

## The relevance of relevance: A review of Charles Forceville's *Visual and Multimodal Communication: Applying the Relevance Principle*

Daniel Barratt

Relevance theory (RT), originally proposed by Sperber and Wilson (1986/1995), is a theory of human communication that is based on a general view of human cognition. In the classic version of RT, the paradigmatic example of human communication is a verbal exchange between a single speaker and a single hearer who are occupying the same space and time. In the book *Visual and Multimodal Communication: Applying the Relevance Principle*, Charles Forceville (2020) has two main goals: first, to make classic RT accessible to a wider (and non-expert) audience; and second, to adapt RT in order to account for other forms of communication beyond the purely verbal – namely, visual and multimodal communication. In the second half of the book, Forceville demonstrates how this adapted version of RT can be applied in a series of case studies on “static visuals”. In this extended review, I attempt to do justice to Forceville’s ambitious project and to outline the key concepts and terminology involved. In the discussion, I present some criticisms and questions, while offering some suggestions as to how adapted RT can be developed even further in order to account for moving images and film.

**Keywords:** relevance theory, pragmatics, visual communication, multimodality, static images, moving images, cognitive film theory

### 1. Introduction

In their influential book *Relevance: Communication and Cognition*, Sperber and Wilson (1986/1995) proposed a theory of human communication that is based on a general view of human cognition. The central idea of *relevance theory* (RT) is that in any communicative situation between a speaker and a hearer, there is always a gap between the speaker’s literal meaning (what they actually say) and the speaker’s intended meaning (what they really mean). While Sperber has French nationality, both Wilson and the current reviewer originally hail from the United Kingdom. Taking inspiration from the London Underground, the catchphrase that always springs to my mind when thinking about RT is “mind the gap”: mind the gap between the train and the platform, mind the gap between the United Kingdom and the European Union, and mind the gap between the speaker’s literal meaning on the one hand and the speaker’s intended meaning on the other. How, then, to mind the gap without falling into the void in between (and without breaking your leg and missing the proverbial train)? As we will see, when it comes to issues of communication, the gap in question must be minded (and bridged) by the *Principle of Relevance* and by the process of cognitive inference, a process that involves both background knowledge and situational context.

RT was proposed as a radical alternative to a more traditional way of thinking about communication. According to Sperber and Wilson (1987, p. 697), this traditional way of thinking comes under the umbrella of the *code model of communication*. In the précis of their book, they state (albeit controversially; see Keller, 1998):

From Aristotle through to modern semiotics, all theories of communication [have been] based on a single model, which we call the *code model*. A *code* is a system which pairs internal messages with external signals, thus enabling two information-processing devices (organisms or machines) to communicate.

An example of a simple code is Morse code, where the signal of three dots, three lines, and three dots is paired with the message “SOS”, a message which is universally recognised as a distress signal but which, contrary to popular belief, does not stand for the phrase “save our souls” or “save our ship” (these phrases are actually backronyms). Meanwhile, an example of a more complex code are some aspects of the English language, where the phonetic shapes *kaa-fee*



/'kɑ:fi/ or *kaw-fee* /'kɔ:fi/ are paired with the meaning COFFEE which, in turn, can be combined with a variety of other morphemes/words to form a phrase or sentence.<sup>1</sup>

For Sperber and Wilson, the paradigmatic example of human communication is a *verbal exchange* between a single speaker and a single hearer (or a single sender and a single addressee) who are occupying the same space and time. Consider the following dialogue between the characters of Mary and Peter (Sperber & Wilson, 1987, p. 699; Forceville, 1994, p. 97):

- (1) *Peter*: Do you want some coffee?  
*Mary*: Coffee would keep me awake.

The key question here is: Does Mary actually want some coffee and why? According to Sperber and Wilson, the traditional code model of communication describes the decoding process which takes us from *phonetics* (the basic speech sounds that Mary makes) all the way to *semantics* (Mary's literal meaning, out of context): "Drinking a cup of coffee would keep me, Mary, awake." However, a sizeable gap still remains. In the words of Sperber and Wilson (1987, p. 697), this gap must be minded (and bridged) "not by more coding, but by inference".

This is where the so-called *inferential model of communication* fits into the picture: this model describes the inferential process which takes us from semantics all the way to *pragmatics* (Mary's intended meaning, in context). For the time being, two points should be noted. First, the inferential process hinges on the aforementioned Principle of Relevance. Of course, Mary could have answered Peter's question with a simple "Yes (please)" or "No (thank you)". For pragmatic reasons, however, Peter must assume that Mary, by choosing to say what she does, is trying to be optimally relevant: that is, she is trying to provide some additional information that is worth the extra processing effort. Second, the inferential process involves the integration of information that is external to the linguistic code: namely, background knowledge (e.g., coffee contains caffeine, caffeine is a stimulant) and situational context (e.g., it is late in the day, Mary wants to sleep).

But here comes the crux. If RT is proposed as a theory of (all) human communication, then it should be capable of accommodating types of communication beyond the verbal "Mary-and-Peter" variety originally described by Sperber and Wilson. First, it should be able to account for what is arguably the main form of human communication along with language: namely, *visual communication*. And second, it should be able to account for the possibility of two or more "modes" of communication working together: that is, *multimodal communication*.<sup>2</sup> These two questions bring us to the subject of this extended review, Charles Forceville's *Visual and Multimodal Communication: Applying the Relevance Principle*, published in the year 2020 by Oxford University Press. Forceville, who originally hails from the Netherlands, is Associate Professor of Media Studies at the University of Amsterdam. He has written extensively on the topics of pictorial (or visual) metaphors in advertising and the importance of RT as a theoretical model for understanding human communication. He is the author of *Pictorial Metaphor in Advertising* (1996) and the co-editor of *Multimodal Metaphor* (Forceville & Urios-Aparisi, 2009), *Creativity and the Agile Mind* (Veale, Feyaerts, & Forceville, 2013), and *Multimodal Argumentation and Rhetoric in Media Genres* (Tseronis & Forceville, 2017).

---

<sup>1</sup> Another example of a code model is the flow diagram which was originally drawn by Shannon and Weaver (1949) and which has been described as "the mother of all communication models". Given their engineering and mathematical background, Shannon and Weaver were primarily concerned with "level A" technical problems ("How accurately can the symbols [codes] of communication be transmitted?") rather than "level B" semantic problems or "level C" effectiveness problems (Fiske, 2011, p. 6).

<sup>2</sup> As Zlatev et al. (2023) point out, the term "multimodality" is ambiguous between denoting the combination of different sensory modalities (like vision and hearing) and the combination of different semiotic systems (like words and images). Therefore, a more precise term for what is discussed here might be *polysemiotic communication*. In this extended review, however, I will continue to use the term "multimodal communication" as this is the term that Forceville himself uses.



In the first half of his new (2020) book, Forceville gives an eloquent and engaging account of RT as originally conceived by Sperber and Wilson – where the paradigmatic example of human communication is a verbal exchange between a single speaker and a single hearer who are occupying the same space and time – and then embarks on the ambitious and important project of adapting RT in order to accommodate visual and multimodal communication. In the second half of the book, he demonstrates how this adapted version of RT can be applied in a series of detailed and illuminating case studies on different types of “static visuals” (p. 2): pictograms, traffic signs, and logos; advertising; political and non-political cartoons; comics; and controversial communication. In the following review, the more descriptive part will be organised by the main chapters of Forceville’s book (Sections 2 to 5), while the more critical and reflective part will be organised by the key arguments and questions that the reading of these chapters has inspired (Section 6). To use the language of RT, the overarching goal is to produce a review that does justice to Forceville’s book by achieving optimal relevance; that is, a review that hopefully generates the most cognitive (and emotional) effects for the least processing effort.

## 2. Sketching the general background

In Chapter 1, Forceville “sketches the general background against which RT is presented as a theory that can account for all forms of communication” (p. 4). His overarching goal is to demonstrate the following three points. First, that RT can be developed into a “comprehensive” theory of communication that includes the visual and multimodal communication of the book’s title; second, that RT is compatible with other “cognitivist” theories of communication; and third, that RT is based on “biological, Darwinian principles” and thus consistent with Darwin’s theory of evolution by natural selection. To cut a long and complex story very short, humans (and other primates) are intentional beings with shared goals; these goals ultimately relate to the “Darwinian” goals of survival and reproduction. Achieving these shared goals requires cooperation, and cooperation, in turn, requires such (behavioural/cognitive) capacities as joint action, joint attention, and, more often than not, some form of communication.

Forceville begins his sketch of the general background by looking at the intentional and cooperative nature of human communication (Section 1.2). This brings us to the work of Gibbs (1999) who “sees linguistic communication as subservient to intentional behavior, rather than the other way round” (p. 8). In other words, intentionality comes first and communication second.<sup>3</sup> As humans, furthermore, we are constantly monitoring our environment for intentional signals. Another of Gibbs’s main ideas is based on Grice’s (1957) distinction between “natural meaning” and “non-natural meaning”. To adapt one of Forceville’s examples, a spontaneous yawn during a lecture may indicate that a student is bored (a case of natural meaning), whereas a deliberate and exaggerated yawn can be interpreted as an intentional signal that the lecturer needs to up their game (a case of non-natural meaning). Only the latter case can be classified as an instance of true communication; the former case should be regarded as an external “symptom” of the student’s internal feelings. Forceville also considers the work of Clark (1996) and notes that “[j]ust as Gibbs considers linguistic communication as subservient to intentionality, so Clark considers it as subservient to joint action” (p. 11).

Although Forceville is primarily concerned with what he calls “static visuals” (p. 2), he also considers cognitive theories of communication that were originally proposed to account for the phenomenon of moving images (Section 1.3). This brings us to the field of cognitive film theory. For example, Bordwell (1989) proposes that, when viewers and critics try to “make sense of a film”, there are four possible types of meaning available: first, *referential meanings* which

---

<sup>3</sup> Note that the term “intentional” is being used here in the analytical philosophy or cognitive psychology sense of “purposeful”, rather than the sense used in phenomenology and cognitive semiotics of the directedness of consciousness beyond itself, according to which all mental acts, including those of perception, imagination, and belief, are intentional.



pertain to the “referents” (objects, characters, events, and so forth) depicted in a film’s story; second, *explicit meanings* which pertain to the overt “message” conveyed by a film and which do not require further interpretive elaboration; third, *implicit meanings* which are “implied” by aspects of a film’s story and which are not reducible to its referential or explicit meanings; and fourth, *symptomatic meanings* which are “symptomatic” of hidden (or “repressed”) motivations with respect to the filmmaker. Similarly, Grodal (2009) argues that film viewers process films using the same perceptual, cognitive, and emotional faculties that they use in real world interactions; furthermore, these faculties evolved by the process of natural selection to solve the problems encountered by our hunter-gatherer ancestors (also see Grodal’s PECMA flow model).

Having considered some of the insights from cognitive film theory, Forceville turns to the genre of visual art, a genre which he admits to largely ignoring in the rest of the book (Section 1.4). The rationale here is that even “high-art” images, which seem to be far removed from everyday life, have their roots in real-world principles of cognition and communication. For example, the work of Arnheim (1969) demonstrates that we make sense of visual art by employing the same Gestalt principles that we employ when looking at visual scenes in the real world, while the work of Gombrich (1999) suggests that even the highest and most aesthetic forms of visual art have “communicative as well as expressive intentions” (p. 21).

In the final part of the chapter (Sections 1.5 and 1.6), Forceville considers the evolutionary (and developmental) origins of communication. The story begins with Tomasello’s (2008, 2019) research on the similarities and differences between humans and apes. From an evolutionary perspective, this research shows that both human infants and apes are capable of requesting something by means of performative gestures. Developmentally, however, humans depart from apes by becoming capable of sharing information; this human capability requires joint attention and eventually the capacity for verbal communication. Interestingly, then, non-verbal communication seems to precede verbal communication both phylo- and ontogenetically. This story is largely corroborated by the work of De Waal (2009, 2016), the only significant point of disagreement being De Waal’s assertion that non-human primates are also capable of joint attention. To conclude, Forceville refers to a volume edited by Boyd, Carroll, & Gottschall (2010), the key arguments being that the humanities should be informed by and consistent with evolutionary theory, and that many humanistic phenomena (such as art and storytelling) can be partly explained in evolutionary terms.

### 3. Understanding relevance theory

Having sketched the general background of RT, Forceville devotes Chapter 2 to outlining the basic principles of the theory itself. His intended audience comprises non-experts and scholars working in the fields of visual and multimodal communication. Throughout the chapter, Forceville adopts Sperber and Wilson’s convention of citing examples in which the sender of the message is usually female and the addressee is usually male. Many of these examples involve our good friends Mary and Peter who, as Forceville later acknowledges (p. 49), seem to know each other extremely well and are probably married.<sup>4</sup> The original version of RT consists of a number of concepts and terms, the most important of which I will attempt to outline in the following section.

The story of RT begins with the most basic principles underlying human communication and Grice’s Cooperative Principles (Section 2.2). For example, Grice (1975) proposed that successful communication is dependent on both the sender and the addressee observing four

---

<sup>4</sup> Although in the first example cited in this review, Peter asks a question and ostensibly assumes the role of the sender, it is Mary’s reply (and Peter’s interpretation of it) that is of primary importance. Thus, as in any naturally flowing conversation between two people, the roles quickly reverse and Mary becomes the sender and Peter the addressee.



“maxims” of conversation (or, more precisely, four “categories” of maxims). Following Forceville’s lead, it is worth repeating those maxims again here:

- (1) the Maxim of Quantity counsels: make your contribution optimally informative;
- (2) the Maxim of Quality counsels: don’t lie, don’t say things for which you have insufficient evidence;
- (3) the Maxim of Relation counsels: be relevant;
- (4) the Maxim of Manner counsels: be perspicuous, avoid obscurity, ambiguity, and prolixity, and be orderly (see Sperber & Wilson, 2012, p. 3).

According to Sperber and Wilson, Grice’s four (categories of) maxims can be reduced (or boiled down) to just one: the Maxim of Relation or what they call the *Principle of Relevance*. Forceville observes that Grice’s maxims are prescriptive in nature (written in a “thou shalt (not)” form, p. 55), whereas Sperber and Wilson’s relevance principle is said to be “hardwired” in the human brain (p. 34). In the postface to the second edition of their book, Sperber and Wilson (1995, p. 260) make a distinction between two Principles of Relevance (also see Sperber & Wilson, 2008, p. 89). The *Cognitive Principle of Relevance* stipulates that “Human cognition tends to be geared to the maximisation of relevance”, while the *Communicative Principle of Relevance* stipulates that “Every act of [ostensive-inferential] communication communicates a presumption of its own optimal relevance”. Forceville notes that the relevance principle of RT is based on the second (communicative) principle which, in turn, has its roots in the first (cognitive) principle. In addition, Forceville argues that the latter is based on the “Darwinian” goals of survival and reproduction and has been adequately covered in Chapter 1.

The next part of the RT story concerns the intentional nature of human communication (Section 2.3). Sperber and Wilson describe the verbal exchange between a sender and an addressee as an instance of *ostensive-inferential communication* and make an important distinction between two types of intention. As in the case of the conversational maxims, it is worth repeating the key definitions here:

*Informative intention*: to make manifest or more manifest to the audience a set of assumptions (Sperber & Wilson, 1995, p. 58).

*Communicative intention*: to make it mutually manifest to audience and communicator that the communicator has this informative intention (Sperber & Wilson, 1995, p. 61).

It is surprisingly difficult to describe these two intentions in a concise and simple way (take it from someone who has repeatedly tried and failed). Tackling them in reverse order, Forceville notes that the communicative intention is “the sender’s intention to engage an envisaged addressee in communication”, while the informative intention relates to the informational content of the message (or “the contents of the assumptions in the message”) that the sender is trying to communicate (p. 35). In light of this formulation, Forceville argues that the communicative intention “logically precedes” the informative intention.<sup>5</sup>

In most cases of communication, both the communicative and informative intentions are satisfied (that is, recognised and fulfilled) simultaneously. In some cases, however, the two intentions become disconnected. Forceville cites five different scenarios: (1) communicative intention not recognised; (2) communicative intention recognised but not fulfilled; (3) communicative intention both recognised and fulfilled; (4) informative intention recognised but not fulfilled; and (5) informative intention both recognised and fulfilled. For the sake of

---

<sup>5</sup> I am not quite sure about this point. In most cases of communication, for example, one could argue that one must be in possession of a piece of information first before one can pass that piece of information on (communicatively) to someone else. In the case of *phatic communication* (e.g., Jakobson, 1960), on the other hand, the desire to communicate may come first and prompt the sender to search for a suitable topic of conversation (that is, information).



simplicity, I will mention just two adapted examples here (and replace the verbs “recognised” and “fulfilled” with the single term “satisfied”). If Mary is trying to talk to Peter but cannot be heard because of background noise, then the communicative intention has been satisfied but the informative intention has not. If, on the other hand, Mary accidentally includes Peter in an email correspondence (by clicking “reply to all” rather than “reply to sender”), then the informative intention has been satisfied but the communicative intention has not.

What is possibly missing from Forceville’s account is a straightforward definition of the term *ostension*. For our purposes, this can be defined as the act or process of pointing to, or drawing attention to, something. Through the very act of uttering a sentence, the speaker is drawing the hearer’s attention to something which needs to be processed. A non-verbal equivalent might be the speaker coughing in a deliberate and pointed way.

Once a communication has been initiated, the addressee must assume that the sender is trying to be optimally relevant (Section 2.4). According to RT, the relevance principle can be formally defined as a function of *cognitive effects (benefits)* and *processing effort (costs)*: the greater the number of cognitive effects and the smaller the processing effort, the greater the relevance (cf. cost-benefit analysis; Sperber & Wilson, 1995, p. 123). How, then, does this formal definition guide the addressee’s interpretation of the sender’s message? Forceville notes that the addressee can stop “their search for relevance” when they have arrived at an interpretation that yields the greatest number of cognitive effects for the least processing effort. To develop this account, the basic idea here can be illustrated by returning to example (1). Question: Why does Mary choose to reply “Coffee would keep me awake” when it requires more processing effort than the simpler options of either “Yes (please)” or “No (thank you)”? Answer: Because it potentially yields a greater number of cognitive effects (or implicatures, see below) such as “It is late in the day” and “Mary wants to go to sleep soon”, or alternatively, “It is late in the day” and “Mary needs to pull an all-nighter and study for an important exam”.

Once a communication has been initiated and relevance has been assumed, the addressee’s main task is to recover the meaning of the sender’s message (Section 2.5). This brings us back to the code model of communication. If the message is delivered in a familiar language, then Forceville notes that the addressee is able to *decode* much of the meaning “thanks to their knowledge of vocabulary and grammar” (p. 42). One of the key points, however, is that even the most comprehensive (linguistic) utterances are “underdetermined”. We can think of the process of linguistic decoding as being performed by a language processing module in the mind/brain (p. 43; see Section 6). The input of linguistic decoding can be regarded as the sender’s message *qua ostensive stimulus*, while the output can be described as a *logical form* (or “assumption schema”) which can be either “fully propositional” (that is, assessable in terms of truth and falsity) or not fully propositional and thus in need of further processing (p. 44).

Following Forceville, let us consider a second Mary-and-Peter example:

- (2) (a) *Mary*: John will be here soon.
- (b) *Mary*: He went to the bank.

In order to make sense of utterance (2a), it must be developed into a logical form which is fully propositional: for example, “John Carpenter is going to be in our house in about half an hour.” This, in turn, requires one or more of the following three procedures to be performed (examples adapted from Forceville). First, *reference assignment*: When decoding a typical noun phrase, the noun or pronoun at the root of that phrase must be “assigned” to a specific “referent”; that is, an identifiable object or person. In example (2a), Peter must decide that the noun *John* refers to their friend John Carpenter. Second, *disambiguation*: If a given word has several possible referents or several possible meanings, then the intended referent or meaning must be “disambiguated”. For example, Peter must decide that the noun *John* refers to their friend John Carpenter rather than their other friend John Smith (or, say, John Carpenter, the American director of the 1978 horror film *Halloween*). Third, *enrichment*: If a given word or phrase is suitably vague, then its meaning can be “enriched”. For example, Peter can extrapolate that the adverb



*soon* refers to a timescale of minutes (“in about half an hour”) rather than days and weeks. Another informative example is the sentence in (2b). In this case, reference assignment may determine that the pronoun *he* refers to a man called Simon Lewis, while disambiguation may determine that the homonym *bank* refers to a financial institution rather than a riverside.<sup>6</sup>

The next step involves the extraction or derivation of different types of information (Section 2.6). Following RT, the most important distinction here is that between *explicatures* and *implicatures*. An explicature is the information that is explicitly communicated by a message, while an implicature is the information that is implicitly communicated. To return to Mary and Peter, the complete formulation of “John Carpenter is going to be in our house in about half an hour” can be regarded as an explicature. The story does not end there, however. This explicature can be combined with contextual information (for example, the knowledge that John has been invited for dinner) to yield certain implicatures: for example, “I, Peter, need to stop reading the newspaper and start laying the table.” Another important distinction is that between *strong* and *weak implicatures* (and between “strong” and “weak communication”). The stronger the implicatures of a message (and the stronger the communication), the more their derivation is the sender’s responsibility; the weaker the implicatures (and the weaker the communication), the more their derivation is the addressee’s responsibility.<sup>7</sup>

Here endeth the main part of the RT lesson. Having outlined the basic principles of RT, Forceville uses the next two paragraphs (Section 2.7) to note one of the central tenets: that is, *relevance is always relevance to an individual*. In the paradigmatic example of verbal communication, a single sender (Mary) must make sure that her utterance is of optimal relevance for a single addressee (Peter). Since Mary and Peter seem to know each other extremely well and are probably married (as noted earlier), their cognitive environments are similar in terms of both background knowledge and situational context. My own way of thinking about this scenario is in the form of a Venn diagram with Mary and Peter’s cognitive environments being represented by two largely overlapping circles.

One of the trickier and more technical parts of the chapter (Section 2.8) concerns the distinction between *descriptive and interpretive uses of utterances*. The following account is my own interpretation of Forceville’s interpretation of Sperber and Wilson (1995, pp. 224-237). To begin with, it should be noted that Sperber and Wilson (1995, p. 231) assume that “every utterance is an interpretive expression of a thought of the speaker’s” (where a thought can be defined as a mental representation with a propositional form). Significantly, a thought/utterance can be used either descriptively or interpretively. A thought/utterance is used “descriptively” when it is a description of an actual state of affairs in the world: to adapt one of Forceville’s examples, the utterance “Willem-Alexander is King of the Netherlands” is literally true at the time of writing. Conversely, a thought/utterance is used “interpretively” when it is an interpretation of another thought/utterance: for example, Mary’s utterance “Robert is a bulldozer” is not literally true but metaphorically reflects Mary’s (more literal) thought that Robert is a tactless person.

In the penultimate part of the chapter (Section 2.9), Forceville returns to the notion of *symptomatic communication*. Following Bordwell (1989), Forceville proposes that “symptomatic meaning” is equivalent to Gricean “natural meaning” and cites the additional examples of “red spots signaling measles” and “dark clouds spelling rain” (p. 52). Crucially, such phenomena

---

<sup>6</sup> Where does decoding end and inference begin? At first glance, it seems that the operations of reference assignment, disambiguation, and enrichment require something akin to cognitive inference; that is, the activation and integration of information (regarding background knowledge and situational context) which is *external* to the linguistic code. It is plausible, however, that these operations can be explained in simpler, associative terms. For example, the disambiguation of the homonym *bank* can be explained in terms of associative relations and semantic priming: in a certain context, it is probable that the node for “bank as financial institution” will be activated and reach its threshold before the node for “bank as riverbank” (Morton, 1969; cited by Kintsch, 1988, p. 163).

<sup>7</sup> Also see Forceville’s discussion of “implied premises” and “implied conclusions” (pp. 47-48), and “implications” (pp. 135-136).



(while potentially meaningful to the observer) are non-ostensive in nature and thus do not fit with the ostensive-inferential account of communication proposed by RT. Forceville acknowledges that the question of where symptomatic communication (*qua* natural meaning) ends and ostensive-inferential communication begins is a “thorny issue”. Ultimately, however, the human tendency to interpret symptoms in meaningful ways is dependent on the first, Cognitive Principle of Relevance, rather than the second, Communicative Principle of Relevance.

In the final part of the chapter (Section 2.10), Forceville discusses “other points for consideration”. To begin with, he addresses some of the potential objections to RT – as originally presented in *Behavioral and Brain Sciences*, 10(4), 1987, and *Language and Literature*, 6(2), 1997. How, for example, can RT cope with such potentially problematic issues as *loose talk* and *lying*? On the issue of how RT deals with loose talk, Forceville notes that imprecise expressions can be partly explained in terms of minimising processing effort. When asked for the time, for example, we tend to round up or down to the nearest quarter hour (unless, say, the questioner is rushing to catch an unusually punctual train). On the issue of how RT deals with lying, however, Forceville’s response is not entirely convincing (RT simply “has no independent maxim or convention of truthfulness”; Wilson & Matsui, 2012, p. 209) and could be developed much further – I will return to this issue in Section 6. Another important point concerns the nature of *cognitive rewards*. To recap, Sperber and Wilson (1986/1995) formally define relevance as a function of *cognitive effects* and processing effort. A number of theorists (e.g., Wharton & Strey, 2019) have suggested that the notion of cognitive effects should be broadened to include *emotional effects*. As we will see, this broadening is significant when it comes to explaining both the cognitive and emotional impact of persuasive media such as advertising and film.

#### 4. Adapting relevance theory to accommodate visual and multimodal communication

In classic RT, the paradigmatic example of human communication is a verbal exchange between a single sender (Mary) and a single addressee (Peter) who are occupying the same space and time, and who have largely overlapping cognitive environments (as in my Venn diagram analogy). How, then, can classic RT be adapted in order to accommodate visual and multimodal communication? This is the subject of Chapters 3 to 5. Many aspects of classic RT – for example, the formal definition of relevance in terms of cognitive effects and processing effort – can be transferred to the adapted version with relatively few modifications. In this section, therefore, I will focus on those modifications which I judge to be of most importance when it comes to adapting what has already been described above.

Moving to Chapter 3, the most obvious challenge in adapting classic RT to accommodate visual communication lies in shifting from the verbal modality to the visual. The first tricky question to consider, then, is: *What is multimodality*? Different theorists have different ways of answering this question and different ways of slicing up the multimodal cake. One way of organising the different “modes” is in terms of the sensory channel (or “input system”), in which case the cake can be sliced into approximately six pieces: one for each of Aristotle’s classic five senses (sight, hearing, touch, taste, and smell) and one for language (cf. Fodor, 1983, 1985). Another way of organising the different “modes” is by semiotic resource, in which case we can make a distinction between, for example, so-called “visuals” and bodily expressions even though both of these resources are perceived via the eyes. Following Bateman (2016), Forceville tentatively subscribes to the following list of “modes”: (1) visuals; (2) written language; (3) spoken language; (4) bodily behaviour (which includes gestures, postures, and facial expressions); (5) sound; (6) music; (7) olfaction; (8) taste; (9) touch (p. 67).<sup>8</sup>

If visuals are a legitimate “mode” of communication, then the second tricky question is: *How can visuals be understood in terms of the traditional code model*? Forceville states: “unlike verbal utterances, *visuals have neither a vocabulary nor a grammar* in the strict sense in which this holds for language.” (p. 75). This is consistent with Messaris (1997) who argues that images

<sup>8</sup> For a more comprehensive answer to the question of “multimodality”, see Zlatev et al. (2023).



are “propositionally indeterminate”: unlike natural languages, images lack the symbolic and syntactic devices necessary for making explicit propositional arguments. For example, Messaris argues that images are good at conveying spatial and temporal relationships, but poor at conveying relationships such as comparison/analogy, contrast, causality, and generalisation.

To continue with the code model, the main idea that Forceville borrows from semiotics is Peirce’s distinction between symbol, icon, and index (p. 79) and Chandler’s (2017, p. 41) somewhat simplified definition of these terms. Here is my own take, in a slightly revised order. In the case of *icons*, there is a degree of resemblance between the sign and the object: for example, a picture of a tree looks like an actual tree in a significant number of respects. In the case of *indexes*, on the other hand, there is a contiguous relationship between sign and object: for example, smoke is caused by and thus an indication of fire, while a pointing gesture is connected with its referent in terms of space and time. In the case of *symbols*, on the other hand, the relationship between sign and object is conventional (that is, not motivated by reality in terms of either resemblance or contiguity): for example, using the word *tree* to refer to a tree is “a matter of convention, agreement, or rule” (Fiske, 2011, p. 45).

As we have seen, the process of “decoding” verbal information requires the procedures of reference assignment, disambiguation, and enrichment. The next question, then, is: *Do each of these procedures have an equivalent when it comes to the “decoding” of visual information?* Finding a visual equivalent for reference assignment is relatively straightforward. Given the Peircean definition of iconicity just cited, an image of a person (e.g., a satirical cartoon of Barack Obama; see Forceville, 2014) can be connected to the actual person (referent) on the basis of visual resemblance. Finding a visual equivalent for disambiguation, on the other hand, is more problematic. In a brief discussion, Forceville states that the closest potential candidates are ambiguous figures such as Arcimboldo’s “fruit-and-vegetable faces”. (Other notable examples of ambiguous figures, not cited by Forceville, are Jastrow’s duck-rabbit and Rubin’s face-vase.) In such cases, however, Forceville argues that “we are not expected to resolve the ambiguity but to relish it” (p. 84); therefore, he reaches the provisional conclusion that there is no straightforward visual equivalent for disambiguation in language. I think that Forceville is too hasty here and in fact some convincing examples can be found later in the book.<sup>9</sup>

Moving to Chapter 4, in the paradigmatic example of verbal communication (involving Mary and Peter), there is a single addressee: namely, Peter. In the case of *mediated-mass communication*, however, there are multiple addressees, potentially occupying different spaces and times, and potentially having different (or non-overlapping) cognitive environments. How, then, can classic RT be adapted to account for multiple addressees given the cognitive differences potentially involved?

Proceeding in small steps, Forceville returns to the Mary-and-Peter dialogue previously mentioned (2a; “John will be here soon”), but adds one extra person, a neighbour called Adrian, to the equation. Adrian has made an unexpected visit and, crucially, has a different agenda and a different cognitive environment to that of Peter. The consequence is that Mary must adjust her original statement (2a) to something like (3):

(3) John will arrive in about fifteen minutes or so for dinner.

This additional information requires extra processing effort on the part of Peter, but allows Adrian to derive the implicature “Adrian, you need to go home now” without Mary being unnecessarily impolite and without Adrian taking unnecessary offense. Forceville contends that this example demonstrates that RT is potentially capable of accommodating “one-to-two”

---

<sup>9</sup> For example, in the Tintin panel considered in Section 5 (Figure 4), one could argue that the lines to the right of Snowy can be disambiguated as blades of grass, while the entrance to the hut can be disambiguated retrospectively (on the basis of the subsequent panel) as a curtain rather than a door.



communication, although one could also argue that the role of politeness here is a possible reason for reinstating Grice's Maxim of Manner – I will return to this point in Section 6.<sup>10</sup>

The next small step is to introduce mediation to the equation. Following Yus (2011), Forceville briefly considers different types of mediated one-on-one exchanges. His main observation, however, is that: "Once the number of addressees in an audience goes beyond a certain number, communication must be somehow mediated." (p. 103) After all, even live communication is often mediated by means of microphones/loud speakers and cameras/video screens. To this observation, I would add a simple formula: *The greater the number of addressees (or the larger the "mass" of the audience), the greater the probability that the corresponding communication is mediated in some way.*

Talk of mediated communication leads to the tricky question: *What is a medium?* This is just as tricky as the question: *What is multimodality?* Forceville considers a few possible ways of answering this question, but does not reach a definitive conclusion. And perhaps it does not really matter for our current purposes. The crucial point with respect to RT is that mediated communication, however it is achieved, allows a given message to reach an increasing number of addressees in increasingly divergent spaces and times. Here, I would add another simple formula: *The greater the divergence in space and time, the greater the potential divergence in the addressees' cognitive environments.*

In the remainder of the chapter, Forceville touches on a few ways in which these potential differences can be constrained: for example, the notion of a "target audience" suggests that the individuals in question have certain properties (demographic and otherwise) in common. Similarly, individuals can be collected into formal or informal groups who share certain knowledge domains and the like. Whatever the number of addressees and whatever the potential cognitive differences, Forceville asserts that one of the central tenets of RT – that "relevance is always relevance to an individual" – still holds (p. 111).

In Chapter 5, Forceville advances an argument for the importance of *genre*. Even though individual viewers may differ from each other in terms of their cognitive environments, the type of genre involved functions to constrain those cognitive environments by providing a set of conventions which influence the viewer's expectations and interpretations in quite specific and reliable ways. Later in the chapter, Forceville points out that Sperber and Wilson have nothing to say about genre, and speculates that this is because they focus their attention on Mary-and-Peter conversations where genre is not an issue (p. 125).

The first question to consider is an ontological one: *What is genre?* Following the French translation, Forceville begins by noting that he is using the term in the broad sense of "type" or "kind" (p. 118; Neale, 2000, p. 9). Turning to the work of Altman (1984), Forceville then considers the (non-static) medium of film and the categorical world of film genres. In short, the film genre of the Western gives the viewer an understanding in terms of both semantics (semantic elements; what the basic building blocks are) and syntax (syntactical structures; how those basic building blocks fit together). To quote Forceville, the semantic elements of the Western include "cowboys, Indians, saloons, revolvers, lassos", and so forth (p. 118). Meanwhile, and to describe my own stereotypical image of the Western, an example of a syntactical structure would be the lone cowboy riding into a dusty town on his faithful horse and then striding through the swinging doors of the aforementioned saloon armed with the aforementioned revolver (and all the events that happen thereafter). Following Lakoff (1987) (who was inspired by the earlier work of Rosch), Forceville proposes that the categories of genres (film and otherwise) can be understood in terms of prototype theory.

The second question to consider is: *What role does genre play?* While the context of a given discourse is highly variable, genre is relatively stable and can be identified surprisingly quickly, often in a matter of milliseconds. One of the litmus tests for this is to try "channel hopping" on television and see how quickly you can identify the genre of a given programme before hopping

---

<sup>10</sup> Forceville also considers an example of "one-to-many" communication in which "Mary is a professor delivering a lecture to a group of 100 students" (p. 102).



to the next channel. Significantly, Forceville proposes that genre functions as “an interface between discourse and context” (p. 121). This is an intriguing idea and can be tentatively formulated as follows:

$$\begin{aligned} \text{discourse (fixed) + context (variable)} &= \text{interpretations X, Y, Z} \\ \text{discourse (fixed) + genre (stable) + context (variable)} &= \text{interpretation X} \end{aligned}$$

## 5. Applying relevance theory to visual and multimodal communication

In the second half of the book, Forceville gives examples of how an adapted form of relevance theory (RT) can be applied to a series of case studies on various examples of visual and multimodal communication: pictograms, traffic signs, and logos (Chapter 6); advertising (Chapter 7); political and non-political cartoons (Chapter 8); comics (Chapter 9); and controversial communication (Chapter 10). It should be noted that Chapters 6 to 9 are organised by genre (p. 131) and ordered in terms of complexity, while Chapter 10 (on the subject of controversial communication) is independent of both factors.

In Chapter 6, Forceville begins the case studies by examining arguably the simplest types of visuals: namely, *pictograms*, *traffic signs*, and *logos*. These visuals can be thought of as predominantly symbolic or conventional signs in public spaces. According to Forceville, the meaning of these visuals is “decoded rather than inferred” (p. 133), hence the later term “coded visuals”. His main claim here is that if these visuals are used ostensively, then they give rise to explicit information or what he terms “(visual) explicatures”. By way of background, Forceville cites one of the few times that Sperber and Wilson (1995, p. 175) comment on the visual domain: “To varying degrees, all non-verbal communication is weak communication.” Forceville argues that this claim does not apply to the case of coded visuals which can be regarded as instances of “strong communication”.

The first pictograms that Forceville analyses are those for “save”, “exit”, “start”, and “gents”. As evidenced by the previous sentence, Forceville notes that each of these pictograms can be substituted for (or is equivalent to) a single word. To put it another way, each pictogram can be straightforwardly “decoded” into a single word (and its associated concept). Furthermore, Forceville suggests that this decoding can be developed (or elaborated) into a fully propositional form by adding something like “here is ...” on the basis of the situational context: for example, “here is the save button”. Although pictograms primarily operate as Peircean symbols, Forceville observes that they are also iconically motivated and thus not arbitrary: for example, the “save” pictogram visually resembles the old-fashioned computer diskettes that were originally used for storing (or “saving”) electronic data (see Figure 1a).<sup>11</sup>

Another issue to consider is whether or not pictograms can be classified as ostensive stimuli in the RT sense. To develop Forceville’s line of reasoning, if the designer of a given software deliberately uses the “save” pictogram to indicate the save button, then one could argue that the roots of both the communicative and informative intentions are in place (at least at the sender side of the equation). The question of whether these intentions are recognised and/or fulfilled, however, depends largely on the knowledge and the needs of the user (at the addressee side).

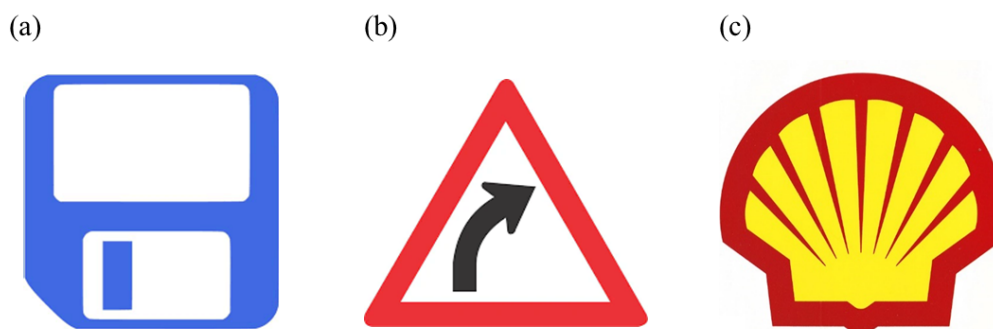
For those fans of good hotels and restaurants (who also have an interest in visual communication), Forceville concludes his analysis of pictograms by looking at the visual methods for signalling comfort and quality in the celebrated *Michelin Guide*. In the remainder of the chapter, Forceville takes a look at traffic signs and logos. Traffic signs can be regarded as a more advanced form of pictograms. While some aspects of traffic signs are iconically motivated in the Peircean sense, other aspects are symbolically coded in terms of both colour and form. For

---

<sup>11</sup> Interestingly, many of the pictograms and icons used in today’s computers and smartphones are based on increasingly old technologies which raises the question of how young people will learn to recognise them in the future.



example, “warning signs” tend to be red and have a triangular shape (Figure 1b). Meanwhile, the logos for companies or brands can be regarded as combinations of Peircean symbols, icons, and (in some cases) indexes. For example, the widely recognised logo for Shell (symbol) visually resembles a shell (icon) and, if placed next to a Shell station, also points to a location where the addressee can potentially refill their car with gas/petrol (index) (Figure 1c).



**Figure 1.** (a) “Save” pictogram; (b) “Curve to the right ahead” traffic sign; (c) Shell logo.

In Chapter 7, Forceville moves from coded visuals to the types of stimuli used in *advertising*. He begins by noting four important differences between the two: first, the types of stimuli used in advertising usually combine the visual and verbal modes (images and words); second, the interpretation of adverts requires inferring (in addition to coding); third, the interpretation of adverts is less dependent on location and more dependent on internal elements; and fourth, the analysis of adverts can be subdivided into various (visual and verbal) “units” (p. 149).

Regarding the genre of advertising, Forceville quotes from his 1996 book: “The single most important genre convention of (commercial) advertising is that it presents a product or service in a positive light.” (Forceville, 1996, p. 104) In terms of RT, the advertiser has done their job if the addressee: (1) notices and (2) pays attention to the advertisement (communicative intention both recognised and fulfilled), and (3) understands and (4) accepts the claim made about the advertised product (informative intention both recognised and fulfilled) (p. 151). Interestingly, Forceville demonstrates the power of both genre and RT by analysing a Chinese advert, the key point being that a surprisingly large amount of meaning can be extracted even before the Chinese text has been translated. In the remainder of the chapter, Forceville analyses a range of interesting cases in order to illustrate both different types of advertising and different aspects of RT: Samsonite (an advert based on a visual/multimodal metaphor); KitKat (an example of “outdoor advertising”); KiKa (an advert for a Dutch not-for-profit organisation, “kids free of cancer”, rather than a commercial product); Assim Saúde (an advert for a Brazilian medical insurance company sponsoring a rock festival in Rio de Janeiro); and finally, Coca-Cola (an advert with four bottles and the tagline “Good things come in fours”, thus subverting the well-known phrase; Figure 2).





Figure 2. Billboard © Coca-Cola (2015).

In Chapter 8, Forceville moves from the genre of print advertising to the genre of *political (and non-political) cartoons*. This time, he points out the similarities between the two genres rather than the differences (here presented in a slightly revised order in order to rhyme with the above): first, both genres usually combine the visual and verbal modalities (images and words); second, both genres require decoding and inferring; third, both genres can be separated to a certain extent from time and place (cartoons “allow for a certain de- and re-contextualization”, p. 167); fourth, both genres can be subdivided into units; and fifth, both genres are capable of making persuasive claims about the “real” world and thus “feed into” visual argumentation theory (VAT).

Regarding the issue of genre, Forceville notes that whereas the goal of (commercial) advertising is to present products in a positive light, the goal of political cartoons is to make “negative and critical claims about a public figure or a state of affairs in the world” (p. 168). In terms of RT, Forceville makes a number of interesting observations: for example, cartoons rely on caricatures and stereotypes in order to facilitate the process of reference assignment; cartoons rely on the “loose use” of visuals in order to reduce processing effort; and cartoons present verbal information in either text balloons inside the panel or in quotation marks below the panel. Forceville analyses a variety of cartoons, including a cartoon by Luojie featuring Walt Disney’s version of Snow White (which is described by Forceville as “truly multimodal”, p.173; Figure 3) and a cartoon by the Netherlands’ very own Peter van Straaten (which requires background knowledge of both the Dutch monarchy and the Dutch national elections). In the summary to the chapter, Forceville acknowledges that a full analysis of cartoons requires additional theories (pp. 183-184): semiotics (intertextual references; Chandler, 2017), blending theory/conceptual integration theory (integrating two or more input spaces to create a blended space; Fauconnier & Turner, 2002), and conceptual metaphor theory (identifying the relevant source and target domains; Lakoff & Johnson, 1980; also see Forceville, 2008).





**Figure 3.** Cartoon © Luo Jie (21 April, 2011). Licensed by: [www.cartoonstock.com](http://www.cartoonstock.com)

While Forceville focuses in Chapters 6 to 8 on stand-alone visuals or “word-and-image texts”, he turns in Chapter 9 to sequences of visuals or “word-and-image texts”, otherwise known as *comics* (described by Eisner, 2006/1985, as the “sequential art”). Although comics are commonly perceived as a 20th century phenomenon, there are some notable historical precursors. For example, Forceville mentions pre-historic cave paintings and the medieval paintings of biblical scenes presented in churches. Another potential example, not mentioned by Forceville, is the famous Bayeux Tapestry which has been described as an early incarnation of the comic strip. While the original target audience for comics was children and teenagers, the more recent development of graphic novels has seen comics become a legitimate form of entertainment for adults.

Turning to the genre of comics, Forceville acknowledges the theoretical work of McCloud (1993) and others, and notes that comics are comprised of eight different types of elements: (1) depictions of objects, characters, actions, and events; (2) pictograms; (3) movement and emotion lines (“pictorial runes”); (4) speech and thought balloons; text blocks with narratorial comments; (6) onomatopoeia; (7) navigation from one panel to another; and (8) multiple relationships between panels. Having outlined the basic building blocks, Forceville analyses a variety of comics which involve either some or all of the aforementioned elements. The first example, a panel from the Tintin comics album *Tintin and the Picaros* (Hergé, 1976), is of particular interest as it touches on the basic definitions of “visual explicatures” and other key terms from RT (Figure 4) – I will return to this example in Section 6. Other notable examples include the “visual instructions” for an IKEA product, 12 panels from the graphic novel *The Arrival* (Tan, 2006) which rely solely on visuals, and seven panels from the superhero comics album *Watchmen* (Moore & Gibbons, 1986) for which Forceville provides a “painstaking (but nonetheless non-exhaustive) analysis” (p. 213) of the visual and verbal information contained within. One thing that strikes me when reading this chapter is the similarity between comics and another “sequential art”, film: both arts involve “storyboards” (in the pre-production phase in the case of film) and both rely on the viewer employing schemas and making basic assumptions in order to fill in the gaps (see Bordwell, 1985).





**Figure 4.** Panel 546 from *Tintin and the Picaros* (Hergé, 1976, p. 45).

The final case study chapter, Chapter 10, concerns a type of communication that Forceville refers to as *controversial communication*, which is independent of both genre and complexity. All of the examples discussed in this chapter have one thing in common: the sender intends to deceive the addressee in some way. Thus, all of the examples are concerned with the issue of truth and truthfulness (or, more precisely, the lack of it). Forceville states: “A first vital point, made earlier in Chapter 2, is that RT does not have a principle of truthfulness. The RT model can thus accommodate the fact that there are many less than fully trustworthy communicators as well as completely immoral frauds [...]” (p. 218) While Forceville regards this accommodation as a potential strength of RT, I would argue that it could be also construed as a potential weakness – I will return to this point in Section 6.

The first example of controversial communication analysed by Forceville is a cartoon by the aforementioned Dutch cartoonist Peter van Straaten which was repurposed (and thus recontextualised) in 2006 by the University of Amsterdam in order to promote a teaching conference. The second example is an advert for the Danish political party Venstre (2001) which features a photograph of a group of young men (who were classified as immigrants, Palestinians, and Muslims) leaving a courthouse after receiving a sentence that was considered by many Danes to be “much too lenient” (Kjeldsen, 2007, p. 127). Interestingly, Forceville observes here that “[c]hoosing the visual mode rather than the verbal mode for the representation of the people” (p. 227) provided the sender with a get-out clause when it came to facing potential accusations of bias or prejudice (cf. Messaris, 1997). Furthermore, Forceville provides a definition of *covert communication*: “the intention of the speaker is to alter the cognitive environment of the hearer, i.e. to make a set of assumptions more manifest to her, without making this intention mutually manifest” (Tanaka, 1994, p. 41; cited by Forceville, p. 227). In other words, covert communication depends on the informative intention, but not the communicative intention, being both recognised and fulfilled. Other notable examples analysed by Forceville include a film still of the Dutch filmmaker Sunny Bergman being “photoshopped” (2007) so that the subject looked more beautiful, and the infamous mug shot of O. J. Simpson being artificially darkened for the cover of *Time* magazine (1994) so that the subject looked more like a criminal.



## 6. Discussion

In the final chapter, Chapter 11, Forceville summarises the main claims about relevance theory (RT) that he has made over the course of his book. To begin with, Forceville states: “These claims have been formulated in rather strong terms, not because I am completely convinced of their correctness but because in this form they will hopefully invite enthusiastic endorsements, critical comments, suggestions for refinements – or downright falsification.” (p. 240) Accepting Forceville’s invitation, I would like to begin with an enthusiastic endorsement of his project, both as a whole and in most of its details. In the academic spirit of taking a good song and making it even better, however, I will use the remainder of this review to address the second and third items on Forceville’s to-do list. Let me start then by outlining some critical comments (or doubts) which, in all fairness, probably apply as much to classic RT as to Forceville’s adapted version.

*Is relevance relevant or redundant?* If relevance is assumed across the board and regarded as a constant, then does it in some way become redundant? If, for example, we assume that both utterances X and Y are optimally relevant, then what is it that distinguishes the meaning of utterance X from the meaning of utterance Y? How does the relevance principle help us to distinguish one from the other? In the end, does it come down to sensitivity to things that are external to the linguistic code: that is, background knowledge and situational context? A defender of RT could argue that the trade-off between cognitive effects and processing effort provides a formal (or computational) criterion for deciding which is the best interpretation, just as an internet search engine needs some sort of criterion for deciding when to terminate a given search and present the most “relevant” results to the user. But it is still not clear how this criterion actually works in practice.

*Is RT too general?* The main goal of Forceville’s book is to contribute to the project of developing RT into a comprehensive (or overarching) theory of communication that is capable of incorporating verbal, visual, and multimodal discourses. At the very beginning of the book, though, Forceville states:

However, it needs to be acknowledged straightaway that RT’s status as no less but also no more than a communication *model* means that being an RT expert does not help one very much in the analysis of specific discourses and texts as such. RT describes and defines the pertinent dimensions of communication; it can demonstrate only in very general terms how specific discourses are interpreted, and it is rarely capable of providing novel or intriguing insights into these discourses. To find tools and guidelines for the analysis of discourse, including their ideological aspects, sources in other disciplines – such as rhetoric, stylistics, cognitive linguistics, narratology, and semiotics – are indispensable complements to RT. (pp. 7-8)

Halfway through the book, Forceville reiterates the point about RT’s status by stating: “The interpretations of the discourses analyzed in the following chapters themselves do not depend on RT.” (p. 132) At the very end of the book, Forceville reiterates this point one final time (see p. 258). These admissions, while admirably honest and typical of Forceville’s engaging writing style, are in danger of undermining the whole project and, in my view, sell the project short. To be sure, the gap between theory and application is another gap that needs to be minded. In some RT-inspired analyses that I have read, there is a danger of losing sight of the theory altogether: what you are effectively left with are close textual analyses (that make reference to background knowledge and situational context). But I do not think that this is the case here. In Forceville’s book, the RT-inspired analyses are genuinely insightful. In any case, perhaps the key to minding the gap in question (and a recommendation for future studies) is to make sure that any RT-inspired analysis is consistently supported with the relevant RT concepts and terminology (cf. Sperber & Wilson, 2008): for example, the different inferential steps taken by the viewer can be spelled out in terms of explicatures (verbal or visual) and implicatures (strong or weak).

*A truth principle?* As we have seen, Sperber and Wilson propose that Grice’s four maxims (or categories) can be reduced (or boiled down) to just one: namely, the Maxim of Relation or



what we can more simply call the “relevance principle”. In this “post-truth” era of misinformation, disinformation, “fake news”, and photoshopped manipulations, however, should Grice’s Maxim of Quality not be somehow reinstated? That is, do we need to supplement the relevance principle with some kind of “truth principle”? Forceville notes that RT is based on “biological, Darwinian principles” and thus consistent with evolutionary theory. In light of this, does the field of evolutionary psychology have anything useful to say about the issue of detecting liars and cheaters (for example, see the account of the “cheater detection module”; Cosmides & Tooby, 1992). A similar argument could be made for a “politeness principle”. In example (3) in Section 4, the role of politeness is a possible reason for reinstating Grice’s Maxim of Manner.

Having outlined some critical comments (or doubts), I would like to conclude by offering some suggestions for refinements (or further developments). I will do so by asking four main questions.

### ***Question 1: How can RT be understood in terms of cognition?***

Given that RT places such an emphasis on the cognitive process of inference – the gaps in human communication must be minded (and bridged) “not by more coding, but by inference” (Sperber and Wilson, 1987, p. 697) – one way of developing Forceville’s project is by considering relevant accounts of cognition (see Table 1).

One of the most influential ways of thinking about cognition is in terms of *modularity*, an idea resurrected by Fodor (1983, 1985). In short, Fodor proposes that the mind is comprised of two kinds of system: *input systems* which are modular and *central systems* which are not. The gist of the modularity thesis is encapsulated in the following quote: “It is, no doubt, important to attend to the eternally beautiful and to believe the eternally true. But it is more important not to be eaten.” (Fodor, 1985, p. 4) In short, the capacity to successfully navigate the world and to avoid being eaten is the domain of the modular input systems, while the capacity “to attend to the eternally beautiful” and the like is the domain of the non-modular central systems. For our purposes, the modularity thesis is of particular significance because Sperber and Wilson (1987, p. 701; 2002, p. 4) regard linguistic decoding as being modular in nature and explicitly attribute the inferential processes involved in verbal comprehension to “Fodorian central processes”.

Following Fodor (1983, 1985), one way of approaching the subject of *input systems* is to begin by asking how many input systems there are. A simple answer is around six: that is, five perceptual systems corresponding to the classic five senses of sight, hearing, touch, taste and smell, plus one system for the processing of language (as noted in Section 4).<sup>12</sup> Input systems are modular in the sense that they have a certain set of properties which can be described in functional, neurobiological, and evolutionary terms. A key functional property is that input systems are “informationally encapsulated” or “cognitively impenetrable” with respect to our beliefs, expectations, and background knowledge. Like Fodor, Sperber and Wilson (1987) regard linguistic decoding as being modular in nature: for example, basic word recognition (e.g., recognising the word *coffee*) in the cases of speech perception or reading can be conceived of as a relatively automatic process akin to basic object recognition (e.g., recognising an actual cup of coffee) in the case of vision. Unlike Fodor, they regard “the inferential tier of verbal comprehension as the application of unspecialised central thought processes” (p. 701). In short, the modular account of input systems gives us an idea of how the code model of communication may be implemented in the mind/brain, while the distinction between different types of input system (linguistic and visual) gives us an idea of how the “decoding” of verbal and visual information is more or less equivalent.

According to Fodor (1983, 1985), *central systems* are responsible for what he describes as the non-demonstrative “fixation of belief”, or, to use a more everyday term, thinking. Other examples of central processing include judgement, reasoning, problem solving, and decision

---

<sup>12</sup> Fodor (1983, p. 47) actually conceives of modules as existing at a more fine-grained scale: within the visual system, there may be mechanisms for such tasks as shape perception and colour perception.



making. Central systems are non-modular in the sense that they have the *opposite* set of functional, neurobiological, and evolutionary properties to those cited previously. A key functional property is that of “global sensitivity”; that is, sensitivity to the entire range of our beliefs, expectations, and background knowledge. For Fodor, the paradigmatic example of a central thought process is scientific theorising, a process which is susceptible to what AI researchers have called the “frame problem” (also see Chiappe & Kukla, 1996). In contrast, Sperber and Wilson (1996) argue for a more ecological conception of rationality, where cognitive effects are balanced by processing effort. For example, the information that is currently activated in Peter’s memory is more likely to play a role in the inferential process of deciding what Mary actually means. In short, the non-modular account of central systems gives us an idea of how the inferential model of communication may be implemented in the mind/brain.

Some evolutionary psychologists have proposed that certain cognitive faculties are modular in nature (for an overview, see Barratt, 2014). For example, it is plausible that there is a folk physics module which enables us to take a “physical stance” to moving objects, a folk biology module which enables us to take a “biological stance” to plants and animals, a folk psychology or “theory of mind” module which enables us to take an “intentional stance” to fellow human beings (e.g., Baron-Cohen, 1995; Dennett, 1987), and a variety of additional modules (e.g., the “cheater detection module”; Cosmides & Tooby, 1992) which enable us to take some sort of “social stance” to significant exchanges with fellow human beings. Significantly, Sperber and Wilson (2002, p. 4) suggest that the “theory of mind” module may play a role in the recognition of a speaker’s intentions: that is, attributing intentional mental states on the basis of behaviour. And so it remains open as to what extent both modular and non-modular systems are involved in the relevance-driven inferential processes described by both the classic and adapted versions of RT.

**Table 1.** Types of cognitive processing. The dual-systems distinction is based on Kahneman (2011), while the dual-types distinction is based on Evans and Stanovich (2013). The notion of “System 1.5” (or “Type 1.5”) is inspired by Boghossian (2014).

	Input systems		Central systems	
<b>Modularity</b>	modular		non-modular	
<b>Domain</b>	visual (= visual decoding)	verbal (= linguistic decoding)	cognitive (= inferential model)	
<b>Dual-systems</b>	System 1	System 1	“System 1.5”	System 2
	<b>Type 1</b>	<b>Type 1</b>	<b>“Type 1.5”</b>	<b>Type 2</b>
<b>Properties 1</b>	sub-personal	sub-personal	personal	personal
<b>Properties 2</b>	automatic	automatic	automatic	<u>reflective</u>
<b>Outputs 1</b>	shallow	shallow	deep	deep
<b>Outputs 2</b>	<u>visual</u> explicatures	<u>verbal</u> explicatures	implicatures	implicatures

### ***Question 2: Can pictures have explicatures?***

In an earlier paper, Forceville (2014, pp. 60-61) suggests that it is legitimate to speak of pictures as having (or “inviting”) certain explicatures. In the current book, Forceville (2020, p. 89) tentatively coins the term “visual explicatures” which can be distinguished from “verbal explicatures”. In order to investigate this question further, let us take another look at the panel from the comics album *Tintin and the Picaros* (Hergé, 1976; see Figure 1). Forceville (2014, pp. 60-61) argues that our interpretation of the Tintin panel begins with us “recognising” certain patches of colour as certain objects, and suggests that this process of object recognition is equivalent to the decoding stage of verbal comprehension. While acknowledging that Sperber and Wilson’s explicatures are propositions that are assessable in terms of truth and falsity, Forceville suggests that it is legitimate to say that the panel “invites” explicatures such as “these are a young



man and a dog” or “these are Tintin and Snowy”; and/or “conveys multiple propositions”. Along similar lines, Forceville (2020, p. 192) proposes that the panel “urges us to derive” the explicature “Tintin and Snowy walk [in a certain way] in the direction of a hut in the wood/jungle.”

I propose that the argument for the existence of visual explicatures can be developed by considering two accounts of the human visual system. First, Fodor (1983, pp. 86-97) argues that the human visual system (as an example of an input system) produces *shallow outputs*. These outputs are not as shallow as what the vision scientist Marr (1982) would describe as a “primal” or “2.5-D sketch” of an object, but not as deep as a scientific classification of that object; rather, these outputs take the form of “basic categorizations”. The notion of basic categorizations can be understood by considering a category hierarchy such as: *fox terrier, dog, mammal, animal, physical object, thing* (example adapted from Fodor, 1983, pp. 94-95; also see Rosch et al., 1976). In this hierarchy, all other things being equal, the “middle level” category of dog has greater salience than, say, the category of fox terrier or thing; that is, this “middle level” is most likely to be the level at which we fundamentally perceive the world. Basic categorizations are reliably predicted by the visual properties of the respective stimuli and “phenomenologically given” (Fodor, 1983, p. 96). Second, in a review of the research on the human visual system, Peissig and Tarr (2007, p. 76) argue that the “primary end state” of visual perception is object recognition; the latter is “a critical precursor to interacting with and reasoning about the world”. In short, we need to be able to successfully recognise (and localise) objects in order to interact with those objects in an appropriate way.

If these two accounts of the human visual system (Fodor, 1983; Peissig & Tarr, 2007) are approximately correct, then it could be argued that it is more likely that we perceive a picture at the level of the basic categorizations of objects, rather than at the lower level of features or the even lower level of pixels (cf. Currie, 1995, p. 130). To develop Forceville’s (2014, 2020) interpretation of the Tintin panel, it is plausible that all viewers who have a functioning visual system, who have been brought up in a regular world (rather than a darkened room), and who have been taught the names and identities of regular objects, will perceive a redheaded young man and a white dog walking towards a hut in a forest (cf. the level of “denotation”; Barthes, 1977). Furthermore, those viewers who have a knowledge of Herge’s work will perceive the redheaded young man and the white dog as the characters of Tintin and Snowy respectively, and those viewers who have a knowledge of different breeds of dog may perceive Snowy as a fox terrier (cf. the level of “connotation”; Barthes, 1977).

In summary, if we adhere to the strict argument that pictures do not have specific atoms of meaning or specific syntactical structures for putting such atoms together, then a picture cannot be described as having explicatures (see Currie, 1995, pp. 113-137). If, however, basic object recognition is sufficient for extracting explicatures, then a picture containing recognisable objects (that is, a figurative representation as opposed to an abstract one) can be said to have “visual explicatures”. It is misleading, however, to describe visual explicatures in verbal terms. The fact that pictures can be verbally described in multiple ways, and potentially yield multiple propositions, gives the impression that visual explicatures are vague and ambiguous (and non-explicit) in nature. Instead, visual explicatures should be thought of as having a visual (or analogical) format that gives precise information about the identities (and locations) of objects.

### ***Question 3: How can RT be adapted to account for multiple senders?***

In the paradigmatic example of verbal communication (involving Mary and Peter), there is a single sender: namely, Mary. In the case of a mediated-mass communication such as film, however, there are usually multiple senders. As evidenced by the increasingly extensive credits that usually scroll at the end of a film, the process of filmmaking is typically a collaborative enterprise involving the contributions of hundreds (and sometimes thousands) of people. How, then, can RT be adapted to account for multiple senders? While Forceville accounts for the issue of multiple addressees by writing a whole chapter on the role of genre (Chapter 5), he only



mentions the issue of multiple senders in passing (pp. 114-115). This, then, is an area of RT that needs to be developed further.

To begin with, we need an account of authorship. In a discussion of film and interpretation, Currie (1995, pp. 247-249) argues that the text itself can only ever contain explicit, literal meaning. Significantly, in order to bridge the gap between explicit, literal meaning and implicit, non-literal meaning, we (the reader) need to refer to something *outside* of the text and that something is intentions. Currie (1995, p. 247) states: "What we call implicit in the text is what we think of as intended (perhaps, indeed, unconsciously intended) by an agent who hopes to convey by his words more than the words themselves literally mean." These intentions can be attributed to either a "real author" or an "implied author". Given, however, the potential problems with what Currie calls "real author intentionalism" (RAI), the alternative position of "implied author intentionalism" (IAI) is sufficient for interpretational purposes.

Following Currie (1995, pp. 258-259), I would argue that the notion of an "implied author" is sufficient for the interpretation of film. Even though filmmaking is typically a collaborative enterprise, potentially involving the contributions of thousands of people, the so-called "final cut" of a film arguably goes through the filter of a single implied author (or "maker"). This argument does not commit us, however, to accepting a version of the "auteur theory" where the director is regarded as the sole author. Take Alfred Hitchcock, for example, the celebrated British director and so-called "master of suspense". According to IAI, we do not have to assume that all Hitchcock films have a common authorial voice; rather, we only have to assume that any single Hitchcock film has a single authorial voice (which is not necessarily that of Hitchcock).

When considering the collaborative nature of filmmaking, Currie talks briefly about the possibility of multiple implied authors (or "makers") with multiple, and potentially conflicting, intentions. Here, I would depart from Currie's account: for pragmatic purposes, the film viewer must usually assume a single implied author (or "maker") with a single, or unified, set of intentions in order to get their comprehension/interpretation of the film off the ground. Indeed, the very notion of arriving at multiple interpretations contradicts one of the central ideas of RT. Sperber and Wilson (2002, p. 19) state: "It is not compatible with the presumption of relevance for an utterance to have two alternative co-occurring interpretations, either of which would be individually satisfactory, since this would put the hearer to the unnecessary extra effort of trying to choose between them." Furthermore, making sense of the implied author's intentions requires the reader/viewer to use their "theory of mind" module, or, to put it another way, it requires them to take an "intentional stance" (e.g., Baron-Cohen, 1995; Dennett, 1987).

#### ***Question 4: How can RT be adapted to account for moving images?***

In the second half of the book, Forceville demonstrates the applicability of RT by providing case studies of various examples of "static visuals" (p. 2) such as pictograms and comics. How, though, can RT be adapted to account for the "moving images" of film?

In the final chapter, Forceville includes a brief discussion of the medium of film (pp. 249-251) and begins by citing the aforementioned Hitchcock who, in RT terms, was "very confident about his ability to achieve relevance to his envisaged audience". Forceville proceeds to note two major characteristics that distinguish film from static visuals. First, film is a *dynamic (or temporal) medium* in the sense that "images unfold and develop over time" (p. 249). This characteristic raises the question of which unit of time (shots, scenes, sequences) is the most appropriate unit of analysis; the longer the unit of time, the greater the potential number of explicatures and implicatures. Second, film is a *multimodal medium* in the sense that it draws on a greater number of semiotic modes than static visuals. In addition to visuals and written language, these modes include bodily expressions, spoken language (dialogue), music, and sound effects.

Although film is primarily thought of as a visual medium, Forceville devotes most of his discussion to the roles of sound and music (possibly to redress the balance with the focus on the visual throughout the book). For instance, Forceville observes that many sounds operate in a coded way, and cites examples such as closing doors, train whistles, and thunder. His argument



here can be developed as follows. If it is legitimate for us to say that the *sight* of a closing door is decoded in terms of object recognition by the visual system, then it is also legitimate for us to say that the *sound* of a closing door is decoded in terms of the auditory equivalent. In addition, Forceville observes that some music also operates in a coded way, and cites the examples of national anthems and famous songs such as “Happy Birthday”. To these various examples, we can add the following. From the commercial world, if a brand logo can be classified as a “coded visual” (that is roughly equivalent to a single word), then something like the Nokia ringtone can be classified as a “coded sound”. And from the filmic world, musical leitmotifs may function as coded ways of representing characters. Interestingly, although music is used in an intentional way by the filmmaker, often it is not “consciously registered” by the viewer: a possible example of covert communication, where the communicative intention is not recognised/fulfilled.

As Forceville suggests, adapting RT to accommodate the “moving images” of film is a large research project that is in need of further development. In the following, I will outline three possible sub-questions that can be investigated further (also see Barratt & Romao, in progress). The related questions of how RT can be adapted to account for both multiple senders and multiple addressees have been considered above.

The first sub-question is: *Can films have explicatures?* This relates to the question considered above: Can pictures have explicatures? To the extent that the explicit meaning of a shot or scene is dependent on dialogue or intertitles, the explicatures in play can be classified as “verbal explicatures”. Conversely, to the extent that the explicit meaning of a shot or scene is not dependent on dialogue or intertitles, the explicatures in play can be classified as “visual” or “auditory explicatures”. In addition, the dynamic and multimodal nature of film gives it two advantages over static visuals in terms of potential explicatures. First, films are capable of providing more precise information about the contours and spatial locations of objects via the depth cue of “motion parallax” (see Messaris, 1994, 2012). Second, films are capable of presenting multiple modes of information (see the notions of “redundancy” and “entropy”, Fiske, 2011; “informational redundancy”, Bordwell, 1985; “redundant emotion cues”, Smith, 2003).

Even if we accept Sperber and Wilson’s (1995, p. 175) claim that non-verbal communication lacks the explicitness of language, it is possible that a number of non-verbal modes operating *in combination* may give rise to a relatively specific meaning or at least a consensus of opinion amongst viewers (cf. the notion of “referential” or “explicit meanings”; Bordwell, 1989). In other words, what films potentially lack in terms of explicitness, they more than make up for in terms of redundancy. This point can be formulated as follows: *The greater the number of modes of information indicating (or pointing in the direction of) meaning X, the greater the probability that the viewer will extract meaning X.* Consider, for example, the notorious shower scene from Hitchcock’s *Psycho* (1960) in which the character of Marion Crane is attacked by a knife-wielding assailant (Figure 5). This scene presents a number of different modes: namely, visuals (the montage of images of Marion being attacked), bodily expressions (Marion’s facial expressions of fear and her defensive movements), sound (Marion’s screams), and music (Bernard Herrmann’s screeching violins). Given that these modes are all pointing in the same direction, it is plausible that they will lead to visual and auditory explicatures along the lines of “Marion is being attacked by a knife-wielding assailant” and “Marion is in a state of fear”. This corresponds approximately to Forceville’s observation that “often reference assignment to an event is achieved by combining information from different modes” (p. 250).





**Figure 5.** Frame stills from *Psycho* (Hitchcock, 1960).

The second sub-question is: *How should we understand film in terms of processing effort?* While the verbal dialogues between Mary and Peter involve a single mode of communication (spoken language) which can be explained by referring to a single type of cognitive processing, the medium of film involves multiple modes which require an alternative, or more nuanced, type of explanation. In light of this, it is helpful to make a distinction between at least two different types of cognitive processing. Conscious (or attentive) processing is thought to operate in a serial fashion and to be capacity limited, whereas non-conscious (or automatic) processing is thought to operate in a parallel fashion and to be capacity unlimited (see Eitzen, 1993). The implications for film are as follows. In theory, any given mode of information presented by film may move in or out of conscious attention at any given moment; the remaining modes can be processed simultaneously and in parallel by different systems of the viewer's mind/brain. This corresponds to Forceville's observation that a non-consciously registered mode such as music may contribute to the overall relevance but operate in a "covert" fashion (p. 250).

The third and final sub-question is: *How should we understand film in terms of cognitive effects?* In a dialogue with the screenwriter Ernest Lehman, Hitchcock talks about the filmmaker's capacity to elicit certain "reactions" from the audience. It is reasonable to assume that the types of reactions that Hitchcock is thinking of here are largely emotional in nature; that is, they can be described more specifically as "emotional reactions" or "emotional responses". Therefore, we need to refer to the idea, discussed by a number of RT theorists (e.g., Wharton & Strey, 2019) and mentioned above, that the notion of cognitive effects should be broadened to include emotional effects. Indeed, we can even say that the primary goal of film is to elicit certain cognitive and emotional effects in the mind of the viewer.

## 7. Conclusion

Why is relevance theory (RT) relevant? In his groundbreaking book, Forceville (2020) achieves two major goals: first, he makes the general model of human communication originally proposed by Sperber and Wilson (1986/1995) accessible to a wider (and non-expert) audience; and second, he demonstrates convincingly how that model can be adapted to account for other forms of communication beyond the purely verbal – namely, visual and multimodal communication.



Although the gap may be wider in the case of visual as opposed to verbal communication, it can be minded (and bridged) by cognitive processes which are more or less equivalent. Throughout this review, I have attempted to do justice to Forceville's ambitious project and to give the reader an idea of the richness of the various examples and case studies he provides.

In the discussion, I have argued that the adapted version of RT proposed by Forceville can be developed even further by addressing four main questions (see Table 2). *First, how can RT be understood in terms of cognition?* By referring to the modularity thesis, we can gain a greater understanding of how the decoding and inferential processes are performed by (modular) input systems and (non-modular) central systems respectively. *Second, can pictures have explicatures?* By referring to the outputs of different types of input system (linguistic and visual), we can make a tentative distinction between "verbal explicatures" and "visual explicatures". *Third, how can RT be adapted to account for multiple senders?* In accordance with "implied author intentionalism" (IAI), we can argue that the multiple senders (or makers) of a film must be pragmatically reduced to a single "implied author" with a single, unified set of intentions. *Fourth, how can RT be adapted to account for moving images?* Here, we can refer to the dynamic and multimodal nature of film and consider the implications with respect to explicatures, processing effort, and cognitive effects.

**Table 2.** Classic RT vs. adapted RT

	<b>Classic RT</b>	<b>Adapted RT</b>	<b>Key questions</b>
<b>Modality</b>	verbal	visual / multi-modal	<i>Question: Can pictures have explicatures? (see Q2)</i> <i>Answer: Verbal explicatures vs. visual explicatures</i>
<b>Sender(s)</b>	single	multiple	<i>Question: Multiple senders? (see Q3)</i> <i>Answer: Single implied author + intentional stance</i>
<b>Addressee(s)</b>	single	multiple	<i>Question: Multiple addressees? (see Ch. 5)</i> <i>Answer: Cognitive differences constrained by genre</i>

By following Forceville's fine example, and by trying to find answers to the four questions just cited, it is hoped that relevance theory can be developed into a theory that has even greater relevance than it had before; that is, a theory that generates even more cognitive (and emotional) effects for even less processing effort.

## Acknowledgments

I would like to thank the editors Jordan Zlatev and Alexandra Mouratidou for their detailed comments and suggestions, and Tico Romao for an ongoing dialogue about relevance theory and film metaphor.

## References

- Altman, R. (1984). A semantic/syntactic approach to film genre. *Cinema Journal*, 23(3), 6-18.
- Anderson, J. D. (1996). *The reality of illusion: An ecological approach to cognitive film theory*. Carbondale: Southern Illinois University Press.
- Arnheim, R. (1969 [1954]). *Art and visual perception*. London: Faber & Faber.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and "theory of mind"*. Cambridge: MIT Press.
- Barratt, D. (2014). The geography of film viewing: What are the implications of cultural-cognitive differences for cognitive film theory? In T. Nannicelli & P. Taberham (Eds.), *AFI Film Reader in Cognitive Media Theory* (pp. 62-82). New York: Routledge.
- Barratt, D., & Romao, T. (in progress). Understanding film metaphor: A renewed case for relevance.



- Barthes, R. (1977). *Image-music-text*. London: Fontana.
- Bateman, J. A. (2016). Methodological and theoretical issues in multimodality. In N.-M. Klug & H. Stöckl (Eds.), *Handbuch Sprache im multimodalen Kontext* (pp. 36-74). Berlin: De Gruyter.
- Boghossian, P. (2014). What is inference? *Philosophical Studies*, 169, 1-18.
- Bordwell, D. (1985). *Narration in the fiction film*. London: Routledge.
- Boyd, B., Carroll, J., & Gottschall, J., Eds. (2010). *Evolution, literature, and film: A reader*. New York: Columbia University Press.
- Chandler, D. (2017). *Semiotics: The basics* (3rd ed.). London: Routledge.
- Chiappe, D. L., & Kukla, A. (1996). Context selection and the frame problem. *Behavioral and Brain Sciences*, 19(3), 529-530.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163-228). New York: Oxford University Press.
- Currie, G. (1995). *Image and mind: Film, philosophy, and cognitive science*. Cambridge: Cambridge University Press.
- Dennett, D. C. (1987). *The intentional stance*. Cambridge: MIT Press.
- De Waal, F. (2009). *The age of empathy: Nature's lessons for a kinder society*. New York: Harmony.
- De Waal, F. (2016). *Are we smart enough to know how smart animals are?* London: Granta.
- Eisner, W. (2006 [1985]). *Comics and sequential art: Principles & practice of the world's most popular art form*. Paramus: Poorhouse Press.
- Eitzen, D. (1993). Attending to the fiction: A cognitive account of cinematic illusion. *Post Script: Essays in Film and the Humanities*, 13(1), 46-60.
- Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223-241.
- Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. New York: Basic Books.
- Fiske, J. (2011). *Introduction to communication studies* (3rd ed.). London: Routledge.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge: MIT Press.
- Fodor, J. A. (1985). Précis of *The modularity of mind*. *Behavioral and Brain Sciences*, 8(1), 1-5.
- Forceville, C. (1994). Pictorial metaphor in billboards: A relevance theory perspective. In J. E. Müller (Ed.), *Towards a pragmatics of the audiovisual, Vol. 1* (pp. 93-113). Münster: Nodus Publikationen.
- Forceville, C. (1996). *Pictorial metaphor in advertising*. London: Routledge.
- Forceville, C. (2008). Metaphor in pictures and multimodal representations. In R. W. Gibbs Jr. (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 462-482). New York: Cambridge University Press.
- Forceville, C. (2014). Relevance Theory as model for analysing visual and multimodal communication. In D. Machin (Ed.), *Visual communication* (pp. 51-70). Berlin: De Gruyter Mouton.
- Forceville, C. (2020). *Visual and multimodal communication: Applying the relevance principle*. New York: Oxford University Press.
- Forceville, C., & Urios-Aparisi, E., Eds. (2009). *Multimodal metaphor*. Berlin: De Gruyter Mouton.
- Gibbs, R. W. Jr. (1999). *Intentions in the experience of meaning*. Cambridge: Cambridge University Press.
- Gombrich, E. H. (1999). *The uses of images: Studies in the social function of art and visual communication*. London: Phaidon Press.
- Grice, P. (1957). Meaning. *The Philosophical Review*, 66(3), 377-388.



- Grice, P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics*, Vol. 3 (pp. 41-58). New York: Academic Press.
- Grodal, T. (2009). *Embodied visions: Evolution, emotion, culture, and film*. Oxford: Oxford University Press.
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin Books.
- Keller, R. (1998). *A theory of linguistic signs*. Oxford: Oxford University Press.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163-182.
- Kjeldsen, J. E. (2007). Visual argumentation in Scandinavian political advertising: A cognitive, contextual, and reception oriented approach. *Argumentation and Advocacy*, 43, 124-132.
- Jakobson, R. (1960). Closing statement: Linguistics and poetics. In T. Sebeok (Ed.), *Style in language* (pp. 350-370). Cambridge: MIT Press.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.
- McCloud, S. (1993). *Understanding comics: The invisible art*. New York: Paradox.
- Messaris, P. (1994). *Visual "literacy": Image, mind, and reality*. Boulder: Westview Press.
- Messaris, P. (1997). *Visual persuasion: The role of images in advertising*. Thousand Oaks: Sage Publications.
- Messaris, P. (2012). Visual "literacy" in the digital age. *Review of Communication*, 12(2), 101-117.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76(2), 165-178.
- Neale, S. (2000). *Genre and Hollywood*. New York: Routledge.
- Peissig, J. J., & Tarr, M. J. (2007). Object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology*, 58, 75-96.
- Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.
- Smith, G. M. (2003). *Film structure and the emotion system*. Cambridge: Cambridge University Press.
- Sperber, D., & Wilson, D. (1995 [1986]). *Relevance: Communication and cognition* (2nd ed.). Oxford: Blackwell.
- Sperber, D., & Wilson, D. (1987). Précis of *Relevance: Communication and cognition*. *Behavioral and Brain Sciences*, 10(4), 697-710.
- Sperber, D., & Wilson, D. (1996). Fodor's frame problem and relevance theory. *Behavioral and Brain Sciences*, 19(3), 530-532.
- Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language*, 17(1-2), 3-23.
- Sperber, D., & Wilson, D. (2008). A deflationary account of metaphors. In R. W. Gibbs Jr. (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 84-105). New York: Cambridge University Press.
- Sperber, D., & Wilson, D. (2012). Introduction: Pragmatics. In D. Wilson & D. Sperber (Eds.), *Meaning and relevance* (pp. 1-27). Cambridge: Cambridge University Press.
- Tanaka, K. (1994). *Advertising language: A pragmatic approach to advertisements in Britain and Japan*. London: Routledge.
- Tomasello, M. (2008). *Origins of human communication*. Cambridge: MIT Press.
- Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Cambridge: Belknap Press of Harvard University Press.
- Tseronis, A., & Forceville, C., Eds. (2017). *Multimodal argumentation and rhetoric in media genres*. Amsterdam: John Benjamins.



- Veale, T., Feyaerts, K., & Forceville, C., Eds. (2013). *Creativity and the agile mind: A multi-disciplinary study of a multi-faceted phenomenon*. Berlin: De Gruyter Mouton.
- Wharton, T., & Strey, C. (2019). Slave of the passions: Making emotions relevant. In K. Scott, B. Clark, & R. Carston (Eds.), *Relevance, pragmatics and interpretation* (pp. 253-266). Cambridge: Cambridge University Press.
- Wilson, S., & Matsui, T. (2012). Recent approaches to bridging: Truth, coherence, relevance. In D. Wilson & D. Sperber (Eds.), *Meaning and relevance* (pp. 187-209). Cambridge: Cambridge University Press.
- Yus, F. (2011). *Cyberpragmatics: Internet-mediated communication in context*. Amsterdam: John Benjamins.
- Zlatev, J., Devylder, S., Defina, R., Moskaluk, K., & Andersen, L. B. (2023). Analyzing polysemiosis: Language, gesture, and depiction in two cultural practices with sand drawing. *Semiotica*, 253, 81-116.

*Author address*

[db.msc@cbs.dk](mailto:db.msc@cbs.dk)

*About the author*

Daniel Barratt is an Associate Professor of Visual Communication at Copenhagen Business School, Denmark. He has a background in the fields of film, philosophy, and cognitive science. His research interests include visual communication and cognitive film theory, visual attention and eye movements, theories and models of emotion, and evolutionary and cultural influences on cognitive processes.