: Palgrave MacMillan.
- Fokkema, Douwe, and Elrud Ibsch (2000). *Knowledge and Commitment: A Problem-Oriented Approach to Literary Studies*. Amsterdam: Benjamins.
- Forceville, Charles (1996). *Pictorial Metaphor in Advertising*. London: Routledge.
- Forceville, Charles (1999). “Educating the eye? Kress and Van Leeuwen's *Reading Images: The Grammar of Visual Design* (1996).” *Language & Literature* 8(2): 163-178.
- Forceville, Charles (2005). “Addressing an audience: Time, place, and genre in Peter Van Straaten’s calendar cartoons.” *Humor* 18(3): 247-278.
- Forceville, Charles (2006). “Non-verbal and multimodal metaphor in a cognitivist framework: Agendas for research.” In: Gitte Kristiansen, Michel Achard, René Dirven, and Francisco Ruiz de Mendoza Ibáñez (eds), *Cognitive Linguistics: Current Applications and Future Perspectives* (379-402). Berlin: Mouton de Gruyter.
- Forceville, Charles (2009). “Relevanz und Prägnanz: Kunst als Kommunikation.“ Trans. by Martina Plümacher. *Zeitschrift für Semiotik* 31(1-2): 47-63.
- Goffman, Erving (1974). *Frame Analysis: An Essay on the Organization of Experience*.

- New York: Harper & Row.
- Grice, H. Paul (1975). "Logic and conversation." In: Peter Cole and Jerry L. Morgan (eds), *Syntax and Semantics 3: Speech Acts* (41-58). New York: Academic Press.
- Jewitt, Carey, ed. (2009). *The Routledge Handbook of Multimodal Analysis*. London: Routledge.
- Kress, Gunther (2010). *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge.
- Kress, Gunther, and Theo van Leeuwen (1996). *Reading Images: The Grammar of Visual Design*. London: Routledge.
- Kress, Gunther, and Theo van Leeuwen (2006). *Reading Images: The Grammar of Visual Design* (2nd, revised edition). London: Routledge.
- Lakoff, George (1987). *Women, Fire and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press.
- Pateman, Trevor (1983). "How is understanding an advertisement possible?" In: Howard Davis and Paul Walton (eds), *Language, Image, Media* (187-204). Oxford: Blackwell.
- Pilkington, Adrian (2000). *Poetic Effects: A Relevance Theory Perspective*. Amsterdam: Benjamins.
- Scannell, Paddy (1996). *Radio, Television & Modern Life: A Phenomenological Approach*. Oxford: Blackwell.
- Sperber, Dan, and Deirdre Wilson (1986). *Relevance Theory: Communication and Cognition*. Oxford: Blackwell.
- Sperber, Dan, and Deirdre Wilson (1995). *Relevance Theory: Communication and Cognition* (2nd ed.). Oxford: Blackwell.
- Tanaka, Keiko (1994). *Advertising Language: A Pragmatic Approach to Advertisements in Britain and Japan*. London: Routledge.
- Tomasello, Michael (2008). *Origins of Human Communication*. Cambridge Mass.: MIT Press.
- Unsworth, Len, and Chris Cléirigh (2009). "Multimodality and reading: The construction of meaning through image-text interaction." In: Carey Jewitt (ed.), *The Routledge Handbook of Multimodal Analysis* (151-163). London: Routledge.
- Waal, Frans de (2009). *The Age of Empathy: Nature's Lessons for a Kinder Society*. New York: Harmony.
- Wilson, Deirdre, and Dan Sperber (2004). "Relevance Theory." In: Laurence R. Horn and Gregory Ward (eds), *The Handbook of Pragmatics* (607-632). Malden MA: Blackwell.
- Wilson, Deirdre, and Tomoko Matsui (2012). "Recent approaches to bridging: Truth, coherence, relevance." In: Deirdre Wilson and Dan Sperber (eds), *Meaning and Relevance*. (187-209). Cambridge: Cambridge University Press.
- Yus, Francisco (2008). "Inferring from comics: A multi-stage account." In: Pelegrí Sancho Cremades, Carmen Gregori Signes, and Santiago Renard (eds), *El Discurs del Comic* (223-249). Valencia: University of Valencia.