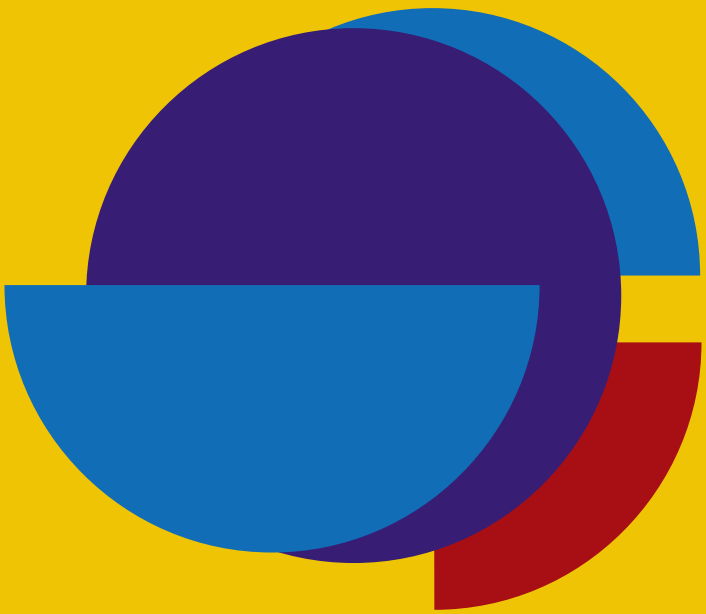
Visual Metaphor

Structure and process

Edited by
Gerard J. Steen



CONVERGING EVIDENCE IN LANGUAGE AND COMMUNICATION RESEARCH 18

JOHN BENJAMINS PUBLISHING COMPANY

Visual Metaphor

Converging Evidence in Language and Communication Research (CELCR)

ISSN 1566-7774

Over the past decades, linguists have taken a broader view of language and are borrowing methods and findings from other disciplines such as cognition and computer sciences, neurology, biology, sociology, psychology, and anthropology. This development has enriched our knowledge of language and communication, but at the same time it has made it difficult for researchers in a particular field of language studies to be aware of how their findings might relate to those in other (sub-)disciplines.

CELCR seeks to address this problem by taking a cross-disciplinary approach to the study of language and communication. The books in the series focus on a specific linguistic topic and offer studies pertaining to this topic from different disciplinary angles, thus taking converging evidence in language and communication research as its basic methodology.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/celcr

Volume 18

Visual Metaphor. Structure and process
Edited by Gerard J. Steen

Visual Metaphor

Structure and process

Edited by

Gerard J. Steen
University of Amsterdam

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/celcr.18

Cataloging-in-Publication Data available from Library of Congress.

ISBN 978 90 272 0151 5 (HB)
ISBN 978 90 272 6347 6 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of contents

Preface

This book presents the major outcomes of the NWO Vrije Competitie program 'Visual Metaphor: A Psycholinguistic Approach' (360–70–390). We are grateful for NWO's financial support and encouragement and hope that these findings will inspire future research on visual metaphor from an interdisciplinary perspective.

All the pictures reproduced in this book are property of their respective owners. We do not hold any copyright about these pictures. In cases where copyright was clear, we obtained permission from copy-right holders for reproduction for scientific purposes – for which we would like to express our gratitude here. All pictures in chapter five were reproduced on the basis of the permission granted for display in the VisMet corpus. In a very small number of cases, we made substantial efforts to track down possible copyright holders but were unsuccessful. All other pictures have been collected from different public sources including different websites, considered to be in public domain. If there still are objections to reproducing any specific picture, please contact the editor.

Gerard J. Steen

CHAPTER 1

Introduction

Wilbert Spooren
Radboud Universiteit Nijmegen

Metaphor has long been approached as a mere figure of speech. Recently it has been reconceptualised, however, and metaphor is now generally considered to be fundamental to the human conceptual system, language being merely one possible modality of expression of metaphorical cognition. Philosophers, linguists, psychologists, and other researchers now agree that metaphor may be found in various modalities and media of communication, including gesture and body language, sound and music, and film and visuals (e.g. Ortony 1979/1993; Forceville 2006; Forceville & Urios-Aparisi 2009; Gibbs 2008).

In spite of this theoretical landslide, however, the wide range of nonverbal metaphor and its processing has neither been empirically investigated on the same scale nor with the same rigour as metaphor in language. The overarching goal of this book is to report on the findings of an interuniversity research program sponsored by the Dutch Organization for Scientific Research, NWO, aimed at exploiting the vast psycholinguistic expertise on metaphor in language for a new, behaviourally founded approach to the structure (Chapters 2, 3, and 4) and processes (Chapters 5, 6 and 7) of metaphor in one of these nonverbal manifestations, static visuals. Its results should, on the one hand, further the development of an encompassing and robust cognitive-scientific theory of metaphor by including visual metaphor, and, on the other hand, enrich our understanding of the communicative possibilities and effects of visual metaphor in multimodal discourse.

The analysis of visual metaphor creates a number of challenges compared to the verbal mode. By way of demonstration, consider Figure 1.1, which is a cartoon about the problems with the Amsterdam to Brussels high speed train Fyra. The train was put into service in December 2012, and met with so many problems that is was taken out of service after only six weeks, resulting in political scandal and lawsuits between the Italian manufacturer of the train AnsaldoBreda and the Dutch national railway organization NS.

https://doi.org/10.1075/celcr.18.01spo

Figure 1.1 "Half of travelers in Fyra train delayed". Cartoon by Peter Boersma (via http://www.peter-b.nl/photos/files/page6-1006-full.html, last accessed April 16, 2013). The text in the speech balloon translates as "THAT F***ING TRAIN GOT STUCK AGAIN!"

The process of interpreting this complex text depends on an impressive number of factors. Some of these involve:

- The genre of the text: We recognize it as a cartoon, a genre in which caricature is used to mock some situation in the news. Our schematic genre knowledge (Steen 2011) tells us that in this genre cartoonists typically want to make a statement, i.e. that they express a standpoint in favour of or against a certain position, which may or may not be backed up by arguments.
- The incongruity of the image: We see a frustrated train driver in a high-tech train cabin. All screens display the notorious "Windows blue screen of death", which signals a fatal error of a computer. There is one exception: one screen shows the Windows98 opening screen. The state-of-the-art nature of the cabin is at odds with the outdated computer technology.
- The resolution of the incongruity: For the text to function as a successful communicative document we must assume that there is a coherent interpretation for the entire text. That is, there is an interpretation of the text in which all of the elements in the text can be linked coherently to a central topic and point. That implies that we must find a resolution for the incongruity. A possible resolution is that the image must be interpreted 'non-literally', and that it makes a comparison in the form of a metaphor, i.e. a link between a non-literally used source domain and the literal target domain.
- The differentiation between source and target domain: The viewer needs to determine what parts of the image correspond to the target domain and what parts correspond to the source domain. In language we can make use of metaphor flags such as *is like* or special grammatical constructions like the Subject-Copula-Predicate construction where targets typically are expressed in subject position and sources in predicate position (Steen et al. 2010). In

the visual domain it is not at all obvious how to find these elements. It can be expected that the incongruity of the image is essential to signal source and target. Also, issues of quantity may play a role. In the cartoon example, several cues indicate that the statement of the cartoonist is *about* the Fyra high speed train, and that hence the Fyra high speed train is target in the metaphor: The cartoon's caption is about the Fyra train and the dominant part of the cartoon displays the train situation. From that we can infer that the source domain is Windows98.

– The reliance on background knowledge about the source and the target domain: Essential background about the source domain involves the fact that at the time of creation of the cartoon (2012) the operating system Windows98 was already obsolete and that the Windows blue screen of death signals a fatal error of the computer system, an error dreaded by all computer users as it signals that the computer really got stuck (*vastgelopen*) and that technical support is needed. Essential knowledge from the target domain is that the introduction of the Fyra train was fraught with problems (the introduction was delayed with a number of years because the manufacturer could not deliver on time, test runs showed all sorts of problems, the safety system required by the EU posed serious problems for the train developers, and when the Fyra train was finally introduced it immediately showed serious delays).
– The interplay of word and image: The metaphorical interpretation of the text ("the operating system of the Fyra train is like the operating system Windows98") does not only rely on the image in the cartoon but also on the words that help readers select the appropriate source and target domain. For the target domain we get verbal cues in the caption and the word on the backside of the driver's seat. For the source domain the verbal cue is in the words on the computer screen. But there is also a more subtle verbal cue available: the use of the word *vastgelopen* ("stuck") is typical in a context of computers (*de computer is vastgelopen*, "the computer is stuck") but is odd in the context of trains (#*de trein is vastgelopen*, "the train is stuck"). As a consequence, the reader is led to believe that part of the balloon text is about computers (the source domain) and not about trains (the target domain).
– The cartoon and its rhetoric: The standpoint of the cartoonist is negative: The Fyra high speed train is to be criticized because it is like a PC with the computer operating system Windows98 and such an operating system is obsolete.

Describing all of these structural aspects in visuals and examining how they work in psychological processing is a hugely complex undertaking.

The complexity of interpreting visual metaphor is even greater when cultural considerations come into play. Consider Figure 1.2, also discussed in Chapter 4.

Figure 1.2 *Het islamdebat* ("the islam debate"). Cartoon by Tom Janssen (via http://tweedekamer.blog.nl/politieke-prenten/2008/02/03/politieke-prent-het-islamdebat-geert-s-regie, last accessed May 11, 2017. The chair shows the word *REGIE* ("DIRECTION")

What we see here is someone sitting in a director's chair (presumably the Dutch far-right-wing politician Geert Wilders, given his characteristic build and hairstyle), with a megaphone next to him. The metaphor expressed here is something like "The Islam debate is like a film set with Geert Wilders as film director". A possible and coherent interpretation of the topic of this image is that Wilders is controlling the Islam debate in the Netherlands. Yet the point of this image is arguably strongly dependent on all sorts of cultural assumptions. Were this cartoon to be published in a right-wing magazine, then the standpoint expressed is probably something like "It is a good thing that Wilders is controlling the Islam debate in the Netherlands", whereas in a left-wing magazine the opposite interpretation is more likely. If the figure in the chair is not recognized as Geert Wilders, it becomes utterly unclear what the metaphor and hence the point of the image will be. Presumably, the metaphor will then become something like "The Islam debate is like a film set with a director". Obviously, it becomes more difficult to distil a specific standpoint from such an interpretation.

To add to the challenges involved in interpreting complex texts containing visual metaphors, consider the example in Figure 1.3. Here we see a hand held as if it is playing a piano. In this print ad for J&B whisky, a hand hovers above a pedestrian's crossing in what is obviously a night city scene. The text in the top left corner reads "Join the city remix", the slogan in the bottom right corner reads "start a party". The position of the hand is that of a piano player. The zebra crossing reminds us of a piano keyboard. On the basis of our genre knowledge we know that the argumentation in this add can be summarized as "You should buy J&B whisky because

Figure 1.3 Print ad 'J&B: Crossing'. Advertising agency: KesselsKramer, Amsterdam, The Netherlands. Via http://adsoftheworld.com/media/print/jb_crossing, last accessed May 11, 2017

if you do that you can start a party". We can assume that the metaphor in this case is "Crossing a zebra is like playing the piano", where "crossing a zebra" is associated with "taking part in city life" and "playing the piano" is associated with "being in a bar". The comparison between these two domains of city life and drinking seems farfetched, but the text becomes more comprehensible once we realize that this ad is part of a much wider J&B cross media campaign "start a party", in which J&B has organized parties all over the world (cf. http://www.jbscotch.com/parties.html, last accessed May 11, 2017). It takes knowledge of a complete campaign to be able to understand the metaphorical comparison in Figure 1.3.

The chapters in this book present the results of different studies into the areas of visual metaphor. It is the outcome of a research program "Visual metaphor: A psycholinguistic perspective" (NWO grant PR-08-06), which is a collaboration of researchers from three Dutch universities, VU University Amsterdam, Tilburg University, and Radboud University Nijmegen. We employ a broad notion of visual metaphor and include multimodal aspects such as verbal text within or on the side of visuals or knowledge of prior, related visuals that may include such multimodal information (cf. Forceville & Urious-Aparisis 2009).

The goal of the book is to take our current state of knowledge with respect to the structure and processing of verbal metaphor as a starting point and to see how far we can get to analyse visual metaphor. The program aimed to develop and test a method for the identification and description of the structure of metaphorical visuals 'in the wild'. Against the general background of a semiotics of figurative meaning invited by visual incongruities, the project then extends the psycholinguistically inspired method for verbal metaphor description developed by Steen and colleagues into the visual realm (Pragglejaz Group 2007; Steen 2007; Steen et al. 2010). These chapters address the first research question of the book: "How is the structure of visual metaphors identified and described?" (Chapters 2 through 4).

For our second topic we use the Career of Metaphor Theory (Bowdle & Gentner 1999, 2005) as the most encompassing theory of metaphor processing in language today. The Career of Metaphor Theory predicts that a metaphor's structure, including its formal type and conceptual conventionality, determine its processing. Since metaphor is nowadays seen as a conceptual phenomenon that is relatively independent of the modality in which it is expressed, psycholinguistic insights into verbal metaphor processing provide new and motivated hypotheses for the processing of visual metaphors and the influence of conceptual conventionality (including frequency) and expressive type (including external resemblance) in that modality. Therefore the second research question of the book is: "How are visual metaphors processed?" In different subprojects the various steps of processing as well as the influence of outward resemblance and conventionality on processing was researched (Chapters 5, 6 and 7).

Part I deals with the structure of visual metaphor and opens with a chapter by Joost Schilperoord, who explores the way in which visual incongruities invite meaning construction. He analyses what it is in metaphorical images that makes them potential conveyors of meaning 'beyond' their denotation, resulting in a visual semiotics for the resolution of incongruities. His chapter is an analysis of the structures of visual incongruity in relation to their postulated functions for metaphorical cognition in visuals.

Against this background, Ester Šorm and Gerard Steen introduce an encompassing and explicit method in Chapter 3 for systematically identifying visual metaphors, called VISMIP. This is an entirely structural method that is based on the MIPVU stepwise procedure for identifying verbal metaphor in natural discourse (Steen et al. 2010), which in itself goes back to MIP (Pragglejaz Group 2007). The chapter goes into many of the issues mentioned in this introduction: what are the units of analysis? how do we find incongruity in the image? and how do we test whether a cross-domain comparison is made? The chapter also describes detailed issues of operationalization for each of these issues, along with an application of the procedure.

Chapter 4 then introduces VisMet, an annotated corpus of visual metaphors that is available to the metaphor community online. VisMet was collected as a testing ground for developing VISMIP but also generated new structural descriptions of some 350 visuals that can be used for further research, including calibration and formal testing of VISMIP. The authors, Marianna Bolognesi, Romy van den Heerik and Esther van den Berg, describe the background of the collection and of the annotation system, but also the website, its limitations and ideas for future work.

Part II deals with aspects of visual metaphor processing. Chapter 5, authored by Romy van den Heerik, Ester Šorm and Gerard Steen, presents think-aloud data regarding visual metaphors in a small corpus of cartoons, art works and ads. The authors argue that these processing findings can be used as evidence for the overall approach as well as details of the development of the visual metaphor identification procedure (presented in Chapter 3).

Chapter 6 is another behavioral study, focusing on how similarity in visual shape (visual metaphor structure) affects the conceptualization of visual metaphors during psychological processing. Lisanne van Weelden, Fons Maes and Joost Schilperoord describe two studies that offer experimental evidence for the idea that shape similarity of juxtaposed objects can be seen as a visual template which facilitates the construction of metaphoric thought. The studies raise interesting issues about the similarities and differences between visual and verbal metaphoric processing.

In Chapter 7, Didier Hodiamont, Hans Hoeken and Margot van Mulken delve into another factor important for the processing of visual metaphor, namely the conventionality of the image (another structural property). The chapter describes an experiment in which metaphors were presented visually or verbally and rated on conventionality. The authors discuss the issue to what degree conventionality is independent of the modality in which the metaphor is presented.

The concluding chapter summarizes the research findings and presents prospects for future research. The latter is done from the perspective of new developments in metaphor research which have added new perspectives to the psycholinguistic model as presented in the Career of Metaphor Theory. These new perspectives, focusing on deliberate versus non-deliberate metaphor use, may also be fruitfully applied to visual metaphor research, and a number of clear connections are made explicit. With this chapter, we present a provisional conclusion to our program on the structure and process of visual metaphor.

References

Bowdle, B. & Gentner, D. (1999). Metaphor comprehension: from comparison to categorization. In M. Hahn & S. C. Stoness (Eds.) *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society* (pp. 90–95). Hillsdale, NJ: Erlbaum.

Bowdle, B. F. & Gentner, D. (2005). The career of metaphor. *Psychological Review*, 112(1), 193–215. https://doi.org/10.1037/0033-295X.112.1.193

Forceville, C. (2006). Non-verbal and multimodal metaphor in a cognitivist framework: Agendas for research. In G. Kristiansen, M. Achard, R. Dirven & F. Ruiz de Mendoza (Eds.), *Cognitive Linguistics: Current Applications and Future Perspectives* (pp. 379–402). Berlin etc.: Mouton de Gruyter.

Forceville, C., & Urios-Aparisi, E. (Eds) (2009). *Multimodal Metaphor*. Berlin etc.: Mouton de Gruyter. https://doi.org/10.1515/9783110215366

Gibbs, R. W. (Ed.) (2008). *Cambridge Handbook of Metaphor and Thought*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511816802

Ortony, A. (1979). *Metaphor and Thought* (2nd ed.). Cambridge: Cambridge University Press.

Pragglejaz Group (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1), 1–39. https://doi.org/10.1080/10926480709336752

Steen, G. (2007). *Finding Metaphor in Grammar and Usage: A Methodological Analysis of Theory and Research*. Amsterdam etc.: John Benjamins. https://doi.org/10.1075/celcr.10

Steen, G. (2011). Genre between the humanities and the sciences. In: M. Callies, W. R. Keller, & A. Lohöfer (eds.), *Bi-directionality in the Cognitive Sciences: Avenues, Challenges, and Limitations* (pp. 21–42). Amsterdam etc.: Benjamins. https://doi.org/10.1075/hcp.30.03ste

Steen, G. J., Dorst, L., Herrmann, B., Kaal, A., Krennmayr, T., & Pasma, T. (2010). *MIPVU: A Method for Linguistic Metaphor Identification in Natural Discourse*. Amsterdam etc.: Benjamins. https://doi.org/10.1075/celcr.14

PART I

Structure

CHAPTER 2

Ways with pictures

Visual incongruities and metaphor

Joost Schilperoord
Tilburg University

2.1 Introduction

The basic tenet of conceptual theories of metaphor is that metaphor is the cognitive process humans engage in when they (re)conceptualize a concept from a target domain in terms of another, apparently dissimilar concept from a source domain (Bowdle & Gentner 2005; Croft & Cruse 2004; Gibbs 1994; Lakoff & Johnson 1980; Steen 2007). Under this view, an expression is considered 'metaphorical', not because metaphor is some inherent property of it but because it is used by speakers to invoke particular mental states which are characterized by mappings between domains. When *visual* language is used to perform these functions, the expression is called a pictorial or visual metaphor (cf. Forceville 1996, 2008; Kaplan 2005; Maes & Schilperoord 2008; McQuarrie 2008; Phillips & McQuarrie 2004; Šorm & Steen 2013; Van Enschot 2006; Van Mulken 2003; Yus 2009). Identifying and describing the expressive tools visual language employs to invite metaphor is one of the central goals of research into visual metaphor. This chapter teams up with this line of research: How can we characterize the structure of images capable of alerting viewers to a metaphorical interpretation?; and How and when does the structure of such images invite metaphoricity?

To start, the notion 'pictorial metaphor' is actually a misnomer. There are no pictorial metaphors; there are only pictures capable of bringing about metaphorical states of mind in viewers. To better understand the properties of those pictures, this chapter proposes to distinguish rigidly between the structure of images on the one hand and metaphorical interpretation induced by them on the other. The focus will be on the conditions under which one leads to the other. As we shall see, especially the role of the topic addressed by those images is critical in understanding the connection between structure and interpretation – metaphorical or otherwise.

An influential hypothesis is that some sort of anomalousness is a necessary feature of a pictorial metaphor. Images capable of promoting metaphorical

https://doi.org/10.1075/celcr.18.02sch

construal usually display some departure from viewers' expectations and their understanding of reality. Such images have been given names like 'ambiguous images' (Jakesch et al. 2013), 'ad hoc pointers' (Yus 2009), 'open' images (Ketelaar et al. 2008) or 'incongruent' images – the term to be used in this paper (cf. Callister & Stern 2008; Kaplan 2005; Michelon et al. 2003; Schilperoord & Maes 2009). In a similar vein Forceville (1996, p. 109) states that there should be 'something odd' present in an image to incite metaphorical responses, while McQuarrie & Mick (1999, p. 38) characterize them as 'artful deviation relative to audience expectation' (ibid, p. 38), which functions as a 'stimulus of incongruity' (ibid, p. 40).[1] A generalized account of the incongruity trope and the conceptual effects it may bring about is offered by Kenneth Burke and famously called *perspective by incongruity*: "(…) extending the use of a term by taking it from the context in which it was habitually used and applying it to another' (Burke 1954, p. 89)". Burke's assumption is that in the hands of the competent communicator the resulting incongruity offers a new and often surprising perspective that may challenge and even alter an audience's preconceptions and existing perspectives. Briefly put, incongruities offer perspective because they impact on an audience's 'sense of what properly goes with what' (Burke 1954, p. 74). To see how the notion of incongruity may apply to visual expressions, consider Figure 2.1, an advertisement for a brand of custard.[2]

---

1. Note that these characterizations rule out the possibility of *using* images metaphorically. For example, if an advertisement for an adult education organization shows an athlete holding a vaulting-pole, metaphoricity is arguably aspired. However, outside this context such an image allows straightforward literal interpretation.

2. All examples of visual incongruities discussed in this paper come from a sample of about three thousand advertisements I have collected together with Fons Maes. The advertising genre is methodologically advantageous for at least three reasons. First, advertisements are in all respects relevant messages, with clearly stipulated contents and pragmatic intentions. If an incongruity is present, it's there on design and any notion of an 'error' can be dismissed: designing these messages is far too expensive for that. Second, designers of such messages often operate at the forefront of exploring expressive possibilities and communicative potential of visual messages. Third, because most existing research on visual metaphor focuses on advertising, using this genre allows me to situate this chapter's content to previous work in this area. The images that appear here are reproduced for purely scientific reason and discussion and were all found in public domains, i.e. dedicated websites on the internet. I have done my best to locate all copy right holders via those websites and ask them for permission to reprint the images, but to little avail. I offer my apologies should I unknowingly have infringed the rights of copy right holders I haven't been able to locate properly.

Figure 2.1 ("Our new custard has a soft bouquet")

The depicted wine box will lead most viewers to expect a bottle of wine as the box's contents; not a pack of custard. Because some unexpected object appears to substitute the expected object an experience of incongruity is brought about (cf. '…taking it from the context in which it was habitually used and applying it to another …'). Note that the depicted pack of custard and the visual context mutually affect each other to create incongruity. On the one hand, the expected 'bottle of wine' is rendered salient because a pack of custard takes its place; without it, one would just see an empty wine box. On the other hand, it is because of the visual context that the pack of custard is experienced as an incongruity; without it, one would just see a pack of custard.

What are viewers supposed to take from this message? First, a general point. The ability to comprehend images like Figure 2.1 is critical, not only to understand advertisements, but also for a variety of other genres, visual, verbal and multimodal. Incongruities are used to convey a broad range of contents and to perform communicative functions like persuasion, explanation, informing and instructing.[3] To comprehend visual incongruities the viewer should first and foremost be able to recognize them for what they are and aspire to: to serve expressive rather than purely descriptive or iconic purposes. Incongruities thus have to be acknowledged as artful instead of some occasional error. Instead of asking 'what do I see here?' viewers are supposed to ask 'why do I see what I see here?'.

Second, comprehension calls for the ability to mentally disentangle the various components that package visual incongruities and to connect them in meaningful ways to each other and to the message's topic and intentions. So, to return to the question how metaphorical meaning might be invited by Figure 2.1, the idea put forward by current accounts is that substitute-based incongruities are able to do so because (i) they call to mind two objects stemming from disparate domains, (ii) one of the objects – usually but not necessarily the one visually represented and the ad's topic – can be assigned the metaphorical role of target while the second, non-represented object may serve as source; and (iii) the established similarity CUSTARD AS WINE allows the mapping of wine-related connotations like exquisite taste, prestige and preciousness; all excellent features for an advertiser to promote his product (cf. especially Forceville 1996; Yus 2009).[4] Hence, the relatively simple manipulation of substituting one entity by another may have a significant impact on how viewers interpret the image and how metaphor affects their assessment of the advertised product (see also Šorm and Steen, this volume).

Most previous accounts of visual incongruities have – albeit under different names – focused almost exclusively on their ability to convey metaphorical meaning. As we shall see, however, visual incongruities can be put to use to express various other conceptual structures, such as range, order, transition, negation and even duration. So, the question still facing researchers is not only when and how

3. Their ubiquity is testified by studies of editorial cartoons (Schilperoord & Maes 2009), comics (Cohn 2007; 2013), graphics (Hariman & Lucaites 2008) infographics (Quispel 2016), photographic design (Kulvicky 2003), games (Van Lankveld et al 2010) and arts (Burford et al. 2003).

4. In the visual metaphor literature incongruities based on Substitute are called 'MP1 metaphors' (Forceville 1996, p. 109), 'replacements' (Phillips & McQuarrie 2004, p. 117) and 'contextual metaphors' (Forceville 2008, p. 194). Examples are analyzed in Forceville (1996, pp. 109–116) and in Phillips & McQuarrie (2004, p. 117). Similar lines of reasoning are offered for incongruity types like 'fusion', see for example Forceville (1996, p. 126) and Phillips & McQuarrie (2004, p. 116).

metaphorical meaning is intended, but also when and how other types of meaning might be at stake. A fuller understanding of these matters calls for a further exploration of at least two issues. The first one pertains to classifying incongruent images. Next to substitute-based incongruities (as in Figure 2.1), several other incongruity types have been isolated, notably object-fusion and juxtaposition. Current practices in this field however appear to rely rather heavily on simply adding new possibilities to existing lists of attested examples (cf. McQuarrie 2008). What is needed is a unified account of the components that produce visual incongruities and their principles of assemblage.

The second, closely related issue concerns the conditions under which metaphorical meaning is conveyed by incongruities like Figure 2.1. As we shall see, this in particular requires us to probe much deeper into the role of the topic of an advertising message in making sense of incongruent images. In his discussion of similar examples, Forceville (1996) puts the issue in straight terms: "(I)n a pictorial metaphor where only one of the two terms is visually represented while the identity of the other is suggested by the pictorial context, it is the visually represented term that is the metaphor's primary subject" (ibid, p. 127). The problem with this 'rule' is that its validity rests fully on equating the advertisement's topic with the metaphor's primary subject. What actually constitutes a necessary condition for substitute-based incongruities to allow metaphorical interpretation (the topic is identical to the visually represented (target) term) is treated as if it were one of its defining characteristics. This situation is also testified by the terms Forceville and others use to refer to substitute-based incongruities: 'MP1 metaphors' and 'contextual metaphors'. But what if this condition – topic is target – is not met? I argue that the way the topic relates to the terms involved in the incongruity had better be treated as a variable, one that determines what kind of meaning construal regarding a certain incongruity is called for, metaphorical or otherwise. In other words, to properly address what conditions determine the 'metaphorical impulse' and what types of incongruity promote what types of meaning construal, we need a methodology that treats incongruity as an independent property of an image and one that explores the role of topic in more detail than existing proposals have hitherto managed to do.

This chapter offers such a methodology by developing an incongruity-based approach to visual metaphor. Sections 2.2 and 2.3 propose a unified account of the structure of visual incongruities by identifying the two constituting components together with several principles of assemblage underlying incongruities. Section 2.4 addresses interpretation as the process of *incongruity resolution*. It identifies eight 'resolution templates', i.e., interpreting heuristics viewers may try out to make sense of a given visual incongruity. Each will be defined in terms of the isolated structural

components of incongruity types on the one hand, and the type of conceptual structure that might resolve the incongruity on the other. Crucially, for each template incongruity structure and resolution will be linked in terms of the way the message's topic relates to the terms or objects involved in the incongruity.

2.2 Visual incongruities: Three examples

A visual stimulus counts as incongruent if a literal reading of it yields an anomaly; it counts as potentially meaningful if the incongruity is 'artful', that is, a deliberate violation of viewers' understanding of reality. With this in mind, consider Figure 2.2, an advertisement for a brand of automobile tyres.

Figure 2.2 ('Anti slip')

Theories of visual cognition account for people's ability to instantly recognize the contents of this image as, say, a 'country road scene' by assuming that they have available some cognitive model of it. Such models represent knowledge about the objects and background elements that populate a scene like this and the ways these entities are spatially arranged and otherwise related (cf. Biederman 1981; Henderson & Ferreira 2004). Sensory input is encoded by comparing it to the relevant model, and will be perceived as 'coherent', 'realistic' or 'congruent' to the extent that it matches the model's premises. This is why most viewers will perceive the drain appearing at the bottom right hand side of the image as an incongruity as it ostensibly conflicts with the model's specifications. A visual incongruity can thus be defined as a stimulus that defies to be accommodated by the cognitive model of the context in which that stimulus occurs.

Previous research on incongruities has documented their ability to capture attention and their recallability (cf. the 'bizarreness effect'; Gounden & Nicolas 2012;

McDaniel & Einstein 1986; Krebs et al. 2009; Michelon et al. 2003; Westerbeek 2016). Apart from such immediate processing effects, incongruities also have been shown to incite elaborated processing efforts in people aimed at resolving the incongruity rather than simply dismissing it as nonsensical (cf. Forabosco 2008; Ritchie 1999).[5] Any account of such explanatory responses should acknowledge the importance of genre. The mere fact that the incongruity in Figure 2.2 appears as part of an advertisement suffices to explain why people would want to resolve it in the first place. The pragmatics of the genre motivates them to consider the incongruity as artful, which will have a significant impact on the kinds of resolutions they will go for.

One promising way to to resolve the incongruity in Figure 2.2 way may be to align roads on the one hand and sinks on the other as places where water may be present that one wants to get of rid of. This ad hoc analogy (cf. Hoffman et al 2009, p. 132) allows the incongruity to be meaningfully related to the recommended product: this tyre removes waste water from a road just as effectively as a drain removes water from a sink. Because this resolution clearly fits this goal of advertisers to make positive remarks about the product they want to sell, the incongruity seems sufficiently explained.

Finding resolutions like the one above arguably calls for elaborated levels of visual competence, i.e., the ability to 'read' expressions cast in visual language, to mobilize relevant knowledge and to connect their contents to memory. It also calls for recipients to be willing (and capable) of working up the required effort. These costs however are outweighed by the benefits: for advertisers because incongruent images attract attention, are recalled better and add rhetorical strength to their messages; for recipients because research indicates that they like ads using incongruent images better and find them more pleasurable than 'what-you-see-is-what-you-get' advertisements (cf. Van Enschot 2006).

This liking effect, however, crucially depends on recipient's ability to actually resolve the incongruity. After all, construing a metaphorical alignment like TYRE AS DRAIN is but one way of resolving incongruities. In many cases different kinds of resolutions are called for. This even holds for instances of incongruity based on the same underlying template: inserting an unanticipated element into the depiction of some scene. Consider for example Figure 2.3, an advertisement for a car.

---

5. As an explanatory concept *Incongruity resolution* has been adapted to account for such diverse phenomena as humor resolution and joke assimilation (cf. Attordo & Raskin 1991; Forabosco 2008; Ritchie 1999), scientific discovery (De Mey 2005), creativity (Koestler 1970), and belief and attitude change (Festinger 1957).

Figure 2.3 ('everything seems older')

Just like Figure 2.2, incongruity is created by inserting an entity – a present-day car – into an unlikely environment – a nineteenth-century street scene. Unlike Figure 2.2, however, the incongruity here cannot be resolved by construing the metaphor TOPIC (?) AS CAR. Instead, it appears that some sort of contrast is implied between the advertised car's advanced technology and that of its competitors (cf. 'everything seems older'). So if indeed this image and the one in Figure 2.2 both employ the visual template of inserting an entity into an 'alien' context, what needs to be explained is why only one of them can be resolved metaphorically while the other defies such a reading.

A critical factor turns out to be the way the topic of both advertisements relates to the entity causing the incongruity. While in Figure 2.2 incongruity could be resolved by relating topic and incongruent entity analogically (TYRE AS DRAIN), the topic of Figure 2.3 actually *is* the entity causing incongruity; i.e. the advertised car. We may thus tentatively hypothesize that this particular feature, i.e., topic (tyre) is not identical to incongruent entity (drain), constitutes a necessary condition for metaphoricity to resolve insert-based incongruity. Put more generally, the variable 'relation between topic and incongruous entity', with possible values 'analogy' or 'identity', may determine the kind of resolution incongruities call for.

Consider yet another example, an ad for a company specialized in sound design (Figure 2.4).

Figure 2.4 ('music and sound design')

In the visual rhetoric literature images like Figure 2.4 have been discussed under the names of 'fusions' (Phillips & McQuarrie 2004), 'MP2-images' (Forceville 1996), 'pictorial metaphor' (Teng & Sun 2002) and 'hybrid metaphor' (Forceville 2008). Like we saw earlier, current analytical practice is thus to conflate this type of visual incongruity with metaphors. However, construing metaphorical meaning (cf. HAMMER AS FLUTE or vice versa) does not resolve the incongruity in Figure 2.4. Instead, the image appears to invite construing a relation of range: 'this company designs every conceivable type of sound, from (the sound of) a hammer to (that of) a flute'. Like before, this example warrants hypothesizing that also merge-based incongruities can be metaphorically resolved only if one of the entities actually is the topic of the message; a condition that is clearly not met in Figure 2.4.

Before exploring the issues raised here in more detail, we first have to determine, in a more principled manner, what visual incongruity is and what types we may encounter. This is the subject matter of the next section.

2.3 Visual incongruities defined

Since deviation from expectation is a defining feature of incongruity, an important question which incongruity theories have to answer concerns the origins of expectancy. Although a variety of sources have been identified,[6] most accounts posit some

6. See for example Brône & Feyarts 2003; Callister & Stern 2008; De Mey 2005; Giora *et al.* 2004; Heckler & Childers 1992; Ludden *et al* 2009; Mandler 1982; Raskin 1987; Ritchie 1999; Vandaele 2002. These authors mention, among other sources, linguistic incongruities, pragmatic incongruities; narrative incongruities, parody and irony, and incongruities raised by knowledge of advertisements, products, brands or media.

version of a cognitive model as source of expectancy, i.e., a '(…) representation of a possibility that has a structure and content that captures what is common to the different ways in which the possibility might occur' (Johnson-Laird et al. 1999, p. 66). A stimulus, then, is experienced as incongruent when it conflicts with or cannot be anticipated from the cognitive model entertained by the viewer. In this chapter the focus is on cognitive models which result from visual perception, i.e., the locus of incongruity should be in the image.[7] In addition, the focus is on knowledge-based expectancy, i.e., on expectations warranted by some cognitive model of 'reality'. Hence, the drain-in-road (Figure 2.2) presents an incongruity because of what is actually shown – a drain in a road – and because no cognitive model of country scenes or roads exists that is capable of predicting the drain's appearance.

All this boils down to incongruity being a parasitic notion. No entity is capable of producing incongruity in isolation; there should always be input which activates a certain cognitive model with which the stimulus conflicts. From here on, I will refer to these input types as input functioning as the *establisher* of congruency (abbreviated *E*), and input functioning as the *violator* of congruency (abbreviated *V*). *E*-input activates a certain cognitive model, while *V*-input violates one of this model's premises. In Figure 2.2, all input activates the country road model, which therefore functions as *E* with the exception of the inserted drain, which functions as *V*. Any visual incongruity can accordingly be captured as a visually disposed combination of some *E* and some *V*: *E(V)*.

The notions Establisher and Violator allow for a unified account of visual incongruities in terms of (i) the cognitive models they employ, and (ii) the assemblage principles by which they are visually integrated. It then turns out that two general types of incongruities can be distinguished: two-model incongruities and one-model incongruities. Each of these can be further classified in terms of more specific principles of assemblage or operations. Here, the discussion will be confined to five two-model operations (Insert, Substitute Merge, Ambiguate and Transmute) and two one-model operations (Distort and Erase). There is a possibility that we can find other operation types, but the claim is that such alternatives always employ either two models or one.

7. What this boils down to is that the experience of incongruity would be the same had the image been used in a completely different context; say at an art exhibition. The restriction also effectively rules out *text-image* incongruities (see for example Van Enschot 2006), or the range of expectancy sources discussed in Callister & Stern (2008).

2.3.1 Two-domain incongruities

The first type of visual incongruity employs two cognitive models: one functioning as *E* and a second one functioning as supplier of *V*, with the latter entity graphically integrated in a visual evocation of the first domain. Figure 2.5 shows a schematic version of this type of incongruity. Ovals are used to signify cognitive models; *CM* = cognitive model, *me* = model entity, *E* = establisher, and *V* = violator.

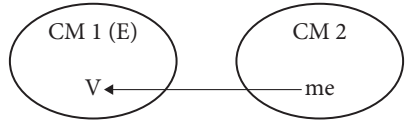


Figure 2.5 Two-domain incongruity: Insert

Note that *me* is indeed parasitic in that it functions as *V* only in reference to a cognitive model (*CM 1*) which contains no slot for it. Figure 2.2 exemplifies this type of visual incongruity: the drain is an *me* defined by a 'sink-model' (*CM 2*). By importing it into a visual evocation of the first model (the road-model *CM1* which functions as *E*), it comes to function as *V*. The same applies to Figure 2.3.

What Figures 2.2 and 2.3 have in common is that the incongruities are constructed by applying a particular assembling operation, one that will be termed Insert: take an entity *me* from some *CM 2* and insert it into a visual evocation of some *CM 1* which cannot accommodate *me*. From this it follows that the crucial property of the Insert assembling operation is that it produces incongruities which relate *E* and *V* in terms of probability: the likelihood for a certain objects to occur in a certain scene (Biederman 1981).

Next to Insert, there is a second assembling operation which applies to two-model incongruities: Substitute. Like Insert, Substitute also takes an *me* from some *CM 2* and inserts it into a visual evocation of some *CM 1*, but the operation additionally erases a model entity which is defined in terms of *CM 1*. Figure 2.6 shows the schematized nature of Substitute, with double strike-through marking an erased *me*1.

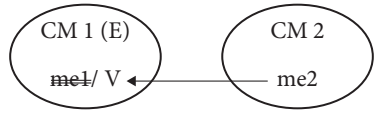


Figure 2.6 Substitute

An example of Substitute is Figure 2.1; a second one appears in Figure 2.7, where a pair of trousers substitutes a new born baby.

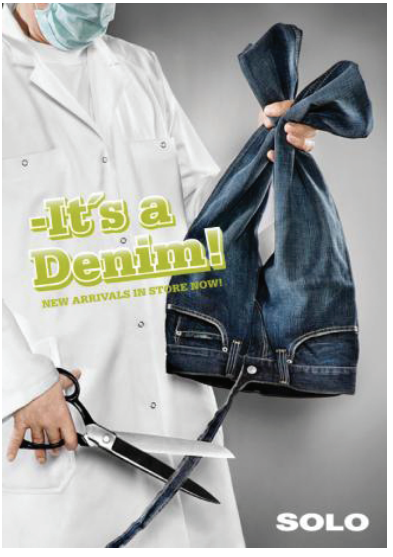


Figure 2.7 ("It's a Denim!")

Like Insert, Substitute produces incongruities which relate V and E in terms of probability. The model functioning as E in Figure 2.7 – the cognitive event model 'child delivery' – cannot accommodate the presence of the pair of trousers, at least not the way that object is depicted here. However, while Insert primarily relates an me to the E-model 'as a whole' (i.e. a drain in relation to a country road), Substitute directs the viewer's attention not only to the inserted me, but at the same time also to the entity erased from the evocation of CM 1: a new born baby. Substitute-based incongruities thus always concern two me's, while Insert concerns only one me. As we shall see farther on, this difference may have important consequences for ways of resolving Insert versus Substitute-based incongruities.

One possible objection against this analysis may be that the Insert-case in Figure 2.2 also employs 'erase' because the entire sink appears to be erased.[8] This objection is unwarranted as long as we assume that operations like Insert and Substitute (and the other types yet to be discussed) always apply only to the image as a visual representation of the cognitive model that functions as Establisher. A litmus-test to decide whether a certain incongruity is Insert-based or Substitute-based, is to render explicit what has to be done to 'correct' the incongruity at hand. In case of

8. This objection was actually put forth by one reviewer.

Figure 2.2, correcting the incongruity only requires removing the Violator from the image. Once the Inserted entity – the Violator- has been removed again, the image is ok. However, to correct Figure 2.7, one not only has to remove the pair of trousers, but also re-insert the entity which most viewers will experience as ostensibly absent: a baby. Therefore, Figure 2.7, but not 2.2, is a Substitute-based incongruity.

A third type of operation which employs two models is called Merge.[9] Instead of inserting model entities into an E-model or substituting an entity from it, Merge graphically integrates two entities stemming from two distinct models. Figure 2.8 shows the schematic version of this type of incongruity (the symbol ~ signifies the 'merging' of model entities me 1 and me 2).

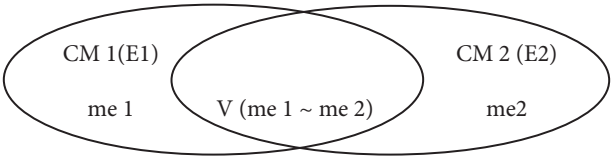


Figure 2.8 Merge

An example of this type of incongruity already appeared in Figure 2.4. The object shown there takes elements from model entities defined either in CM 1 or CM 2, a 'flute-model' and a 'hammer-model', which are conflated to the effect of creating a hybrid object that is at the same time neither a flute nor a hammer as well as both these objects. To account for their violator status, one needs both input models, one representing knowledge of flutes and the other one of hammers. So instead of one, there are two Establisher models here. The hammer-part of the hybrid object cannot be predicted from the flute-model (E 1), while the flute part of it cannot be predicted from the hammer-model (E 2). Another example is Figure 2.9, where Merge has been applied to create a hybrid scene, one in which it is both day and night.

A fourth type of operation which employs two models is called Ambiguate. When this operation is applied, an object defined in CM 1 is depicted such that its perceptual input completely fits a second object, defined in CM 2. The ambiguate effect is created by causing the visual system to make contradictory assumptions about what is where (cf. Chambers & Reisberg 1985). An example of Ambiguate is Figure 2.10, where the white shape on the table is at the same time both a sink and a tissue.

9. Cf. Forceville's (1996) *MP2 metaphors* and Phillips & McQuarrie's (2004) *fusions*.

Figure 2.9 ('With wide-angle lens')

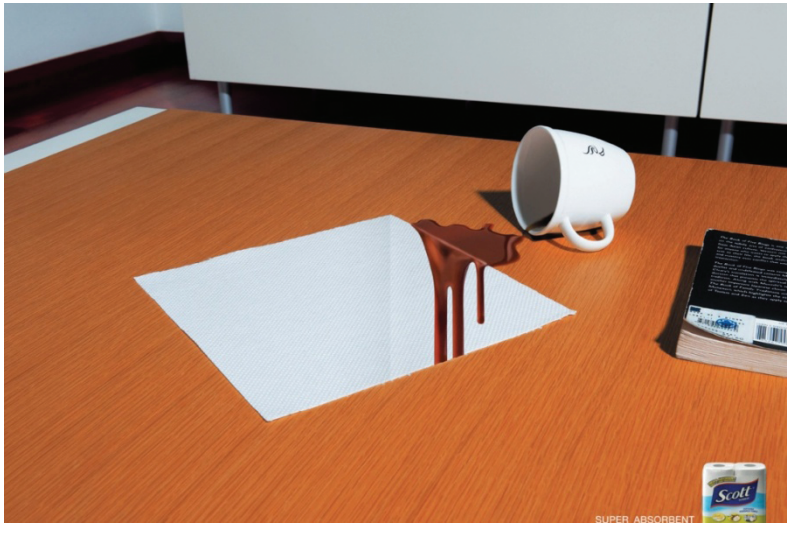


Figure 2.10 ("Super absorbent")

Figure 2.11 shows the schematic version of this type of incongruity, which again takes two establisher models E1 and E2.

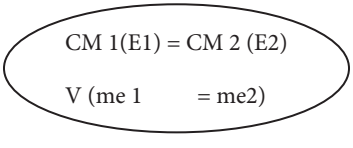


Figure 2.11 Ambiguate

Ambiguate can be regarded as Merge driven to extremes: instead of merging parts of two me's but leaving others unaffected, an ambiguated image fits the visual parameters of two me's in their entirety.

One final two-model operation may be provisionally called Transmute. This operation has an object defined in some CM 1 metamorphosed into an object defined in a different model. I have not been able to find an example but Forceville (2008, p. 189) discusses this possibility. His Figure 9.7 (ibid, p. 189) depicts a vase with tulips that seem to have been transmuted into several drooping wine glasses.

To round off this section, one concern needs to be addressed.[10] One may doubt whether Merge, Ambiguate and Transmute are actually separate operations. I admit that I am not entirely sure whether we actually need distinguishing them. One may support a conceptual criterion and see whether Merge- Ambiguate- and Transmute-based incongruities may call for different underlying resolution processes (see Section 2.4). On the other hand, if we apply a purely graphic criterion the proposed distinction seems supported. The operations produce markedly different kinds of incongruities. Merge produces hybrid objects that are 'this *and* that', i.e. a hammer *and* a flute; Ambiguate produces objects that are 'either this *or* that', i.e. either a sink *or* a tissue; while Transmute produces objects that are 'this *as* that', i.e. tulips *as* wine glasses or wine glasses *as* tulips.

2.3.2 One-domain incongruities

A second broad class of visual incongruity employs one model which functions as E. The defining property is that the incongruity is constructed by targeting on a particular model entity me which is actually defined in terms of the model functioning as E, hence one that can be predicted from E. Figure 2.12 shows the schematic version of this type of incongruity, where the symbol ≈ signifies a graphic operation applied to me in order to make it function as V.

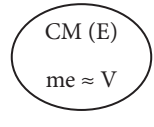


Figure 2.12 One-model incongruity

Two graphic operations will be discussed here that produce one-model incongruities. Either an me is Erased from a visual evocation of CM, or an me is visually Distorted.

---

10. This issue was raised by the same reviewer (see footnote 9).

Erase produces V's by erasing a predictable me from the depiction of the model functioning as E. An example is Figure 2.13.

Figure 2.13 ('Don't be without a topic. Read books')

A crucial prerequisite for detecting an absent entity as V, i.e., as one that is indeed erased instead of being merely absent, is that it concerns a so-called diagnostic model entity (cf. Oversteegen & Schilperoord 2014). A model entity is diagnostic with respect to its defining model, if and only if the two are mutually predictable. Although in Figure 2.13 attributes like a pair of glasses, wrinkles or warts can be defined in reference to the human head model, their absence will not be experienced as resulting from Erase because of their non-diagnostic nature. The entity 'mouth' on the other hand is diagnostic: if there is a face, there also is a mouth (at least, if the face is seen up front). So subjecting this me to Erase yields a true V.

Distort is an umbrella term encompassing a range of operations which all produce visual incongruities by somehow distorting a predictable model entity's manner of appearance. The operation may target, for example, on an object's regular size or the size of one its attributes; to its shape; the material it is made of; its location in a given scene, and so forth. In Figure 2.14 for example, Distort has been applied to turn the depicted object, a tea cup, inside out. Although all attributes, notably the ear and the ornament, are predictable by the tea cup model, this particular visual evocation is clearly incongruent.

Figure 2.14 ('Makes cleaning easy')

Distort-based incongruities may also results from changing object parts or attributes, like the material it is made of or its shape, or from moving an object to an unlikely or impossible location. In Figure 2.15 incongruity is created by relocating the man and his vacuum cleaner. The clever detachment of the objects and their cast shadows suggests both entities to float.[11] Another frequently applied operation enlarges or reduces an object´s proportion relative to other objects or the size of one of its parts; an incongruity exemplified by Figure 2.16.[12]

Note that Distort differs from Transmute in that the latter operation fundamentally alters the ontology of an object (hence the two-model incongruity underlying this operation), while Distort leaves the ontology unaffected.

Summing up, we have identified two general types of visual incongruities in terms of the number of cognitive models that they employ: one- versus two-model incongruities. In addition, seven types of assembling operations to construct visual incongruities have been identified: Insert, Substitute, Merge, Ambiguate, Transmute, Erase and Distort.[13] Table 2.1 summarizes the two types and the operations that have been discussed.

---

11. For discussion of visual incongruities employing depictions of cast shadows, see Schilperoord & Van Weelden (2018).

12. See Schilperoord & Maes (2010) where this type of incongruity is discussed as visual hyperboles.

13. Note that Phillips and McQuarrie's operation Juxtaposition ('similes' in Forceville's (1996) terminology) is not included here. The reason is that juxtapositions do not generate visual incongruities as defined in this article, but instead yield images called 'object alignments'. As argued in Schilperoord et al. (2009), juxtapositions constitute the visual counterpart of rhetorical *schemes* in that they employ gestalt principles and object-constitutive features like shape and size (see Teng & Sun 2002; Van Weelden 2013; Van Weelden et al. this volume).

Figure 2.15 ('Repairs hands that work for a living')

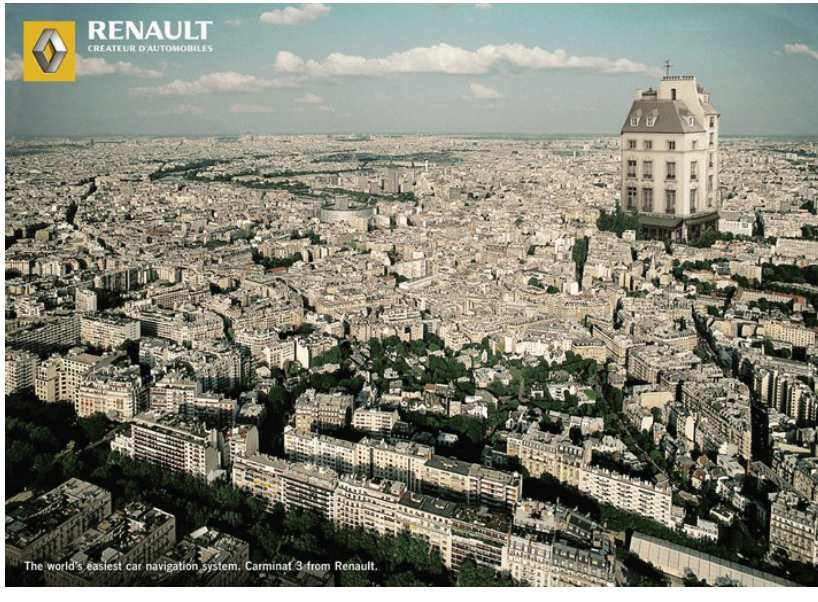


Figure 2.16 ('The world's easiest car navigation system')

Table 2.1 Incongruity types

Number of cognitive models	Assembling operation	V and E
2	Insert	V = me from CM2 inserted into CM1(=E)
	Substitute	V = me2 from CM 2 inserted into CM1(=E) + me2 substitutes me1
	Merge	V = 'parts-of-me' from CM1/E1 and CM2/E2 merged
	Transmute	V = entire me from CM1 (=E1)transmuted into entire me from CM 2 (=E2)
	Ambiguate	V = entire me from CM 2 (E1) = entire me from CM 1 (E2)
1	Erase	V = 'diagnostic' me from CM (=E) erased
	Distort	
	(Enlarge, Reduce, Relocate,, Change…)	V = me from CM (=E) distorted (size, shape, location,…)

We now turn to modeling the ways the various incongruity types may be resolved, i.e., the meaning construal operations they may invoke.

2.4 Incongruity resolution

Upon encountering an incongruent stimulus, humans will engage in (often) effortful and elaborative processing to explain the deviant contents of that stimulus and to find reasons why it is communicated. This process is called Incongruity resolution (cf. Forabosco 2008) and the way a given incongruity is resolved determines how a person interprets the incongruity. A resolution to an incongruity is a conceptual structure that is pasted around the incongruent stimulus but also adds meaning elements which lack an overt correspondence with the (visual) input. A resolution is adequate if it explains the incongruity in a way that fits all types of knowledge involved.

The resolution TYRE AS DRAIN for the incongruity in Figure 2.2 is adequate for three reasons. First because it provides an explanation for the anomalous depiction of 'a drain in a road' in terms of a conceptual structure that is both meaningful and relevant in the given context. Second because the resolution grants the image's referential claim ('it is about reality') while acknowledging its expressive (rather than iconic) purposes ('the tyre functions as if it is a drain in the road'). And third, because it accords with the viewer's knowledge about the particulars of this sort of message ('It is an advertisement') and how s/he is supposed to respond to it ('It tries to persuade me').

This section analyzes several examples of visual incongruities in advertisements with the goal of identifying what will be called *resolution templates*. These templates should, on the one hand, be specific enough to apply to actual cases and to provide relevant resolutions, preferably ones reflecting the designer's intent, while on the other hand they should be general enough to grant types of resolutions for a range of visual incongruities, thereby abstracting away from the specifics of the inputs functioning as E and V. As noted at the outset, each template will be characterized in terms of three related but independent variables. The first variable, to be abbreviated *Incon*, concerns the type of incongruity: two-model versus one-model incongruities which are further regimented in terms of the applied assembling operation. In addition, Incon identifies the components functioning as V and E. The second and third variables will be specified in tandem as an If-Then constellation. The If-part accounts for the role of context in resolving incongruities by identifying a limited set of relations between the topic of the message and the entity functioning as V (abbreviated topic-to-V). The Then-part specifies a particular conceptual structure, one taken from a limited set of possibilities, which resolves the incongruity in a way that fits the specifications of Incon and the If-variables.

Identifying resolution templates for incongruities marks in particular how the methodology developed here differs from previous ones. It supersedes the common practice of conflating incongruity type and resolution type (cf. 'MP1 metaphor'; 'integrated metaphor') by modeling resolution templates as interpretative heuristics which viewers may attempt in order to resolve a given incongruity. In addition, instead of the common practice of fixing the topic-to-V relation (i.e., topic = V), this relation will be treated as a variable, and an especially decisive one for that matter. The difficulty with most existing accounts of pictorial metaphor (especially Forceville 2008, 182ff and Yus 2009, 162ff) is that they envisage only one way of relating topic to incongruity and only one way to resolve the latter: the topic is the metaphor's target concept. However, two incongruities with identical Incon-specifications may call for different resolution types depending on how the If-part specifies topic-to-V. This is why it is crucially important to keep the structural integrity of an incongruity apart from the resolution it promotes. If an incongruent image invites metaphoricity, it is because the latter happens to resolve the former, and not because metaphorical meaning is wired-in, so to speak.[14]

14. The visual modality lacks the kinds of expressive devices language users have at their disposal to mark a text's topic, such as 'This paper *addresses the issue* …', or 'This book *is about*…'. Therefore, to identity a visual expression's 'aboutness' we have to use contextual and pragmatic cues, such as the fact that advertisements are about the product or service they recommend. The issue 'how do viewers decide what a certain image is about?' is indeed a wildly complex one – one that this chapter will not address.

Two general topic-to-V relations will be identified: either the topic is identical to V or to an entity defined in terms of E, or it is related otherwise to the incongruity, for example as set-member or as cause-and-effect. Consequently, establishing this relation is the task set to the recipient and it is what resolving incongruities boils down to.

2.4.1 Substitute

The typical Substitute-based incongruity is Figure 2.1. The context, a wine box, functions as Establisher, while the substituting object – a pack of custard – functions as Violator (see also Forceville (1996, 115ff), and Phillips & McQuarrie (2004, p. 117)). The incongruity can be resolved by establishing metaphorical alignment between V and the replaced bottle of wine. In this way, mappings are invited of wine-related features like 'exquisite taste' and 'special, upscale' onto the target.

To explain how metaphorical construal resolves the incongruity, it is critical to take into account that the object functioning as V is identical to the ad's topic. This topic-is-identical-to-V relation, or 'topic ID V' for short, is a necessary condition for metaphorical relations to resolve Substitute-based incongruities. The first resolution template can thus be given as (1).

(1) Type 1 resolution
Incon 2-model; Substitute; V and E (me) have been identified
If Topic ID V or Topic ID me[15]
Then Resolve incongruity by construing metaphorical meaning *V AS ME* or *ME AS V*.

The Incon-variable specifies the nature of the incongruity: the establisher model E, the identity of the violator V and of the substituted object me which is defined in terms of E. The If-part specifies the condition for the resolution to work in terms of the topic-to-V/me relation: Topic ID V or Topic ID me. The Then-part specifies the conceptual structure which resolves the incongruity: the metaphor *V AS ME* or *ME AS V*.[16] The full specification for Figure 2.1 is (1a).

(1a) **Incon** 2-model; Substitute; V = pack of custard, E = wine box, me = bottle of wine
If Topic ID V = pack of custard.
Then Resolve incongruity by construing a metaphorical meaning *CUSTARD AS WINE*

---

15. In some cases this type of identity, i.e. Topic ID me, is required (cf. Forceville 1996, 122ff and his Figure 6.9). For convenience, both possibilities will be simply called topic ID V.

16. To mark metaphorical relations, we follow the common practice of giving the components in small capitals connected by 'as': X AS Y.

Type 1 resolutions work for a lot of Substitute-based cases. For Figure 2.7, the pair-of-delivered-trousers, the relation topic ID V applies while construing V (=trousers) AS ME1 (=new born) promotes matching the predicates deliver and manufacture and placing their arguments in correspondence with each other: doctor – manufacturer; child – pair of jeans. Mapping properties from 'child' to the advertised product, such as a new-born's delicacy and unique value and the kind of expert care taken to deliver it, warrants the product claim "This jeans is not simply invented but realizes a 'jeans conception' that has been developing all along and instead of simply manufacturing jeans, this jeans is 'delivered' with the utmost specialized care and expertise".

Let us now consider a Substitute-based incongruity for which Type 1 resolutions fail to account. The example is Figure 2.17, an advertisement for the men's magazine *Playboy*.

Figure 2.17 ('Netherlands could use a bit more *Playboy*')

The point is to demonstrate how the topic-to-V variable causes an incongruity with Incon specification identical to the one specified in (1) to call for a resolution other than metaphor, hence with a different If-Then specification. The image shows a typical Dutch snack automaton containing several equally typical Dutch snacks – croquettes. One of them is however substituted by a well-made lobster which hence causes incongruity. Despite lots of possible connotations, the ad arguably claims its product to possess exceptional status: it is an 'upscale' and 'sophisticated' magazine. Our question hence is: how is this meaning visually evoked and what topic-to-V relation is involved in resolving the incongruity?

With regard to the Incon-part, the incongruity is again Substitute-based. The pictorial context of the snack automaton functions as E, while the lobster, substituting a croquette, functions as V. However, applying a Type 1 resolution yields LOBSTER AS SNACK which obviously fails to offer a sensible resolution. Closer inspection of the If-part specification explains the reason why type 1 resolutions fail

to offer a way out: this ad is neither about lobsters nor about snacks. So fixing the topic-to-*V* relation as topic ID V is not possible here.

To identify the type of resolution that is called for, let us, for the moment, specify the If-part as NOT topic ID V and see what impact this has on the Then-part. To this end, I adopt the notion of categorical domain which has been expounded by Shen (1999) and applied to pictorial metaphor in Maes and Schilperoord (2008).

Shen's critical theoretical contribution concerns what he calls a Hybrid Model of metaphor interpretation which differentiates between schematic and categorical source domains (Shen 1999, 1632). Consider again the delivered-trousers example in Figure 2.7. According to Shen's theory, 'delivering a new-born' represents a schematic domain of a certain event which captures attributes like baby, doctor, scissors and so on, and connections between attributes specified as relations of contiguity: thematic, causal, spatial and temporal (cf. Shen 1999, 1636). For example, scissor relates instrumentally to doctor; the latter using the former as an instrument to cut the umbilical cord. In schematic domains components therefore relate to the entire schema as parts-to-whole. The metaphor TROUSERS AS BABY hence aligns a schematic domain 'delivering a new born' which functions as source and is evoked by the visual context (i.e., E), and a second schematic domain 'manufacture', which functions as target and is suggested by the topic of the message, trousers (i.e., V). The relevant parts of the two domains are given in (2), where "AS" marks analogical mappings between domains.

(2) doctor → SCHEMATIC DOMAIN ['DELIVERY-OF-NEW-BORN']
Attributes: <baby, doctor, …>
Relations: baby is conceived
baby develops
doctor delivers baby
expert care to perform delivery
…

trousers → SCHEMATIC DOMAIN ['MANUFACTURE']
Attributes: <manufacturer AS doctor, trousers AS baby, …>
Relations: trousers manufactured AS 'conceived'
trousers invented AS 'baby develops'
manufacturer produces trousers AS 'doctor delivers baby'
expert care AS 'expert care to perform delivery'
…

By contrast, components of a categorical domain do not relate to the domain as parts-to-whole, but instead as members-to-set. Together they constitute what is called a taxonomic category (cf. Shen 1999, 1639), the members of which are characterized by taxonomic similarity. Crucially, the meaning of any member is

determined by its position relative to other members. This relative position is determined by a given member's 'diagnostic properties' (ibid, 1639) which are rendered by the message. Under this view, Violators stand out as excellent devices to provoke a 'natural superordinate category' and to promote some of its properties as diagnostic. In Shen's own example, 'USA Today is the hamburger of American newspapers', the incongruent entity 'hamburger' invites the category 'types of food' by virtue of its taxonomic similarity with other types of food, while its position within that set is determined by the salient property 'a popular mass product'. As such, this particular member's property not only distinguishes it from other members within the set, but even puts it in contrast with these other members.

Applied to Figure 2.17, the depicted entities evoke the taxonomic category 'types of food', while the 'position' of the inserted lobster within that set is determined by its salient properties 'sophisticated', 'upscale' and 'for the happy few'. This puts the lobster in contrast with the second salient member snacks, the latter coming to symbolize properties like 'ordinary', 'for the masses' and 'cheap' (see (3)).

(3) Lobster (=V) → CATEGORICAL DOMAIN 'types of food'
Members <lobster, snacks>
Relation lobster contrasts with snacks: 'sophisticated, …' vs. 'ordinary, …'

Because (3) gets rid of the nonsensical LOBSTER AS SNACK, it offers a suitable specification of the Then-part of the desired resolution. But how can we connect this to the topic of Figure 2.17 (i.e. the advertised magazine)? To get the If-part of the resolution in place, metaphorical meaning construal is required. The topic of this ad calls to mind a second categorical domain 'magazines', which is analogical to the *V*-incited categorical domain 'types of food', that is, including the taxonomic contrast (see (4)).

(4) Lobster (=V) → CATEGORICAL DOMAIN 'types of food'
Members <lobster, snacks, …>
Relation lobster contrasts with snacks: 'sophisticated, …' vs. ordinary, …'
AS
Topic → CATEGORICAL DOMAIN 'types of magazines'
Members <topic AS lobster, ?? AS snacks>
Relation topic contrasts with ?? AS lobster contrasts with snacks: 'sophisticated, …' vs. 'ordinary, …'

The entity marked by the question marks in (4) allows to be specified as, say, 'other magazines', notably the competitors of the advertised magazine. In this way we get 'Playboy is the lobster of men's magazines', or, spelled out, 'the position of the

topic relative to other members within the category 'magazines' is analogical to the position of 'lobster' relative to the other members within the category of foods". This particular topic-to-V relation is where the analogy needed to resolve the incongruity should be located. In other words, instead of topic ID V the resolution to Figure 2.17 rests upon fixing this variable as topic AS V.

Generalizing, we can specify the Then-part of the resolution as setting up a categorical domain which is suggested by the entity functioning as V. The resulting Type 2 resolution signifies an analogy between contrasts, and is detailed in (5), where the taxonomic contrast between V and m is abbreviated V *co* m.

(5) Type 2 resolution

Incon	2-model; Substitute; V, E(m)		
If	NOT Topic ID V → Topic AN V		
Then	Resolve incongruity by construing categorical domains		
	V-domain	Members	<V, ….m1, …>;
		Relation	V *co* m
	Topic-domain	Members	<topic AS V, …m2 AS m1,…>
		Relation	(topic AS V) *co* (m2 AS m1)

Note that the 'If-part' in (5) not only marks the difference between Type 1 and Type 2 resolutions, but additionally takes care of the desired analogical alignment of the two categorical domains, including the taxonomic contrast.

The analysis of the lobster-incongruity, resulting in a proposal for Type 2 resolutions, is no singularity. The exact same resolution serves to resolve the incongruity in Figure 2.18, an advertisement for a delivery van.

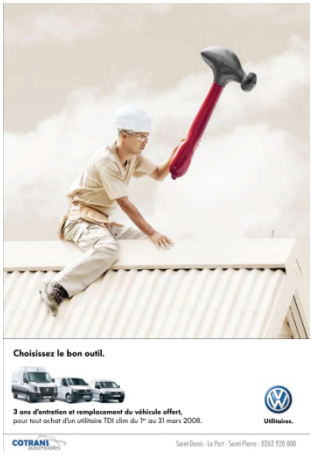


Figure 2.18 ("choose the right tool")

Again, we have a case of NOT Topic ID V (= the inflatable hammer the carpenter attempts to hammer a nail with) as the topic is the recommended van. Resolving the incongruity proceeds by setting up a taxonomic domain 'types of hammers', which is analogically aligned with the domain 'types of vans'. The position of the inflatable hammer within the set is determined by the salient property 'toy', or 'surrogate', with associated notion 'playing', properties which put it in contrast with other members, notably 'real' hammers having properties like 'tool' and 'working'.

Summing up, type 2 resolutions involve a visually evoked metaphor which, to my knowledge, has hitherto not been observed in the visual metaphor literature: target contrasts with X like source contrasts with Y. The examples discussed suggest the existence of a multi-modal construction which carries content comparable to the verbal analogy between contrasts construction 'X is the Y of Z' ("Afghanistan has become the Vietnam of the Russian Army"). Here Y evokes a taxonomic contrast that is claimed to apply also to the position X occupies in category Z.

2.4.2 Merge

Merge employs object features like shape, color, size, or material; or scene characteristics like proportionality, likelihood of entities, support or arrangement; and merges them together into one 'hybrid', non-existing object or scene. The relevant example is Figure 2.4, where the shape of flutes and hammers and the materials these object are made of are merged to produce an image which depicts neither a hammer nor a flute and at the same time both these objects.

As for substitutions, the visual metaphor literature considers Merge-based incongruities to be excellent templates for evoking metaphoricity. Again, the terms by which they are referred to suggests an equal footing of structure and meaning: 'MP2 metaphors' (Forceville 1996, p. 126), 'hybrid metaphors' (Forceville 2008, p. 194) and 'integrated metaphors' (Teng & Sun 2002, p. 304). Existing analyses of Merge-based images warrant a third resolution type (see (6)).[17]

(6) Type 3 resolution

Incon	2-model; Merge; CM1(me1) and CM2 (me2) have been identified
If	Topic ID me1 or Topic ID me2
Then	Resolve incongruity by construing a metaphorical meaning: ME1 AS ME2 or ME2 AS ME1

17. See among others Forceville (1996, 126ff); Maes & Schilperoord (2008, p. 240); Phillips & McQuarrie's (2004, p. 123); Teng & Sun (2002, p. 305).

Note that except for the specific nature of the components involved in the Merge-based incongruity, Type 3 resolution is strikingly similar to Type 1 resolution. Their specifications of the If-Then variables are identical. This is as it should be, given that both incongruity types can be resolved by constructing metaphorical meaning. In addition, resolution templates similar to this one can be constructed for Transmute and Ambiguate-based incongruities. As long as it is possible to equate one of the incongruity components in cases of Transmute or Ambiguate with the message's topic, metaphor serves to resolve it.

To apply (6) to Figure 2.4 it should be possible to identify one of the two entities – hammer or flute – as topic. However, the reason for none of these resolutions to yield sensible results is because the actual topic of this message is a company in 'sound design'. This raises an obvious problem: the incongruity cannot be resolved by mapping features from flutes to hammers or vice versa. To make sense of cases like this, that is where NOT Topic ID me1 or me2, an alternative relation between topic and me1 and/or me2, should be established. One resolution readily presenting itself follows from considering the nature of the product the advertising company provides: sounds. In fact, a salient attribute of the merged objects concerns the sounds they make. The meaning of the ad then becomes something like 'This Company produces every conceivable sound, from flutes to hammers'. The question then is; how can we account for the way the Merge-based incongruity contributes to this claim?

A solution is offered, again, by the concept of categorical domains. The topic represents a set which motivates setting up a categorical domain 'sounds', of which (the sounds produced by) the objects hammer and flute are members. Topic and m1/m2 hence relate to each other as set-member. Again, the meaning of the domain attributes hammer and flute follows from taxonomic similarity and is determined by their positions relative to each other. However, unlike the lobster example above, neither of these objects denotes some salient prototype, like a lobster representing the ultimate type of luxurious food, and therefore instead of a relation of contrast, the objects relate in terms of a gradient scale or range of which they represent equally 'extreme' points. This warrants Type 4 resolutions (see (7)).

(7) Type 4 resolution
Incon 2-model; Merge; CM1(me1) and CM2(me2) have been identified
If Topic = set, me1/me2 = members
Then Resolve incongruity by construing categorical domain
Members < me1, me2 >
Relation Range me1, me2

Note that according to this resolution there is no object entity functioning as either primary or secondary object, and so it is impossible to identify what functions as

V and what as E. The resolution therefore specifies a range of the sounds produced by a hammer in relation to that of a flute and vice versa.

The problem Type 3 resolutions (see (6)) pose can be put in further perspective by imagining the image as part of a message advertising, say, a construction company. Although strictly speaking the topic of this ad is not identical to one of the merged model entities, it still would be possible to construct a near-identical, metonymic relation between topic and 'hammer'. In that case, a Type 3 resolution applies: the incongruity can be resolved by constructing the metaphor HAMMER AS FLUTE, inviting viewers to conceptualize the kind of construction work provided by the advertiser as 'a work of art'; with 'the hammer used as if it were a musical instrument'. Obviously, in the present context this resolution fails to resolve the incongruity.

To conclude, just like Substitutions, Merge-based incongruities are not inherently metaphorical. Instead, only if the condition Topic ID me1/me2 is met, does setting up a metaphorical relation resolve the incongruity. If the identity relation is not possible, other resolutions have to be tried out, like the set-member resolution type specified in (7).

2.4.3 Insert

In contrast to Substitute- and Merge-based incongruities, Insert-based incongruities have received little attention in the visual metaphor literature. One reason may be that most theoretical and analytical research in this area has been mainly preoccupied with 'possible ways of arranging two visual elements on a page' (McQuarrie 2008, p. 263), whereas perceptually Insert concerns only one object. Nevertheless, as the drain-in-road and other examples suggest, Insert yields two-model incongruities: one functions as E while V-entities are defined in terms of a second model. So the possibility of metaphoric resolutions for Insert incongruities should be seriously considered.

In what follows, resolutions will be examined for two types of Insert-based incongruities; the difference between them again based on the topic-to-V relation. We discuss cases with topic identical to V and cases with topic analogical to V. As we will see, metaphorical resolutions are only valid in the second cases.

The relevant Insert-based example with topic ID V is Figure 2.3. The present-day car we see driving around in a nineteenth-century street scene yields a visual anachronism. Importantly, because the message is about the car, car and Violator coincide. Assuming the message transmitted here is something like "this car's technology is so advanced, it is as if all others come from past times", the analytical issue is to explain the incongruity's contribution to this claim.

In my view, it can be resolved by setting up a categorical domain 'vehicles', similar to the sort of Substitute cases exemplified by the lobster-image (Figure 2.17).

Taxonomic similarity together with the car's contextually highlighted property 'modern technology', a contextually highlighted 'diagnostic property', puts it in contrast to the other members of the category: nineteenth-century carriages representing the property 'old-fashioned'. This type of resolution is summarized in (8).

(8) Type 5 resolution
Incon 2-model; Insert; E and V have been identified
If Topic ID V
Then Resolve incongruity by construing categorical domain
Members <V, ….me, …>;
Relation V CO me

Note that the member me has to be inferred because, unlike for example in Figure 2.17, it is not visually represented (the inserted car does not actually substitute other members of the categorical domain). It can be identified once again as the advertised product's competitors which 'are like' yesterday's carriages (cf. the tag line 'everything seems older'). In addition, since V and me are both members of the same taxonomic category, and since topic ID V, resolving this incongruity does not call for metaphoricity. The Insert template is employed here merely to suggest a relation of contrast, that is, the advanced technology of the recommended car versus that of its competitors.

Insert-based incongruities where the topic cannot be equated with V require an alternative resolution. The relevant example is the drain-in-road image in Figure 2.2. The resolution is (9).

(9) Type 6 resolution
Incon 2-model; Insert; *V, E*
If NOT Topic ID *V* → Topic AN *V*
Then Resolve incongruity by construing metaphoric meaning topic AS *V*

The drain (=V) motivates setting up a schematic source domain 'sink' which captures certain entities like a container, a faucet and a drain, and certain relations between those entities. For example, the entity drain relates to the entity container in that the latter is used to contain liquids which can be removed by the drain. In addition, it relates to the entity container in terms of spatiality and part-whole: a drain is located 'at the bottom' and 'in' a container and, as such, it is a part of it. Each of these components hence relates to the entire schema, i.e., 'sink' as parts-to-whole. In addition, the topic (tyre), which is defined in terms of E, motivates a second schematic domain which capture attributes and relations that can be analogously aligned to attributes and relations of the source domain: tyre as drain, concrete as container, and so on. The specified resolution type 6 for Figure 2.2 is (9).

(9a) **Incon** 2-model; Insert; V = drain, E = road
If Topic AN V
Then Resolve incongruity by construing metaphoric meaning topic AS V
V SCHEMATIC SOURCE DOMAIN ['SINK']
Attributes: <container, drain, …>
Relations: Container contains liquids
Drain serves to remove liquids from container
Topic SCHEMATIC SOURCE DOMAIN ['ROAD']
Attributes: <concrete (=container), tyre (=drain), …>
Relations: Concrete (=container) contains liquids
Tyre (=drain) serves to remove liquids from
Concrete (=container)

2.4.4 Distort and Erase

A reasonable hypothesis about Distort and Erase-based incongruities is that because they employ one cognitive model, metaphorical construal cannot be expected to function as resolution, as it necessarily involves two models. And in fact, none of the attested cases of Distort and Erase require metaphorical resolutions. Instead, the overarching characteristic of one-model incongruity resolutions is that they establish some sort of within-model contrast, the nature of which is determined by the operation applied to create the incongruity. We briefly discuss two cases: the mouthless face (Figure 2.13), a case of Erase and the blown-up building (Figure 2.16), a case of Distort/Enlarge.

The Erase-based incongruity can be resolved by construing a so-called bounded contrast, that is, the presence versus absence of the relevant model entity (with the incongruity depicting the absence-situation). The topic 'read books' relates to V in terms of (near) identity as the incongruity visually represents the situation when the recommended behavior (to read books) is not honored: to be without a topic is as having no mouth. The generalized resolution template is (10).

(10) Type 7 resolution
Incon 1-model; Erase; V (me) = mouth, E = human face (CM)
If Topic ID V or Topic CE V
Then Resolve incongruity by construing bounded contrast V (absent) vs. 'corrected V'(present)

Distort-based incongruities can be resolved by construing a graded contrast, cf. soft/hard, big/small, high/low, and so on. The disproportionate-building-in-city scene (Figure 2.16) is a case in point. The violator is constructed by providing it an illegitimate value for the model parameter size (cf. Biederman 1981). Comparing V

to corrected V (a normally proportioned building) produces a gradual contrast of magnitude with the 'permitted' value functioning as criterion. In the context of the topic, a car navigation system, the contrast 'legitimate vs. illegitimate size' projects onto a functional scale ranging from 'hard to find' to 'easy to find'. Hence, the topic relates to V in terms of cause-and-effect. The generalized resolution template is (11).

(11) Type 8 resolution
Incon 1-model; Enlarge; V = building, E = city scene
If Topic CE V
Then Resolve incongruity by construing graded contrast V) vs. 'corrected *V*'(normal)

This resolution warrants the claim that the recommended product makes it very easy to find one's destination as if the destination were this big.

2.4.5 Summary

Table 2.2 summarizes the components of the eight resolution templates that this section has identified. It combines all possibilities for the three components of each template: the nature of the incongruity, the topic-to-V relations and the types of conceptual structures able to resolve any found incongruity.

Table 2.2 Resolution templates

N_{model}	Assembling operation	Resolution types	topic to V
2	Insert	Analogy	
	Substitute	– categorical domain	
	Merge	– schematic domain	
	Transmute	Range	Identity
	Ambiguate	Contrast	Cause-and-effect
		…	Analogy
1	Erase	Contrast, bounded	…
	Distort	Contrast, graded	
		…	

This summary suggests three types of resolutions which involve metaphorical construal: (i) the topic of the message may relate to *V* in terms of analogy (cf. Figure 2.2 as a case in point); or (ii) involve an analogy between contrasts (Figures 2.17 and 2.18); or (iii) the resolution type itself involves cross domain mapping (cf. Figure 2.7 as a case in point). In the latter case, the topic-to-V relation should be one of identity. Note that previous accounts of visual metaphor have almost exclusively dealt only with this latter type.

2.5 Closing remarks

As already testified by previous scholarship, the defining characteristic of images capable of inviting metaphor is indeed anomalousness. However, in contrast to earlier accounts, metaphor is not a property of such images. Instead, it involves the construction of a conceptual structure that is capable of resolving an incongruent image. This chapter has attempted to identify the circumstances for metaphorical meaning to actually resolve incongruities.

It is the variable topic-to-V that turns out to be critical. Being able to identify a topic provides an incongruent image with 'aboutness' and 'agency' that it would otherwise lack. Crucially, however, only when a topic can be identified with one of the components involved in an incongruity, usually V, will metaphorical meaning resolve the incongruity. In case topic relates to V otherwise, different kinds of conceptual structure are called for to resolve a given incongruity. Further research should inform us about the possibility of other templates, either in terms of other topic-to-V relations and resolution types or in terms of particular combinations of the three components. The claim, however, is that whatever alternative possibilities may be identified, they will adhere to the three-part variable structure exemplified by the proposed resolution types.

In order to identify overlapping concerns between the methodology advocated in this chapter with other contributions in this book, I focus on Šorm and Steen's Chapter 3. For a start, the aims of the two chapters differ: while the VISMIP procedure described by Šorm and Steen aspires to offer tools and analytical categories to identify visual metaphors, this chapter has focused on circumstances under which metaphorical meanings resolve incongruities. However, what is shared is the prominence of topical awareness in what Šorm and Steen refer to as 'defining the image's meaning'. Like this chapter, Šorm and Steen consider the topic of the message as the 'dominant conceptual domain of the image'. Despite terminological differences, Šorm and Steen's levels of meaning and the resolution templates distinguished here critically include topic (cf. their 'primary subject matter') and the nature of visual input. Overlapping aspirations also emerge in Šorm and Steen's call for an 'operational definition' of visual incongruity. Their 'topic incongruities' are essentially similar to what I call topic-to-V relations while their 'property incongruities' can be specified in terms of conceptual relations between Establisher and Violator as specified in Then-parts of resolution templates. The semantic categories of Agent, Action, Object and so forth Šorm and Steen propose to provide 'structural descriptions' can be seamlessly incorporated as specifications of Vs. The incongruity in Figure 2.18, the man hammering a nail with an inflatable toy hammer, can be specified as a deviant value assigned to the semantic category 'object' involved in a 'hammering event'. Cognitive theories of object and scene knowledge

(cf. Biederman 1981; Humphrey & Forde 2001) can be adopted in similar fashion to add necessary detail to the nature of a given incongruity.

Finally there are a couple of distinctions proposed here which are not addressed by Šorm and Steen, but which in my view are well worth incorporating in VISMIP. First, the distinction between one- versus two-model incongruities might help to explain when metaphorical meaning of an incongruity can be at stake. Second, it would be rewarding to incorporate the distinction between categorical and schematic domains. Its benefits have been substantiated, I hope, by the analyses of the lobster image and the one with the present-day car driving around in a nineteenth-century street. Third, as demonstrated in the former section, the incongruity approach promises to develop a unified account of the notion 'multimodality', which might also be a valuable addition to VISMIP. And it would tell when a visual incongruity is metaphorical.

References

Attardo, S. & Raskin, V. (1991). Script theory revis(it)ed: Joke similarity and joke representation model. *Humor* 4, 293–347. https://doi.org/10.1515/humr.1991.4.3-4.293

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.) *Perceptual Organization* (pp. 213–263). Hillsdale, NJ: Lawrence Erlbaum.

Bowdle, B. F. & Gentner, D. (2005). The Career of Metaphor. *Psychological Review*, 112, 193–216. https://doi.org/10.1037/0033-295X.112.1.193

Brône, G. & Feyaerts, K. (2003). *The cognitive linguistics of incongruity resolution: Marked reference-point structures in humor*. Document retrieved 04032016, wwwling.arts.kuleuven.be.

Burford, B., Bricks, P. & Eakins, J. P. (2003). A taxonomy of the image: on the classification of content for image retrieval. *Visual Communication* 2, 123–161. https://doi.org/10.1177/1470357203002002001

Burke, K. (1954). *Permanence and Change*. Indianapolis.

Callister, M. A. & Stern, L. A. (2008). Inspecting the Unexpected: Schema and the Processing of Visual Deviations. In E. F. McQuarrie & B. J. Phillips (Eds.) *Go Figure; New Directions in Advertising Rhetoric*, (pp. 137–159). New York, London, Sharpe.

Chambers, D. & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception and Performance*, 11, 317–328.

Cohn, N. (2007). A visual lexicon. *Public Journal of Semiotics*, 1, 53–84.

Cohn, N. (2013). Visual narrative structure. *Cognitive Science*, 37, 413–452. https://doi.org/10.1111/cogs.12016

Croft, W. & Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511803864

De Mey, T., (2005). Tales of the unexpected. Incongruity-resolution in humor comprehension, scientific discovery and thought experimentation. *Logic and Logical Philosophy*, 14, 69–88. https://doi.org/10.12775/LLP.2005.006

Enschot, R. Van (2006). *Retoriek in Reclame. Waardering voor schema's en tropen in tekst en beeld* (Rhetorics in Advertisement. Appreciation of schemes and tropes in text and image). Doctoral Dissertation, Nijmegen University.

Festinger, L., (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.
Forabosco, G., (2008). Is the Concept of Incongruity Still a Useful Construct for the Advancement of Humor Research? *Lodz Papers in Pragmatics*, 4, 45–62. https://doi.org/10.2478/v10016-008-0003-5
Forceville, C. J. (1996). *Pictorial Metaphor in Advertising*. London/New York, Routledge. https://doi.org/10.4324/9780203272305
Forceville, C. J. (2008). Pictorial and Multimodal Metaphor in Commercials. In E. F. McQuarrie & B. J. Phillips (Eds.), *Go Figure; New Directions in Advertising Rhetoric* (pp. 178–205). New York, London, Sharpe.
Gibbs, R. W. , Jr. (1994). *The poetics of mind: figurative thought, language and understanding*. Cambridge University Press.
Giora, R., Fein, O., Kronrod, A., Elnatan, I., Shuval, N. & Zur, A. (2004). Weapons of Mass Distraction: Optimal Innovation and Pleasure Ratings. *Metaphor and Symbol*, 19, 115–141.
Gounden, Y. & Nicolas, S. (2012). The impact of processing time on the bizarreness and orthographic distinctiveness effects. *Scandinavian Journal of Psychology*, 53, 287–94. https://doi.org/10.1111/j.1467-9450.2012.00945.x
Heckler, S. E. & Childers, T. L. (1992). The Role of Expectancy and Relevancy in Memory for Verbal and Visual Information: What is Incongruency? *Journal of Consumer Research*, 18, 475–492. https://doi.org/10.1086/209275
Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action* (pp. 1–58). New York: Psychology Press.
Hariman, R. & Lucaites, J. L. (2008). Visual Tropes and Late-Modern Emotion in U.S. Public Culture. *Poroi*, 5, 47–93. https://doi.org/10.13008/2151-2957.1015
Hoffman, R. R., Eskridge, T. & Cameron, S. (2009). A Naturalistic Exploration of Forms and Functions of Analogizing. *Metaphor and Symbol*, 24, 125–154. https://doi.org/10.1080/10926480903028094
Humphreys, G. W. & Forde, E. M. E. (2001). Hierachies, similarity, and interactivity in object recognition: 'Category-specific' neuropsychological deficits. *Behavioral and Brain Sciences*, 24, 453–509.
Jakesch, M., Leder, H. & Forster, M. (2013). Image Ambiguity and Fluency. *PloS ONE*, 8, 1–15. https://doi.org/10.1371/journal.pone.0074084
Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S. & Caverni, J. (1999). Naïve Probability: A mental model theory of extensional reasoning. *Psychological Review*, 106, 62–88. https://doi.org/10.1037/0033-295X.106.1.62
Kaplan, S. (2005). Visual metaphors in print advertising for fashion products. In: K. Smith, S. Moriarty, G. Barbatsis & K. Kennedy (Eds.) *Handbook of Visual Communication: Theory, Methods, and Media*. Mahwah, NJ: Lawrence Erlbaum. Pp. 167–177.
Ketelaar, P., Grinsbergen, M. S. van & Beentjes, J. W. J. (2008). The Dark Side of Openess for Consumer Response. In E. F. McQuarrie & B. J. Phillips (Eds.) *Go Figure; New Directions in Advertising Rhetoric*. New York, London, Sharpe. Pp. 114–137.
Koestler, A., (1970). *The Act of Creation*. London: Pan Books.
Krebs, R. M., Schott, H. B., Schütze, H., & Düzel, E. (2009). The novelty exploration bonus and its attentional modulation. *Neuropsychologica*, 47, 2272–2281. https://doi.org/10.1016/j.neuropsychologia.2009.01.015
Kulvicki, J. (2003). Image Structure. *Journal of Aesthetics and Art Criticism*, 61, 323–340. https://doi.org/10.1111/1540-6245.00118

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago University Press.

Lankveld, G. Van, Spronck, P. & Van den Herik, J. (2010). Incongruity-Based Adaptive Game balancing. *Advances in Computer Games*, 68, 208–220 https://doi.org/10.1007/978-3-642-12993-3_19

Ludden, G. D. S., Schifferstein, H. N. J. & Hekkert, P. (2009). Visual-Tactual Incongruities in Products as Sources of Surprise. *Empirical Studies of the Arts*, 27, 61–87.

Maes, A. & Schilperoord, J. (2008). Classifying Visual Rhetoric: Conceptual and Structural Heuristics. In E. F. McQuarrie & B. J. Phillips (Eds.) *Go Figure; New Directions in Advertising Rhetoric* (pp. 227–257). New York, London, Sharpe.

Mandler, G., (1982). The structure of value: Accounting for taste. In: M. S. Clark & T. Fiske (Eds.). *Affect and Cognition* (pp. 3–36) Hillsdale, NJ: Erlbaum.

McDaniel, M. A. & Einstein, G. O. (1986). Bizarre imagery as an effective memory aid: The importance of distinctiveness. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 12, 54–65.

McQuarrie, E. F. (2008). A Visit to the Rhetorician's Workbench: Developing a Toolkit for Differentiating Advertising Style. In E. F. McQuarrie & B. J. Phillips (Eds.) *Go Figure; New Directions in Advertising Rhetoric* (pp. 257–277)New York, London, Sharpe.

McQuarrie, E. F. & Mick, D. G., (1999). Visual rhetoric in advertising: Text-interpretive, experimental and reader-response analyses. *Journal of Consumer Research*, 26, 37–54. https://doi.org/10.1086/209549

Michelon, P., Snyder, A. Z., Buckner, R. L., McAvoy, M. & Zacks, J. M. (2003). Neural correlates of incongruous visual information: An event-related fMRI study. *NeuroImage*, 19, 1612–1626. https://doi.org/10.1016/S1053-8119(03)00111-3

Mulken, M. Van (2003). Analyzing Rhetorical Devices in Print Advertisements. *Document Design*, 4, 114–128. https://doi.org/10.1075/dd.4.2.02mul

Oversteegen, E. & Schilperoord, J. (2014). Can pictures say no or not? Negation and denial in the visual mode. *Journal of Pragmatics*, 67, 89–106. https://doi.org/10.1016/j.pragma.2014.03.009

Phillips, B. J. & McQuarrie, E. F. (2004) Beyond Visual Metaphor: A New Typology of Visual Rhetoric in Advertising. *Marketing Theory*, 4, 113–126. https://doi.org/10.1177/1470593104044089

Quispel, A. (2016) *Data for all. How designers and laymen use and evaluate popular information visualizations*. Doctoral Dissertation, Tilburg University.

Raskin, V. (1987). Linguistic heuristics of humor: a script-based semantic approach. *International Journal of the Sociology of Language*, 65, 1987, 11–25.

Ritchie, G. (1999). Developing the incongruity-resolution theory. *Proceedings of the AISB Symposium on Creative Language*, (78–85). Edinburgh.

Schilperoord, J. & Maes, A. (2009) Visual metaphoric conceptualizations in editorial cartoons. In: Forceville, C. J. & E. Urios-Aparisi (Eds.) *Multimodal Metaphor* (pp. 213–243). Berlin/New York, Mouton de Gruyter.

Schilperoord, J. & Maes, A. (2010). Visuele hyperbolen (Visual hyperboles). *Tijdschrift voor Taalbeheersing*, 32, 74–95. https://doi.org/10.5117/TVT2010.2.VISU373

Schilperoord, J., Maes, A. & Ferdinandusse, H. (2009). Perceptual and conceptual visual rhetoric: the case of symmetric object alignment. *Metaphor and Symbol*, 24, 155–174. https://doi.org/10.1080/10926480903028110

Schilperoord, J. & Van Weelden, L. (2018). Rhetorical shadows: The conceptual representation of incongruent shadows. *Spatial Cognition & Computation*, 18, 97–114.

Shen, Y. (1999). Principles of Metaphor Interpretation and the Notion of Domain: A Proposal for a Hybrid Model. *Journal of Pragmatics*, 31, 1631–1653. https://doi.org/10.1016/S0378-2166(99)00035-1

Šorm, E. & Steen, G. J. (This volume.) VISMIP: Towards a method for visual metaphor identification.

Šorm, E. & Steen, G. (2013). Processing visual metaphor; A study in thinking out loud. *Metaphor and the Social World*, 3, 1–34.

Steen, G. (2007). *Finding metaphor in grammar and usage: A methodological analysis of theory and research*. Amsterdam: Benjamins. https://doi.org/10.1075/celcr.10

Teng, N. Y. & S. Sun. (2002). Grouping, simile, and oxymoron in pictures: A design-based cognitive approach. *Metaphor and Symbol*, 17, 295–316. https://doi.org/10.1207/S15327868MS1704_3

Vandaele, J. (2002). Narrative Humor (II): Exit Perspective. *Poetics Today*, 33, 59–125.

Weelden, L. Van (2013). *Metaphor in Good Shape*. Doctoral Dissertation, Tilburg University.

Weelden, L. Van, Maes, A. A., & Schilperoord, J. (This vol.). How visual form affects metaphorical conceptualization: The role of shape similarity.

Westerbeek, H. (2016). *Visual realism*. Doctoral Dissertation, Tilburg University

Yus, F. (2009). Visual metaphor versus verbal metaphor: A unified account. In: Forceville, C. J. & E. Urios-Aparisi (Eds.) *Multimodal Metaphor* (pp. 147–173). Berlin/New York, Mouton de Gruyter.

CHAPTER 3

VISMIP

Towards a method for visual metaphor identification

Ester Šorm and Gerard J. Steen
VU University Amsterdam / Universiteit van Amsterdam

3.1 Introduction

Consider the political cartoon in Figure 3.1, which is an illustration of visual metaphor. This cartoon invites us to compare the male figure (presumably the Dutch right-wing politician Geert Wilders), which functions as a target, to a film director, functioning as a source. Since this illustration meets the theoretical definition of metaphor as a form of cross-domain mapping across two conceptual domains (Steen et al. 2010), the cartoon shows how metaphor may also be realized by visual communication. The problem addressed in this chapter is how such a metaphor may be identified in the wild by means of an explicit procedure.

There are different methods of identifying and analysing visual metaphor around (Andriessen, Kliphuis, McKenzie, & Van Winkelen 2009; Forceville 1996, 2002, 2008; Kaplan 2005). They add a valuable contribution to the field, because they aim to provide independent criteria for what counts as metaphor, enabling scholars from various academic disciplines (e.g., linguistics, cognitive and social sciences, cultural studies) to do methodologically sound research into visual metaphor. However, the state of the art lacks convergence. It is the goal of the present chapter to bring together their most important insights as well as insights from verbal metaphor identification in order to make a substantial advance in visual metaphor identification.

A first account of identifying pictorial metaphor was proposed by Forceville (1996, 2002, 2008), particularly in the context of advertising. He claims that three questions should be answered "for the identification and interpretation of anything that purports to be a metaphor" (1996, p. 201): (1) What are the two terms (or domains) of the metaphor? (2) What is the primary subject (or target domain) and what is the secondary subject (or source domain)? (3) Which features of the secondary subject (or source domain) must be projected upon the primary subject (or target domain)? These questions quite clearly lead to an *analysis* of the pictorial metaphor.

https://doi.org/10.1075/celcr.18.03sor

Figure 3.1 Cartoon by Tom Janssen (via http://tweedekamer.blog.nl/politieke-prenten/2008/02/03/politieke-prent-het-islamdebat-geert-s-regie, last accessed April 15, 2013). The chair shows the word 'DIRECTION' (translated by the author from the Dutch word 'REGIE'). The words 'THE ISLAM DEBATE (translated by the author from the Dutch 'HET ISLAMDEBAT') are written at the top-centre of the cartoon

It cannot be argued, however, that these questions lead to *identification*, as identification of the metaphor is already assumed in the first question.

Forceville's (1996, 2002, 2008) intentions are in line with Kaplan's (2005), who presents an "outline of a process for identifying and analyzing visual metaphors" (p. 171), with a focus on those in advertising. Indeed, his procedural steps are relevant to identification ("Note the Presence of a Metaphor", p. 172) as well as analysis ("Designate the Two Terms of the Metaphor" and "Identify the Properties That Are Transferred to the Product", p. 173). Andriessen et al. (2009), focusing on visual metaphor in knowledge management, present their method as being a metaphor *analysis* tool rather than an *identification* tool. Their procedural steps, however, could also pertain to visual metaphor identification ("Identify dominant metaphors", p. 409, and "Identify metaphor used", p. 410). These existing methods of visual metaphor identification and analysis are highly suitable to get a grip on visual metaphor, but two additional aspects need to be taken into account to get an even firmer grip on the phenomenon.

First of all, existing methods appear to aim at two goals simultaneously: *identifying* metaphorical visual expression and *analysing* visual metaphor in terms of its target domain, source domain, and the set of mappings between them. Relying on insights gained in linguistic metaphor research, however, there are good reasons to keep visual metaphor identification and visual metaphor analysis separated (Steen 2007; Steen 2009; Steen et al. 2010). As identifying conceptual structures of visual metaphor is probably less simple than determining whether or not a visual element

is used metaphorically, methods deferring conceptual metaphor identification would lead to higher reliability than methods including it. In addition, identifying conceptual structures of visual metaphor is not necessary for an analyst to identify a visual as metaphorically used. It is sufficient to say that there are two domains and that they may be related by some cross-domain mapping (see also Steen 2007, p. 285). Therefore, it is desirable and feasible to separate visual metaphor identification from analysing conceptual structures.

Secondly, existing methods of visual metaphor identification and analysis were developed independently of MIP and MIPVU, two procedures for identifying metaphor in verbal discourse (Pragglejaz Group 2007; Steen et al. 2010). MIPVU is a refinement and extension of MIP for various reasons that will not all be mentioned here (see Steen et al. 2010, pp. 220–221). One of the advantages of MIPVU that is relevant to the current study, however, is that it captures direct expressions of metaphor as well as indirect expressions of metaphor, whereas MIP was restricted to finding the latter form. We hold that MIPVU provides new insights into visual metaphor, in terms of the conceptualization of the phenomenon, but also in terms of the operational steps of metaphor identification, since many visual metaphor seems to relate more closely to direct rather than to indirect metaphor, as we will suggest.

This chapter aims at the development of a Visual Metaphor Identification Procedure (hereafter called VISMIP).[1] Our concern is finding visual forms of metaphor. We will not yet be concerned with ways in which we should connect those visual forms to conceptual metaphor. In addition, the focus is on single static images in persuasive genres, namely political cartoons and advertisements. Our final goal is to develop a procedure that applies to a much wider ranges of images but this is not the ambition of the current study. This means that small aspects of the procedure may have to be adjusted or generalized later on, as we shall indicate.

It should also be made clear from the outset that VISMIP does not necessarily describe the mental operations of those who are confronted with visual metaphor. We are careful to make a distinction between research into the structures and functions of metaphor, on the one hand, and the way metaphors are processed in people's individual minds, on the other. These are two domains of study that require the collection and analysis of entirely different data. At most, the development of VISMIP was given a behavioural boost by insights into visual metaphor processing (Šorm & Steen 2013; Van den Heerik et al. this vol.). Also, VISMIP is intended to be aligned with psycholinguistic insights into verbal metaphor processing that led to the design of the verbal metaphor identification procedures MIP (Pragglejaz Group

1. This work is part of the research programme *Visual metaphor: a psycholinguistic perspective*, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO). We thank all team members for their helpful comments on an earlier draft of this chapter.

2007) and MIPVU (Steen et al. 2010). It is the job of the next section to proceed from this starting point in formulating related required properties for a visual metaphor identification procedure.

The outline of this chapter is as follows: Section 3.2 offers a conceptual analysis of visual metaphor identification. We explore how insights about verbal metaphor identification expressed in MIPVU may be adapted to the study of visual metaphor, and identify ten distinct issues that should be addressed to make adaptation possible. Having identified these fundamental conceptual issues in visual metaphor identification, we address them in operational terms Section 4.3, so that we can proceed with specifying the various stages of visual metaphor identification. We present the steps and instructions of VISMIP, and demonstrate them through exemplary analyses in Section 3.4.

3.2 VISMIP: Conceptualization

This section develops ideas for a new method for visual metaphor identification (VISMIP). First, we briefly go into different types of metaphorical language use (3.2.1). In Section 3.2.2, we present the part of the MIPVU procedure that pertains to indirect metaphor. In Section 3.2.3, we present the part of the MIPVU procedure that pertains to direct metaphor. In the sections that follow (3.2.4–3.2.6), we formulate the issues that are to be addressed when adapting the MIPVU procedure for finding metaphor in language to finding metaphor in images.

3.2.1 Linguistic classes of metaphor: Indirect and direct metaphor

In studies of linguistic metaphor identification, different classes of metaphor have been discovered (e.g., Herrmann 2013; Kaal 2012; Krennmayr 2011; Steen 2008; Steen et al. 2010). Three clear categories of linguistic metaphor have been distinguished, one of which, implicit metaphor (Steen et al. 2010, pp. 39–40), will not be discussed here as it does not clearly relate to images. In order to develop a method for identifying visual metaphor, the first step should be to explore if the other two classes could also apply to visual metaphor.

The first class of metaphor is called as 'indirect metaphor'. In this type of metaphor, metaphorical meaning is expressed by indirectly used language. Language is said to be indirectly used when its meaning in context can be contrasted with a more basic meaning when it is used in other contexts (Steen et al. 2010). To illustrate this class of metaphor, Steen et al. (2010, p. 11) use the example *If you use that strategy, he'll wipe you out*, where *wipe out* are indirectly used words that have to be interpreted by comparison to a more basic sense (related to cleaning).

The second class of metaphor is called 'direct' metaphor, where 'direct' refers to the direct language by which metaphorical meaning is expressed. Direct metaphors occur in the form of similes, analogies, and other non-literal comparisons. Steen et al. (2010, p. 11) illustrate this class of metaphor by Shakespeare's renowned line *Shall I compare thee to a summer's day*? As readers of this line, we do not need to contrast contextual and basic meanings of words, as in indirect metaphor, but the source domain language (*summer's day*) is used directly to trigger a cross-domain mapping. Both the target domain and the source domain are introduced explicitly, and we are explicitly invited to construct a comparison between the two domains.

Direct metaphor is not just linguistically different from indirect metaphor, but also communicatively different (Steen 2008). Where most indirect metaphor gives the addressee some freedom whether to pay explicit attention to the source domain and then build a cross-domain mapping or not, direct metaphor gives the addressee no other option than to perform those mental actions (Steen 2008). In the Shakespearean line mentioned above, the alien source domain ('summer's day') is deliberately introduced into the dominant target domain ('thee') so that the addressee must concentrate on the source domain and perform a mapping to solve the arisen tension. Direct metaphor is thus called deliberate metaphor: a metaphor that is used *as* metaphor in order to alter the addressee's perspective on a current discourse referent or topic (Krennmayr 2011; Pasma 2011; Steen 2008, 2010, 2011, 2015).

3.2.2 Finding metaphor in the Method for Linguistic Metaphor Identification (MIPVU)

MIPVU is based on the theoretical idea that linguistic metaphor works on the basis of incongruity. It entails the introduction of a foreign source domain into the dominant target domain of the discourse, violating conceptual coherence. This has to be resolved by a mapping from the source domain to the target domain (Steen 2007; Steen et al. 2010, pp. 11–12).

As noted, incongruity in linguistic metaphor may take two different forms, depending on the type of metaphor (Steen et al. 2010). In indirect metaphor, the basic sense of a word, functioning as source domain, is said to be incongruous with the meaning of the surrounding co-text. For example, the basic sense of *valuable* ('worth a lot of money') in incongruous within the sentence *She is doing valuable work in school.*[2] As there is contrast as well as comparison between the contextual ('very useful and important') and the basic meaning ('worth a lot of money'),

2. Example borrowed from the VU Amsterdam Metaphor Corpus website, http://www2.let.vu.nl/oz/metaphorlab/metcor/documentation/relation_to_metaphor.html (last accessed January 27, 2015).

valuable is given a metaphorical meaning. For direct metaphor, the incongruity is of another form: it is the newly introduced, literally expressed source domain that is said to be incongruous with the rest of the text. In *He's like a ferret*[3], for instance, the source domain term *ferret* is used directly on the linguistic level, creating incongruity with the rest of the text. Metaphorical meaning arises through contrast and comparison between two different domains (people and animals)that function as distinct and different types of referents in the meaning of the utterance.

The next step is to examine the part of MIPVU that captures the two categories of linguistic metaphor. The Pragglejaz Group (2007) developed MIP, a procedure for identifying metaphorically used words in discourse, which in MIPVU are called indirect metaphor. We will not spell out each single step constituting the MIP procedure. For the purposes of the present study, it suffices to summarize the procedure: having established a general understanding of the discourse (step 1), the discourse should be divided into lexical units (step 2). For each unit, its contextual meaning should be established (step 3a), followed by its more basic meaning in other contexts than the one observable (step 3b). Then the analyst should decide whether the meanings established under the previous steps are sufficiently distinct yet sufficiently similar (step 3c). If so, the unit is marked as metaphorical (step 4).

We now cite the procedural steps of MIPVU that should lead to identification of directly used lexical units. MIPVU assumes that before executing these steps, the analyst has already completed step 1 and step 2 of MIP: (1) getting a general impression of the discourse as a whole and (2) dividing the entire discourse into lexical units. The MIPVU procedure targeted at direct metaphor looks like this:

1. "Find local referent and topic shifts.
   - Good clues are provided by lexis which is "incongruous" (Cameron 2003; Charteris-Black 2004) with the rest of the text.
2. Test whether the incongruous lexical units are to be integrated within the overall referential and/or topical framework by means of some form of comparison.
   - Good clues are provided by lexis which flags the need for some form of similarity or projection (Goatly 1997).
3. Test whether the comparison is nonliteral or cross-domain.
   - Cameron (2003, p. 74) suggests that we should include any comparison that is not obviously non-metaphorical, such as *the campsite was like a holiday village*. Consequently, whenever two concepts are compared and they can be constructed, in context, as somehow belonging to two distinct and contrasted domains, the comparison should be seen as expressing a cross-domain mapping. Cameron refers to these as two incongruous domains.

3. See the previous note.

4. Test whether the comparison can be seen as some form of indirect discourse about the local or main referent or topic of the text.
   – A provisional sketch of a mapping between the incongruous material functioning as source domain on the one hand and elements from the co-text functioning as target domain on the other should be possible.
   If the findings of tests 2,3, and 4 are positive, then a word should be marked for direct metaphor ('MRW, direct')." (literally copied from Steen et al. 2010, pp. 38–39; see also pp. 14–15).

Next, for each of the steps cited above we will discuss (a) what adaptation to visual metaphor entails, (b) whether existing methods of visual metaphor identification and analysis have already incorporated such a step and (c) what issues should be addressed to establish adaptation. This should lead to an appropriate and well-motivated conceptualization of each phase of visual metaphor identification.

3.2.3 Understanding the image's meaning and unitizing: Conceptualization

Adaptation of MIPVU to VISMIP

The first step in MIP and MIPVU is: "Read the entire text/discourse to establish a general understanding of the meaning" (Steen et al. 2010, p. 5). This step is indispensable in linguistic metaphor identification. Without it, one cannot establish the meaning of a lexical unit in context (step 3a in MIP). Without knowing the contextual meaning of a lexical unit, identifying the indirect form of linguistic metaphor is impossible. Understanding the general meaning of the text is also necessary to identify the direct form of linguistic metaphor: without a general understanding of the entire text, it is impossible to know the topic of the text. Without knowing the topic of the text, it is impossible to find referent and topic shifts (step 1 in MIPVU).

The essential condition of a general understanding of the entire materials also applies to visual metaphor. If we do not reach a general understanding of an image, it will be impossible to find incongruous visual elements. After all, we need some degree of understanding to judge the coherence of the image. If we cannot detect visual incongruity, we will not be able to identify visual elements that may be metaphorically used (see also Schilperoord this vol.). So, VISMIP should include a step leading to a general understanding. That would then be the visual counterpart of the following step: "Read the entire text/discourse to establish a general understanding of the meaning" (Steen et al. 2010, p. 5).

The second step in MIP and MIPVU is: "Determine the lexical units in the text/discourse" (Steen et al. 2010, p. 5). This step presupposes that each analyst applying the procedure is supposed to make the following decision: what counts as a stable unit of analysis (Steen et al. 2010, p. 13)? Linguistic metaphor may be found at many

different levels, varying from morphemes to longer stretches of discourse (Steen 2007; Steen et al. 2010). According to Steen (2007), then, studies concerned with finding metaphor, in whichever area of research, will have to specify which unit is being selected for analysis. MIP and MIPVU have chosen the word as the unit of analysis, which is called lexical unit for theoretical reasons (Pragglejaz Group 2007; Steen et al. 2010). In order to improve the explicitness and reliability of the method, guidelines have been set up for deciding what counts as a lexical unit (Steen et al. 2010, pp. 27–32).

Clearly, MIP and MIPVU are explicit where in language metaphor is to be found. Similarly, VISMIP should be explicit about where in the image metaphor is to be found. If analysts study a visual and decide on coding it as related to visual metaphor, they have to be sure about where exactly in the image they looked for finding visual metaphor. This is in line with the generally accepted idea that in content analyses, the units under study should be carefully selected (Krippendorff 2004; Neuendorf 2002; Riffe, Lacy & Fico 2005). VISMIP will therefore include a step leading to a division of images into separate units. That would then be the visual counterpart of the following step: "Determine the lexical units in the text/discourse" (Steen et al. 2010, p. 5).

Understanding and unitizing in existing methods

Kaplan (2005) and Forceville (1996, 2002, 2008) do not explicitly propose a step in which the analyst tries to establish a general understanding of the picture. Instead, in the first steps of their methods, the analyst is supposed to signal pictorial elements that seem distorted or out of place (Kaplan 2005), or to indicate the two terms of metaphor (Forceville 1996, 2002, 2008). It is possible that Kaplan (2005) and Forceville (1996, 2002, 2008) assume a general understanding of the visual in their methods, and that they just decided not to include a separate, explicit step to establish that. Andriessen et al. (2009) do seem to have included a step that approximates the first step of MIP and MIPVU: "Identify dominant metaphor(s) in the visuals by looking with a birds-eye view at the visuals" (p. 409). The activity of "looking with a birds-eye view at the visuals" (p. 409) could be considered as the visual counterpart of "Read the entire text/discourse" (Steen et al. 2010, p. 5), although it remains speculative what it means to look with a birds-eye view, and what it should result in.

How do existing methods of metaphor identification and analysis deal with unitizing? Kaplan (2005) and Forceville (1996, 2002, 2008) appear to have taken the entire visual as a unit of analysis, examining whole pictures for metaphorical use. Since they do not explicitly address the selected unit of analysis, however, this cannot be concluded with certainty. Furthermore, if it is indeed the case that they

have selected the whole picture as a unit of analysis, it remains unaddressed why they have chosen to do so, and why other options, such as coding parts of the image, have not been selected. Andriessen et al. (2009) appear to have chosen a different unit of analysis, looking at each visual element individually and asking themselves whether or not the visual element is used metaphorically. What exactly counts as a visual element, however, remains somewhat unclear.

Raising issues

We have identified four issues related to understanding and unitizing the image, two concerning general understanding and two concerning unitizing. We hold that a general understanding of the visual materials is necessary for finding visual incongruity, or, to formulate it less strictly, that a general understanding would at least facilitate this. This involves two problems.

First of all, we need to know more exactly what it means to have a general understanding of the meaning of an image. 'Meaning' is an umbrella term that may cover a number of dimensions in relation to images, including the things that are depicted, the notions they bring to mind, and the communicative intentions conveyed. In order to develop the first step of VISMIP, we need to become clearer on what should be understood by the meaning of an image.

It has become clear that we need to know what it means exactly that analysts arrive at a general understanding of an image's meaning. It is also important that analysts arrive at a general understanding in a systematic and reliable manner. Ideas about what a visual shows and what it is about may differ between analysts. If these ideas differ too strongly, there is no point in continuing the procedure of identification. Therefore, it is important to develop explicit instructions for establishing a general understanding of the image.

In order to develop a procedure that leads to a unitized image, two other issues are to be addressed. First of all, we have to be sure where visual metaphor can be found. Do we want to find it in the image as a whole, similar to Kaplan (2005) and Forceville (1996, 2002, 2008), or in the parts constituting the image, similar to Andriessen et al. (2009)? Second, once we have defined our unit of analysis, we have to make sure that these units are identified in a reliable manner. If unitizing reliability is lacking, it is pointless to determine whether or not analysts agree on metaphorical and non-metaphorical cases: if analysts do not agree on the cases that they identify as metaphorical, it could be possible that lack of agreement is due to prior disagreement about the components of the cases that are to be categorized as metaphorical or non-metaphorical. Setting up guidelines for deciding on visual units should increase unitizing reliability.

All in all, this leads to the following questions:

1. How may the meaning of an image be defined?
2. What procedural step(s) should be taken in order to reach a general understanding of an image's meaning?
3. How do we define our unit of analysis?
4. Having defined our unit of analysis, how do we handle it in applying the procedure?

To be able to answer these questions, we need more insights into the meaning of images, and we need to explore the ways in which images can be decomposed into units. This will happen in Section 3.3.1.

3.2.4 Finding incongruity and testing need for comparison: Conceptualization

Adaptation of MIPVU to VISMIP

If we assume that visual metaphor also works on the basis of incongruity (cf. Schilperoord, this vol.), VISMIP should include a step leading to the discovery of visual incongruity. In this context we need to distinguish between direct and indirect metaphor. For indirect metaphor identification, we need to see whether any word in an utterance has a contextual meaning that is distinct from but comparable to a more basic meaning in other contexts. This may apply equally well to any visually depicted object in an image, where the object has a contextual 'meaning' which is different than its basic 'meaning' in other contexts. When turning to direct metaphor in language, we need to look for a visual counterpart of the following step in the MIPVU procedure, which focuses on direct metaphor: "Find local referent and topic shifts" (Steen et al. 2010, p. 38).

In finding direct linguistic metaphor, it should also be tested whether the incongruous lexical units are to be incorporated within the topic of the discourse (Steen et al. 2010). The following question should help the analyst: are there any lexical signals present which indicate the need for comparison? MIPVU focuses on various signals of potential cross-domain mapping, including *like*, *as*, *compare*, *similar*, and *imagine* (see Steen et al. 2010, pp. 40–41 for a first overview of markers; cf. Goatly 1997). Similar to linguistic metaphor identification, it is important for visual metaphor identification to test the need for comparison. Detecting visual incongruity is necessary for analysts to identify a case of visual metaphor, but it is not sufficient. They also need clues that the visual incongruity is to be resolved by comparison. We should find the visual equivalent of the following step in the MIPVU procedure: "Test whether the incongruous lexical units are to be integrated within the overall referential and/or topical framework by means of some form of comparison" (Steen et al. 2010, p. 38).

Finding incongruity and testing need for comparison in existing methods

In his method for analysing visual metaphor, Kaplan (2005) hints at finding incongruity under the first step of noting the presence of metaphor:

> The first hint that some feature (or combination of features) is intended to evoke a metaphorical response in the viewer is the presence of images or pictorial elements that seem distorted or out of place. Often, the cue is a violation of the viewer's understanding of physical reality. (2005, p. 172)

Kaplan (2005) continues to elaborate on the cues that should help the analyst to find incongruity. He mentions some of the techniques by which the viewer's understanding of physical reality can be violated: modification of physical characteristics, inappropriate setting or depicted function, and juxtaposition. In contrast to Kaplan (2005), Forceville (1996, 2002, 2008) does not propose an explicit step pertaining to the discovery of incongruous elements, nor do Andriessen et al. (2009). Finding incongruity has hence been incorporated in Kaplan's (2005) method of visual metaphor analysis, but existing methods of visual metaphor identification and analysis have not included a separate explicit step in which the analyst decides whether the discovered incongruity is to be resolved by means of comparison.

Raising issues

We need to address four issues before we can incorporate the steps of finding incongruity and testing the need for comparison in VISMIP. First of all, we need an operational definition of visual incongruity. If we conceive visual incongruity as visual elements that seem distorted or out of place, like Kaplan (2005) appears to do, we should at least think about how to define distortion or out-of-placeness. We should also consider the possibility that visual incongruity takes another form than a violation of physical reality, and that more than one type of visual incongruity thus exists.

Secondly, we need a clear set of guidelines for deciding about when exactly visual incongruity occurs. Kaplan (2005, pp. 172–173) explains how images are made incongruous by describing several techniques, including modification of physical characteristics, inappropriate setting or depicted function, and juxtaposition, but there may be other techniques to be explored.

Thirdly, we should think about how we can decide on the units that could be the candidates for some form of comparison. In some direct verbal metaphor, like *A is B* constructions, it is clear which units are the candidates for comparison: both the word(s) expressing the dominant conceptual domain (target domain) and the word(s) expressing the alien conceptual domain (source domain) are given, for instance in the sentence *He is like a drug to me.* This may also happen in visual metaphors, as when a Gibson guitar and a nuclear explosion are placed side by

side. In many visual metaphors, however, the candidates may be less clear as a result of distortion or out-of-placeness (see Schilperoord, this vol., for examples). In order to decide about the unit that could function as source or target domain, a process of reasoning seems necessary: first to start thinking about what exactly causes the inappropriateness of a unit within the image, and then thinking about what would be a more appropriate unit instead. For instance, in Figure 3.1, it can be said that Geert Wilders appears out of place within the film set location. The concept that would be in place ('film director') is not expressed with all its usual characteristics – only the characteristic of being located in a director's chair on a film set are expressed – so that requires more reasoning compared to a visual form where both Geert Wilders and the film director would be placed side by side *(*the visual equivalent of the linguistic metaphor *Geert Wilders is like a film director*).

The fourth issue refers to the decision about whether some form of comparison is needed. In linguistic metaphor identification, metaphor flags are sometimes available for motivating this decision (Steen et al. 2010), for instance *resembling* in the sentence *Poplar leaves have an elegant outline resembling that of an Arab minaret*.[4] The obvious question is: is there a visual equivalent of metaphor flags that can be used in VISMIP? Are there any signals alerting us that some form of comparison may be at play between two units? If not, we may need other ways to decide if the incongruous visual units are to be integrated within the overall referential and topical framework by means of comparison.

The issues can be summarized as follows:

5. How may visual incongruity be operationally defined?
6. How can we find visual incongruity?
7. How can we decide on the units that could be the candidates for some form of comparison?
8. How can we decide whether some form of comparison is needed?

We will address these issues in Section 3.3.2.

3.2.5 Testing cross-domain-ness and indirectness: Conceptualization

Adaptation of MIPVU to VISMIP

As metaphorical use depends on a contrast between concepts, MIPVU instructs the analyst to test whether the comparison between meanings is cross-domain. For indirect metaphor, the operational criterion for deciding on sufficient contrast is

4. Example borrowed from the VU Amsterdam Metaphor Corpus website, http://www2.let.vu.nl/oz/metaphorlab/metcor/documentation/signals.html (last accessed January 27, 2015).

whether the contextual sense and the basic sense are listed as two separate (numbered) descriptions in the dictionaries (Steen et al. 2010, p. 54). Such a sharp operational criterion is lacking for direct metaphor, as this form of metaphorical meaning does not arise from contrast between basic and contextual senses, but rather from a contrast between two conceptual domains in the context of an utterance. To distinguish non-metaphorical from metaphorical similarity, MIPVU suggests that two domains need to cause topical incongruity, at least to some extent (Steen et al. 2010, p. 96).[5] This criterion also underlies the more specific contrast between two senses in indirect metaphor (Steen et al. 2010, pp. 11–12).

Assuming that visual metaphor depends on a contrast between two domains, similar to linguistic metaphor, an equivalent of the 'nonliterality test'[6] will have to be included in VISMIP. That would then most clearly be the visual counterpart of the following step in the MIPVU procedure pertaining to direct metaphor: "Test whether the comparison is nonliteral or cross-domain" (Steen et al. 2010, p. 39).

The final test for identifying direct metaphor is: does the comparison convey a meaning about the referent or topic of the text (Steen et al. 2010, p. 39)? In order to answer this question, the analyst will have to be able to relate the source domain to the target domain. (This, again, is the more general variant of the more specific case in indirect metaphor, where the basic meaning is the basis for comparison with the intended contextual meaning while the latter is also meant as pertaining to the local referent or topic of the utterance.) Precisely pinpointing connections between the domains can be postponed until a conceptual analysis takes place. For identification of metaphor, it is sufficient to execute cross-domain mapping in a sketchy way (Steen 2007; Steen et al. 2010).

The sketching of cross-domain mappings in MIPVU should be adapted to VISMIP. It is not sufficient to say that two visually expressed concepts belong to different domains. The next step would be to test whether the comparison has something relevant to say about the topic of the image. We should find the visual equivalent of the following step in the MIPVU procedure for finding direct metaphor: "Test whether the comparison can be seen as some form of indirect discourse about the local or main referent or topic of the text" (Steen et al. 2010, p. 39).

---

5. This rule is not completely satisfying, as topical incongruity should already be established under step 1 of MIPVU. So, if under step 1 topical incongruity is established, then under step 3 the comparison is automatically nonliteral according to the rule. In other words, a positive finding for test 1 (test on topic shift) implies a positive finding for test 3 (test on nonliterality), making test 3 redundant.

6. The term 'cross-domain-ness' is used for VISMIP instead of 'nonliterality', as it is more appropriate to visual materials.

Testing cross-domain-ness and indirectness in existing methods

Kaplan (2005) recognizes the importance of cross-domain-ness and cross-domain mapping when he discusses 'the threshold question' (2005, p. 174), referring to the question as to *when* to classify a picture as metaphorical. As a solution to this question, he proposes the following rule:

> To classify a picture as metaphorical there should be enough similarities between the two elements with respect to physical form or abstract qualities as to permit a tentative analogy, yet enough differences between them to cue the viewer that the artist did not intend the depiction or representation to be taken as literally true.
> (2005, p. 174)

When Kaplan (2005) says that there should be "enough differences between [the two elements] to cue the viewer that the artist did not intend the depiction or representation to be taken as literally true", he proposes the same criterion as in the nonliterality test as included in the MIPVU procedure. When he says there should be "enough similarities between the two elements with respect to physical form or abstract qualities as to permit a tentative analogy", he refers to a similarly required comparison between the two domains of meaning. In contrast to Kaplan (2005), Forceville (1996, 2002, 2008) and Andriessen et al. (2009) do not explicitly address the problem of deciding on sufficient distinctness in their procedures.

Existing methods of visual metaphor identification and analysis have incorporated a step aimed at conceptual mapping. Forceville (1996, 2002, 2008) proposes the question: which features of the secondary subject (or source domain) must be projected upon the primary subject (or target domain)? Similarly, Kaplan (2005) advises to "Identify the Properties That Are Transferred to the Product" (p. 173), where the product corresponds to the target domain. Andriessen et al. (2009) present the question "What is the meaning of the visual element when applied to the target domain?" (p. 409, Table 2), also referring to the mapping from source domain to target domain. All three questions, however, seem to ask for a rather precise description of the connections between the source domain and the target domain instead of giving the major point(s) of similarity.

Raising issues

In line with MIPVU, and following Kaplan (2005), VISMIP should include a step to check whether there are enough differences between elements involved in a comparison. It would also be helpful, however, to become clear on what exactly counts as enough. In addition, VISMIP should include a step that tests if a mapping between domains can be sketched. So, the issues related to cross-domain-ness and indirectness can be formulated as follows:

9. How can we decide whether two concepts belong to two distinct and contrasted domains?
10. How can we decide whether the comparison can be seen as some form of indirect discourse about the topic of the image?

Issues 9 and 10 will be handled in Section 3.3.3.

3.2.6 Summary: Ten issues in visual metaphor identification

This section has presented a brief conceptual analysis of visual metaphor identification. It explored how the MIPVU procedure for linguistic metaphor identification can be adapted to the VISMIP procedure of visual metaphor identification. Furthermore, we raised several issues that need to be addressed in the process of adaption. The ten issues are summed up here:

1. How may the meaning of an image be defined?
2. What procedural step(s) should be taken in order to reach a general understanding of an image's meaning?
3. How do we define our unit of analysis?
4. Having defined our unit of analysis, how do we handle it in applying the procedure?
5. How may visual incongruity be operationally defined?
6. How can we find visual incongruity?
7. How can we decide on the units that could be the candidates for some form of comparison?
8. How can we decide whether some form of comparison is needed?
9. How can we decide whether two concepts belong to two distinct and contrasted domains?
10. How can we decide whether the comparison can be seen as some form of indirect discourse about the topic of the image?

The next section aims to present possible solutions to these issues.

3.3 VISMIP: Operationalization

This section addresses the issues of visual metaphor identification that were presented in Section 3.2, taking us to the operationalization of VISMIP. First of all, we deal with the problem of reaching a general understanding of image's meaning and then discuss the way in which visual data can be unitized in VISMIP (Section 3.3.1). We then explore how the general notion of visual incongruity can be found for each

unit of analysis and also how the need for comparison may be tested (Section 3.3.2). Section 3.3.3 ends with a discussion on how it may be checked whether the domains involved in the comparison are distinct, and whether there is indirect discourse about the image's topic.

3.3.1 Understanding the image's meaning and unitizing: Operationalization

In Section 3.2.4, we raised the following four issues:

1. How may the meaning of an image be defined?
2. What procedural step(s) should be taken in order to reach a general understanding of an image's meaning?
3. How do we define our unit of analysis?
4. Having defined our unit of analysis, how do we handle it in applying the procedure?

Having the answers to questions 1 and 2 should enable us to adequately adapt MIPVU's first procedural steps "Read the entire text/discourse to establish a general understanding of the meaning" (Steen et al. 2010, p. 5) to visual materials. Addressing the issues 3 and 4 should help us adapting the step: "Determine the lexical units in the text/discourse" (Steen et al. 2010, p. 5). First, we deal with the issue of defining the image's meaning.

Addressing issue 1: Defining the image's meaning

To get more insight into the meaning of images, we need to look at three approaches to visual analysis: visual semiotics, iconography, and visual argumentation. For our discussion of visual semiotics and the iconographical method of analysis, we primarily rely on Van Leeuwen (2001). Semiotics is the study of both linguistic and non-linguistic signs, where signs usually refer to the basic units of symbolic communication (Jensen 2008; Lyne 2008). Where De Saussure and Peirce were the main figures in developing semiotics in logic and linguistics, Barthes contributed greatly to the consolidation of the field in communication research (Jensen 2008). A semiotic analysis in a visual communication context means using a refined conceptual apparatus for describing "the exact ways in which the meanings of an image are produced through that image" (Rose 2005, p. 72).

According to Van Leeuwen (2001), the essence of visual semiotics along the lines of Barthes is "the layering of meaning" (p. 94). The first layer is called *denotation* and refers to the recognition of visual elements in the image. Denotative meaning answers questions such as: What is the subject (person or thing) in the image? What action is being performed? Where is the subject?

The second layer is referred to as *connotation* and refers to what the recognized visual elements stand for. Connotative meanings are broader concepts associated

with represented people, places, or things, including cultural meanings and ideologies (Van Leeuwen 2001; Moriarty 2005; Penn 2000). They may be constructed through the cultural associations that are attached to the represented visual elements, or through the way in which they are represented, such as style or technique.

The distinction between denotation and connotation can be illustrated by Van Leeuwen's (2001, p. 93) analysis of a textbook image. The image depicts three women wearing headscarves. On the first (denotative) level of meaning, the headscarves form a denotative signifier (i.e. in the visual realm, the depiction that refers to a concept). The denotative signified (i.e. the concept to which the depiction refers), then, is 'immigrant women'. On the second (connotative) level of meaning, particular techniques of photography (angle and framing) are the connotators (or: the signifiers for connotation). The fact that we see the women in a long shot in a profile angle is taken as a signifier for 'detachment' and 'socially distant from the viewer' (Van Leeuwen 2001, p. 99), which together form the depiction's connotation.

Iconography is a qualitative method of visual content analysis and interpretation, which has been refined and popularized by art historian Panofsky (Müller 2008). Like visual semiotics, iconography is characterized by layering of meaning. Where visual semiotics distinguishes two layers, however, iconography distinguishes three: representational meaning, iconographical symbolism, and iconological symbolism.

According to Van Leeuwen (2001), representational meaning is comparable to the notion of denotation in visual semiotics, as it involves a description of who or what is depicted, what action is performed, where the action is performed, etcetera. Identification of this 'primary subject matter' may be established on the basis of the title of the work, personal experience, background research, visual or verbal intertextuality (Van Leeuwen 2001, following Hermeren 1969).

Iconographical symbolism refers to the ideas or concepts connected to the first layer of meaning (i.e. representational meaning). It involves generally accepted conventions, which the artist would be probably aware of and would consciously use. As an example of iconographical symbols, Van Leeuwen (2001) mentions fruit as a racist iconographical symbol of black people's 'laziness'. To identify the 'secondary meaning' of a depiction, one should look for arguments that symbolic meanings may be intended. These arguments may be textual, such as a symbolic motif made extra salient by composition, or contextual, for example when a particular symbolic motif is known to be a commonly used symbol in a certain art period (Van Leeuwen 2001, following Hermeren 1969).

Iconological symbolism (or: iconology), the third layer of meaning, establishes the meaning of an image in a wider context, for example historical or social (Clarke 2010). As this type of meaning goes beyond knowledge of symbols and motifs towards ideological meaning, the artist may not even be aware of the iconological

symbolism in his work. Van Leeuwen (2001) suggests that pictorial analysis on this level should lead to interpretation or explanation rather than description: "Iconological analysis (…) draws together the iconographical symbols and stylistic features of an image or a representational tradition into a coherent interpretation which provides the 'why' behind the representations analysed" (p. 116). Whereas representational meaning in iconography corresponds to denotative meaning in visual semiotics, symbolic meaning (both iconographical and iconological) corresponds to connotative meaning (Van Leeuwen 2001).

Within argumentation theory, visual argumentation is a relatively new area of research. Although there have been reservations about whether visuals arguments are actually possible (Birdsell & Groarke 1996; Blair 1996; Fleming 1996; Johnson 2003), there is growing consensus that some visual images may function as arguments intended to persuade people (e.g., Alcolea-Banegas 2009; Kjeldsen 2012; Ripley 2008). Consequently, scholars from various argumentation-theoretical schools have begun to expand their views on verbal arguments to visuals.

Feteris, Groarke and Plug (2011), for example, consider cartoons to be part of a critical discussion where standpoint and arguments are advanced in a visual manner. In order to understand the choices the cartoonist has made in this discussion, they regard reconstruction of the argumentation as a necessary step: "In determining the intended meaning of the cartoon, which often contains a visual metaphor, it must be established to which standpoint and to which arguments the cartoonist has committed themselves" (pp. 62–63).

Similar to cartoons, advertisements are considered to be argumentative. According to Van Eemeren (2010), advertisers are protagonists who try to defend the standpoint that the product or service should be bought, although this standpoint is usually left implicit. So once we have identified an image as being an ad, we will soon be able to infer the standpoint carried by the image (Kjeldsen 2012). In future work, where visual metaphor identification will have to branch out to non-argumentative visuals, this reconstruction of the point of the text may have to be more general, relating to for instance narrative and expository texts as well.

Let us give an explicit answer to the question: how may the image's meaning be defined (question 1)? The visual semiotics and iconography approaches to images teach us that different levels of visual meaning should be distinguished. Pictures are presumed to have a layer of denotative (or representational) meaning referring to the visual elements recognized in a picture, and a layer of connotative (or symbolic) meaning referring to the associated interpretations of these recognized elements. The visual argumentation approach to images indicates another dimension of meaning, which refers to the intended meaning of an image, which can be determined by reconstructing the standpoint and arguments defended.

Since denotation and connotation (as well as representation and symbolism) carry specific associations in linguistics that may be slightly different than in visual semiotics, we are going to adhere to a more simple terminology that is broadly shared, which was also part of the previous discussion as well (see also Van den Heerik et al. this vol.). For denotation (and representation), we are going to utilize the term referential meaning. And for connotation (and symbolism), we are going to employ the term more general and abstract meaning. Moreover, the standpoint of an argumentative text may be seen as the more specific instantiation of the point of any text, a point which is typically made in connection with a dominant topic. This terminology (referential meaning, more general and abstract meaning, point and topic) is not meant to differ from the above terms, generally acceptable, and moreover in line with the way verbal metaphor identification has been handled in MIPVU. These are the terms, then, in which the general understanding of the image will be analysed in VISMIP.

Addressing issue 2: Procedural steps to understanding the image's meaning
The first step in VISMIP is as follows:

1. Look at the entire image, including visual and verbal elements, to establish a general understanding of the meaning.

This instruction is more explicit about what analysts should look at ('including visual and verbal elements'), compared to Andriessen et al.'s (2009) instruction to "look with a birds-eye view at the visuals" (p. 409), but it still leaves room for interpretation. Therefore, four more specific steps were added to guide the analyst in establishing a general understanding. At the same time, the four steps were designed to do justice to the distinction between referential meaning and more general and abstract meaning according to visual semiotics and iconography, and to the notion that an image has a point and a topic. Each of the four steps is now discussed in more detail.

Step 1a: Referential meaning
The first substep of VISMIP instructs the analyst to describe the subject of the image on a referential level: the visual elements that one recognizes in a picture. Developing this step, the field of information science proved to be useful, as this field is concerned with the semantic annotation of images in visual data collections. Annotation of images means that the subjects of the images need to be described, and that these descriptions need to be structured in some manner to enable efficient search behavior. As VISMIP needed an adequate way of describing and unitizing images, potential tools for structuring image descriptions were of interest.

Tam and Leung (2001) present a tool for semantic annotation of visual materials, which appeared useful for VISMIP. They describe the image by using one or more structured statements comprising the elements *Agent, Action, Object, Recipient* and *Setting*. In addition to these elements, optional adjectival and adverbial phrases can be used to describe a given image. Figure 3.2 shows an example of what Tam and Leung (2001) call Structured Annotation.

1. Image 1: *Tall, blond man and spotted dog walk quickly in park*. [Agent(man|tall, blond) Agent(dog|spotted) Action(walk|quickly) Setting(in park)] *Smiling girl gives green apple to frowning boy under tree*. [Agent (girl|smiling) Action(give) Object(apple|green) Recipient(boy|frowning) Setting(under tree)]
2. Image 2: *Fireworks above harbor*. [Object(Fire-works) Setting(above harbor)] *Well-dressed man sings on pier*. [Agent(man|well-dressed) Action(sing) Setting(on pier)] *Well-dressed man dances on pier*. [Agent(man|well-dressed) Action(dance) Setting(on pier)]

Figure 3.2 Example from Tam and Leung (2001, p. 934)

There are two important advantages to the tool: it uses a limited number of components and the components that are used are common to many languages. If a limited number of components are being used, then inter-analyst agreement on the components will probably be increased. And if the components are common to many languages, then analysts will not need specific (linguistic) expertise to structure their image descriptions.

This tool for annotating visual materials can be used for the first as well as the second step of VISMIP. Through this methodology, we can describe an image on the referential level and also structure those descriptions. The first substep of VISMIP, inspired by Tam and Leung (2001), is formulated as follows:

1a. Describe in just a few simple phrases the referential meaning of the image, i.e. what/who is being depicted here, what he is doing, where he is doing it, and so on. For example: 'Tall, blond man and spotted dog walk quickly in park. Smiling girl gives green apple to frowning boy under tree.' If the referential meaning is ambiguous and allows for more than one interpretation, then give alternative descriptions.

Step 1b: Attaching more general and abstract meaning

More general and abstract interpretations come into play in the next substep:

1b. Test whether there are any clues that tell you that more general and abstract meaning should be attached to what is described under step 1a. For example, if the exemplary image described under 1a is accompanied by the caption 'summer joy', we have a clue that the more abstract concept 'summer joy' should be attached to the referential meaning.

The instruction intentionally uses the term 'test', as analysts should look for cues that the sender of the message intended abstract interpretations. VISMIP does not want analysts to start associating without having a clear reason, for this would produce descriptions that are more about the analyst's mind than the picture's meaning. This would probably also reduce inter-analyst agreement.

Step 1c and 1d: Reconstructing the standpoint and deriving the topic
Having described the meaning of the picture on two levels, the referential level (step 1a) and the more general and abstract level (step 1b), it is time to proceed to the next level of meaning. For our materials this is mostly an argumentative level referring to the communicative purpose of the picture, but we can generalize beyond this in the way suggested above. VISMIP wants analysts to answer the question: what point is conveyed by using the picture? So, a third substep was formulated:

1c. Reconstruct the point connected to the image.

Having formulated the point, analysts may continue to step 1d:

1d. Derive from step 1c the topic of the message, i.e. that about which the point is stated.

Tools for these steps come from argumentation analysis and more generally discourse analysis in ways that do not necessarily depend on metaphor.

Addressing issue 3: Defining the unit of analysis
The instrument of Tam and Leung (2001) was used to guide analysts in describing the referential meaning of images in step 1a of VISMIP. It was in addition used to determine what counts as a unit of analysis. To give an explicit answer to the issue of unitizing:

> *In VISMIP, a unit of analysis is defined as a component of a structured description of the referential meaning of an image. A component may receive the label Agent, Action, Object, Recipient, or Setting, following Tam and Leung (2001).*

VISMIP wants to find metaphorical relations between the parts constituting the representational meaning of the image. It is not entirely clear whether this is in line with Kaplan (2005) and Forceville (1996, 2002, 2008), who seem to be marking entire pictures as metaphor, or with Andriessen et al. (2009).

Addressing issue 4: Handling unit of analysis in the procedure
For determining the units of analysis, analysts should use the image description on a referential level, so what they stated under step 1a. Step 2 then is formulated as follows:

2. Structure the descriptive phrase(s) under step 1a. For example: 'Tall, blond man and spotted dog walk quickly in park. [Agent(man|tall, blond) Agent(dog|spotted) Action(walk|quickly) Setting(in park)] Smiling girl gives green apple to frowning boy under tree. [Agent(girl|smiling) Action(give) Object(apple|green) Recipient(boy|frowning) Setting(under tree)'.

To develop guidelines for structuring, we used Tam and Leung's (2001) definitions of each of the description components. The guidelines are shown in Table 3.1.

Table 3.1 Practical guidelines for deciding on step 2 (adapted from Tam & Leung 2001, pp. 933–934)

1. An Agent is a noun, for example, 'man' or 'ball', in the semantic role of a person or thing that is the doer of an action, process or event, for example, 'man walking'. A descriptive phrase may have zero or more Agents. An Agent may be modified by adjectival phrases.
2. An Action is a verb-what the agent is doing, for example, 'walking' or 'rolling'. An Action may be modified by adverbial phrases. A descriptive phrase may have zero or one Action. If an Agent performs multiple Actions, they must be described in different descriptive phrases, for example 'man walking in park; and 'man smoking cigarette'.
3. An Action may have zero or more direct Objects, or zero or more Recipients, for example, 'girl giving apple (Object) to boy (Recipient)'. Objects and Recipients may be modified by adjectival phrases.
4. A Setting encompasses event, place, and time, for example 'during a party' (Event), 'in a park' (Place), 'in the afternoon' (Time). A descriptive phrase may contain several settings, for example 'in a park' (Place), 'under a tree' (Place).

3.3.2 Finding incongruity and testing need for comparison: Operationalization

Section 3.2.4 also raised the following questions:

5. How may visual incongruity be defined?
6. How can we find visual incongruity?
7. How can we decide on the units that could be the candidates for some form of comparison?
8. How can we decide whether some form of comparison is needed?

Having the answers to questions 5 and 6 should enable us to adequately adapt MIPVU's procedural step "Find local referent and topic shifts" (Steen et al. 2010, p. 38) to visual materials. Solving the issues 7 and 8 should help adapting the step:

"Test whether the incongruous lexical units are to be integrated within the overall referential and/or topical framework by means of some form of comparison" (Steen et al. 2010, p. 38). First, a careful attempt is made to solve the issues relating to visual incongruity (issue 5 and 6).

Addressing issue 5: Defining visual incongruity

Each unit of analysis identified under step 2 of VISMIP (the unitizing step) should be systematically tested for visual incongruity. But how can we define visual incongruity? To answer this, the similarities between visual metaphor and verbal metaphor were used as a starting point to reflect on the ways in which visual incongruity may occur. These reflections resulted in two broad types of visual incongruity, which will next be reported.

Visual incongruity type 1: Topic-incongruity

First of all, it seems that incongruity can occur between each of the elements of the representational meaning of the image on the hand, and the topic of the image on the other hand. In that case, incongruity is dependent on the topic of the image ('topic-incongruity'). Following MIPVU's terminology, the topic forms the dominant conceptual domain of the image where the topic-incongruous element expresses an alien conceptual domain within the dominant conceptual domain of the image. This is essentially the same as Schilperoord's (this vol.) two-domain type of incongruity.

Let's say, for example, that the cartoon in Figure 3.1 tries to say something about 'The behaviour of Geert Wilders in the Islam debate', the topic. We could then call Geert Wilders topic-congruous, and the director's chair and the film set topic-*in*congruous. It can be said that by depicting the topic-incongruous director's chair and the film set, an alien conceptual domain is introduced into the dominant conceptual domain of the image.

Visual incongruity type 2: Property-incongruity

Secondly, incongruity can occur independently from the topic of the image, within the elements of the referential meaning. In that case, there is incongruity between the usual properties of an element on the one hand, and the properties as they are represented in the image on the other hand ('property-incongruity'). A case in point is Figure 2.13 in Chapter 2 (Schilperoord, this vol.). It presents a face without a mouth, with the caption 'Don't be without a topic. Read books.' Here incongruity occurs between the usual properties of an element, the human face, on the one hand, and the way in which they are represented in this image on the other.

Here is an explicit answer to issue 5: visual incongruity can be operationally defined as a discrepancy between (a) a unit of analysis on the one hand and the image's topic on the other hand or (b) the depicted properties of a topic-congruous

unit of analysis on the one hand and its usual properties on the other hand (cf. Schilperoord, this vol.)

Addressing issue 6: Finding visual incongruity

Issue 6 is: how can we find visual incongruity? In other words: how do we handle the operational definition of visual incongruity in applying the procedure? The operational definition has led to the following instructions:

3. Find incongruous visual units.
    a. Decide for each unit under step 2 whether it is incongruous with the topic as formulated under step 1d ('topic-incongruous').
    b. Decide for each topic-congruous unit under step 2 whether it shows properties that are incongruous with the properties that are typically true of that unit ('property-incongruous').

For step 3a, analysts should determine in most cases if a unit of analysis is disharmonious with the image's topic or not. This is a difficult issue as long as clear guidelines about demarcations of domains or categories are lacking (Steen 2007). It is up to the analyst in a particular project to come up with criteria for differentiation that can be adhered to in a reliable way. What this boils down to is an operationalization of the issue of two distinct domains discussed by Schilperoord (this vol.) as well.

For step 3b above, some additional guidelines were added for deciding on atypical properties. To decide whether units have incongruous properties, analysts may use Wu and Barsalou's (2009) taxonomy of properties. This taxonomy includes five broad categories, but we only consider the categories of entity properties and situation properties, as these are relevant to the VISMIP procedure. See the Appendix for an overview of these categories and their subcategories.

The taxonomy of Wu and Barsalou (2009) provides insight into the kinds of properties that could be called atypical and thus guides analysts using VISMIP in deciding on step 3b. Following Wu and Barsalou (2009), VISMIP makes a distinction between properties of an *entity* where an entity is a concrete object or a collection of (at least) some concrete entities and properties of a *situation* where a situation typically includes agent(s) performing some action in some setting. So, a particular unit of the image's representational meaning may be found to be incongruous because it shows a property that is an unusual entity property, for instance when an image depicts a banana and this banana is coloured pink. When the very same banana is used by an ape for transportation, eaten by a dog, or found hanging in an apple tree, it is said to be incongruous because it shows situation properties that are atypical.

Addressing issue 7 and 8: Deciding on candidates for comparison and testing need for comparison

Deciding on the units that could be the candidates for some form of comparison depends on the type of visual incongruity that is found under step 3 of VISMIP. Under step 3a, the analyst has to determine whether there are any units of analysis that are incongruous with the topic of the image. If there are any, then the analyst should wonder which replacing concepts could potentially be used for comparison. As we will now illustrate, the decision about potential candidates should be based on two criteria: (a) the replacing concept should be coherent with the image topic and (b) it should be harmonious within the referential meaning of the image.

In case of Figure 3.1, for instance, analysts should think about which concepts, instead of the topic-incongruous concepts *director's chair* and *film set*, would be congruous with the topic ('Geert Wilders' behaviour in the Islam debate') and would be coherent within the description of the representational meaning at the same time ('Geert Wilders sits back in director's chair on film set'). A possible analysis that someone could come up with then is that the concepts *parliament seat* and *House of Representatives* meet these criteria. After all, they are harmonious with the topic and they fit into the representational meaning of the cartoon ('Geert Wilders sits back in a parliament seat in the House of representatives').

Under 3b, the analyst should determine whether there are any topic-congruous units that show unusual properties. If there are any, the analyst should wonder which replacing concepts could be used for comparison. The first question should be: which concept typically owns the incongruous properties? The second question should be, similar to the previous substep: is that concept harmonious within the referential meaning?

In Figure 2.13 from the previous chapter, for example, analysts should reason that 'face with mouth' is the concept that typically owns the incongruous properties depicted here. Then they should also come to the conclusion that this concept would be coherent with the referential meaning of the image where 'face without mouth' is not ('Face with mouth is related to having topics to talk about'). Again, this is fully compatible with Schilperoord (this vol.).

Having thought about potential candidates for comparison, the next step would be to decide whether the comparison is *needed* (issue 8). When under step 3 replacing concepts can be established that are coherent with the main topic (step 3a) or own the atypical properties (step 3b), and at the same time match the referential meaning of the image (step 3a and step 3b), there should be enough reason for comparison. This resulted in the following instructions:

4. Test whether the incongruous units are to be integrated within the overall topical framework by means of some form of comparison.
    a. For each incongruous unit under 3a, determine which replacing unit would be congruous with the topic AND would be coherent with the referential meaning of the image.
    b. For each incongruous unit under 3b, determine which replacing unit would typically own the incongruous properties AND would be coherent with the referential meaning of the image.

In sum, how can we decide on the units that could be the candidates for some form of comparison (issue 7)? In case of topic-incongruity, the topic-incongruous units should be replaced by concepts that *are* congruous with the topic of the image and also fit into the representational meaning of the image. In case of property-incongruity, the incongruous units should be replaced by concepts that *do* own the unusual properties and also fit into the representational meaning of the image. How can we decide whether some form of comparison is needed (issue 8)? If replacing concepts can be found that meet the above-mentioned criteria, there is enough reason for comparison. In other words: whenever suitable candidates for comparison are found, there is a need to compare them.

3.3.3 Testing cross-domain-ness and indirectness: Operationalization

Section 3.2.5 stated that the following issues should be discussed:

9. How can we decide whether two concepts belong to two distinct and contrasted domains?
10. How can we decide whether the comparison can be seen as some form of indirect discourse about the topic of the image?

Addressing these issues would enable the adaption of the following steps in the MIPVU procedure pertaining to direct metaphor: "Test whether the comparison is nonliteral or cross-domain" (Steen et al. 2010, p. 39), and "Test whether the comparison can be seen as some form of indirect discourse about the local or main referent or topic of the text" (Steen et al. 2010, p. 39).

Addressing issue 9: Testing cross-domain-ness
Analysts working with VISMIP may use WordNet (http://wordnet.princeton.edu) as a tool to answer the question of cross-domain-ness for each comparison. WordNet is a lexical database of English words in which words are grouped together based on their meanings. WordNet also states super-subordinate relations, linking more general concepts to increasingly specific ones. The hyperonomy relations

in WordNet proved to be useful for VISMIP's purposes. For every concept to be compared, the analyst should click the option 'inherited hypernym', and then a hierarchy of concepts is displayed. Then analysts should study a number of layers, say three, under a word to see whether any hypernyms overlap. If they do not, then the domains to which the concepts belong can be considered to be distinct. If they do, then the domains are considered to be similar.

Suppose, for example, that for Figure 3.1, three comparisons are involved: those between *parliament seat* and *director's chair*, between *House of representatives* and *film set*, and between *Geert Wilders* and *film director*. Analysts looking for visual metaphor could then find out through WordNet that all three comparisons are cross-domain, with

- the parliament seat belonging to the domain of 'legal right>right>abstraction' (Wordnet, under seat sense #7), and the director's chair belonging to the domain of 'seat>furniture>furnishing' (Wordnet, under chair sense #1),
- the House of representatives belonging to the domain of 'house>legislature>assembly' (WordNet) and the film set belonging to the domain of 'representation>creation>artifact' (WordNet, set sense #4), and
- Geert Wilders belonging to the domain of 'politician[7]>leader>person' (WordNet), and film director belong to the domain of 'film maker>producer>creator' (WordNet).

For the cross-domain-ness test, the MIPVU-instruction was used:

5. Test whether the comparison(s) is/are cross-domain.

In conclusion, issue 9 was: how can we decide whether two concepts belong to two distinct and contrasted domains? We decide on the basis of what WordNet has to say about hyperonomy relations. If the third-level-hypernyms of concepts do not overlap, the concepts belong can be considered to be distinct. If they do overlap, then the domains are considered to be similar. The number of levels may have to be adjusted with increasing experience in use.

Addressing issue 10: Testing indirectness

In MIPVU, the way to decide about indirect discourse is to check whether a provisional sketch is possible between the material functioning as source domain and the material functioning as target domain. For VISMIP, we follow the same line of argumentation. Having tested if the comparisons are cross-domain, which appears to be the case in for instance Figure 3.1, we should continue testing whether the

7. Geert Wilders cannot be found in WordNet, but it is reasonable to classify him as a politician.

comparisons have something to say about the topic of the image by thinking about a rudimentary mapping from the source domain to the target domain. There is one difference with the linguistic situation, however, because there the incongruous material functions as source domain, where in visual metaphor the incongruous material has the potential to function as target domain, depending on the type of incongruity (cf. Schilperoord, this vol.).

Let's say, for instance, that Figure 3.1 involves three different comparisons that are all cross-domain. The next step for an analyst would be to sketch for each comparison a very first mapping from the visual elements functioning as source domain, to the visual elements functioning as target domain. Three different analyses could then be made, listed below.

- The comparison between *parliament seat* on the one hand and *director's chair* on the other can be seen as indirect discourse about the topic of the image ('Geert Wilder's behaviour in the Islam Debate'): it is a place from where someone can control.
- This also goes for the comparison between *House of Representatives* and *film set*: it is a place from where someone can control.
- We can understand *Geert Wilders* in terms of a *film director*'s controlling behaviour: the behaviour of Geert Wilders in the Islam debate is like the behaviour of a film director on a film set.

The visual elements *director's chair* (analysis 1), *film set* (analysis 2) and *Geert Wilders* (analysis 3) are incongruous material in the image; *director's chair* and *film set* for being incongruous with what the image is about and *Geert Wilders* for showing a-typical characteristics. The function, however, varies between these incongruous elements: *director's chair* and *film set* function as source domain, whereas *Geert Wilders* functions as target domain. Where in linguistic metaphor the incongruous elements would function as source domain, this example shows that visual metaphor is different: incongruous material may function either as source domain or target domain, depending on the nature of the incongruity.

VISMIP includes the following instruction to test indirectness:

6. Test whether the comparison(s) can be seen as some form of indirect discourse about the topic as formulated under step 1d.

In conclusion, issue 10 was: how can we decide whether a comparison can be seen as indirect discourse? Similar to MIPVU, we decide on the basis of the possibility of a sketchy mapping between the visual unit functioning as source domain on the one hand and the visual unit functioning as target domain on the other.

3.4 VISMIP: Application

This section demonstrates how VISMIP works through two illustrative applications. The first example illustrates those cases where VISMIP would lead to the decision to code a visual unit as a metaphor (Section 3.4.1). The second example illustrates that the analysis of an image stops whenever a finding of a test is negative (Section 3.4.2). In addition, it illustrates those cases where two domains that are potentially involved in visual metaphor are not sufficiently distinct.

3.4.1 Application where VISMIP leads to marking as metaphor

Let us look at the cartoon in Figure 3.3 and demonstrate the first step of the VISMIP procedure.

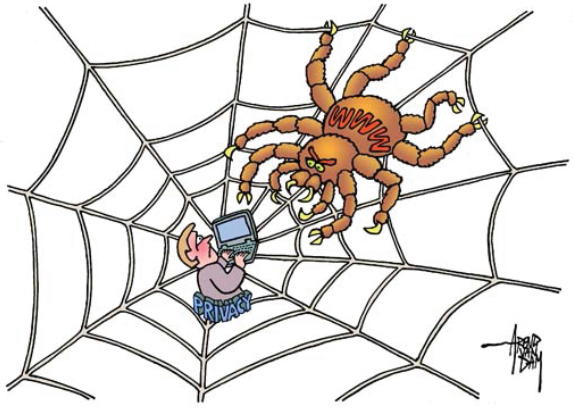


Figure 3.3 Cartoon by Arend van Dam (via http://www.arendvandam.com/kleuren.php?position=365, last accessed April 5, 2013)

1. Look at the entire image, including visual and verbal elements, to establish a general understanding of the meaning.

1a. Describe in just a few simple phrases the referential meaning of the image, i.e. what/who is being depicted here, what he is doing, where he is doing it, and so on. For example: 'Tall, blond man and spotted dog walk quickly in park. Smiling girl gives green apple to frowning boy under tree.' If the denotative meaning is ambiguous and allows more than one interpretation, then give alternative descriptions.

analysis 1a: Spider threatens man and computer in spider web.

1b. Test whether there are any clues that tell you that more general and abstract meaning should be attached to what is described under step 1a. For example, if the exemplary image described under 1a is accompanied by the caption 'summer joy', we have a clue that the more general and abstract concept 'summer joy' should be attached to the referential meaning.

analysis 1b: The man and computer in combination with the verbal text 'privacy' are likely to represent 'online privacy'. The spider accompanied by the verbal text 'WWW' is likely to represent the World Wide Web.

1c. Reconstruct the point underlying the image.

analysis 1c: Trying to reconstruct the point for our exemplary cartoon in Figure 3.3, we arrive at the following reasoning: under step 1b, we claimed that the man and computer in combination with the verbal text 'privacy' are likely to represent 'online privacy'. We also said that the spider accompanied by the verbal text 'WWW' is likely to represent the World Wide Web. As according to the description under step 1a the spider is the one threatening and the man and the computer are the ones being threatened, the picture is probably meant to convey that the World Wide Web threatens our online privacy (and not the other way around). If we assume, however, that the ultimate goal of the message is to warn the viewer for this disadvantage of the World Wide Web, the point should be something like this: 'You should be careful on the World Wide Web in relation to online privacy'.

1d. Derive from step 1c the topic of the point, i.e. that about which the point is stated.

analysis 1d: From the point formulated under the previous step, we can derive the following topic: (privacy on) the World Wide Web.

2. Structure the descriptive phrase(s) under step 1a. For example: 'Tall, blond man and spotted dog walk quickly in park. [Agent(man|tall, blond) Agent(dog|spotted) Action(walk|quickly) Setting(in park)] Smiling girl gives green apple to frowning boy under tree. [Agent(girl|smiling) Action(give) Object(apple|green) Recipient(boy|frowning) Setting(under tree)'.

analysis 2: Spider threatens man and computer in spider web. [Agent(spider) Action(threaten) Object(man) Object(computer) Setting(in spider web)]

3. Find incongruous visual units.

3a. Decide for each unit under step 2 whether it is incongruous with the topic as formulated under step 1d ('topic-incongruous').

analysis 3a: The topic is (privacy on) the World Wide Web. The Action(threatens), the Object(man) and the Object(computer) are congruous with this topic; the other elements are not. We will underline the incongruous elements in our description: Spider threatens man and computer in spider web [Agent(spider) Action(threaten) Object(man) Object(computer) Setting(in spider web)].

3b. Decide for each topic-congruous unit under step 2 whether it shows properties that are incongruous with the properties that are typically true of that unit ('property-incongruous').

analysis 3b: Object(man) – SMALL(a-typical entity property), THREATENED BY SPIDER(a-typical situation property), IN SPIDER WEB(a-typical situation property). Object(computer) – SMALL(a-typical entity property), THREATENED BY SPIDER(a-typical situation property), IN SPIDER WEB(a-typical situation property).

4. Test whether the incongruous units are to be integrated within the overall topical framework by means of some form of comparison.

4a. For each incongruous unit under 3a, determine which replacing unit would be congruous with the topic AND would be coherent with the referential meaning of the image.

analysis 4a: One incongruous element, the Agent(spider), needs to be integrated within the global topic by means of comparison with the following agent: 'World Wide Web'. The other incongruous element, the Setting(in spider web), needs to be integrated within the global topic by means of comparison with the following setting: 'on the Internet'. It would result in the following description: World Wide Web threatens man and computer on the Internet [Agent(spider = World Wide Web) Action(threatens) Object(man) Object(computer) Setting(in spider web = on the Internet)]. Such a description would be coherent AND continuous with the global topic.

4b. For each incongruous unit under 3b, determine which replacing unit would typically own the incongruous properties AND would be coherent with the referential meaning of the image.

analysis 4b: The Object(man) and Object(computer) have properties that are not typical: Object(man) – SMALL, THREATENED BY SPIDER, IN SPIDER WEB; Object(computer) – SMALL, THREATENED BY SPIDER, IN SPIDER WEB. The concept that may typically have those properties is: prey. The Object(man) and Object(computer) need to be compared to prey. It would be coherent with the referential meaning: Spider threatens prey in spider web. [Agent(spider) Action(threaten) Object(man = prey) Object(computer = prey) Setting(in spider web)]

5. Test whether the comparison(s) is/are cross-domain.

analysis 5: Three comparisons should be tested. In WordNet, The World Wide Web falls under 'computer network>network>system', whereas a spider falls under 'arachnid>anthropod>invertebrate'. The Internet falls under 'computer network>-network>system', whereas spider web falls under 'web>object>physical entity'. Under step 1b, it was determined that the man and computer together connote 'online privacy'. 'online privacy' falls under 'reclusiveness>aloneness>disposition' (WordNet, privacy sense #1,), whereas prey falls under 'animal>organism>living thing' (WordNet, prey sense #2). So, these comparisons are all cross-domain.

6. Test whether the comparison(s) can be seen as some form of indirect discourse about the topic as formulated under step 1d.

analysis 6: The comparison between the World Wide Web and the spider can be seen as indirect discourse about the topic of the image: the World Wide Web is as threatening to online privacy as a spider to its prey.

The comparison between the Internet and the spider web can also be seen as indirect discourse about the topic of the image: we should understand the Internet in terms of the spider web's potential to entrap.

Under step 1b, it was determined that the man and computer together connote 'online privacy'. Last, the comparison between online privacy on the one hand and prey on the other can be seen as indirect discourse about the topic of the image: we should understand online privacy in term of prey's threatened position.

The final step is deciding on metaphorical use:

7. If the findings of tests 4, 5, 6 are positive, then a visual unit should be marked for metaphor.

analysis 7: The findings of tests 4, 5, and 6 are positive, so the topic-incongruous Agent(spider) and Setting(in spider web), and property-incongruous Object(man) and Object(computer) should all be marked for metaphor.

3.4.2 Application where VISMIP stops due to same-domain-ness

Figure 3.4 Print ad by Etcetera, Amsterdam, the Netherlands. Creative Directors: Edward Bardoul, Raymond van Schaik. Art Director: Chris Sant. Copywriter: Dieuwer Bulthuis. Photographer: Simon Warmer. Published: January 2011. Via http://adsoftheworld.com/media/print/bolletje_toasted_bread, last accessed April 18, 2013). The package says 'I want Bolletje toasted bread' (translated by the author from the Dutch 'Ik wil Bolletje geroosterd brood')

The advertisement in Figure 3.4 demonstrates that VISMIP stops when cross-domain mapping is not possible. To make the example clear, the complete analysis is presented. To enhance conciseness, the VISMIP instructions are omitted in this example but can be consulted in the previous section.

analysis 1a: Description 1: Slices of toasted bread pop up from Bolletje toasted bread package. Description 2: Slices of toasted bread in Bolletje toasted bread package.

According to description 1, we recognize slices of toasted bread doing something (popping up), and a place (Bolletje toasted bread package) from where the slices pop up. At the same time, we are unsure whether the slices of bread are in fact popping up; they might as well just be in the package, not doing any activity. In the latter case, we recognize an object being located somewhere. It can be said that the first meaning reflects a process, where the second meaning reflects a state. We number the descriptions to indicate that the representational meanings are alternatives.

Besides the verbal text on the package ('I want Bolletje toasted bread'), which does not appear to signify any more general and abstract meaning associated with the descriptions under 1a, there are no verbal elements present in the image. So:

analysis 1b: No, nothing points to more general, symbolic interpretations.

Having decided that nothing abstract should be added to the denotative level of meaning formulated under step 1a, we reconstruct the point and any related arguments forwarded by the image:

(analysis 1c:) You should buy Bolletje toasted bread.

Obviously, the advertiser's ultimate aim is to convince the viewers that they should buy Bolletje toasted bread (1).

analysis 1d: From the point 'You should buy Bolletje toasted bread', it can be derived that the topic is 'Bolletje toasted bread'.

Having established an understanding of the image's meaning, we take the alternative descriptions as stated under step 1a and divide them into units.

analysis 2:
Description 1: Slices of toasted bread pop up from Bolletje toasted bread package [Agent(slices of toasted bread) Action(pop up) Setting(from Bolletje toasted bread package)]. Description 2: Slices of toasted bread in Bolletje toasted bread package [Object(slices of toasted bread) Setting(in Bolletje toasted bread package)].

The next step is concerned with detecting incongruous elements. One or more elements of the representational meaning under step 2 may be discontinuous with the topic, but this appears not to be the case for the Bolletje ad in Figure 3.4,

analysis 3a: First we execute this step for description 1. According to step 1d, 'Bolletje toasted bread' is the topic. The Setting(from Bolletje toasted bread package) is clearly congruous with this topic. It is open for discussion whether the Agent(slices of toasted bread) and the Action(pop up) are congruous with the topic, but they are in any case not clearly *in*congruous with the topic. Conclusion: for description 1 there is no incongruity between any of the units under step 2 and the topic.

Then we execute this step for description 2. According to step 1d, 'Bolletje toasted bread' is the topic. The Setting(in Bolletje toasted bread package) is clearly congruous with this topic. We have already decided for description 1 that 'slices of toasted

bread' are not incongruous with the topic. So, it also goes for description 2 that there is no incongruity between each of the units and the topic.

Having looked at incongruity with the topic, we continue to look at incongruity with typical properties.

analysis 3b: Description 1: Agent(slices of toasted bread) – pop up from BOLLETJE TOASTED BREAD PACKAGE(situational property, a-typical location). Description 2: Object(slices of toasted bread) – IN BOLLETJE TOASTED BREAD PACKAGE (situational property, a-typical location).

analysis 4b: In description 1, the Agent(slices of toasted bread) shows a property that is not typical: it pops up from a Bolletje toasted bread package. The Object(slices of toasted bread) in description 2 is a-typically located in a Bolletje toasted bread package. The concept that may typically have these properties is: Bolletje toasted bread. Agent(slices of bread) and Object(slices of bread) need to be integrated within the topical framework by comparison to Bolletje toasted bread. It would create coherence in each of the alternative referential meanings:

Description 1: Bolletje toasted bread pop ups from Bolletje toasted bread package [Agent(slices of toasted bread = Bolletje toasted bread) action(pop up) Setting(from Bolletje toasted bread package)].

Description 2: Slices of toasted bread in Bolletje toasted bread package [Object(slices of toasted bread = Bolletje toasted bread) Setting(in Bolletje toasted bread package)].

analysis 5: The comparison between Bolletje toasted bread on the one hand and slices of toasted bread on the other is not cross-domain, with both concepts belonging to 'baked goods>food>solid' (WordNet, bread sense #1).

The analysis stops here. The findings of tests 5 are negative. This is the conclusion: there are no elements to be marked for metaphor.

3.3.5 Conclusion

The aim was to develop a Method for Visual Metaphor Identification (VISMIP). The procedure is an application to visual metaphor of the Method for Linguistic Metaphor Identification (MIPVU) developed by Steen et al. (2010). The procedural steps lead to a decision about whether visual units within an image are metaphorically used. An overview of the procedure can be found in Table 3.2.

Table 3.2 Set of instructions in the Visual Metaphor Identification Procedure (VISMIP). The instructions can be used to identify visual units that are related to metaphor

1. Look at the entire image, including visual and verbal elements, to establish a general understanding of the meaning.
   - 1a. Describe in just a few simple phrases the referential meaning of the image, i.e. what/who is being depicted here, what he is doing, where he is doing it, and so on. For example: 'Tall, blond man and spotted dog walk quickly in park. Smiling girl gives green apple to frowning boy under tree.' If the referential meaning is ambiguous and allows more than one interpretation, then give alternative descriptions.
   - 1b. Test whether there are any clues that tell you that more general and abstract meaning should be attached to what is described under step 1a. For example, if the exemplary image described under 1a is accompanied by the caption 'summer joy', we have a clue that the more abstract concept 'summer joy' should be attached to the referential meaning.
   - 1c. Reconstruct the point underlying the image.
   - 1d. Derive from step 1c the topic of the point, i.e. that about which the point is stated.
2. Structure the descriptive phrase(s) under step 1a. For example: 'Tall, blond man and spotted dog walk quickly in park. [Agent(man|tall, blond) Agent(dog|spotted) Action(walk|quickly) Setting(in park)] Smiling girl gives green apple to frowning boy under tree. [Agent(girl|smiling) Action(give) Object(apple|green) Recipient(boy|frowning) Setting(under tree)'.
3. Find incongruous visual units.
   - 3a. Decide for each unit under step 2 whether it is incongruous with the topic as formulated under step 1d ('topic-incongruous').
   - 3b. Decide for each topic-congruous unit under step 2 whether it shows properties that are incongruous with the properties that are typically true of that unit ('property-incongruous').
4. Test whether the incongruous units are to be integrated within the overall topical framework by means of some form of comparison.
   - 4a. For each incongruous unit under 3a, determine which replacing unit would be congruous with the topic AND would be coherent with the referential meaning of the image.
   - 4b. For each incongruous unit under 3b, determine which replacing unit would typically own the incongruous properties AND would be coherent with the referential meaning of the image.
5. Test whether the comparison(s) is/are cross-domain.
6. Test whether the comparison(s) can be seen as some form of indirect discourse about the topic as formulated under step 1d.
7. If the findings of tests 4, 5, 6 are positive, then a visual unit should be marked for metaphor.

There are at least two challenges to VISMIP. The biggest challenge will be establishing the topic of the image in a reliable manner, as generally spoken, artworks are known for attracting multiple interpretations on different levels of understanding (e.g., describing what you see and attaching more general and abstract meaning). VisMet 1.0, a corpus of annotated images from different genres, may be used to refine and extend VISMIP in such a way that it captures images from which the 'final' message cannot easily be reconstructed (see Chapter 5, Bolognesi et al. this vol.).

Secondly, in order to monitor the quality and the reliability of the proposed procedure, it is necessary to test whether or not the technique leads to sufficient agreement among analysts. Metaphor analyses conducted in the humanities, whether in linguistics, discourse analysis, or semiotics, need to show that they are reliable, not researcher-dependent interpretations, in order to be taken seriously as providing valid input into metaphor studies in the cognitive and social sciences, where the processing and effect of linguistic and multimodal discourse stimuli is researched on the behavior of language users and discourse participants. The wish to estimate reliability is in line with previous studies conducted by Steen et al. (2010) to examine the reliability of their linguistic metaphor identification procedure, which lie at the foundation of the current study.

References

Alcolea-Banegas, J. (2009). Visual arguments in film. *Argumentation*, 23, 259–275. https://doi.org/10.1007/s10503-008-9124-9

Andriessen, D., Kliphuis, E., McKenzie, J., Van Winkelen, C. (2009). Pictures of knowledge management, developing a method for analysing knowledge metaphors in visuals. *Electronic Journal of Knowledge*, 7(4), 405–415. Retrieved from http://www.ejkm.com/volume7/issue4/p405, last accessed February 2, 2015.

Birdsell, D. S., & Groarke, L. (1996). Toward a theory of visual argument. *Argumentation & Advocacy*, 33, 1–10.

Blair, J. A. (1996). The possibility and actuality of visual arguments. *Argumentation and Advocacy*, 33, 23–29.

Cameron, L. (2003). *Metaphor in educational discourse*. London and New York: Continuum.

Charteris-Black, J. (2004). *Corpus approaches to critical metaphor analysis*. London: Palgrave Macmillan.

Clarke, M. (2010). Iconology. In *The Concise Oxford Dictionary of Art Terms* (2nd ed) (p. 125). Oxford: Oxford University Press. https://doi.org/10.1093/acref/9780199569922.001.0001

Eemeren, F. H. van (2010). *Strategic manoeuvering in argumentative discourse: Extending the pragma-dialectical theory of argumentation*. Amsterdam: John Benjamins. https://doi.org/10.1075/aic.2

Feteris, E., Groarke, L. & Plug, J. (2011). Strategic maneuvering with visual arguments in political cartoons: A pragma-dialectical analysis of the use of topoi that are based on common cultural heritage. In E. Feteris, B. Garssen & F. Snoeck Henkemans (Eds.), *Keeping in touch with pragma-dialectics: In honor of Frans H. van Eemeren* (pp. 59–74). Amsterdam: John Benjamins. https://doi.org/10.1075/z.163.05fet

Fleming, D. (1996). Can there be Visual Arguments? *Argumentation & Advocacy*, 33, 11–22.

Forceville, C. (1996). *Pictorial metaphor in advertising*. Londen: Routledge. https://doi.org/10.4324/9780203272305

Forceville, C. (2002). The identification of target en source in pictorial metaphors. *Journal of Pragmatics*, 34, 1–14. https://doi.org/10.1016/S0378-2166(01)00007-8

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). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511816802.028

Goatly, A. (1997). *The language of metaphors*. London: Routledge.

Herrmann, J. B. (2013). *Metaphor in academic discourse: Linguistic forms, conceptual structures, communicative functions and cognitive representations*. LOT Dissertation Series, 333. Utrecht: LOT.

Hermeren, G. (1969). *Representation and meaning in the visual arts*. Lund: Scandinavian University Books.

Jensen, K. B. (2008). Semiotics. In W. Donsbach (Ed.), *The International Encyclopedia of Communication*. Blackwell Publishing, Blackwell Reference Online. Retrieved from http://www.communicationencyclopedia.com/subscriber/tocnode?id=g9781405131995_yr2010_chunk_g978140513199524_ss31-1, last accessed September 16, 2011.

Johnson, R. H. (2003). Why ''visual arguments'' aren't arguments. In: *IL@25. A Conference Celebrating the Twenty-Fifth Anniversary of the First International Symposium on Informal Logic*. Retrieved from http://web2.uwindsor.ca/faculty/arts/philosophy/ILat25/edited_johnson.doc, last accessed August 22, 2014).

Kaal, A. (2012). *Metaphor in conversation*. Oisterwijk: Uitgeverij BOX Press.

Kaplan, S. J. (2005). Visual metaphors in print advertising for fashion products. In K. Smith, S. Moriarty, G. Barbatsis & K. Kenney (Eds.), *Handbook of visual communication: Theory, methods, and media* (pp. 167–177). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kjeldsen, J. E. (2012). Pictorial argumentation in advertising: Visual tropes and figures as a way of creating visual argumentation. In F. H. van Eemeren & B. Garssen (Eds.), *Topical themes in argumentation theory: Twenty exploratory studies* (pp. 239–256). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-4041-9_16

Krennmayr, T. (2011). *Metaphor in newspapers*. LOT Dissertation Series, 276. Utrecht: LOT.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Leeuwen, T. van (2001). Semiotics and iconography. In T. van Leeuwen & C. Jewitt (Eds.), *Handbook of Visual Analysis* (pp. 92–118). Londen: Sage.

Lyne, J. (2008). Rhetoric and Semiotics. In W. Donsbach (Ed.), *The International Encyclopedia of Communication*. Blackwell Publishing, Blackwell Reference Online. Retrieved from http://www.communicationencyclopedia.com/subscriber/tocnode?id=g9781405131995_yr2010_chunk_g978140513199523_ss78-1, last accessed September 16, 2011.

Moriarty, S. (2005). Visual semiotic theory. In K. Smith, S. Moriarty, G. Barbatsis, & K. Kenney (Eds.), *Handbook of visual communication: Theory, methods, and media* (pp. 227–241). Mahwah, NJ: Lawrence Erlbaum.

Müller, M. G. (2008). Iconography. In W. Donsbach, (Ed.). *The International Encyclopedia of Communication*. Blackwell Publishing, Blackwell Reference Online. Retrieved from http://www.communicationencyclopedia.com/subscriber/tocnode?id=g9781405131995_yr2010_chunk_g978140513199514_ss1-1, last accessed 16 September 2011.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications.

Pasma, T. (2011). *Metaphor and register variation: The personalization of Dutch news discourse*. Oisterwijk: BOXpress.

Penn, G. (2000). Semiotic analysis of still images. In M. W. Bauer & G. Gaskell (Eds.), *Qualitative researching with text, image and sound: A practical handbook* (pp. 227–245). London: Sage.

Pragglejaz Group (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1), 1–39. https://doi.org/10.1080/10926480709336752

Riffe, D., Lacy, S., & Fico, F. G. (2005). *Analyzing media messages: Using quantitative content analysis in research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ripley, M. L. (2008). Argumentation theorists argue that an ad is an argument. *Argumentation*, 22: 507–519. https://doi.org/10.1007/s10503-008-9102-2

Rose, G. (2005). Visual methodologies. In G. Griffin (Ed.), *Research methods for English studies* (pp. 67–90). Edinburgh: Edinburgh University Press.

Šorm, E., & Steen, G. J. (2013). Processing visual metaphor: A study in thinking out loud. *Metaphor and the Social World*, 3(1): 1–34. https://doi.org/10.1075/msw.3.1.01sor

Steen, G. J. (2007). *Finding metaphor in grammar and usage: A methodological analysis of theory and research*. Amsterdam: John Benjamins. https://doi.org/10.1075/celcr.10

Steen, G. J. (2008). The paradox of metaphor: Why we need a three-dimensional model of metaphor. *Metaphor and Symbol*, 23(4), 213–241. https://doi.org/10.1080/10926480802426753

Steen, G. J. (2009). From linguistic form to conceptual structure in five steps: analyzing metaphor in poetry. In G. Brone & J. Vandaele (Eds.), *Cognitive poetics: Goals, gains and gaps* (pp. 197–226). Berlin/ New York: Mouton de Gruyter.

Steen, G. J. (2010). When is metaphor deliberate? In N. -L. Johannesson, C. Alm-Arvius, & D. Minugh (Eds.), *Selected Papers from the Stockholm 2008 Metaphor Festival* (pp. 43–63).

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU*. Amsterdam: John Benjamins. https://doi.org/10.1075/celcr.14

Steen, G. J. (2011). The contemporary theory of metaphor – now new and improved! *Review of Cognitive Linguistics*, 9(1), 26–64. https://doi.org/10.1075/rcl.9.1.03ste

Steen, G. J. (2015). Developing, testing and interpreting Deliberate Metaphor Theory. *Journal of Pragmatics*. https://doi.org/S0378216615000946.

Tam, A. M., & Leung, C. H. C. (2001). Structured natural language descriptions for semantic content retrieval of visual materials. *Journal of the American Society for Information Science and Technology*, 52(11), 930–937. https://doi.org/10.1002/asi.1151

Wu, L. -l., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica* 132, 173–189. https://doi.org/10.1016/j.actpsy.2009.02.002

Appendix. **Wu and Barsalou's (2009)** taxonomy of properties (see Section 3.3.2)

Table 3.3 Taxonomy of properties that function as practical guidelines for deciding on VISMIP's step 3b (copied from Wu & Barsalou 2009, p. 187). In the examples, lowercase text refers to concepts, whereas uppercase text refers to a property of the concept

Entity Properties (E). Properties of a concrete entity, either animate or inanimate. Besides being a single self-contained object, an entity can be a coherent collection of objects, or an institution, if it consists of at least some concrete entities (e.g., forest, government, and society).
Larger whole (E_W). A whole to which an entity belongs (e.g., window – HOUSE; apple – TREE).
Spatial relation (E_S). A spatial relation between two or more properties within an entity, or between an entity and one of its properties (e.g., car – window ABOVE door; watermelon – green OUTSIDE).
External surface property (E_{SE}). An external property of an entity that is not a component, and that is perceived on or beyond the entity's surface, including shape, color, pattern, texture, size, touch, smell, and taste (e.g., watermelon – OVAL; apple – RED; car – STINKS).
Internal surface property (E_{SI}). An internal property of an entity that is not a component, that is not normally perceived on the entity's exterior surface, and that is only perceived when the entity's interior surface is exposed, including color, pattern, texture, size, touch, smell, and taste (e.g., apple – WHITE, watermelon – JUICY).
External component (E_{CE}). A three-dimensional component of an entity that, at least to some extent, normally resides on its surface (e.g., car – HEADLIGHT; tree – LEAVES).
Internal component (E_{CI}). A three-dimensional component of an entity that normally resides completely inside the closed surface of the entity (e.g., apple – SEEDS; jacket – LINING).
Systemic property (E_{SYS}). A global systemic property of an entity or its parts, including states, conditions, abilities, and traits (e.g., cat – ALIVE; dolphin – INTELLIGENT; car – FAST).
Entity behavior (E_B). An intrinsic action that is characteristic of an entity's behavior, and that is not an entity's normal function for an external agent, which is coded as SF (e.g., dog – BARKS; children – PLAY).
Associated abstract entity (E_{AE}). An abstract entity associated with the target entity and external to it (e.g., computer-SOCIETY; transplanted Californian-RELIGIOUS AFFILIATION).
Quantity (E_Q). A numerosity, frequency, or intensity of an entity or its properties (e.g., jacket – an ARTICLE of clothing; cat – FOUR legs; tree – LOTS of leaves; apple – COMMON fruit; watermelon – USUALLY green; apple – VERY red).
Situation Properties (S). A property of a situation, where a situation typically includes one or more participants, at some place and time, engaging in an event, with one or more entities (e.g., picnic, conversation, vacation, and meal).
Participant (S_P). A person in a situation who typically uses an entity or performs an action on it and/or interacts with other participants (e.g., toy – CHILDREN; car-PASSENGER; furniture – PERSON).
Location (S_L). A place where an entity can be found, or where people engage in an event or activity (e.g., car – IN THE GARAGE; buy – IN A STORE).

Table 3.3 (*continued*)

Spatial relation (S_S). A spatial relation between two or more things in a situation (e.g., watermelon – the ants crawled ACROSS the picnic table; car – drives ON the highway; vacation – we slept BY the fire).

Time (S_T). A time period associated with a situation or with one of its properties (e.g., picnic – FOURTH OF JULY; sled – DURING THE WINTER).

Action (S_A). An action that a participant performs in a situation (e.g., shirt – WORN; apple-EATEN).

Associated entity (S_E). An entity in a situation that contains the focal concept (e.g., watermelon – TABLE; cat – LITTER).

Function (S_F). A typical goal or role that an entity serves for an agent (e.g., car – TRANSPORTION; clothing – PROTECTION).

Quantity (S_Q). A numerosity, frequency, or intensity of a situa- tion or any of its properties except of an entity, whose quanti- tative aspects are coded with E_Q (e.g., vacation – lasted for EIGHT days; car – a LONG drive).

Manner (S_M). The manner in which an action or behavior is performed (e.g., watermelon – SLOPPY eating; car – FASTER than walking).

Event (S_{EV}). An event or activity in a situation (e.g., watermelon – PICNIC, car – TRIP).

State of the world (S_{SW}). A state of a situation or any of its components except entities, whose states are coded with ESYS (e.g., mountains – DAMP; highway – CONGESTED).

Origin (S_{OR}). How or where an entity originated (e.g., car – FACTORY; watermelon – GROUND).

CHAPTER 4

VisMet 1.0

An online corpus of visual metaphors

Marianna Bolognesi, Romy van den Heerik
and Esther van den Berg
University of Oxford / Universiteit van Amsterdam / Independent

4.1 The VisMet corpus

4.1.1 The need for a corpus of visual metaphors

Even though in recent years visual metaphor research has attracted the attention of a variety of scientific communities, there is still a lack of consensus among scholars about several fundamental questions related to visual metaphors. One of these is the question of how to identify a visual metaphor "in the wild". Without an answer to this question, we cannot claim to have a precise view of what exactly a visual metaphor is.

Tentative answers to this question are presented in this book, including a taxonomy to detect and categorize visual metaphors according to different types of incongruences that violate the expected visual patterns within the image (Schilperoord this vol.) and a procedure for identifying visual metaphors (Šorm and Steen this vol.). Previous studies in visual metaphor identification and analysis, however, often have the drawback that they are the result of the empirical observation of a limited number of images, often representing a single researcher's view on the materials and clarified with a selection of examples that have been handpicked to illustrate what has been assumed. These approaches have not been applied to larger samples of images or to images that belong to different genres. The lack of shared resources explains why there is no benchmark against which different proposed models can be tested, compared, and possibly integrated. This obstructs efforts to provide useful evaluations of different models, making it impossible to proceed towards a consensus on the identification and analysis of visual metaphors.

VisMet 1.0 is the first online corpus of images that allow a metaphorical interpretation. It constitutes the first step toward gathering a substantial amount of data

https://doi.org/10.1075/celcr.18.04bol

in order to test hypotheses on the structures, functions and processes of visual metaphors. It is our aim to use this shared resource as a benchmark and as a set of test cases which can employed for the development of new models and methodologies, for the testing of existing intuitions and for the expansion of established theories to include metaphorical images from more genres. All of this will eventually improve our understanding of visual metaphor.

4.1.2 The VisMet project

The primary aim of the VisMet project is to create a corpus of annotated images that display visual metaphors, to be used by students and researchers from different academic fields interested in visual rhetoric. The first stage of the project, the pilot phase, has produced VisMet 1.0, an online corpus of roughly 350 images, which was released in June 2014 (http://www.vismet.org/VisMet/). An important aspect of the VisMet project is its ambition to provide a "democratic", scalable, and dynamic corpus, where the most recent debates, insights, and analyses of visual and multimodal metaphor can be shared online. The most recent application of VisMet is a dataset of crowdsourced tags, elicited from hundredths of online workers, who were exposed for different time durations (1s, 5s, 15s, 20s) to the images in the corpus. Participants were instructed to tag the images, i.e. to produce keywords that would describe salient aspects of the visual input. A semantic analysis of the tags produced for the different time exposures shows that concrete depicted entities are tagged mainly in short time exposures, while abstract concepts evoked by the images and contextual information required for the metaphorical interpretation are produced only when participants are exposed to the images for longer time durations (Bolognesi et al. accepted). This addition to the corpus contributes to advance knowledge about non-academic people's views of metaphor use in visuals. The VisMet corpus is also envisioned as a platform that will attract a variegated community of scholars who will provide their own contributions to the examination of visual metaphors.

In this chapter, we will discuss the relevance of the VisMet corpus and explain how we collected the images of VisMet 1.0. We will also describe how we proceeded to analyze and annotate them, making explicit various problems and obtaining new insights along the way. We will then show how we designed and built the database as well as the website that hosts VisMet. We conclude the chapter with an overview of how the corpus may be enlarged in the future and how it can be put to use in future research.

4.1.3 The construction of the VisMet corpus

The VisMet 1.0 corpus contains images that present visual incongruences (i.e. deviations from reality, as defined in Schilperoord this vol.), as well as images in which the incongruence is not strictly perceptual (i.e., just visual), but instead is derived from a relation between linguistic anchors, e.g., a title or caption, and the visual representation (see Schilperoord this vol., on the relation between visuals and topics). In this respect, the corpus includes both monomodal pictorial metaphors, where both metaphor domains are visually represented, and multimodal metaphors of the verbo-pictorial variety, where one of the two domains is defined or cued verbally (cf. Forceville 2006). This second type of metaphor is frequently used in advertising (where the textual slogan or the company logo often helps us identify the target domain and topic of the message – the product or service to be sold), political cartoons (where the textual anchors outside the image help frame the topic and direct our attention to the right target domain), and art (where the title of the artwork or its caption often define the abstract concept that plays the target domain role). We believe that linguistic anchors within and outside the image are often crucial to identify the source and target domain, as they let the necessary mappings emerge and help with the interpretation of the visual metaphors – as described in more detail by Schilperoord (this vol.), Van den Heerik et al. (this vol.), and Šorm and Steen (this vol.).

The images were selected from a variety of online databases (collections of advertisements, galleries of cartoons, artists' personal portfolios, etc.). Explicit permission to reproduce the images online was requested and obtained from the authors/copyright holders of the images by e-mail. The images that were eventually included in the VisMet 1.0 corpus can be categorized into the following genres: advertising (46 commercial and 18 social campaigns), artworks (97 illustrations, 3 photographs, 5 graffiti), political cartoons (183 in all), and a category called "other", which includes other images that have not been classified as belonging to one of the other three categories.

Subsequently, the images were first annotated with formal metadata such as title, year of production, author (see Section 4.3), as well as a label for the genre. Type of content was then further categorized by means of a taxonomy classifying visual metaphors according to two dimensions of meaning: the expression dimension, which contains the referential meaning of the visual expression (see Šorm and Steen this vol.), and the conceptual dimension, from which the more general and abstract meaning is derived (see Šorm and Steen this vol.). Analyses which illustrate the application of this taxonomy are given in Section 4.1.

The website also allows for the appending of notes to images. These notes have been used to explain the rationale for certain annotations, but these notes are not

meant to fix the interpretation of the images and exclude alternative readings. We are aware that further analysis of the images, in particular the application of VisMIP to the corpus, may lead to alternative or contrasting interpretations of the metaphors, and thus to a potential re-evaluation of the annotations or even the taxonomy itself. This openness to new developments is crucial to the vision of VisMet as presenting a dynamic, democratic corpus.[1]

4.2 Theoretical background and related work

4.2.1 Vismet and contemporary models of visual metaphor

Contemporary research has given definitions of visual metaphor which vary from that of a broad category of visual rhetorical tropes to that of a very restricted class of images (see Forceville 2008 for a detailed overview). For example, Kennedy (1982) broadly describes visual metaphors as intended violations of standard modes of depiction that are universally recognizable. Kennedy's (1982) model sketches 25 types of visual metaphor including a variety of other tropes, such as metonymies and litotes. Whittock (1990) proposes a similar model which distinguishes between 10 types of visual metaphor in cinema (see Forceville 2008, p. 468). In recent years, models of visual metaphor have also been proposed in the social sciences (e.g., management, business), and they have been applied to narrower sets of images (see Phillips and McQuarrie 2004; Gkiouzepas and Hogg 2011), in particular to advertisements. The most prominent model for visual metaphor identification and analysis in advertising, however, is that of Forceville (1996; 2002).

Forceville recognizes four (formerly three) types of visual metaphor, based on the formal manifestation of the source and the target domain:

1. in *hybrids* source and target are artificially integrated within a single non-compossible gestalt;
2. in *integrated* metaphors (which are not strictly pictorial, as they can be realized in 3D objects) source and target are integrated in a compossible fashion;
3. in *similes* source and target are aligned and fully represented;
4. and in *contextual* metaphors only the target is fully represented, while the source is cued by the visual context.

1. In this respect, we are grateful to the postgraduate students of the Metaphor Lab Amsterdam winter school 2017 for discussing and providing additional notes to the analyses, during the workshop on visual metaphor.

Forceville's classification has become a cornerstone of visual metaphor studies and provides useful parameters of metaphorical expressions in advertisements. The model seems broadly applicable within its genre, where the target domains can very often be defined as the product or service and the source domain is often concrete.

Yet Forceville's typology is less easy to apply to other genres, such as political cartoons or works of art. The reason for this is that the classification applies to the level of referential meaning. It thus fits best those advertisements for which referential meaning and more general and abstract meaning overlap. In Figure 4.1, for example, the target domain is the product to be sold (the mouthwash), which has been presented as similar to the source domain (a grenade). The mapping that is activated in the interpretation of the image is the destructive power of the grenade, which is compared to the ability of the mouthwash to kill bacteria. What we see is what we get: referential and more general and abstract meaning overlap: MOUTHWASH-IS-GRENADE.

Yet not all advertisements contain such a direct link between target domain and source domain. Consider for example the ad in Figure 4.2. An analysis of this image that is analogous to the previous example would yield the verbal schematization "IRIS-IS-JUNGLE". Unlike the verbalization of the metaphor of Figure 4.1, this verbalization does not capture the more general and abstract meaning of the image, nor does it activate all the necessary mappings. A next step is needed in which we move from the referential meaning of the image, i.e., a hybrid representation of an iris with a jungle, to a more general and abstract level via a chain of metonymies, taking into account any necessary contextual cues (as also pointed out in Koller 2008; cf. Forceville 2009). The conclusion would then be that the eye symbolizes a camera and the jungle the surrounding environment. Integrating this conceptual structure with the contextual cues, we eventually understand the intended message: the Land-Rover 5 camera system (product to be sold) is a human eye, observing the surrounding environment. A possible mapping of this metaphor projected from source to target would be that the camera (target) has the reliability of a human eye (source), i.e. delivers images that are as detailed and of high quality as those provided by the human eye.

This type of complexity can be found elsewhere too. For instance, visualized political messages frequently require mappings to concepts as abstract as, for example, the European economic crisis (Bounegru & Forceville 2011) or the phenomenon of immigration to Europe (El Rafaie 2003). These concepts are often expressed metaphorically or metonymically in ways that are typical of this genre, that is, through familiar symbols (see below), as shown in Figure 4.3.

Figure 4.1 Advertisement for mouthwash (author and copyright owner: Ferdi Rykziyanto)

Figure 4.2 Advertisement for a Land Rover camera system (author and copyright owner: Y&R)

Figure 4.3 Political cartoon published in 2012, about the dynamics among EU state members (author and copyright owner: Silvan Wegmann)

Here, arguably, the two prime ministers metonymically symbolize their countries. The identification of Sarkozy is facilitated by the colors of his hat, which resemble the French national flag; the cape worn by Angela Merkel stands for the European Union symbolized by the EU flag, and the EU's member states are represented by little men hidden under the cape. This exemplifies the kind of complex chain of relations that may be required during the interpretation of a political cartoon.

A final challenge for advertisement-based theories is the interpretation of visual metaphors in artistic imagery in paintings, digital illustrations and photographs. Works of art can clearly stimulate the viewer to build a metaphor, yet the

mappings projected from source to target domain often remain vague. Unlike a creative agency trying to sell a product, an artist may perceive ambiguity in their work as a strength rather than a weakness.

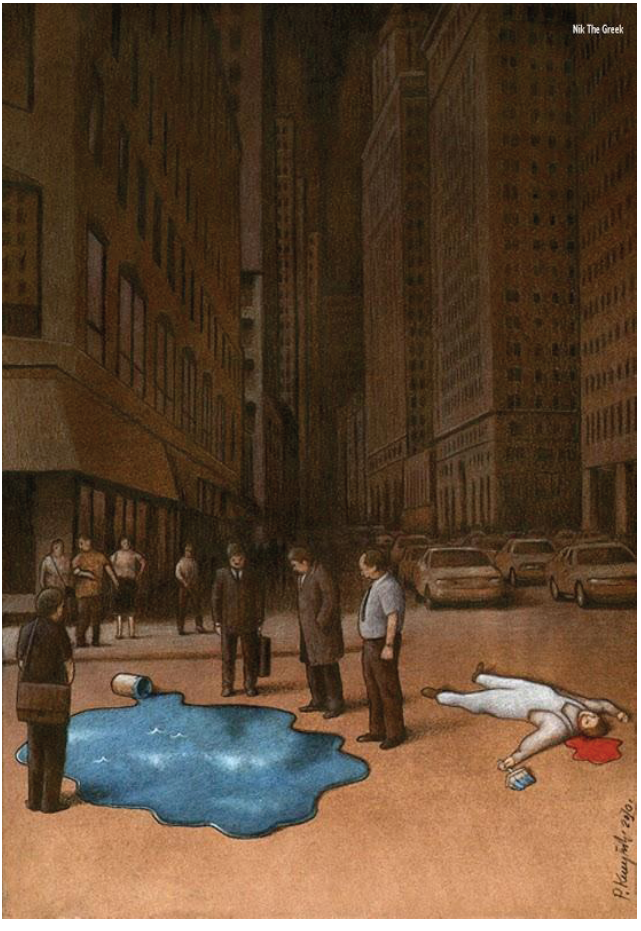


Figure 4.4 Illustration, without title (author and copyright owner: Pawel Kuczynski)

Figure 4.4 illustrates the ambiguity that is typical of artistic visual metaphors. We see a stain of paint that resembles the sky, but this incongruity allows more than one solution in the conceptual dimension. As a matter of fact, the author (the Polish artist Pawel Kuczynski) posted this image on his Facebook page and hundreds of people commented on it, providing different interpretations. All of them included some form of metaphor. One is reported in Table 4.1 below.

The challenge posed by visual metaphors in different genres suggests that a broader theory of visual metaphor is needed which incorporates existing ideas

and integrates them into a more encompassing model that accounts for a larger variety of visual metaphor and that captures mappings in both dimensions (i.e., expression and conceptualization). These dimensions are largely compatible with the referential level of meaning and the more general and abstract level of meaning in VisMIP (Šorm and Steen this vol.) and with Schilperoord's (this vol.) types of incongruity and resolution templates. It is one aim of this volume as well as of the Vismet corpus to promote further discussion about these various perspectives in order to integrate them into an encompassing, explicit framework.

4.2.2 The semiotic dimensions of visual metaphor

In our view, visual and linguistic metaphors have modality-specific inner variability that can be quite reliably compared on the dimensions of expression and conceptualization, but they substantially differ on the dimension of communication. The communication dimension includes the contrast between the deliberate versus non-deliberate use of a metaphor (Steen 2013). We believe that deliberateness is a typical feature of visual metaphors. This perspective is supported by Kennedy (1982), and highlighted by Forceville and Urios-Aparisi in the following terms: "a phenomenon to be labeled a visual metaphor should be understandable as an *intended* violation of codes of representation, rather than as being due to carelessness or error." (2009, p. 23). As for the conceptualization of visual metaphors, we focused on the well-known conceptual contrast between *conventional* and *novel*. To characterize the type of expression in the metaphorical images, we labeled them as containing either a *familiar* or a *creative* expression of a metaphor. The next section explains and provides examples for the precise definitions of these labels.

4.3 Analyzing images in Vismet

4.3.1 Analyzing different dimensions of visually encoded meaning

Our practice in analyzing the visual metaphors in our corpus was as follows. We first took a first look at all the information available in and about the image: context (when was the image published and why?), style (what do other images of the same author look like?) and author (what is his background?). Subsequently we focused on the visual representation and tried to verbalize one or more incongruities and correspondences between domains, based on the depicted or visually cued elements, i.e., the referential meaning of the image (cf. Šorm and Steen this vol.). If the referential meaning was ambiguous and allowed for more than one interpretation,

we provided alternative descriptions. The result of this process captures the *expression dimension* of the image.

The next step was to provide a verbalization of one or more of the correspondences between domains on a more abstract level to determine the more general meaning of the image (cf. Šorm and Steen, this vol.). In order to do this, we started from the metaphorical correspondences regarding the referential meaning, and reached the more general and abstract one through cognitive operations such as chains of metonymies.

The final step is to describe the overall communicative message of the metaphor. It captures the point and the topic of the message (cf. Šorm and Steen, this vol.). In advertising, we believe that this is the dimension in which the brand name should appear.

Table 4.1 shows how Figures 4.1 to 4.4 can been analyzed using this three-tiered schema.

Table 4.1 Sample analyses of four visual metaphors for three semiotic dimensions

Dimension	Figure 4.1	Figure 4.2	Figure 4.3	Figure 4.4
Expression	*Mouthwash is grenade*	*Iris is jungle*	*Merkel is queen, Sarkozy is jolly, EU is a cape*	*Paint is sky Paint is color*
Conceptualization	*mouthwash is grenade*	*camera system is human eye watching over the surrounding environment*	*Germany is the European leader, France is Germany's funny servant, the EU is a cape that covers/protects its members*	*Sky is freedom Color is life/ excitement*
Communication	*Listerine mouthwash is as powerful in killing bacteria as a bomb*	*The Land Rover 5-camera system provides a reliable view on the surrounding environment as a human eye would*	*Germany is leading Europe with a sort of help from France, in a historical moment where the other state members are weak and incapable of reacting.*	*In a dark and monotonous city the man who tries to bring color and freedom is destined to die*

4.3.2 Expression

The *expression dimension* focuses on what can be seen in the image. This includes two variables. The first is *realization* and concerns the way in which two domains are visually arranged in the image. This has been analyzed according to the types identified in Phillips and McQuarrie's model (2004), which is based on three categories: juxtaposition, fusion, and replacement.

The second is *symbolism*. It conveys whether the domains of the metaphor that expresses the referential meaning of the image are represented through well-known symbols or not. We mark a visual metaphor as involving symbolism when either the source, or the target domain, or the necessary context (to understand the message) identified at the expression dimension is represented through a symbol, i.e., a conventionalized visual sign that would trigger that specific concept also in absence of other contextual cues. Many symbols are exploited in visual metaphors (especially in political cartoons) for they convey in a familiar and short-hand way abstract concepts (see Figure 4.5). Other symbols (such as cultural icons) are exploited because they evoke un-represented meaning, which the viewer must unravel by understanding the visually cued reference (see Figure 4.10). A non-exhaustive repertoire of symbols is listed below.

- http://en.wikipedia.org/wiki/List_of_symbols
- cultural icons (famous photographs, movie posters, famous paintings, fairy tale figures)
- graphs, charts, pie charts, words
- shapes of countries, flags/persons that represent countries, currency for countries
- dove for peace, cupid for love, green reaper for death, red flag for revolution

Examples of visual symbolism can be found in the examples in Figure 4.5: an 'at'-sign may refer to the internet or technology in general; a dollar sign may represent the American economy; a circle of stars may refer to the European Union; the bar code may stand for a controlled market that may be tracked.

During the annotation of the images we also encountered other symbols that need to be added to this list. These include a *clock* or *hourglass* as a symbol for *time*, a *mouse*, *phone* or *tablet* as a symbol for *modern technology*, and a *band-aid* as a symbol for *fixing*. A broad selection of these symbols is reported in the list of 'potential visual familiarity markers' (see the VisMet website: http://www.vismet.org/VisMet/index.php). With a democratic VisMet corpus, we would like to discuss these markers with more experts and non-experts.

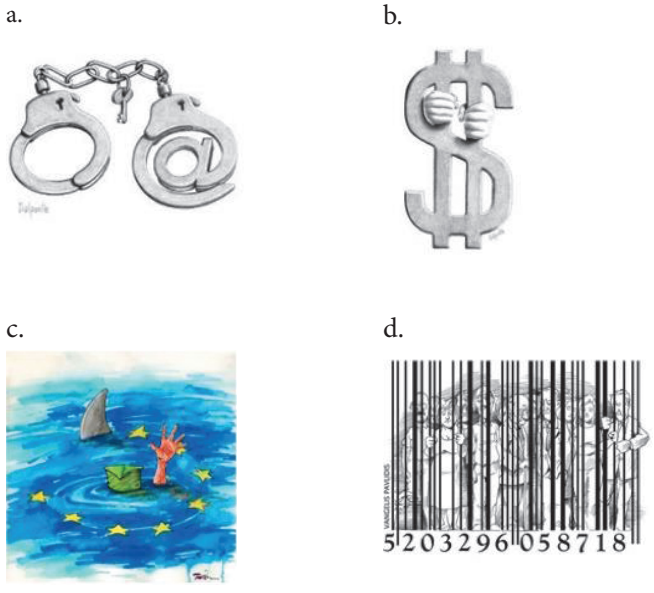


Figure 4.5 Examples of visual familiarity markers: Illustrations and political cartoons extracted from the VisMet corpus

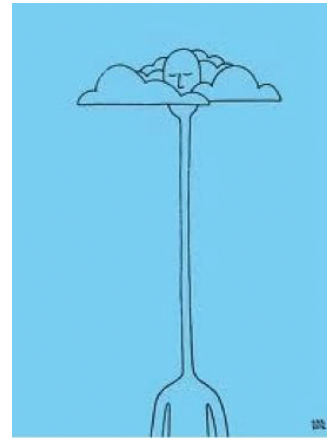
Linguistic familiarity There is a specific group of images whose metaphoricity cannot be understood without including allusion to a metaphorical linguistic expression. Figure 4.6 contains some clear-cut cases of such linguistic familiarity.

These images are straightforward representations of the non-metaphorical meaning of metaphorical expressions: "walking in his father's footsteps", "one's head being in the clouds", and "being packed like sardines". In other images, such linguistic expressions are not the central topic of the visual metaphor, but are instead exploited to convey a specific message. This happens mainly in advertisements. For example, "give a hand" to WWF, in Figure 4.7, is a linguistic expression that helps draw the intended conclusion about coming to the animal realm's aid.

We annotated these images as *linguistically familiar*. Such familiarity belongs to the expression dimension. To record this particular property, we first named the incongruous and the most salient units in the image (in the examples in Figure 4.10: *footsteps*, *head* and *cloud, sardines*) and then queried them in the following dictionary of idioms: http://www.thefreedictionary.com/. The idioms returned were then compared with the image, and if they proved crucial for parsing, we marked the image for linguistic familiarity.

a.

b.



c.

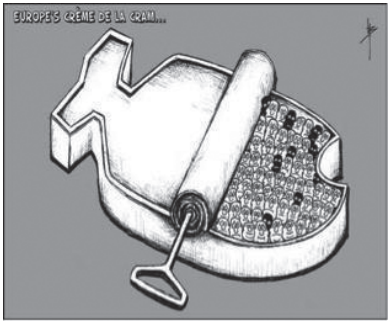


Figure 4.6 Examples of images annotated for linguistic familiarity: Illustrations and political cartoons extracted from the VisMet corpus

Figure 4.7 WWF social campaign; slogan within the image: "Give a hand to wildlife" (author and copyright owner: Saatchi & Saatchi)

4.3.3 Conceptualization

According to the contemporary cognitivist approach launched by Lakoff and Johnson (1980), metaphors are figures of thought because they manipulate concepts. It follows that they should be independent from the expression in one specific modality. Therefore, conventional metaphors that underpin linguistic expressions (e.g., LIFE IS JOURNEY) should also be found in visual rhetoric. As a result, scholars have attempted to analyze still images which they claim contain such conventional metaphors (recently, Ng & Koller 2013). The difficulty with such analyses, however, is that they are strongly dependent on personal intuitions about what constitutes the deepest layer of the conceptual metaphor. It is often unclear how these personal intuitions can be derived from the observation of the image. For this reason, the labeling of an image for conceptual conventionality or novelty can be a highly subjective decision.

In an attempt to overcome this unreliability, we suggest a simple procedure for identifying the presence of conventional metaphors that have already been established in the literature on linguistic metaphors in images. This procedure, exemplified through the analysis of four images on the VisMet website (http://www.vismet.org/VisMet/annotation.php), is based in four steps:

1. Look at the image, search for the incongruous units, and list those entities that visually violate an expected scene. Such entities are candidates which may play a role in the construction of a metaphor.
2. Collect from the electronic lexical database WordNet (Miller 1995) synonyms, hypernyms and related terms for each of the incongruous units detected and listed in step 1. This step allows the annotator to cover a larger spectrum of possible verbalizations of the incongruous units that have been listed in step 1.
3. Compare the list of terms obtained from step 2 with a comprehensive archive of conventional metaphors (see the VisMet website for suggestions). Each word in the list should be used as a keyword to retrieve conventional metaphors containing that specific word as a source or target domain.
4. For each conceptual metaphor returned by step 3, decide whether the metaphor applies to the image or not (i.e., whether the metaphor helps the viewer to understand the image). If it does not apply it should be noted as checked but irrelevant and disregarded. If at least one conventional metaphor applies to the image, the conventional metaphor should be noted and labeled as *conventional*. Otherwise it should be labeled as *novel*.

In this way, we proceeded to annotate the images collected for VisMet on the conceptual dimension. The disadvantage of this procedure is that it excludes any conventional metaphors that are *not* listed in the existing archives. For this reason, we maintained a list of 'potential conventional metaphors' (see the VisMet website:

http://www.vismet.org/VisMet/index.php) for any conceptual metaphors that seemed to be unjustly lacking from the list of conventional metaphors.

4.3.4 The annotation process

During the Spring of 2014 three annotators annotated the images of the VisMet corpus following the annotation scheme described above. The three annotators first labeled a sample of 10 images on all the dimensions described above, and then proceeded to annotate batches of 50 images each, until full coverage of the corpus was reached. For each batch, the annotators started annotating the images individually. After that, the annotations were compared and discussed together, to address problems and tailor the annotation scheme where needed. The inter-rater agreement for all the dimensions of meaning (prior to discussion, so comparing the annotations provided individually by the annotators), per batch, can be found in Table 4.2.

Table 4.2 Reliability scores based on the annotations of all the categories (conceptualization, expression realization, symbolism, and presence of linguistic expressions)

Batch	Reliability (Fleiss' kappa)
1 (10 images)	.36
2 (50 images)	.27
3 (50 images)	.54
4 (50 images)	.54
5 (50 images)	.76
6 (50 images)	.66
7 (50 images)	.70
8 (33 images)	.75

Although the annotation remains open to subjectivity, the inter-coder agreement gradually increased during the annotation process from low agreement ($k = 0.36$) in the first batch to a substantial agreement ($k = 0.75$) in the last batch. This shows how the annotators slowly moved towards more consensus with regard to their annotations and the annotation procedure became clearer. However, every batch had its share of difficulties, as we will see in the next section.

4.3.5 Annotation problems

Images that have been dropped

Some images that were collected because of a perceived perceptual incongruity, were eventually not included in the corpus because the incongruity did not lead to the identification of metaphorical correspondences (cf. Schilperoord, this vol.). These images were placed under the lens of the VisMIP test and analyzed step by step through this detailed procedure, in order to decide whether they should be acknowledged as visual metaphors, or rejected. An image that after the VisMIP test was eventually rejected is shown in Figure 4.8.

Figure 4.8 Political cartoon about the dynamics between rich and poor, no title (author and copyright owner: Jaksa Vlahovic)

A schematic report of the results of the VisMIP test applied to Figure 4.8 is illustrated in Table 4.3.

Table 4.3 Application of the VisMIP test to an image that was eventually rejected

1. **Look at the entire image, including visual and verbal elements, to establish a general understanding of the meaning.**
1a. Describe in just a few simple phrases the referential meaning of the image, i.e. what/who is being depicted here, what he is doing, where he is doing it, and so on. *A hand drops coins into a beggar's hat, which has a hole in it, and the coins end up in the nice hat of the man who was seemingly being charitable.*
1b. Test whether there are any clues that tell you that more general and abstract meaning should be attached to what is described under step 1a. *No verbal anchors. The hat with a hole might stands for poverty and the top hat for wealth (metonymies)*
1c. Reconstruct the point of the image *Charity looks good but serves selfish goals, because money eventually returns to the hands of the rich.*

Table 4.3 (*continued*)

1d. Derive from step 1c the topic of the point.
Charity

2. **Structure the descriptive phrase(s) under step 1a.**
Agent(hand1 | with nice sleeves)
Agent(hand2 | with not nice sleeves)
Agent(hand3 | with nice sleeves)
Action(dropping coins | done by hand 1)
Action(holding broken hat | done by hand 2)
Action(holding top hat | done by hand 3)
Object (coins | falling through the broken hat into the top hat)
Object (begger's hat | with hole)
Object (top hat | nice and clean)

3. **Find incongruous visual units**

3a. Decide for each unit under step 2 whether it is incongruous with the topic as formulated under step 1d ('topic-incongruous').
Agent(hand3 | with nice sleeves) incongruous!
Action(holding top hat | done by hand 3) incongruous!
Object (coins | falling through the broken hat into the top hat) incongruous!
Object (top hat | nice and clean) incongruous!

3b. Decide for each topic-congruous unit under step 2 whether it shows properties that are incongruous with the properties that are typically true of that unit ('property-incongruous').
Object (begger's hat | with hole) property incongruous!

4. **Test whether the incongruous units are to be integrated within the overall topical framework by means of some form of comparison.**

4a. For each incongruous unit under 3a, determine which replacing unit would be congruous with the topic AND would be coherent with the referential meaning of the image.
Agent(hand3 | with nice sleeves) should not be there at all.
Action(holding top hat | done by hand 3) should not happen at all.
Object (coins | falling through the broken hat into the top hat) substituted with coins that stay in the begger's hat.
Object (top hat | nice and clean) should not be there at all.

4b. For each incongruous unit under 3b, determine which replacing unit would typically own the incongruous properties AND would be coherent with the referential meaning of the image.
Object (begger's hat | with hole) substituted with a hat without holes

5. **Test whether the comparison(s) is/are nonliteral or cross-domain.**
There are no comparisons or, if there are, they are within the same domain (metonymies beggar stands for poverty in general, top hat stands for wealthy person doing charity)

6. **Test whether the comparison(s) can be seen as some form of indirect discourse about the topic as formulated under step 1d.**
No

7. **If the findings of tests 4, 5, 6 are positive, then a visual unit should be marked for metaphor.**
This is not a metaphor.

Problems with realization

The main problem encountered when annotating images for *realization* was lack of clarity about the difference between Forceville's *contextual* and *integrated* images, even for advertisements. Figure 4.9, for example, displays an advertisement for a coffee roasting company. In this image, the espresso cup seen from above resembles a switch, turned onto "max". One could argue that the coffee cup and the switch are *integrated* in a compossible fashion, thanks to the perspective chosen by the artist, which makes the two shapes identical. On the other hand, without the linguistic anchors provided by the context "min" and "max" would the coffee cup still resemble a switch? If not, then the contextual cues are crucial to evoke the source domain (the switch), which is not represented, and thus the metaphor should be annotated as *contextual*. For many images, it was possible to defend different values for the *realization* parameter emphasizing different properties of elements within the image.

Figure 4.9 Advertisement for Black and Blaze coffee roasting company (author and copyright owner: Inhalt&Form Werbeagentur BSW)

Problems with symbolism

While annotating images for symbolism we stumbled upon the problem of having conventional symbols that were not recognized by all the annotators because the conventionality of these symbols is only recognized by certain subcultures or demographics. Figure 4.10, for example, displays a bottle of sparkling wine (target domain and product to be sold) depicted as a rocket that lands on the moon.

Only one annotator recognized the moon and the bottle in its eye as a symbol that refers to the first known science fiction film 'Le Voyage dans la lune' (A Trip to the Moon), from 1902, in which a spaceship lands in the moon's eye. At this point, the question is whether we should call the moon and its rocket/bottle a symbol or not. This in turn depends on whether the focus is on the perception of a general audience, or the likely intention of the author, an issue which merits future discussion and research.

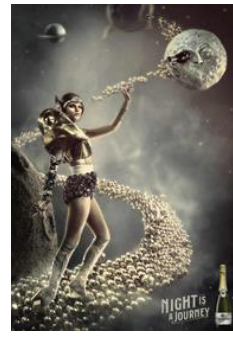


Figure 4.10 Advertisement for Italiana sparkling wine brand, slogan within the image: "Night is a journey" (author and copyright owner: Milk, Lithuania)

Problems with the linguistic familiarity

The recurrent question that arose while annotating for linguistic familiarity was whether images which can be related to a certain linguistic expression also *depend* on that expression for their meaningfulness. There exists a class of images which seem to exploit the meaning of a familiar linguistic expression to reinforce the message, but which can be interpreted successfully by the viewer even if they fail to detect this reference.

In Figure 4.7, the linguistic expression 'give a hand', reported also verbally in the slogan, is crucial for the interpretation of the WWF campaign conveyed by the image, and therefore it was annotated as linguistic familiar. In Figure 4.9, instead, the linguistic expression "turn somebody on" can be attributed to the image but it is not crucial to understand the conveyed message, because the metaphor works on the conceptual dimension through the conventional metaphor PEOPLE ARE MACHINES. Because this metaphor is conceptual the understanding of the image should succeed independently from the access to the specific linguistic expression "turn somebody on". Similarly, in political cartoons such as the one in Figure 4.11, we do not need to know the linguistic expression "blind someone to something" or "being blind about a situation" in order to understand the message conveyed by the image, because we process it on a conceptual level, through the conventional metaphor KNOWING IS SEEING.

The problem remains unsolved for those images that make use of linguistic expressions conveyed in a language that is not familiar to the annotators. An open and democratic platform could help significantly in recognizing expressions and help us identify foreign expressions that are present in the pictures of the corpus, as well as help us improve the methodology to present a more nuanced view of the possible relations between metaphorical linguistic expressions and visual metaphors.

Figure 4.11 Political cartoon about Obama's decisions in relation to Iraq war (author and copyright owner: Bernardo Erlich)

4.4 The VisMet website

Whoever wishes to sample the VisMet corpus need only visit http://www.vismet.org/VisMet/, a website that is the result of a collaborative effort by the Metaphor Lab Amsterdam, Metaphor Geeks (International Center for Intercultural Exchange) and the Network Institute (VU University Amsterdam). Members of the Metaphor Lab Amsterdam and of Metaphor Geeks determined the ideal database structure and interface based on the corpus' intended functions. This input was then translated to a technical implementation in the form of a website by Dr. Marco Otte, member of the Tech Lab (Network Institute, Amsterdam). Technical details about the implementation of the website are provided in the coming section, while a description of the filters that allow to browse the corpus is provided in Section 4.2.

4.4.1 Implementation details

The VisMet 1.0 website was built using a combination of php, html, Javascript and a MySQL database to hold the references to the images. The database holds the location of the thumbnail and full size images on the server and all the meta information about the image, e.g., the author, or year of publication. On the back-end of the web site, secured by password, a simple editor interface allows administrators of the website to add, change and delete images in the system.

The home page of the VisMet website uses a php script to determine which of a predefined set of large sample images will be presented to the current user. Each time the page is refreshed a new image is randomly selected. Title, subtitle and credits are also extracted from this file and displayed with a short delay to allow for a focus on the main image before the explanatory title is shown.

The browse page, also coded in php, uses an open source Javascript called Isotope (version 1.5 by Metafizzy, licensed under GPL v1.3 license)[2] to interactively show images based on a set of filters. As for the image profiles, these were created by the Lightbox Javascript tool (Lightbox version 1.8 by Lokesh Dhakar).[3]

4.4.2 Filtering images with the side bar

The browse page, by default, is set to display the full corpus in the form of thumbnails which can be clicked to raise an embedded page containing detailed information about an image (the image profile, see Section 4.4.3). In order to facilitate usage of the corpus, the website provides an interactive display which can modulated by the toggling of filters by using the sidebar on the left, in order for the user to select only those images that are of particular interest to them.

The filtering functionality works as follows. Each parameter can be set to 'any', in which case no images will be filtered out. The parameters themselves are not mutually exclusive, meaning that it is possible to combine the filters to select, for example, only paintings with novel metaphors, or only advertisements with linguistic expression-based metaphors which are conventional. However, within a parameter, the values are mutually exclusive. It is therefore not possible to select images from two genres at the same time, or to view both hybrid and integrated metaphors, but no contextual or aligned ones.

4.4.3 Image profile

Metadata on an individual image can be viewed on the image profiles, which contain the filtering keywords of the annotation scheme, as well as the following metadata: title, timestamp, author, copyright owner, year of publication, country of origin, brand (if applicable), text within the image, text outside the image, source, URL, and notes. This pop-up screen can then be used to group images of the same author ('Show all of this author') or the same year of publication ('Show all of this year'). The final button ('Report abuse') allows users to quickly address complaints to the VisMet team.

---

2. http://isotope.metafizzy.co

3. http://lokeshdhakar.com/projects/lightbox2/

4.4.4 Text pages

The text pages on the website have two purposes: explaining to the user what the VisMet project is about and how it may be used to study visual metaphors, and informing the user about the possibility to utilize the images and contribute by offering new ones. These include: the project description, the annotation scheme, all the legal aspects of national and international copyright law, explained in the Terms of Use page, contacts, and a specific page where the visitor is encouraged to contribute to the corpus by submitting new material and approaching the VisMet research team with questions and comments about the project.

4.4.5 Social networks

To stimulate interaction with both academic and non-academic publics (book illustrators, cartoonists, people interested in creative images and metaphor), we created some pages on social networks. The website homepage links to these social network sites. On Facebook we have a 'VisMet Visual Metaphor' page, where we post visual metaphors and information about conferences and the theory behind visual metaphor. This page is directed at both laypersons and experts and wants to interest people for visual metaphor and to introduce them to the theory behind these metaphors. People are also invited to join discussions about the images of the corpus and to help us analyze the images.

The website links also to a Twitter page and a LinkedIn page. On the Twitter page, information will be shared about new images that will be added to the corpus and news about the project or website will be published here. On LinkedIn, we start discussions about theoretical issues. Here, we would like to start theoretic conversations with a more expert public. This way, with various social media platforms, we hope to reach different people with different interests to eventually interest them for visual metaphor and invite them to join our discussions on the subject. Until now, also a lot of non-academic users have shown interest in the corpus. 'Solving a riddle' seems to be an activity a lot of people enjoy.

4.5 VisMet 1.0 and its future

With the VisMet project, we have shown that creating a corpus of visual metaphor is a necessary step in the blooming field of visual metaphors, as it provides a shared electronic resource that can be used as a benchmark against which different models can be tested and applied. The material shown on the website can be used for

various kinds of research and provides a unique opportunity to take a further look into the analysis of visual metaphors across different genres.

This corpus has also led to a newly developed annotation scheme. Applying the existing scheme to visual metaphors was not easy, but instead raised many issues and relevant points concerning the application of ideas devised for verbal metaphors to a different modality. With VisMet 1.0 we hope to raise scientific awareness about the various methodological issues that have been described in this chapter. In particular, VisMet 1.0's aim is to facilitate the public sharing of insights that will help produce a more reliable procedure in the future.

Our highest priority now is to add new functionalities to the website to facilitate online discussions within a community of experts and non-experts which can help advance the study of visual metaphor. It will be interesting to see, for example, if the 'potential familiarity symbols' and 'potential conventional metaphors' are seen as familiar or conventional by both expert and non-expert others. A new functionality that we are currently implementing is the addition of crowdsourced tags. Tags produced by VisMet users will be a valuable source of information in terms of revealing which features are considered salient. The corpus also needs to become more user-friendly by allowing more effective querying and easier contribution of material and ideas.

The first steps towards raising awareness of the VisMet corpus were taken at the RaAm conference in Cagliari, Italy, June 2014, when the VisMet 1.0 website was launched. Since then, we have received positive reactions from expert from the field, but also from cartoonists, creative agencies and visual metaphor scholars. We hope that with their positive response to the corpus, we can enthuse other authors to give us permission to use their material. Since the corpus is scalable, we also envision to add new images in the near future, collecting permissions from the authors and copyright owners as we have done so far. In this respect, we hope to reach a turning point, by which the authors and copyright owners of the images will ask the VisMet team to include their images in the corpus, rather than the other way around.

We envision the VisMet corpus to be enlarged in the future with new materials, more images and images from other genres. But, of course, our main goal is to provide researchers of visual metaphor with a resource that helps them foster visual metaphor research. We hereby invite you to have a look at the VisMet 1.0 project, join the open, democratic platform and the subsequent discussions on visual metaphor and share your feedback, opinions and insights.

Acknowledgements

We wish to thank Frank van Harmelen, former director of the Network Institute, for sponsoring the realization of the corpus, Marco Otte for implementing the website, Isabel Del Negro Alousque for her advice on the VisMIP procedure, and Gerard Steen for his support in every aspect of the project.

References

Bolognesi, M., Timmermans, B., & Aroyo, L. (accepted). VisMet and the crowd: what does social tagging reveal about visual metaphors. In M. Bolognesi, K. Despot & M. Brdar (Eds.) *Fantastic Metaphors and where to find them*. Amsterdam: Benjamins Publishers.

Bounegru, L. & Forceville, C. (2011). Metaphors in editorial cartoons representing the global financial crisis. *Visual communication*, 10(2), 209–229. https://doi.org/10.1177/1470357211398446

El Refaie, R. (2003). Understanding visual metaphor: the example of newspaper cartoons. *Visual communication*, 2(1), 75–95. https://doi.org/10.1177/1470357203002001755

Forceville, C. J. (1996). *Pictorial Metaphor in Advertising*. London/New York: Routledge. https://doi.org/10.4324/9780203272305

Forceville, C. J. (2002). Further thoughts on delimitating pictorial metaphor. *Theoria et Historia Scientiarum*, 6, 213–27.

Forceville, C. J. (2006). Non-verbal and multimodal metaphor in a cognitivist framework: Agendas for research. In: G. Kristiansen, M. Achard, R. Dirven & F. R. de Mendoza Ibàñez (Eds.), *Cognitive Linguistics: Current Applications and Future Perspectives* (pp. 379–402). Berlin: Mouton de Gruyter.

Forceville, C. J. (2008). Pictorial and multimodal metaphor in commercials. In: E. F. McQuarrie & B. J. Phillips (Eds.), *Go Figure! New Directions in Advertising Rhetoric* (pp. 272–310). Armonk NY: ME Sharpe.

Forceville, C. J. (2009). Metonymy in visual and audiovisual discourse. In: E. Ventola & A. J. Moya Guijarro (Eds.), *The World Told and the World Shown: Issues in Multisemiotics* (pp. 56–74). Basingstoke/New York: Palgrave MacMillan.

Forceville, C. J. & Urios-Aparisi, E. (Eds.). (2009). *Multimodal metaphor*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110215366

Gkiouzepas, L. & Hogg, M. K. (2011). Articulating a new framework for visual metaphors in advertising: a structural, conceptual, and pragmatic investigation. *Journal of Advertising*, 40(1), 103–120. https://doi.org/10.2753/JOA0091-3367400107

Kennedy, J. M. (1982). Metaphor in pictures. *Perception*, 11(5), 589–605. https://doi.org/10.1068/p110589

Koller, V. (2008). "The world in one city": Semiotic and cognitive aspects of city branding. *Journal of Language and Politics*, 7(3), 431–450. https://doi.org/10.1075/jlp.7.3.05kol

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago.

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11), 39–41. https://doi.org/10.1145/219717.219748

Ng, C. J. W. & Koller, V. (2013). Deliberate Conventional Metaphor in Images: The Case of Corporate Branding Discourse. *Metaphor and Symbol*, 28(3), 131–147. https://doi.org/10.1080/10926488.2013.797807

Phillips, B. J. & McQuarrie, E. F. (2004). Beyond Visual Metaphor: A New Typology of Visual Rhetoric in Advertising. *Marketing Theory*, 4, 113–136. https://doi.org/10.1177/1470593104044089
Schilperoord, J. (this vol.). Ways with pictures: Visual incongruities and metaphor.
Steen, G. J. (2013). Deliberate metaphor affords conscious metaphorical cognition. *Journal of Cognitive Semiotics*, 5 (1–2), 179–197.
Šorm, E. & Steen, G. (this vol.). VISMIP: Towards a method for visual metaphor Identification.
Van den Heerik, R., Šorm, E., & Steen, G. (this vol.). Behavioral evidence for VISMIP.
Whittock, T. (1990). *Metaphor and film*. New York: Cambridge University Press.

PART II

Process

CHAPTER 5

Behavioral evidence for VISMIP

Romy van den Heerik[1], Ester Šorm[2] and Gerard J. Steen[1]
[1]Universiteit van Amsterdam / [2]Haagse Hogeschool

5.1 Introduction

This chapter presents a first exploration of the different steps that may be involved in visual metaphor processing, inspired by a behaviorally motivated procedure for visual metaphor identification by analysts (see Šorm and Steen this vol.). Šorm and Steen (2013) used a think aloud task to let people verbalize their thoughts while looking at pictures containing visual metaphors. The resulting data were analyzed for the mental operations participants undertake in visual metaphor processing. In the present chapter, we will analyze these data in relation to the visual metaphor identification procedure presented in Chapter 3: we aim to identify a minimal number of cognitive operations in processing that are also reflected as operational steps in the procedure. We will indicate the similarities and differences between the metaphor processing data and the metaphor identification in Chapter 3 (Šorm & Steen this vol). In this way we will explain how the thinking out loud data can contribute to fine tuning VISMIP, which itself was initially modeled on the verbal metaphor identification procedure MIPVU (Steen et al., 2010). VISMIP is hence also validated against processing behavior for visual metaphor processing by the general public.

5.2 Method

The fundamental idea of thinking out loud is that the information heeded by a participant carrying out a psychological task may be verbalized concurrently or retrospectively and that such a report provides evidence of the information processing that is going on (Steen 1994). Thinking out loud thus gives access to the concepts people attend to when they perform cognitive tasks, making it an appropriate technique for mapping the conceptual operations that may be at work during visual metaphor processing (Ericsson & Simon 1993). It is granted that thinking out loud

https://doi.org/10.1075/celcr.18.05van

may increase consciousness during processing to some extent. This is however not just a hindrance: in this way, one receives optimal and explicit data (Kintgen 1983). Thinking out loud can be considered a valid method and an excellent means to collect data on visual metaphor processing, as the data obtained provide invaluable information about people's concomitant mental activities (Steen 1994).

5.2.1 Participants

Data were collected in December 2009. For this study, 24 visitors of the Public Library in Leiden (The Netherlands) were randomly recruited, 12 men and 12 women. The average age of the participants was 47 years, ranging from 17 to 69. 18 of them participated in higher education. All of the participants had the Dutch nationality and Dutch as their native language. Each participant was tested individually in a separate room in the library and was informed about the financial reward of 10 euros, which they were to receive after completing the test session. All participants were naïve about the purpose of the study.

5.2.2 Material

Twenty different pictures were used of which eight were used for warming-up purposes and twelve for collecting data on visual metaphor processing. The experimental materials represented four different genres: political cartoons (5), magazine advertisements (5), educational illustrations (5) and paintings (5). We selected diverse genres to prevent verbalizations from being too peculiar to a particular category and in order to have available a sufficient range of potentially different verbal reactions.

Each stimulus, except the cartoons, included both verbal and nonverbal elements. It should be noted that the original paintings did not include titles, but that these were added by the authors. Each title was printed under the image (using Arial 16 bold). As the original titles were in English and French, Dutch translations were presented between round brackets. The stimuli were presented on paper and printed in color on separate pages. Each page was put in a ring binder, so that participants could turn the pages.

No more than twenty stimuli were presented, to prevent participants from losing interest and attention. The eight pictures used in the warming-up phase included four different genres, to give participants an idea of the kinds of images they were to expect in the test phase. However, the warming-up images did not contain any visual metaphors, to avoid learning effects. In the test phase, each genre was represented by three different metaphorical pictures to make sure that responses

would not be dependent on one particular picture. The order of presentation of the 20 test images was randomly varied between participants, in order to limit the effects of learning transfer.

5.2.3 Instructions

The participants were informed that ads, cartoons, educational illustrations, and paintings would be presented to them. In addition, they were told that the study was intended to gain insight into the way people process pictures. They also received instructions for thinking out loud, based on Ericsson and Simon (1993, p. 376), Steen (1994, p. 136), and Van den Haak (2008, p. 32). For the sake of consistency, the instructions were read out from a written checklist (cf. Steen 1994, p. 136). A copy of this check-list was shown to the participants, so that they could read along with the oral instructions. The instructions also stated that participants would be reminded to think aloud in case they paused too long, and that they could turn the page if they felt they had said enough about a certain picture. Subsequently, an opportunity was given to ask for clarification of the instructions.

5.2.4 Test session

With the instructions, participants were requested to look at a number of pictures as if they were alone in the room, and to verbalize all thoughts that occurred to them. This warming-up period preceded the actual test, to familiarize the participants with verbalizing their thoughts. Giving participants the opportunity to practice thinking aloud is recommended by, for instance, Van Someren, Barnard and Sandberg (1994, p. 43) and Ericsson and Simon (1993, p. 240–241, p. 257). In the warming-up period, four non-metaphorical images from different genres were presented to the participants. At this stage, the recording equipment was tested. Participants who felt comfortable with thinking aloud immediately proceeded with the test session. For one participant who did not feel comfortable with the task of thinking aloud, the practice phase was prolonged with another set of four images (see Van Someren, Barnard & Sandberg 1994, p. 44). Again, there was one magazine advertisement, one political cartoon, one educational illustration and one painting.

During the test session, any interference was avoided. The only interference occurred when the participant stopped thinking aloud. In that case, the participant was reminded to talk aloud by the researcher, who said 'Keep on talking, please'. Using a brief reminder such as 'Keep on talking, please' has been advised by, e.g., Van Someren, Barnard and Sandberg (1994, p. 44) and Ericsson and Simon (1993, p. 256). In the event, this happened just once. After the test session,

participants were kindly thanked for their cooperation and debriefed. They also received a financial reward (about which they had been informed before participation). One session took about 30 minutes.

5.2.5 Data transcription

The participants' thoughts were recorded on tape and transcribed into written protocols. As the study was conducted with Dutch participants, all protocols were in Dutch (for this chapter, the thinking out loud data have been translated into English by an English native speaker, with near-native proficiency in Dutch). The protocols were segmented into clauses to allow for an analysis of metaphor processes, based on the windows of attention of the participant. Clauses have been used as the basis semantic unit of discourse in quite a few analytical approaches (e.g., Carlson, Marcu & Okurowski 2001; Longacre 1983; Mann & Thompson 1988; cf. Verhagen 1996, p. 250); they are the building blocks of text which are easiest to classify. Therefore, as a rule, the clauses corresponded to the basic units of discourse. When a group of words contained a subject and predicate of its own and formed part of a sentence, this was considered a clause. However, there were also some exceptions to this rule: following Mann and Thompson (1988), and also Carlson and Marcu (2001), clausal subjects, complement clauses, and restrictive relative clauses were considered as part of the clauses in which they were embedded ('host clause units'), rather than as independent units. A set of segmentation guidelines described how the protocols were segmented into clauses, also dealing with incomplete or elliptical clauses.

5.2.6 Empirical evidence for VISMIP

Based on these data, we will in the following paragraphs present how various aspects of VISMIP can be empirically validated in the mental processes of non-experts while processing a visual metaphor. No hard figures will be given concerning the incidence of various findings. Rather, think aloud examples will be discussed in order to compare them to the steps of the procedure. We will divide the data into four major groups: 'General understanding', 'Units of analysis and finding incongruity', 'Testing the need for comparison and testing non-literality' and 'Potential explanation by cross-domain mapping'. These groups form the foundation for the major steps in VISMIP in Chapter 3.

5.3 General understanding

When identifying a visual metaphor, analysts should first try to grasp the general meaning of an image by looking at the entire image (cf. Schilperoord this vol.; Sorm & Steen this vol.). Grasping the general meaning of an image means that visual elements should be recognized: what agents and objects are being depicted, what is happening in the picture, and where and when is it happening? To determine this referential level of meaning, people may use contextual information, such as verbal elements accompanying the picture or previous knowledge on, for instance, the designer of the image or genre (Šorm & Steen 2013). The think aloud data show how during ordinary visual reception, people refer to the title of the work, the genre ('this is a cartoon/an advertisement), the artist ('I think this is Dalí') and the accompanying text (they start reading). Moreover, people often first describe the referential meaning of the image, that is, what/who is being depicted, what he is doing, where he is doing it, and so on. Consider the following two examples:

(1) Participant 5
Oh.
Yes.
Boy.
Little boy.
Pistol.
Pawn.
Strange pawn.
Can also be part of a… staircase.

(2) Participant 17
That is an uh drawing of a boy holding a uh pistol to his head.
With in the background a uh pawn.
And a bare space with uh black and white tiles.

Both are reactions to a painting by Magritte entitled 'Checkmate' (Figure 5.1). In the foreground, a man points a gun at his own head. In the background, a chessboard floor and a pawn are depicted.

First of all, the reactions show that people process pictures in terms of what is actually shown. It can be said that this is the straightforward meaning of a picture. Second, the reactions illustrate that viewers may describe the referential meaning in different manners: in the first example the image is described in single words whereas in the second example the image is described in sentence structures. Finally, Example (1) shows that viewers may give alternative or even competing interpretations to what they see in an image. The participant recognizes a 'pawn', 'but it can also be part of a staircase'. After a first glance at an image, viewers are

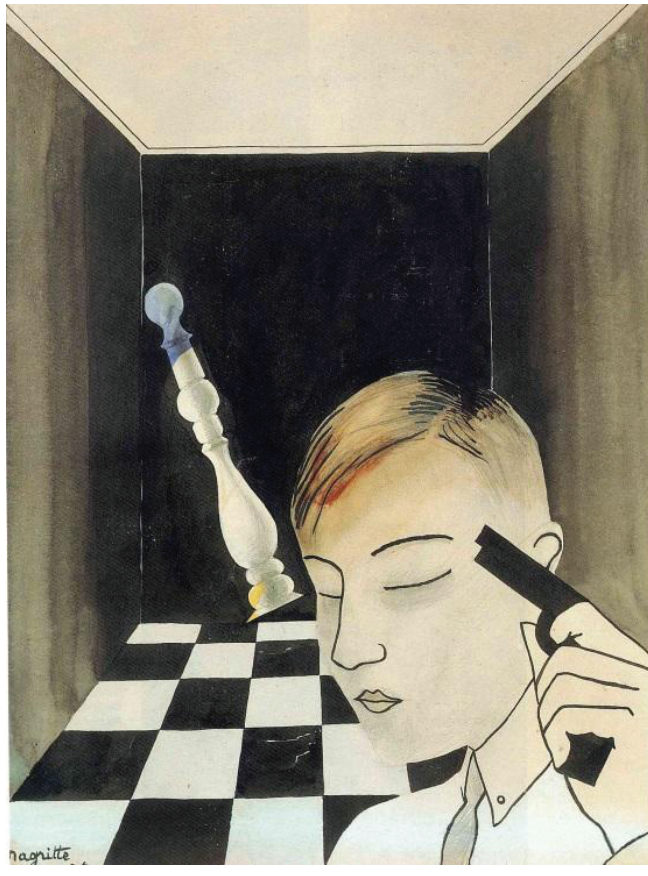


Figure 5.1 'Checkmate' by Magritte

thus not immediately sure how to interpret the different elements of the visual. In the examples, the participants frequently hesitate, including 'er', 'oh' or 'uh', before they categorize or conceptualize their thoughts and verbalize the corresponding word, identifying the various referents of the image.

Consider the relation between these think aloud data and the VISMIP instructions. First of all, there is the choice between image descriptions in single words similar to the description in the first example, and image descriptions in sentence structures as in the second example. Because of their explicit expression of relations between words, or concepts, or referents, descriptions in sentence structures are more preferable compared to using single words, as connections between visual elements then become more visible. Therefore VISMIP has the following instruction: "Describe in just a few simple phrases the referential meaning of the image".

Second, there is the decision on the types of visual elements that should be included in a description. Considering the examples above, viewers may prefer using

only one type of visual element (what/who is being depicted), but they may also use different types. It is a good idea to make analysts aware of the fact that different elements can be used ("what/who is being depicted here, what he is doing, where he is doing it, and so on") without pressuring or obliging them to use all elements. In addition, VISMIP also gives analysts the opportunity to give alternative image descriptions in case of perceived ambiguity ("If the referential meaning is ambiguous and allows more than one interpretation, then give alternative descriptions").

Grasping the general meaning of an image also means that one should check whether other or more abstract meanings should be associated with the referential meaning. The think aloud data showed that not all participants proceed beyond the description of a referential meaning. However, when they do verbalize this, viewers naturally tend to include all kinds of abstract meaning, including emotions, judgments and evaluations. This can be illustrated with the following excerpts that refer to Figure 5.2, showing an advertisement for Del Monte canned fruit.

This advertisement shows a fresh-looking apple with a tab on the side, under the caption 'Only the best fruit', and causes all kinds of associations in participants. Besides descriptions of the different elements of the picture, participants provide us with their (dis)preference for apples ('I don't like apples'), their feelings about canned fruit in general ('It is less healthy and less tasty'), and their opinions about the image in general ('it looks funny' or 'it is a visually attractive image'). This can be illustrated by the following excerpts.

(3) Participant 24
An apple with a Del Monte can.
Yes,
my first association is of: the nice apple is tinned,
which means it is ultimately less nice.

(4) Participant 20
So that that that ring on that, that apple, that reminds me of a nice gla –
uh can of lemonade uh that glug glug
comes out of the fridge,
whereas the cans don't give me that feeling at all.

The first example shows how the participant finds an abstract meaning to be associated with the referential meaning. He first labels what he sees in the image ('An apple with a Del Monte can'), thus giving referential meaning to the advertisement. Subsequently, the process of association is unambiguously explicated by the word 'association': 'my first association is of: …'. The representational meaning is then associated with canned fruit and with what we could call a judgment: 'the nice apple is tinned, which means it is ultimately less nice'.

Figure 5.2 Advertisement for Del Monte canned fruit

Example (4) illustrates a similar process as Example (3), but signaled in a different manner. This participant uses the signal 'that reminds me of' to indicate that he starts associating with this 'ring'. It makes him think of a cold can of lemonade. Moreover, the correction 'gla – uh can of lemonade' shows how participants are constantly (re)interpreting visual information. First, the participant presumably thinks of a nice glass of lemonade, but corrects himself because the tab on the apple

indicates that it is in fact a can. Later on, the participant eventually explicates his association regarding the depicted cans: he does not associate the same (positive) feelings with cans of fruit.

In the think aloud data participants also frequently focus on conventional symbolic meaning when they are confronted with images from different genres. A clear example is the following excerpt:

(5) Participant 1
On the right uh cranes,
Presumably they symbolize western, civilized or industrial society.

This is a response to a cartoon by Tom Janssen, showing the Holy Pope as a tightrope walker (Figure 5.3).

Figure 5.3 Cartoon by Tom Janssen

The participant attaches an abstract meaning to what he recognizes as hoisting cranes, explicitly indicated by the use of 'symbolize'. This clearly signals the viewer's concern with a more abstract meaning, but there are other signals participants use. Other participants for example state that tightrope walking 'means' negotiating and, regarding the Figure 5.1, checkmate 'is' suicide. Consider also the following excerpt referring to the above-mentioned cartoon by Tom Janssen (Figure 5.3):

(6) Participant 5
Uh,
One side a sickle,
Other side a uh I think a star of David.
Two triangles.
And it is standing on the side –
Because the line means before or after death.

In the cartoon, the Pope carries a vertical stick with symbols of Islam, Christianity, and Judaism. First, the participant describes this stick in terms of its components ('sickle', 'star of David', 'two triangles'), which is another illustration of the empirical validity of an identification procedure extracting referential meaning. Then 'the line' – a visual component of the stick that the participant has recognized as such without verbalizing this before – is interpreted in a more abstract way: 'Because the line means before or after death'. With the use of 'because', the participant gives an argument for the position of the visual elements. In this case, the use of 'means' signals the participant's concern with the more general, abstract meaning. The think aloud data thus show that besides the referential meaning of an image, understanding the image also comprises understanding the abstract meaning of an image, sometimes including conventional symbolic meaning.

There are several other examples of participants verbalizing this process, which is a good reason for analysts to do the same when they identify visual metaphors in a set of materials. However, where participants in our think aloud study usually start associating without having clear reasons, analysts should check whether there are any clues that the messenger has in fact intended this association. This will prevent the analyses of an image from becoming too personal, and increase the reliability of the method. An instruction to this effect has been added to the identification procedure presented in Chapter 3.

In VISMIP, having grasped the meaning of an image on a referential and more general and abstract level, analysts should continue to reconstruct the message that the image serves to convey. The think aloud data provide us with evidence that this step also plays a distinct role in the minds of viewers. The following verbalization exemplifies a reconstruction of both the standpoint of the message and the arguments that support it:

(7) Participant 3
I I suspect that this is a uh way to [make clear to] nitwit pleasure boating people who of course haven't got a clue what kind of things can happen eh on the water, to make clear in a friendly way that they have to look out for the commercial traffic…,
because they sometimes have to do complicated maneuvers that as a lay person you don't see coming,
And that can be quite dangerous.

The participant in this example reacts to an illustration from a brochure by *Varen doe je Samen!* on safe recreational boating (Figure 5.4). It shows a cargo ship that pushes away a smaller ship. Both ships have human features (eyes, mouth, arms).

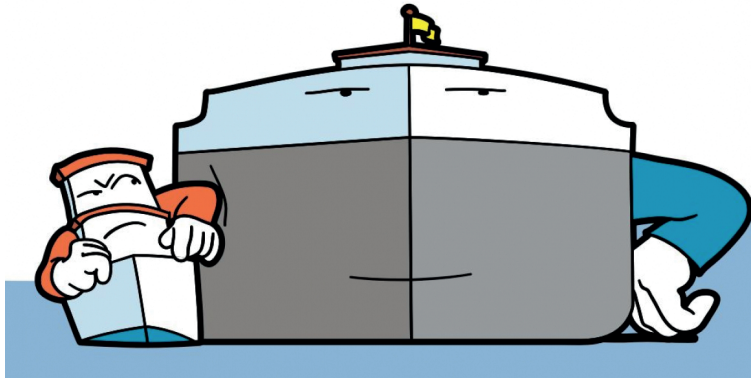


Figure 5.4 Illustration from a brochure by *Varen doe je Samen!* on safe recreational boating

First, the participant reconstructs the standpoint that is being advocated through the image. He does not express the standpoint explicitly as 'pleasure boats have to mind commercial traffic', but it is clear that he is forming a similar interpretation of the messenger's point of view. Then he goes on to reconstruct the argument to make this point of view more acceptable. He notes that 'because they [commercial traffic] sometimes have to do complicated maneuvers that as a lay person you don't see coming' and he continues 'and that can be quite dangerous'. While verbalizing this argument the participant uses hedges like 'I suspect', 'sometimes' and 'quite' to express his uncertainty about how he exactly should interpret the image and the corresponding message. This shows how carefully an interpretation can be constructed.

It should be noted that this participant's analysis is quite a model analysis of the message conveyed by the visual in the sense that both its standpoint and supporting arguments are reconstructed. What also happens is that viewers try to reconstruct only the standpoint, similar to what happens in the following excerpt:

(8) Participant 24
But at any rate that black pig's snout,
Death as a pig in this print,
Perhaps someone meant kinda: have a critical look at the clearing away of animals,
It can also be something like uh: we murder too many pigs,
On the other side the doctor is standing ready with one of those syringes.
Yes.

This verbalization refers to a cartoon by Tom Janssen (Figure 5.5), showing a doctor and a pig dressed as Old Father Time, with their backs turned to each other.

Figure 5.5 Cartoon by Tom Janssen

The participant presents alternative interpretations of the claim that is presumably defended by the cartoonist. At the same time, the example also shows how participants may have great doubts about what they see and how to interpret it. The first interpretation is expressed in 'have a critical look at the clearing away of animals', introduced by 'perhaps someone meant kinda', and the second interpretation can be found in: 'it can also be something like er: we murder too many pigs'. Any difficulties arising with the integration of new material into the current situation may cause the reader to consider explicitly the structure of both the new information and the current state of his or her knowledge (Steen 1994). Moreover, the participant does not try to reconstruct any supporting arguments that may be expressed by the image (e.g., 'being critical of the clearing away of animals prevents the fall of humanity'); at least, such an attempt is not verbalized. This example also shows how the rope and scythe get interpreted as Old Father Time, following conventionalized symbolism, as we saw above.

The think aloud protocols also showed that participants frequently attempted to interpret the intended claim of the message but did not succeed in formulating it. The following verbalization illustrates this:

(9) Participant 6
You can imagine all kinds of things,
but it can also be a plant,
the whole thing a bit photoshopped,
and there'll probably be a message behind it,
but honestly I have to say that it is not entirely clear to me what story the photo is telling.
It does have warm colours.
But it is not entirely clear to me what the message or the mission of the picture is.
I want to leave it at that.

The participant reacts to an ad for a Gibson guitar where the guitar and a nuclear explosion in similar form are juxtaposed (Figure 5.6).

Figure 5.6 Advertisement for Gibson Pure guitars

The participant makes it very explicit that although he assumes the presence of an intended message ('there'll probably be a message behind it'), he does not entirely get it. First, he states: 'but honestly I have to say that it is not entirely clear to me what story the photo is telling'. And then he repeats: 'but it is not entirely clear to me what the message or the mission of the picture is'. In terms of visual metaphor identification and processing: he does not succeed in reconstructing the standpoint

to which the advertiser committed himself. The form of nuclear explosions reminds the participant of a plant and the color of the image attracts his attention, but it does not lead to a further interpretation; 'you can imagine all kinds of things'. This is a perfect illustration of the general point of Schilperoord's Chapter 2, that incongruous structure and incongruity resolution are two separate issues. What is added here is that incongruous structures may be too challenging to some viewers in some circumstances for them to successfully reach some form of resolution. Visual metaphor can thus be risky discourse.

A final observation has to do with the valence of the standpoint. In this regard, the participant may also arrive at a different standpoint than the one intended by the sender of the message. In Examples (3) and (4), we saw that the participants do not verbalize the intended standpoint that the canned apple is just as fresh and tasty as a normal apple. The first participant instead states that 'that nice apple is tinned, which means it is ultimately less nice', nor does the second participant want to ascribe positive evaluations to the canned fruit. This makes it clear that in this case, the participants do not arrive at the intended conclusion. However, for analysts this would not be a problem, as they are instructed to think only about the intended message and since in advertisements the product will always be promoted by the messenger, it will thus be evaluated positively.

These think aloud data hence show at least three things: participants may reconstruct both standpoint and supporting arguments (see Example (7)), they may limit themselves to interpreting the standpoint only (see Example (8)), and they may not at all succeed in reconstructing the argumentation underlying the image (see Example (9)). These observations can also be related to the visual metaphor identification procedure. In VISMIP, analysts ideally are to reconstruct standpoint and supporting arguments, but they can limit themselves to reconstructing the standpoint for practical reasons. What is more, analysts might be unsuccessful in reconstructing any standpoints, in other words, even a simplified version of message analysis would sometimes not be successfully executed. If this happens, the analysis should be stopped. After all, if the standpoint of a message cannot be reconstructed, the topic of the image as a whole cannot be established either. And if the main topic of a message cannot be established, it is impossible to analyze the visual incongruity in terms of what is incongruous with what (cf. Schilperoord this vol.).

The last aspect aimed at grasping the general meaning of an image is hence to determine what the message is about. The think aloud data reveal some clear examples of viewers who try to determine the topic. While some participants start their verbalization by stating the topic of the picture, frequently the topic follows from the description of the image.

(10) Participant 21
The Pope.
Is walking on the tightrope
Between Jerusalem and Palestine.
And it is about the… probably about the building on uh the bank,
The West Bank, I believe.

Example (10) is a response to the cartoon by Tom Janssen, showing the Holy Pope as a tightrope walker (Figure 5.3). The first step of the participant is to describe what he sees in the image ('The Pope. Is walking on the tightrope. Between Jerusalem and Palestine'). This corresponds to the identification of the referents described above. Then the participant starts to think about the topic of the image and concludes that the image is 'probably about the building on the eh the bank'. In terms of our previous discussion, the participant skips thinking about more general and abstract meaning as well as about any underlying argumentation to proceed immediately to the image's topic, indicated by 'And it is about the…'. In a systematic visual metaphor identification procedure, analysts are asked to make explicit each step, not proceeding immediately from a description of the visual to the topic of the image. If the think aloud data are indicative, these intermediate steps may be optional or harder to perform.

The data also show that different participants may derive different topics from one and the same image. This confirms that it is necessary to make the analysis of the topic explicit in the identification procedure. Consider, for instance, the following verbalizations:

(11) Participant 14
Erm…
Is about the banks…
The banking crisis, I think.

(12) Participant 11
This is about uh –
The next one is about uh… a hospital

Both verbalizations refer to a cartoon of a hospital scene by Tom Janssen (Figure 5.7). A patient, lying in a hospital bed, screams 'where is my bonus!!' and the doctor's comment is '…signs of recovery..!', Participants 14 and 11 however disagree about the topic of the cartoon.

The participant in Example (11) thinks that the cartoon is about banks, or the banking crisis, whereas the participant in Example (12) thinks that the cartoon is about a hospital. Their disagreement can be explained by their different use of the

Figure 5.7 Cartoon by Tom Janssen

text in the image: the first participant may have involved the speech balloons in determining the main topic, whereas the second participant may not. It seems as if the second participant has simply derived the topic from the referential meaning (I see a hospital scene, so that is what this cartoon is about). It is also possible, though, that the second participant used the phrase 'is about', but in fact meant 'shows'. In other words, it may also have been a way of talking about referential meaning.

In addition, the first participant may have had the previous knowledge needed to derive the main topic, whereas the second may not (or did not access it on time). This example shows how for one person the visual dominance of the source domain gets the upper hand in his or her interpretation, which is too superficial, whereas another person may see through this. Here, the complexity and risk of visual metaphors in daily, quick use is obvious. Yet this should not be an issue for analysts, as they examine the image as part of a professional, academic task.

Determining the image's topic may also be truly problematic for viewers, as the following excerpt shows:

(13) Participant 8
Yes,
This is a bit more difficult.
Oh,
This is about swine flu, I think.
Or not?
What's this about?
I wouldn't –
Actually I wouldn't know.

This verbalization refers to Tom Janssen's cartoon depicting the doctor and the pig dressed as Old Father Time (Figure 5.5). The participant first says that the image is about swine flu, but is uncertain about this interpretation ('I think', 'or not?', 'what's this about?'). He eventually admits that he does not know what the image is about and does not understand the author's intended message. This example shows how participants frequently search for a confirmation of the first idea they have. When they do not immediately find it, they start doubting and even say that they do not understand the picture. A procedure for metaphor identification, by comparison, can help analysts to identify the topic: it is to be derived from the standpoint. If analysts succeed in reconstructing the point of a message, it should also be possible for them to determine the topic.

Summarizing, participants do not only determine the different visual referents (and their properties) in an image, the more general and abstract meaning, the point of the related message and its support, but they also try to establish the topic. This general understanding forms the basis for finding any incongruous elements – visual elements that do not fit the topic of the image.

5.4 Units of analysis and finding incongruity

In principle, analysts need to divide their description of the image's referential meaning into smaller units. They could for instance write: *Tall, blond man and spotted dog walk quickly in a park.* Or: *Smiling girl gives green apple to frowning boy under tree.* This could be organized even more tightly by the following conventions: *[Agent(man|tall, blond) Agent(dog|spotted) Action(walk|quickly) Setting(in park)].* And: *[Agent(girl|smiling) Action(give) Object(apple|green) Recipient(boy|frowning) Setting(under tree).* For analysis, it is necessary to separate descriptive phrases into smaller parts in order to systematically decide for each of these parts whether they are incongruous or not (see also Schilperoord, this vol.).

However, as might be expected, the think aloud data did not show explicit instances of this way of unitizing. It would be quite unnatural for participants to start parsing the meaning they had just described. Nevertheless, verbalizations most of the time exist of single words or short phrases, naturally dividing descriptions into small units. Moreover, verbalizations sometimes do show a certain structure of the referential description. While Example (1) shows a description in terms of objects only, the reaction in Example (2) includes various visual elements: an agent ('a boy'), an action ('holding a uh pistol to his head'), and a setting against which the action takes place ('With in the background a uh pawn. And a bare space with uh black and white tiles'). This example shows how participants eventually do structure their descriptive phrases, by using subjects, verbs and objects in their sentences.

In other examples, verbs are rare, for example in the next excerpt, about 'Portrait de Mrs. Isabel Styler-Tas', by the Spanish surrealist painter Dalí (Figure 5.8).

Figure 5.8 'Portrait de Mrs. Isabel Styler-Tas' by Dalí

(14) Participant 19
With on the right a woman,
nice and dignified.
What's she got?
A sort of medallion around,
with a tree sticking out.
And on the left you see a kind of image of her,
but made of nature.
She looks like ice,

In this example, we see how the participant describes the painting without using many verbs. Still, we can parse the information and structure the descriptive phrases. The participant indicates settings ('on the right', 'on the left'), agents ('a woman' 'a kind of image of her'), and provides a lot of descriptions ('nice and dignified', 'a sort medallion around', 'a tree sticking out', 'made of nature', 'she looks like ice').

This process of structuring referential meaning is a prerequisite for finding incongruity. However, the think aloud data revealed few verbalizations that related to incongruity identification. This observation was not expected, as it follows from previous research into visual metaphor advertisement processing that consumers would use deviation from expectation as a cue to start thinking about possible metaphorical interpretations (McQuarrie & Mick 1999; Phillips 1997; Phillips 2003; cf. Šorm & Steen 2013, p. 3). The participants in our study may have used deviation

from expectation, but then very few actually verbalized it. In the next verbalizations incongruity was identified:

(15) Participant 3
It is also a little pig with a burka on,
that's totally weird…
What are you supposed to do with that?

(16) Participant 12
A weird picture,
a little pig in uh disguise
and a big uh doctor.

These two verbalizations refer to the cartoon about swine flu (Figure 5.5). In the verbalizations, the qualification 'weird' is given to the pictures. Unfortunately, the participants do not clarify their explicit appreciation any further by what it is exactly that makes the image 'weird'. Is it because what they recognize as a 'burka' (Example (15)) and a 'disguise' (Example (16)) is not in harmony with the topic of the image (presumably 'swine flu'), or simply because pigs do not usually wear a burka or a disguise?

In an identification procedure, this difference must be made explicit. We may assume that visual units may be incongruous for two different reasons: because they do not match the main topic, or because they possess unusual properties. The first type of incongruity corresponds to Schilperoord's (this vol.) two-domain incongruities, and the second to one-domain incongruities. These two particular examples also show that more than one interpretation of an image is possible: processing of one and the same picture by different participants can lead to interpretational differences, turning the cloak into a burka, a disguise or, Old Father Time (Example (8)).

In the next examples, the participants verbalize more information on the incongruity found within the image:

(17) Participant 19
Erm…
Fireplace with an Ambi Pur in it instead of fire.

(18) Participant 20
That thing, that is uh –
Yes,
that's something that you stick in the socket
and then leaves a nice smell or something.
That's in the middle of the… chimney-piece,
the space behind,
where normally the… call it the [sic] hole for a hearth or something like that would be.

These are two different responses to an advertisement for Ambi Pur Aromatic Wood air fresheners. Figure 5.9 shows a fireplace of which the wood and fire are replaced by an air freshener. Recognizing the product seems an important step of understanding the advertisement, making the connection between the scent of a fireplace and the air freshener.

The participant in the first example (Example (17)) establishes that the fireplace shows the promoted product instead of a component that would have been more typical ('fire'). The participant in the second example notices that the air freshener is located somewhere where it is not typically located ('in the middle of the… chimney-piece, the space behind, where normally the hole for a hearth would be'). However, the participant first identifies the location where the product ('that thing') should be placed ('that is something that you stick in the socket'), after which he mentions the use of the product ('and then leaves a nice smell'). In this case the think aloud data clearly show the existing processing problem (Steen, 1994) and reveal aspects of the incongruity structures described in Chapter 2 by Schilperoord (this vol.).

An even more explicit verbalization of incongruity can be found in the next example about the Del Monte advertisement (Figure 5.2).

(19) Participant 12
A really painful picture actually
Because it uh just like a –
Yes,
A strange object is being pushed into something alive

In this example, the participant explicitly marks the incongruity by calling the tab 'a strange object', that 'is being pushed into something alive'. He also describes how the incongruity leads to a 'painful' picture. This way, the participant seems to say that it is not right, or incongruous.

Interestingly, what is found or marked as incongruous seems to depend on the genre of the picture. Example (20) includes a verbalization of the painting by Dalí (Figure 5.8).

(20) Participant 24
There is a tree growing out of her bosom
Erm…
If you look closely,
you see more and more appearing in the drawing.
A kind of rock as well,
that woman that she is talking to,
a kind of road through a rock.

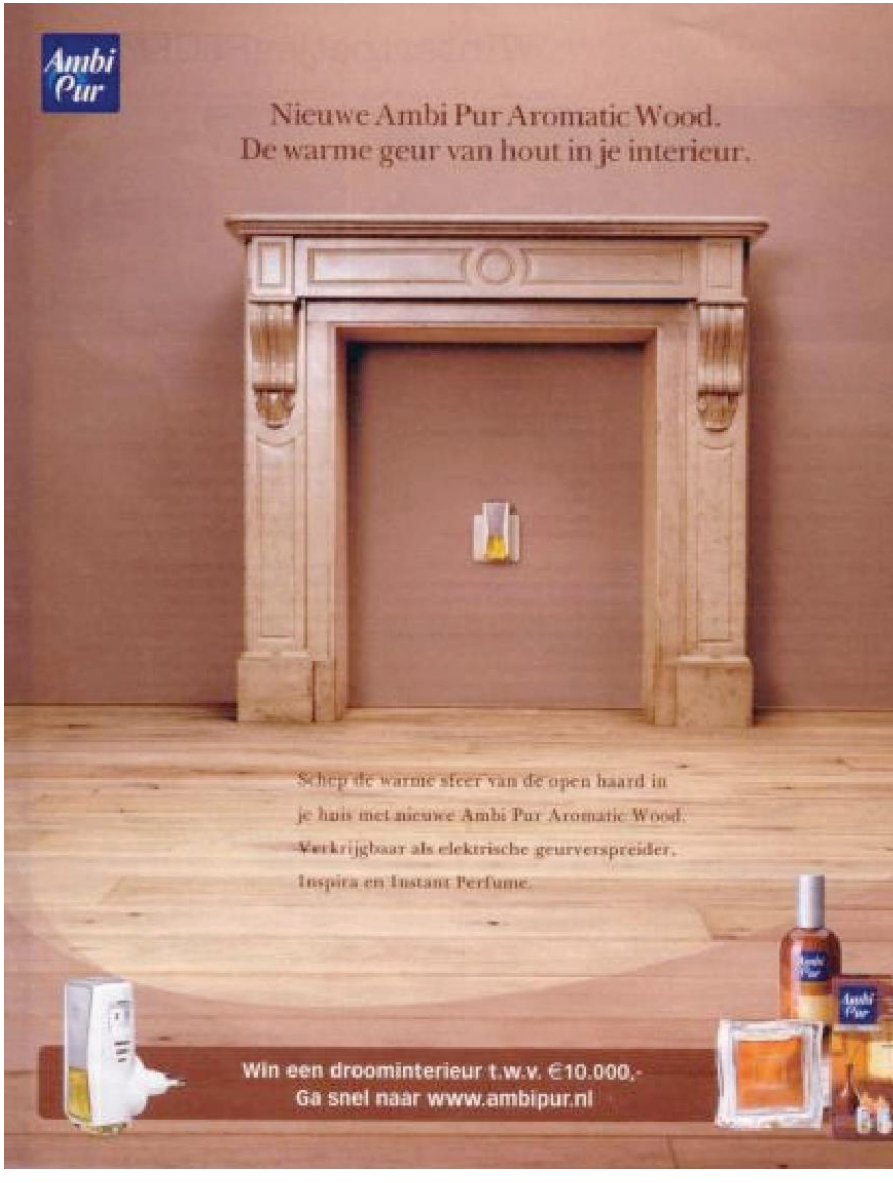


Figure 5.9 Advertisement for Ambi Pur air fresheners

The participant describes the picture and its incongruities: 'There is a tree growing out of her bosom', 'A kind of rock as well, that woman that she is talking to, a kind of road through a rock'. These are all descriptions of properties that are property-incongruous. However, this does not seem to be notable for this picture, since incongruity seems to be an expected characteristic of the surrealist painting. We can thus conclude that incongruity forms part of the mental steps people undertake when they see a visual metaphor, but depending on the participant and the genre, it is also marked as such.

5.5 Testing the need for comparison and testing non-literal meaning

At a certain moment, the participants need to take the next step and resolve the metaphorical puzzle and understand what is actually happening in the picture, frequently referred to as 'what the author means'. The think aloud data show that besides the identification of incongruities, viewers are concerned with the need for some kind of similarity or comparison to help them get the message of the picture. Analysts will eventually have to do this, too (Šorm & Steen this vol.).

In the Ambi Pur example (Example (19)), the participant already states that where the product is located, we would normally find the hole of the fire place. This will have to be included in our metaphor identification procedure, just like the related operation verbalized in Example (21).

(21) Participant 18
And she has a nice compass or something like that around her neck,
and she has probably been a rich woman.
And on the other side of the woman there is a kind of stone image of her.

Here, the participant makes clear how the picture (Figure 5.8) depicts a woman on one side and on the other side 'there is a kind of stone image of her'. Comparing the woman on the left and right and their (incongruous) properties shows how this participant tries to integrate the incongruous units within the overall topic of the picture.

A clear comparison is also made in Example (22), where the participant focuses on the resemblance in shape between the mushroom and the guitar (Figure 5.6) and with 'the nice thing is of course that they look like each other' and 'here of course it is I think in this representation about', he states how comparison is the means through which the message of the picture has to be unraveled.

(22) Participant 14
And uh… the nice thing is of course that they look like each other,
so the contours of a eh… uh a guitar
And and those of uh the mushroom,

The exploding… –
The explosion.
The representation of it.
And uh here of course it is I think in this representation about

These examples show how participants frequently verbalize their process of comparison. The think aloud protocols clarify the strategies and knowledge employed by viewers to arrive at a solution, after shedding some light upon the latter's representation (Steen 1994).

The next step then focuses on how the found comparison eventually leads to a further interpretation of the message of the picture, being nonliteral or cross-domain. This is also what analysts need to do at some point. It is a step that involves a rather straightforward yes or no question for analysts. Participants however do not make explicit whether a comparison is nonliteral or cross-domain. Nevertheless, from the think aloud data we can observe how participants interpret comparisons, by looking at the explanation participants give concerning nonliteral or cross-domain mappings. Interestingly, the next example shows how participants can also interpret images in a sheer literal way.

(23) Participant 16
(Reads title out aloud:) Checkmate,
(Reads title out aloud:) Checkmate.
Well,
this man he uh has lost,
he's checkmate.
And he has probably made a bet.
Like: if I lose,
then I will shoot myself in the head.

The corresponding picture (Figure 5.1) shows different elements that can be related to chess: the tiles, the (falling) pawn and the title of the image 'checkmate'. The incongruous element of the boy with the gun can be made sense of by also connecting it to chess and losing the game. The ultimate step in this argument is either interpreting losing as literally or nonliterally losing the game. The participant in Example (23) makes clear that he opts for the first interpretation: 'he has probably made a bet. Like: if I lose, then I will shoot myself in the head'.

5.6 Potential explanation by cross-domain mapping

The final move in interpreting visual metaphor is to test whether the comparison(s) can be seen as some form of indirect discourse about the topic. This final move shows how a process of metaphor interpretation builds on the findings in the previous steps. Example (24) shows how the participant of Example (22) continues with an interpretation of the found comparison, verbalizing his thoughts on how the comparison forms a cross-domain message about the topic 'guitar music' (Figure 5.6).

(24) Participant 14
That eh… that eh music,
Or perhaps eh guitar music,
that that a uh… a uh… can uh do something to you,
eh.
So that that can cause a sensation which uh… well that sometimes {…} ecstasy in people –
Music can make ecstasy clear in people.
and uh I think that that is kinda the idea behind this image.

Here, the participant uses the comparison to say something about the guitar music in terms of the sensation or ecstasy that it can cause. He clearly understands that the comparison aims at a nonliteral message and verbalizes how the comparison combines the different domains into an indirect discourse about the topic. In the following example about the same picture, the participant does not relate the guitar music to a sensation or ecstasy, but states that the guitar probably has nuclear powers. This is probably the result of a different interpretation of the more general, abstract meaning, the point and support of the message, and its topic; all of this goes to show how previous steps are of great importance for the eventual interpretation of the picture.

(25) Participant 6
That is a photo with a guitar
And uh the model next to it is a uh kind of organic, plantlike thing,
But there is also a mushroom cloud from an atom bomb in it,
So probably it is a guitar with atomic power

As a matter of fact, the final step is not always verbalized by participants. Thinking out loud, participants usually describe the image, give an evaluation, and sometimes come up with a more general, abstract meaning. An explanation or conclusion is not frequently explicated.

An interesting exception can be seen in Example (26), in which the participant describes Figure 5.10. This is an illustration from a brochure by *RIVM* (the Dutch

National Institute for Public Health and the Environment) that describes how bacteria can get resistant to antibiotics.

Figure 5.10 Illustration from a brochure by *RIVM* on antibiotic resistance

(26) Participant 4
and so that is uh a really nice image,
the bacteria as a kind of eh yes –
I actually see it as a kind of soldiers on the attack,
And the antibiotics antibiotic, so that has to b-beat them…,
yes,
they have become resistant,
and can defend themselves against the antibiotics.

Here, after giving an evaluation ('that is uh a really nice image'), the participant expresses an attempt to understand the metaphor in the illustration by using words belonging to both the target domain ('resistance to antibiotics') and the source domain ('a war'): 'the *bacteria* as a kind of eh yes – I actually see it as a kind of *soldiers* on the *attack*'. In this case, the participant lets his creativity run free to integrate the metaphor in his discourse about the topic. It shows how participants think about similarity in conceptual structure between units functioning as source domain and units functioning as target domain. Eventually, he comes to solve the 'puzzle', and formulates the message of the image. This way, thinking out loud sometimes reflects the processing occurring after the 'click of comprehension'.

Most participants do not verbalize an explicit conclusion in which they state if the image includes a metaphor or not, which seems natural in a task that does not explicitly focus on metaphor identification. Still, participants do mark some visual units explicitly for metaphor, however not always understanding what this term exactly involves. Participants mostly talk about metaphors when they think something has a double, 'symbolic' meaning or has a bigger message. Participant 15 for example describes the cartoon from Figure 5.3 as 'a beautiful piece of symbolism' and 'a metaphor for … the whole secularization':

(27) Participant 15
And I think it is a beautiful uh piece of symbolism.
You could also call it a metaphor for uhm that in this society –
Just the whole secularization

In Example (28), we find a similar marking of metaphor. The participant here sees the phrase 'give each other space' as a message applying to the maritime field, but also to everyday life. Since the Netherlands is densely populated, this phrase becomes applicable to this more general situation, too.

(28) Participant 20
And a uh well known maritime term: (reads headline:) 'give each other space'.
Well,
that [sic] can be another message in there,
a kind of metaphor for people in everyday life.
(reads headline:) Give each other space.
In a chock-block country like the Netherlands I presume.

In Example (26), we see all of the previous three steps: Test whether the incongruous units are to be integrated within the overall topical framework by means of some form of comparison"; "Test whether the comparison(s) is/are nonliteral or cross-domain"; and "Test whether the comparison(s) can be seen as some form of indirect discourse about the topic as formulated under step 1d". As a result, the incongruous properties of the bacteria can be integrated within the overall topical framework of 'resistance to antibiotics' by means of comparing it to 'people fighting a war'. This comparison connects the target domain 'resistance to antibiotics' with the source domain 'war' and this can be seen as indirect discourse about 'resistance to antibiotics', bacteria fighting the antibiotics with human characteristics. For analysts, this would lead to the last step of a procedure as the summary of the previous three steps. This makes it possible for analysts to conclude whether they can indeed identify the image as a visual metaphor.

5.7 Conclusion

The findings described in this chapter provide insight into what happens when people attempt to reach an understanding of visual metaphors. The results have implications for building a theory for visual metaphor processing. Moreover, the think aloud data provide behavioral evidence for the various parts of a visual metaphor identification procedure. The steps of this procedure can be related to the mental operations that non-experts undertake in visual metaphor processing of

pictures of various genre. This may offer a valid idea of how visual metaphor may be identified in analysis.

We can conclude that the following steps can be found in the think aloud data. Participants described the referents in the picture, gave a more general, abstract meaning, a point and the support of the point in the message, and the topic of the message. However, they evidently do not formulate this as clearly as an analyst is supposed to do. This also holds for their structuring of the descriptive phrases. The verbalizations do include the required information, but this is not as much of an analysis as would be asked from analysts. From here on, verbalizations are less frequent, suggesting that these steps may also be more difficult for analysts to perform or verbalize. Incongruity is occasionally marked as such and is occasionally integrated with the topic by means of comparison. And there is some indirect discourse about the topic. Yet participants do not mark metaphors explicitly as such, since this was not the aim of their task. They do sometimes talk about 'metaphor', but use this term to conclude that the picture has a message that has a double or bigger meaning.

We also found that some visual metaphors were more problematic for participants than others. One explanation can be found in the difference between the participants regarding extra-pictorial knowledge. Genre variation also plays an important role. Art for example seems to be more difficult to interpret than the other genres, and different interpretations of cartoons are not rare. Concerning advertisements, recognizing the product or the brand seems to significantly help participants formulate the message, as does the text for the illustrations from brochures.

The think aloud technique has served our purpose quite well: it yielded a detailed picture of real life visual metaphor processing. The data show how participants process the visual metaphors reflecting operations that will also have to be included in a visual metaphor identification procedure. However, one of the problems of the thinking aloud procedure is its interruptive and disturbing character. Only eight of the participants stated that they would perform the task exactly the same at home. Participants claimed to be more concentrated and focused during the task than they would be at home, with more eye for detail. However, they also felt less relaxed, found thinking out loud uncomfortable and evaluated it as unnatural. Participants sometimes indicated that they did not understand the image and stated that 'the idea is clear', indicating that they understood the image, but did not further verbalize it. This may also be due to the artificial setting, in which the participants may have had the idea that they were being watched and did not have unlimited time to look at the image as they would have during 'leisurely comprehension'. Moreover, since (some of the) processes may operate simultaneously, not all processes involved may be verbalized and participants may vary in the degree

to which they are aware of these distinct processes. Therefore, it is still open for discussion to what extent think aloud data reflect actual processing.

Apart from this, thinking out loud data may be rich, but they are not the most attractive data to work with from the point of view of reliable analysis. Verbal data in the form of introspective and retrospective reports constitute a controversial case as a form of scientifically acceptable evidence about psychological processes (László, Meutsch & Viehoff 1988). They are out of the control of the researcher, as it is the participant who determines the length, aptness, clarity, and complexity of the goods delivered (Steen 1991). In future research, therefore, the think aloud method may be complemented by online methods, such as eye tracking. This may also provide insight into the order of mental processes when looking at visual metaphor. Since our main purpose was to describe and illustrate the different processes that may validate our procedure for visual metaphor identification by analysts, however, we did not yet grasp the opportunity of systematically analyzing the order in which these processes typically occur.

References

Carlson, L., & Marcu, D. (2001). *Discourse tagging reference manual*. Retrieved from http://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf.

Carlson, L., Marcu, D., & Okurowski, M. E. (2001). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, Aalborg, Denmark. https://doi.org/10.3115/1118078.1118083

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (revised ed.). Cambridge, MA: MIT Press.

Haak, M. van den (2008). *A penny for your thoughts: Investigating the validity and reliability of think-aloud protocols for usability testing* (Doctoral dissertation). University of Twente.

Kintgen, E. R. (1983). *The Perception of Poetry*. Bloomington: Indiana University Press.

Laszlo, J., Meutsch, D., & Viehoff, R. (1988). Verbal reports as data in text comprehension research: An introduction, *Text:* 8(4), 283–294. https://doi.org/10.1515/text.1.1988.8.4.283

Longacre, R. (1983). *The grammar of discourse*. New York: Plenum Press.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8, 243–281. https://doi.org/10.1515/text.1.1988.8.3.243

McQuarrie, E. F., & Mick, D. G. (1999). Visual rhetoric in advertising: Text-interpretive, experimental, and reader-response analyses. *Journal of Consumer Research*, 26, 37–54. https://doi.org/10.1086/209549

Phillips, B. J. (1997). Thinking into it: Consumer interpretation of complex advertising images. *Journal of Advertising*, 26(2), 77–87. https://doi.org/10.1080/00913367.1997.10673524

Phillips, B. J. (2003). Understanding visual metaphor in advertising. In L. M. Scott & R. Batra (Eds.), *Persuasive imagery: A consumer response perspective* (pp. 297–310). Mahwah, NJ: Lawrence Erlbaum Associates.

Someren, M. W., van Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. London: Academic Press.

Šorm, E., & Steen, G. J. (2013). Processing visual metaphor: A study in thinking out loud. *Metaphor and the Social World*, 3(1): 1–34. https://doi.org/10.1075/msw.3.1.01sor

Šorm, E. & Steen, G. (this vol.). VISMIP: Towards a method for visual metaphor Identification.

Steen, G. J. (1991). The empirical study of literary reading: Methods of data collection. Poetics, 20(5–6), 559–575.

Steen, G. J. (1994). *Understanding metaphor in literature: An empirical approach*. London: Longman.

Steen, G. J., Dorst, L., Herrmann, B., Kaal, A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification in natural discourse: From MIP to MIPVU*. Amsterdam: John Benjamins. https://doi.org/10.1075/celcr.14

Verhagen, A. (1996). Tekstsegmentatie, onderschikking en subjectiviteit [Tekst segmentation, subordination and subjectivity]. *Gramma/TTT*, 5(3), 249–268.

CHAPTER 6

How visual form affects metaphoric conceptualization

The role of shape similarity

Lisanne Van Weelden[1], Fons Maes[2] and Joost Schilperoord[2]
[1] Universiteit Utrecht / [2] Tilburg University

6.1 Introduction

A common assumption that guides metaphor research is that metaphors "(…) establish correspondences between concepts from disparate domains of knowledge" (Bowdle & Gentner 2005, p. 193). An expression like 'a cigarette is like a bullet' invites readers to map features from the source domain (bullets) to the target or topic domain (cigarettes), like for example a bullet's potential to harm or even kill an organism. Metaphors thus convey that aspects of the source concept apply to the target concept as well. Compared to literal statements, like 'smoking cigarettes seriously damages your health', metaphorical statements have been shown to be superior "(…) instigators of conceptual change than are their literal counterparts" (Bowdle & Gentner 2005, p. 193). The interpretation of metaphorical statements proceeds by comparing the two entities to discover what meaning supports the correspondences (i.e., what is it about a cigarette that it can be compared to a bullet?) and to subsume the target entity 'cigarettes' under an ad hoc category 'things that kill you' signified by the source domain 'bullets' (Barsalou 1983).

Because metaphor is conceptual in nature, language may well be its most privileged mode of expression, but certainly not the only one. Over the past decades a considerable body of research has focused on possibilities to express metaphors in vision, that is, on visual metaphors (e.g., Forceville 1996; Maes & Schilperoord 2008; McQuarrie & Mick 1999; Phillips & McQuarrie 2004; van Mulken 2003). The goal of this chapter is to contribute to one line of research into visual metaphor: the study of the formal (pictorial, graphic) ways that are employed in order to establish correspondences between concepts from disparate domains. If, like verbal metaphors, visual metaphors strive at inviting viewers to map features from source to target, then how are the two domains combined in terms of pictorial cues?

https://doi.org/10.1075/celcr.18.06van

6.2 Formal differences between verbal and visual metaphor

A straightforward way of verbally combining two metaphoric domains consists in simply juxtaposing the two relevant terms (e.g., 'cigarette – bullet'). Apart from that, specific syntactic constructions seem well suited to express metaphor, in particular the copula ('a cigarette is a bullet'), and the comparative construction ('a cigarette is like a bullet'), called metaphor and simile, respectively. If metaphor can be expressed visually, then what might be the visual counterpart of these syntactic constructions? Different visual structures have been suggested as visual counterparts of the 'X is (like) Y' construction (e.g., Forceville 1996; Maes & Schilperoord 2008; Phillips & McQuarrie 2004; Schilperoord, Maes & Ferdinandusse 2009; Schilperoord, this volume; Van Mulken 2003). Following Phillips and McQuarrie (2004), we see three basic visual structures: juxtapositions (placing objects side-by-side), fusions (merging two objects into one hybrid object) and replacements (displaying one object within a visual context such that it evokes the second object). These three templates are exemplified by Figure 6.1a to c: the juxtaposition of cigarette and bullets (Figure 6.1a), the fusion of a cigarette and a knife (Figure 6.1b) and the replacement of tobacco by bullets (Figure 6.1c).[1]

Figure 6.1a Juxtaposition

Figure 6.1b Fusion

1. The images that appear here are reproduced for purely scientific reason and discussion and were all found in public domains, i.e. dedicated websites on the internet. We have done our best to locate all copy right holders via those websites and ask them for permission to reprint the images, but to little avail. We offer our apologies should we unknowingly have infringed the rights of copy right holders we haven't been able to locate properly.

Figure 6.1c Replacement

Note that all three images basically express the same metaphorical correspondence between cigarettes and weaponry. Interpretative differences, or differences in the amount of conceptual change evoked by these (and other) templates, may be said to be caused by formal differences. This, then, substantiates the idea that structural properties of visual metaphors can be considered independently from their contents.

6.3 Juxtaposition as a visual syntactic template

This chapter focuses on how juxtaposed objects can be visually arranged so as to suggest and support a metaphoric interpretation. Although juxtaposition is a basic template, it allows for considerable variation in how precisely objects from different domains of knowledge are juxtaposed. According to Schilperoord, Maes and Ferdinandusse (2009) object grouping is the psychological principle juxtaposition is based on: two or more entities are abstracted from their normal context and are aligned with respect to perceptual attributes such as size, shape, spatial orientation, and distance. In addition, the objects are usually depicted along a horizontal, vertical, or diagonal axis. Figure 6.1a exemplifies several of these juxtaposing techniques. The bullets and the cigarette are brought together in the artificial context of an advertisement, abstracted away from their natural environment. They are lined up at equal distances from each other, with similar spatial orientation, and projected upon an imaginary horizontal axis. In addition, the objects have more or less similar shape and size. The image furthermore appears to be almost two-dimensional, as all notions of depth have been removed.

Schilperoord et al. (2009) furthermore demonstrate that pictorial juxtaposition techniques comprise two broad classes. The first class involves the manipulation of the visual features of an object (the so-called 'object-constitutive' features, like size, shape, color, texture), while a second class involves the visual organization of objects to the effect of creating similarity, the so-called 'object-depictment' features, like the number of objects, projection onto a horizontal, vertical, or diagonal line, distance, distance from the viewing point, spatial orientation, similarity in shape, size or color. Schilperoord et al. claim that these techniques aim at expressing similarity

which in turn facilitates the mapping of metaphorical or associative relations. In the case of Figure 6.1a, these techniques direct the viewer to conceptualize cigarettes as members of the category of things that kill, typified by bullets. If this is accepted, conceptual effects may differ depending on what juxtaposing techniques are employed. The goal of this chapter is to put this idea to the experimental test for one of the visual techniques, often used in visual metaphor: shape similarity, that is, the manipulation of shape such that objects from different domains look alike.

6.4 Shape Similarity

The experiments reported in this chapter show that shape similarity (e.g. cigarette and bullets having the same shape in Figure 6.1a) suggests conceptual or functional similarity between the objects. The hypothesis that is tested is that shape similarity facilitates metaphoric processing more than when the technique of making objects look similar is not employed. How can we theoretically justify this prediction?

Humphreys and Forde's (2001) Hierarchical Interactive Theory (HIT) offers a detailed cognitive account of the relation between object shape and function. The theory describes how perceptual and semantic (i.e., conceptual) information interact in object identification. The theory posits three types of stored knowledge of objects: structural descriptions, semantic knowledge, and name representations. When we see an object, its visual features activate a structural description that captures information about the object's outer appearance, such as its shape, size, form, but does not include conceptual information. At the same time, structural descriptions of objects that look similar are activated. Thus, several visual features are activated, features that are unique to the object, but also features that the object shares with similarly looking objects. Structural descriptions spread activation to stored semantic knowledge of the object. According to the HIT model, competition arises between semantic knowledge of the object in question and semantic knowledge of similarly looking objects. The main tenet of HIT is that activated semantic knowledge feeds back to structural descriptions. Due to higher activation of semantic knowledge of the object in question, the activation of the correct structural description is reinforced, and suppresses activation levels of competing structural descriptions. So, first, visual information provides access to non-visual semantic information, and second, this semantic information reinforces visual information in object identification.

Although HIT only considers objects in isolation, its basic cognitive mechanism can be used to predict what happens when conceptual or metaphoric relations have to be found between two similarly or dissimilarly shaped objects. For the former case, HIT predicts considerable overlap between the activated structural

descriptions of the two objects. These overlapping visual features are therefore predicted to activate semantic knowledge that is relevant not only to each object separately, but also, reciprocally, to the 'other' object. In Figure 6.1a, structural descriptions of the two objects activate information that is relevant to both cigarettes and bullets. This, in turn, might facilitate finding correspondences based on semantic knowledge. In addition, finding perceptual correspondences might be stimulated as well, as the feedback from the semantic level to the structural description level is likely to reinforce the objects' perceived similarity in shape. If, on the other hand, the cigarettes and bullets had been depicted to look differently, they would have activated different structural descriptions. As a result, HIT would predict less direct activation of shared semantic knowledge, which, in turn might impede finding correspondences based on semantic knowledge. At least we may assume that finding correspondences would not be facilitated by (shared) structural descriptions. Furthermore, feedback from semantic information will reinforce shape dissimilarity, which also may make it harder to find conceptual correspondences. Under this view, perceptual similarity is employed to encourage mappings between objects stemming from disparate conceptual domains, not only 'attributive' mappings highlighted by overlapping structural descriptions, for example, cigarettes and bullets being both cylindrical, but also relational mappings, for example, cigarettes and bullets causing people to die earlier than they might hope (see Gentner & Clement 1988).

This is the general hypothesis that was tested in two studies that will be reported on in this chapter. Both studies aimed at finding evidence for our prediction that, compared to its absence, shape similarity facilitates finding both perceptual and conceptual relations between objects from different conceptual domains (Van Weelden, Maes, Schilperoord, & Cozijn 2011; Van Weelden, Maes, Schilperoord, Swerts 2012; Van Weelden 2013). We define shape as the outline of the picture of a particular object. In both studies we used pairs of line drawings of objects as stimuli.

6.5 Study 1

6.5.1 Experiment 1

The goal of the first set of experiments (reported more fully in Van Weelden et al. 2011) was to find evidence for the idea that shape similarity encourages viewers to find conceptual similarities between depicted objects. In these studies, we used twelve picture sets of utensils like the ones shown in Figure 6.2. Each set consisted of four object pairs. Each pair had an identical target object (shown right), but a different match object. Half of the pairs showed objects with the same shape (S+,

i.e., guitar/banjo and spoon/banjo), whereas the other half of the pairs showed objects with a different shape (S–, i.e., drum/banjo and sunglasses/banjo). In addition, half of the pairs showed objects that had a similar or comparable function, that is, for which a common category is immediately available (F+, i.e., guitar/banjo and drum/banjo), whereas the other half showed pairs of objects that do not have similar or comparable functions – that is, for which no such category is available (F–, i.e., spoon/banjo and sunglasses/banjo). Note that the latter type of object pairs constitutes the kind of images discussed earlier, i.e. objects originating from different domains of knowledge, which may be used in metaphorical relations. The objects' similarity or dissimilarity in shape was measured using a Shape Context Matching program, which uses the contours of shapes to measure their similarity (Belongie, Malik, & Puzicha 2002).

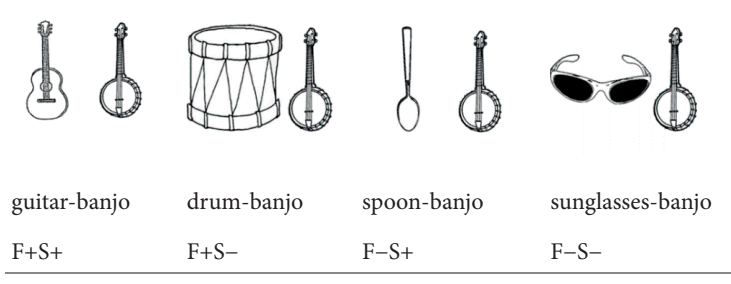


Figure 6.2 Example picture set with four picture pairs: Similar function and shape (F+S+), similar function and different shape (F+S–), different function and similar shape (F-S+), different function and shape (F–S–)

In the first experiment, participants were presented with the picture pairs and were invited to activate functional knowledge of the objects (i.e., a kind of semantic description as posited in HIT) by answering the question "Can you use these two objects for the same purpose?" This way, we intended to encourage participants to focus on the functional correspondences of the objects, without making them aware of any perceptual resemblances. Participants were to give a 'yes' or a 'no' response. The two objects were presented sequentially and the first object was presented for a duration of 50 ms only, just enough time to identify the object. In the analysis, we focused on reaction times and accuracy.

In case of the functionally similar pairs of objects, a common functional category is already available. The prediction based on HIT is that shape similarity will enhance activating this common category, whereas dissimilarity will at least temporarily block this activation. Therefore, we expect more and faster 'yes' responses for the similarly shaped object pairs (F+S+) than for the dissimilarly shaped object pairs (F+S–). For the functionally dissimilar pairs of objects, participants are

presented with a cognitive condition that is basically similar to the one involved in metaphors: two objects from disparate domains have to be functionally compared. 'Yes' responses, although strictly speaking incorrect, are therefore assumed to reflect processing that is similar to what happens in processing metaphors. The HIT-based prediction is that similarly shaped pairs (F−S+) will lead to fewer and slower produced correct 'no' responses as compared to non-similarly shaped pairs (F−S−).

The analysis showed that for the functionally similar object pairs, participants were indeed faster in indicating that the objects could be used for the same purpose when the objects had the same shape (F+S+) as compared to when they had different shapes (F+S−). They also produced fewer erroneous 'no' responses in the same shape condition (F+S+). For the functionally dissimilar object pairs, the results were also in agreement with the HIT-predictions: it took the participants more time to indicate that the objects could not be used for the same purpose when they had the same shape (F−S+) as compared to when they had different shapes (F−S−); and participants gave more incorrect affirmative responses to similarly shaped object pairs. This experiment hence shows that similarity in shape facilitates task performance for the functionally similar pairs, whereas it inhibits task performance for functionally dissimilar pairs. The latter seems to be the result of the participants getting confused by the fact that the objects' shape does not match their functional dissimilarity.

In sum, shape similarity makes it easier to see objects as functionally similar, and more difficult to see them as functionally dissimilar. Applied to 'functionally dissimilar' objects figuring together in the same metaphoric advertisement (as in Figure 6.1a), the results especially for the functionally different objects thus suggest that shape similarity may function as a bootstrap mechanism for the activation of metaphoric relations.

6.5.2 Experiment 2

Comparing functions of objects involves the search for properties that the compared objects have in common. Although the first experiment suggested this process to be stimulated by perceptual similarity, even if no readily common category was available, the experiment is silent to the question what kind of correspondences participants found between the two presented objects. Therefore, in a second experiment we investigated what kind of comparisons people produce when they compare objects that have similar or dissimilar shapes. If finding these correspondences is based on perceptual similarities, do people produce mainly perceptual (i.e. 'structural') correspondences or also relational (i.e. 'semantic') ones? In interpreting verbal metaphors, people are known to prefer structurally rich and systematic mappings over purely attributive ones (e.g., Gentner & Clement 1988). So far, however,

we know little about whether this preference also guides processing in case the output modality is visual. The conceptual nature of metaphor may suggest a similar preference in visual metaphor.

Using a production task, we aimed at getting insight into the mental operations that people undertake when they make comparisons between visually evoked objects. We presented participants with the task to describe one object of an object pair in terms of the second object, in such a way that in a (presumed) follow-up experiment another participant would be able to pick the object from a set of objects. In doing so, participants were only allowed to mention correspondences. The pairs of objects that were used (i.e., the to-be-described object and the source object) were the same as in the first experiment. We measured the number of mentioned (attributive and relational) correspondences, as well as the speech onset times. A correspondence was coded as 'attributive' if it described a similarity in shape or a part-whole structure that could be extracted from the pictures, such as 'both objects are round.' All other correspondences could not be derived from the visual presentation of the objects and were therefore coded as 'relational,' such as 'both objects are used to play sports with.'

The results of the analyses confirmed the expectations. Both functionally similar and dissimilar pairs (F+ and F−) produced larger number of correspondences if the objects had similar shape. Similarity in shape also had an effect on the type of the mentioned correspondences. That is, for the functionally dissimilar pairs (like banjo/spoon), the ones lacking an immediate common functional category, participants produced more attributive correspondences for the similarly shaped pairs, but they also produced more relational correspondences. The latter outcome shows that similarity in shape facilitates finding relational similarities for functionally dissimilar object pairs, like a banjo and a spoon. The fact that we did not find this effect for the functionally similar pairs (F+S+) suggests that perceptual similarity only stimulates finding a conceptual link if this link is not already present. This is precisely what is at stake if metaphorical relations are to be interpreted. Furthermore, we found higher speech onset times in case of the absence of a common category, as compared to when such a category was present. One of the produced correspondences was 'things that wealthy people use' for the picture pair of a hat and a chess tower. It thus follows that the delayed speech onset times we found for these pairs of objects could be the result of a process akin to metaphoric processing. In line with the suggestion that perceptual similarity only stimulates finding a conceptual link if this link is not already present, the analysis of the speech onset times showed reduced speech onset times as a result of shape similarity in case of conceptual dissimilarity. Again, such a result was absent in case of functional similarity.

All in all, shape similarity indeed seems to function as a bootstrap mechanism for the activation of metaphoric relations, between objects that have distinct

functions and for which no ready-made common category is available. That is, for these object combinations, more correspondences were produced (both attributive and relational) and the thinking time to do that decreased as a result of shape similarity. In HIT terminology, this shows that the overlap in the objects' shape, and thereby structural descriptions, activates shared semantic (i.e., conceptual) knowledge. Yet, when the semantic knowledge is already shared, then the additional overlap in shape does not particularly strengthen the semantic relation.

6.6 Study 2

Although the results of Study 1 seem to suggest that perceptual similarity facilitates the processing of visual metaphors, it must be noted that the object pairs that were used in this study were quite different from the object pairs we discussed in the introductory section. Therefore, in a follow-up study (see Van Weelden et al. 2012) we used pairs of objects that were inspired by real advertisements which should lead participants to engage in more natural metaphoric mappings, the kind of mappings known from real life metaphors. This way, we hoped to get results which are more generalizable to the processing of real visual metaphors.

We used fourteen picture sets like the one shown in Figure 6.3. We left the conceptually similar pairs out in this study. Each set consisted of two object pairs, both with the same target and source object, but with the shape of the source object depicted either similar or dissimilar to the target object (e.g., the elephant and the forklift of Figure 6.3). All source objects were natural objects, whereas the target objects were artifacts.

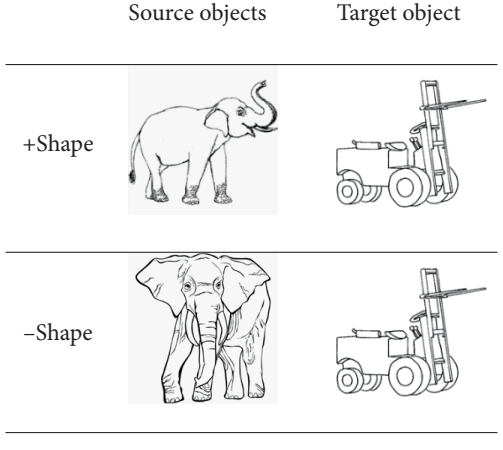


Figure 6.3 Example picture set with two picture pairs: +Shape and –Shape

Each experimental trial first showed participants the source object and then the target object. Participants were furthermore asked to conduct a source task and a target task. For the source task, participants were instructed that when the picture (the source) was presented, they had to produce as many characteristic features as they could within a 15 s time span. Subsequently, the second picture (the target) was presented. For the target task, they were instructed to mention as many correspondences between target and source picture as they could within a 20 s time span. We included the source task in the experiment to check and ensure whether the two shapes of the source pictures would activate the same features, and to investigate on which source features the correspondences were based. Because participants could only notice the shape (dis-)similarity if they were shown the second object, the experimental factor could only affect the target task. Statistical analyses focused on the number of mentioned features (source task) and correspondences (target task), the number of attributive and relational correspondences, and the speech onset times. The coding scheme was slightly different from the previous study. Gentner and Kurtz (2005) propose that entities can be assigned to a relational category or a property (i.e., attributive) category. To describe correspondences between members from a property category, concrete nouns can be used, for instance, 'they both have wings.' Correspondences between members from a relational category can be expressed by verbs, for example, "they both accelerate very fast." The latter contrast is expressed in language in a fairly straightforward way. When expressing a correspondence between two objects, concrete nouns are often preceded by (conjugations of) the auxiliary verbs to be or to have. Accordingly, the presence of auxiliary verbs can serve as a cue for property correspondences. So, correspondences such as "both are round" and "they both have wings" were coded as property (i.e., attributive) correspondences. The presence of action verbs can indicate relational correspondences. Hence, we coded "this actually protects your head and the shell protects as well" and "they both spin" as relational correspondences.

The analysis of source features showed no effect of shape similarity. This indicates that the perceptual difference between the two source pictures did not lead to differences in the number of features mentioned. The analyses showed that participants were faster in finding and producing a correspondence when the objects had similar shapes, as compared to when the objects did not look alike. Similarity in shape resulted in the production of more correspondences between the objects as well. Regarding the nature of the produced correspondences, there were no differences between the proportions of both the attributive and relational correspondences for the similarly and dissimilarly shaped object pairs. Hence, participants found both more attributive and relational correspondences.

Thus, similar to the second experiment of the first study, this study showed that similarity in shape facilitates finding correspondences between two objects from

different domains of knowledge. As the object pairs used in this study were highly alike to the object pairs we encounter in advertisements, our findings suggest that finding a metaphorical relation for the advertisement presented in Figure 6.4a is easier as compared to the one presented in Figure 6.4b.

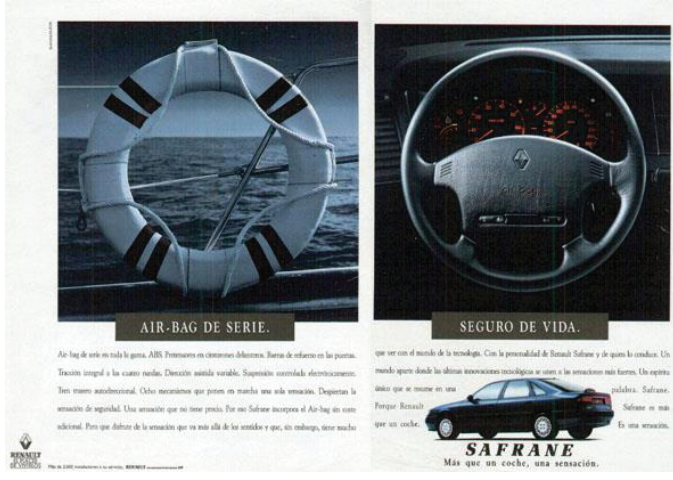


Figure 6.4a An advertisement for a car brand

Figure 6.4b An advertisement for a car brand (adjusted)

In the advertisement presented in Figure 6.4a, a lifebelt is juxtaposed to an airbag-equipped steering wheel to communicate the message that an airbag-equipped steering wheel has the ability to save someone's life in case of an accident, just like

a lifebelt. Our studies show that the fact that both objects are round facilitates finding a metaphorical relation between the lifebelt and the airbag-equipped steering wheel. The message of the advertisement presented in Figure 6.4b is also that an airbag-equipped steering wheel has the ability to save someone's life. However, this time the steering wheel is juxtaposed to a rescue can. As the shape of the steering wheel and the shape of the rescue can are dissimilar, it is more challenging to find a metaphoric relation between the two.

6.7 Conclusion

The two studies offer evidence for the idea that shape similarity of juxtaposed objects can be seen as a visual template which facilitates the construction of metaphorical thought. When asked to carry out a 'literal' categorization task (Study 1) shape similarity appeared to inhibit performance especially in the case in which the two objects were similar in shape but different in function. Apparently, shape similarity of juxtaposed objects provides viewers with an initial bias to consider these objects also as conceptually equivalent. This suggestion is confirmed by the reduced onset time in the production experiment of Study 1: when asked to produce correspondences for functionally dissimilar objects, the onset times were lower for similarly shaped objects, which again suggests that shape similarity functions as a bootstrap mechanism. But the production of correspondences shows that shape similarity has an effect on later cognitive stages as well: respondents were able to produce more correspondences with same shape objects, not only correspondences directly suggested by the shape of the objects, but also conceptually based, relational correspondences. This result accords with the bidirectional relationship between structural and semantic representations as suggested by Humphrey and Forde's Hierarchical Interactive Theory. The results were confirmed when respondents were asked to produce correspondences between two juxtaposed objects for which fairly simple metaphoric mappings could be found (Study 2). Also in this study, shape similarity resulted in faster and more correspondences, attributive and relational ones.

The juxtaposition of similarly shaped objects can be seen as the visual counterpart of the verbal 'X is (like) Y' construction. In the two cases, the interpretation is guided by the detection of a conceptual mismatch between two concepts which are 'syntactically' presented as being identical. In language, identity is suggested by the conventional be-sentence; in the visual modality, shape similarity can be used to that purpose, and the results of the two studies show that shape indeed offers the appropriate metaphoric mindset. As such, it can be seen as contributing to ideas found elsewhere in the literature, for example the idea that visual juxtaposition is

to be seen as variant of a verbal simile (X is like Y) and visual fusion as a variant of a verbal metaphor (X is Y) (Forceville 1996).

Such direct comparisons between the verbal and visual mode should be handled with care, though. The verbal syntax of metaphor and simile carries a much more conventionalized meaning than the visual counterparts. A copular sentence connecting different domain concepts reads as a compelling instruction to create metaphoric mappings. Juxtaposed or fused objects do not do more than suggest such an interpretation.

Furthermore, formally, simile and metaphor statements differ in the use of one single word, all other things being exactly the same. The formal difference between juxtapositions and fusions are huge: juxtaposing objects does not affect the integrity of object representations, fusing them results in two objects being presented only partially, and being integrated into a non-existing new object. These differences have serious processing implications. Juxtaposing objects from different domains does not affect the visual recognition of the individual objects, and allows direct access to the individual representations of the concepts, which can then be conceptually compared. Fused objects are typically subject to the so-called bizarreness effect, or the secondary distinctiveness effect, in which objects are not distinctive with respect to their visual environment ('primary distinctiveness') but with respect to the knowledge we have about objects (Schmidt 1991). Such distinctive objects are known to require additional processing, just to acquire an adequate recognition of the object in relation to existing mental representations of the fused objects (e.g., Gounden & Nicolas 2012; Michelon, Snyder, Buckner, McAvoy, & Zacks 2003). This difference in visual processing, independent of metaphorical conceptualization obfuscates the processing differences suggested or found between metaphorical juxtapositions and fusions in the literature. For example, Phillips and McQuarrie (2004) predict more cognitive elaboration and more or other perlocutionary effects for fusions compared to juxtapositions, but to our knowledge, this prediction has not been not validated in empirical studies (e.g., van Mulken, Le Pair, & Forceville 2010). And even if there would be found a predicted difference in processing complexity, this would not necessarily mean a difference in metaphor complexity.

Verbal and visual metaphoric templates also differ in the scope and bandwidth of concepts they are able to connect. Verbal templates can connect any dyad of concepts and domains as far as they can be molded in the format of a noun phrase. Also, they are perfectly able to connect concrete source and abstract target domains, combinations which are the core of metaphor. Visual representations are more limited and indirect in these capacities.

Finally, metaphor is ubiquitous in language. It is less clear to what extent it is ubiquitous in visual communication as well. Visual metaphor is predominantly studied in well-defined genres, with advertisements and cartoons as cases in point.

Many spontaneous forms of visual communication use visual techniques evoking correspondences based on primary metaphors, such as size differences suggesting differences in importance, spatial orderings suggesting hierarchical or temporal orderings, etc. Still, it is unclear if visual metaphor can be said to be ubiquitous the way metaphor in language is. As far as the use of the conventional templates such as *A is (like) B* is concerned, analytical studies prove that they only represent a small portion of metaphor in language (e.g., Steen 2007). One may think that all manifestations of visual metaphor somehow fit in the three structural types as presented in Figure 6.1. But even if this were the case, it would not be much more than a sketchy, unreliable map in the hands of scholars exploring the rich flora and fauna of visual metaphor in human communication.

References

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, 11(3), 211–227. https://doi.org/10.3758/BF03196968

Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape context. *IEEE Transactions on pattern analysis and machine intelligence*, 24(4), 509–522. https://doi.org/10.1109/34.993558

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, 112(1), 193–215. https://doi.org/10.1037/0033-295X.112.1.193

Forceville, C. (1996). *Pictorial metaphor in advertising*. London, UK: Routledge. https://doi.org/10.4324/9780203272305

Gentner, D., & Clement, C. (1988). Evidence for relational selectivity in the interpretation of analogy and metaphor. In G. H. Bower (Ed.) *The psychology of learning and motivation: Advances in research and theory* (pp. 307–358). New York, NY: Academic Press.

Gentner, D., & Kurtz, K. (2005). Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman & P. W. Wolff (Eds.), *Categorization inside and outside the laboratory* (pp. 151–175). Washington, DC: APA.

Gounden, Y. & Nicolas, S. (2012). The impact of processing time on the bizarreness and orthographic distinctiveness effects. *Scandinavian Journal of Psychology*, 53, 287–294. https://doi.org/10.1111/j.1467-9450.2012.00945.x

Humphreys, G. W., & Forde, E. M. E. (2001). Hierarchies, similarity, and interactivity in object recognition: "Category-specific" neuropsychological deficits. *Behavioral and Brain Sciences*, 24, 453–509.

Maes, A., & Schilperoord, J. (2008). Classifying visual rhetoric: Conceptual and structural heuristics. In E. F. McQuarrie & B. J. Phillips (Eds.), *Go figure new directions in advertising rhetoric* (pp. 227–257). New York/London: Armonk

McQuarrie, E. F., & Mick, D. G. (1999). Visual rhetoric in advertising: Text-interpretive, experimental and reader-response analyse. *The Journal of Consumer Research*, 26 (1), 37–54. https://doi.org/10.1086/209549

Michelon, P., Snyder, A. Z., Buckner, R. L., McAvoy, M., & Zacks, J. M. (2003). Neural correlates of incongruous visual information: An event-related fMRI study. *NeuroImage*, 19(4), 1612–1626. https://doi.org/10.1016/S1053-8119(03)00111-3

Phillips, B. J., & McQuarrie, E. F. (2004). Beyond visual metaphor: A new typology of visual rhetoric in advertising. *Marketing Theory*, 4(1/2), 113–136. https://doi.org/10.1177/1470593104044089

Schilperoord, J. (This vol.). Ways with pictures: Visual incongruities and metaphor.

Schilperoord, J., Maes, A., & Ferdinandusse, H. (2009). Perceptual and conceptual visual rhetoric: The case of symmetric object alignment. *Metaphor and Symbol*, 24, 155–173. https://doi.org/10.1080/10926480903028110

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory and Cognition*, 19(6), 523–542. https://doi.org/10.3758/BF03197149

Steen, G. (2007). *Finding Metaphor in Grammar and Usage*. Amsterdam: John Benjamins BV. https://doi.org/10.1075/celcr.10

Van Mulken, M. (2003). Analyzing rhetorical devices in print advertisements. *Document Design*, 4(2), 114–128. https://doi.org/10.1075/dd.4.2.02mul

Van Mulken, M., Le Pair, R. & Forceville, C. (2010). The impact of perceived complexity, deviation and comprehension on the appreciation of visual metaphor in advertising across three European countries. *Journal of Pragmatics*, 42, 3418–3430. https://doi.org/10.1016/j.pragma.2010.04.030

Van Weelden, L., Maes, A., Schilperoord, J., & Cozijn, R. (2011). The role of shape in comparing objects: How perceptual similarity may affect visual metaphor processing. *Metaphor and Symbol*, 26(4), 272–298. https://doi.org/10.1080/10926488.2011.609093

Van Weelden, L., Maes, A., Schilperoord, J., & Swerts, M. (2012). How object shape affects visual metaphor processing. *Experimental Psychology*, 59(6), 364–371. https://doi.org/10.1027/1618-3169/a000165

Van Weelden, L. (2013). *Metaphor in good shape*. Doctoral thesis, Tilburg University.

CHAPTER 7

Conventionality in visual metaphor

Didier Hodiamont, Hans Hoeken and Margot van Mulken
Independent / Universiteit Utrecht / Radboud Universiteit Nijmegen

7.1 Introduction

Metaphor has long been regarded as a relatively rare rhetorical figure, one that can only be established in language, and as a phenomenon that is mainly used by poets and literary authors. In the early eighties of the previous century, this view was intensively and successfully challenged (see e.g. Lakoff & Johnson 1980; Ortony 1979). Basically, the claim was made that metaphors are pervasive in our thinking and speaking, that language is only one mode in which metaphors can be expressed next to other modes, and that these expressions provide a window on fundamental human thought processes (see Steen 2007, for a review).

This view on metaphor as a cognitive phenomenon rather than a literary one, has inspired both theoretical and empirical research. For instance, the question to what extent metaphors can be expressed in modes other than the verbal mode alone has received a lot of research attention (see Muller & Cienki 2009). This is especially true for the expression and evaluation of metaphors in the visual mode (see e.g. Forceville 1996; Phillips & McQuarrie 2004; Maes & Schilperoord 2008; Van Mulken et al. 2010). The existence of visual equivalents of verbal metaphors provides interesting venues for testing fundamental claims about the mode-independency of the cognitive processing of metaphors.

In this chapter, we focus on one of the factors that has been documented to influence the processing of verbal metaphors, namely the extent to which a metaphor is perceived as conventional. Bowdle and Gentner (2005) have shown that a metaphor's conventionality plays an important role in the processing of metaphors with more conventional metaphors being processed in a qualitatively different way than more novel ones. If the cognitive processing of metaphors is largely independent of the mode in which they are expressed, the processing of visual metaphors should exhibit a similar sensitivity to conventionality. Furthermore, if the metaphorical relation is represented at a deeper conceptual level, then one would expect the perceived conventionality of a verbal metaphor to be similar to that of

https://doi.org/10.1075/celcr.18.07hod

its visual counterpart. This latter hypothesis is tested in a study described in this chapter. First, we will discuss Conceptual Metaphor Theory and the role ascribed to conventionality in the processing of metaphors.

7.2 Conceptual Metaphor Theory and the role of conventionality

Probably the most influential theory representing the cognitive turn in the study of metaphor is Lakoff and Johnson's Conceptual Metaphor Theory (1980; 1999). According to Conceptual Metaphor Theory (CMT), metaphor is primarily a mode of thought. The thought process entails the mapping of two (or more) conceptual domains. For instance, to communicate that a certain make of car is very fast, one may use the metaphor 'this car is a rocket'. In this metaphor, two domains are connected: the domain of the car and the domain of a rocket. The rocket domain acts as a source for characteristics that are relevant to the car domain, therefore the rocket is referred to as the source domain and the car as the target domain. To interpret the metaphor, a characteristic of the source, in this case: 'speed', is mapped onto the target resulting in the interpretation: 'this car is fast'. In more general terms: metaphoric constructions result in our understanding of the target domain becoming suffused with a prototypical characteristic of the source and changes as a result of this mapping process. So in a salient interpretation of this metaphor, this car is not literally a rocket, but it must be faster than just a regular car.

Lakoff (1993, p. 20) claims that metaphor is not primarily a figure of language and therefore it is not solely restricted to the verbal domain. Holyoak and Thagard (1995, p. 218) state in a similar vein that "metaphoric correspondences are not fundamentally between words, but rather between systems of concepts". This implies that the same metaphorical relations could be expressed via different representational systems, a claim which has been put forward by several scholars (see e.g. Forceville 2009; Müller & Cienki 2009).

There is ample evidence that similar metaphorical relations can be expressed both verbally and visually. One very common metaphoric relation is the one between ANGER and the idea of HOT FLUID IN A PRESSURISED CONTAINER[1] (Kövecses 2005). This relation is reflected in verbal expressions such as "he was boiling with anger" or "why don't you cool down a bit". With regard to the visual mode, Forceville (2005) has shown several representations of this same conceptual idea of anger in

1. Conceptual metaphors are traditionally described in capitals, but this does not mean that conceptual metaphors are primarily anchored verbally. These descriptions are only an attempt to capture the meaning of an abstract, conceptual idea into a lexicalised definition.

an Asterix comic book. The image of a person with steam coming out of his ears can be considered as an idiom in cartoon language, but also more challenging manifestations (like bulging eyes or jagged connecting lines) are considered to be motivated by the same conceptual idea of anger. Thus, it seems safe to conclude that the same metaphorical relation can be expressed in a verbal and a visual mode.

Visual metaphors are frequently encountered in both print advertising (Phillips & McQuarrie 2004; Van Enschot et al. 2008) and commercials (Van Enschot & Hoeken 2015). In Figure 7.1 two advertisements can be seen in which a car is compared to a rocket (Figure 7.1a) or to a dolphin (Figure 7.1b).

Hoeken, Swanepoel, Saal and Jansen (2009) argue that the popularity of metaphors (but also of other tropes such as ellipsis, irony) in advertising is the result of the fact that the same message (e.g., this car is fast) has to be communicated over and over again. If the advertisers were to use exactly the same message format again and again (e.g., this car is fast), consumers would grow tired of the message and would ignore it, because it does not contain any new information. By using different tropes (e.g., by comparing the car to a rocket, a leopard, a jet plane), advertisers can communicate the message in different forms thereby prohibiting the consumers to grow tired of the identical message claim.

The frequent occurrence of visual metaphors in genres such as cartoons, comic books and advertising, metaphors for which a verbal equivalent is present, confirms CMT's claim that the same metaphor can be expressed in different presentation modes. However, CMT goes further than that. If metaphor and metaphor processing is to be located at the conceptual level, visual metaphors should be processed in a similar way as verbal metaphors. Obviously, interpreting words is different from interpreting images. However, Quiroga et al. (2005) show that different representations (e.g. pictures of objects, or words referring to these objects) can activate the same subset of neurons, so the activation of these mappings may be relatively independent of the input modality. Likewise, the metaphorical relation between concepts could be established irrespective of whether the original input is in a verbal or in a visual format.

Although CMT claims that metaphors can be expressed in different modalities and processing should proceed modality independent, the empirical research has mainly been conducted employing verbal metaphors. Several scholars (e.g., McGlone 2007; Murphy 1997, p. 99) have called for research on the processing of metaphors in other modalities in order to assess the validity of the idea that manifestations of metaphors in different modalities are a result of the same underlying conceptual process. Visual metaphors provide an excellent opportunity for such a research enterprise.

Figure 7.1a

Figure 7.1b

Figure 7.1 Two car advertisements

7.3 Metaphor processing: The career of metaphor

Bowdle and Gentner's (2005) concept of 'career of metaphor' is an interesting approach to assess the claim that metaphor processing is independent of the modality in which the metaphor is presented. To establish a metaphorical relation, one has to select characteristic(s) in the source domain that are mapped onto the target domain. Two different processes have been suggested: comparison and categorisation. Comparison entails comparing the (relations between) characteristics in the source domain to similar (relations between) characteristics in the target domain. For instance, to interpret the metaphor 'this car is a dolphin', one has to compare characteristics of a dolphin (e.g., grey, smooth, elegant, mammal, intelligent) to what may be relevant characteristics of a car. This comparison process is more likely to result in the selection of 'smooth mode of transportation' or 'intelligent' as relevant characteristics than selecting 'mammal' or 'grey'.

On the other hand, comparing the two domains is not necessary when processing by categorisation. The source domain automatically evokes the relevant characteristic because this characteristic is more or less quintessential for this concept. For instance, one may consider the characteristic 'fast' as being prototypical for a rocket. When interpreting the metaphor 'this car is a rocket' through a categorisation process, no comparison is needed: the rocket more or less automatically evokes the characteristic 'fast'.

Whether the interpretation of a metaphor is the result of a comparison or a categorisation process depends to a large extent on the metaphor's conventionality (see, e.g., Giora 1997; Bowdle & Gentner 1999). Conventionality arises from a person's repeated exposure to, and interpretation of metaphors containing the same concept as the source. If one is being confronted with a certain metaphor for the first time, for instance, 'this helmet is a sea shell', this is considered a novel metaphor. The combination of the two domains is quite novel and unexpected, and one has to compare the two domains involved to reach a satisfying interpretation. A more conventional metaphor involves a concept that has repeatedly been used to communicate the same characteristic. For instance, the concept of 'pearl' is frequently used in advertising as a metaphor to communicate the 'preciousness' of a certain product. Cars, chocolate, liquor have all been compared to a pearl. As experience with the pearl metaphor grows, it becomes easier to interpret the intended message as 'this product is precious'.

Metaphor conventionality is thus not static, but a factor that changes through a cognitive learning processes. When a certain concept is used very often as the source of a metaphor, it can become prototypical for a certain category. The alignment with its abstract, categorical meaning can become salient to the extent that it is included in dictionaries. In English and Dutch dictionaries lemmas like 'pearl'

or 'magnet' are not only described literally, the figurative meaning is mentioned as well. In certain cases, this figurative meaning of the concept may become even more salient than its original meaning. Metaphorical expressions such as 'our relationship has stranded' may no longer evoke thoughts about boats and shallow waters. Such highly conventional metaphoric expressions are often perceived as if they were literal, because people are no longer aware that two different domains are being compared (Lakoff & Johnson 1980, p. 5; Bowdle & Gentner 2005).

In some extraordinary cases, the metaphoric meaning replaces the original literal meaning and becomes a so-called 'dead' metaphor. An example is the 'blockbuster' metaphor (Bowdle & Gentner 2005, p. 209). The literal meaning of a blockbuster used to be "a very large bomb that can demolish an entire city block". Later people started to use this term metaphorically, as "anything that is highly impactful". As generations passed, fewer people knew the original meaning of this word. Bowdle and Gentner (2005) have coined the development of a metaphor from novel, via conventional to dead as the "career of metaphor".

The career of metaphor depicts the development of a metaphor's conventionality on a continuum, but more importantly: it also relates different levels of conventionality to different mapping processes. Metaphors that are processed more frequently can become more conventional. When a metaphor reaches a certain level of conventionality, metaphors can also be interpreted by direct categorisation. For instance, in interpreting a conventional metaphor like 'this car is a rocket', the category 'fast' is (probably) evoked automatically as the ground of the metaphor. This is a result of the repeated metaphoric use of the concept 'rocket' as a source domain, usually chosen to stress the speed of its target. By frequently using this concept in this specific manner, a relation is formed that connects the concept of a 'rocket' directly to a more abstract category of 'speed' or 'something that is fast'. Therefore, metaphors using 'rocket' as a source domain do not need to be understood through comparison, but can be interpreted immediately by categorisation.

In a series of experiments, Bowdle and Gentner (2005) have been able to document this implication of their career of metaphor concept in language use. They were able to manipulate the extent to which a metaphor was conventional by having a group of participants process several metaphors with the same ground (W is an acrobat; X is an acrobat; Y is an acrobat) and then compare the way in which they process a new metaphor (Z is an acrobat), to the way in which another group of participants who had not been exposed to metaphors involving acrobat as the source domain process the same 'Z is an acrobat' metaphor. They were able to show that the group for which the acrobat metaphor had been conventionalised used categorisation to interpret the metaphor, whereas the other group employed a comparison process. As such, they provide compelling evidence for the importance of conventionality for the processing of verbal metaphors. The question is whether conventionality plays an important role in the processing of visual metaphors as well.

7.4 The career of visual metaphor

7.4.1 Introduction

If CMT is correct in claiming that metaphoric processing is largely independent of the mode in which the metaphor is presented, then one would expect the conventionality of a metaphor that is verbally expressed to be correlated to the same metaphor when it is visually expressed. This implies that visual metaphors, similar to verbal metaphors can vary with respect to their conventionality.

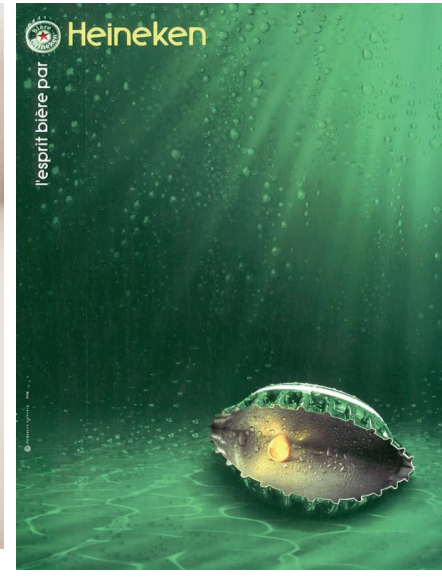
According to Bowdle and Gentner (2005), novel metaphors constitute one end of the continuum. Novel metaphors require people to compare the source and target domain to identify characteristics of the source that should be mapped onto the target domain. An example of such a novel metaphor is the car advertisement in which the car is compared to a dolphin (see Figure 7.1b). To establish the metaphorical meaning of this advertisement, participants need to select one or more characteristics of a dolphin that could be relevant to this car.

The more frequently someone has encountered a metaphor in which a certain source domain is used to communicate a certain characteristic, the more conventional this metaphor becomes. Within the visual domain, certain metaphors are used relatively frequently. Figures 7.2a and 7.2b show two ads in which the advertised product is compared to a pearl (while the context of a pearl is still available, the pearl itself is replaced by the product). Instead of referring solely to the valuable and rare gem that grows inside oysters, the pearl has come to stand almost exclusively for preciousness and pricelessness, as if the object were a concept or an abstraction. The pearl has undergone this evolution both in the visual and in the verbal domain. Also in words, it is possible to refer to the great value of something by making use of a pearl metaphor, in expressions such as 'he has some pearls of wisdom to offer'. Apparently, both visual and verbal pearls have undergone the same evolution. And although with pearls the visual metaphor has probably been inspired by the pre-existing verbal variety, this order of things is not mandatory.

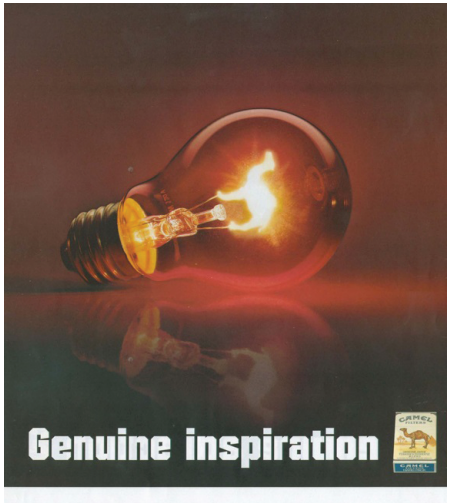
For light bulbs (another source domain that is used quite frequently in advertising), it is less common to come across verbal equivalents. Still, everybody knows that the depiction of a light bulb only rarely refers to a glass bulb that has to be inserted into a socket, but more probably refers to the concept of a bright idea (see Figures 7.2c–d). Here too, the concrete object has come to stand for an abstraction.

a.

b.



c.



d.

Figure 7.2 Illustrations of advertisements

At the other end of the continuum, Bowdle and Genter (2005) position "dead" metaphors, that is, metaphors for which the literal meaning of the source concept is no longer activated, but only the figurative meaning is. It is questionable whether such visual images can become "dead" in this sense because of their iconic nature. For activating the concept, the image has to be recognized as such. That is, to visually express the fact that one's relation has stranded, a boat stranded on a beach has to be depicted and needs to be recognized as such in order for the metaphoric relation to be processed. The verbal expression will most likely not activate such images (anymore). Still, there may be some visual metaphors that no longer activate the relevant concept. The ❤ image is used abundantly to express that someone or something is loved. It is possible that people no longer identify this image as "heart" but immediately activate its intended figurative meaning "loves". The same might be true for the light bulb mentioned above. Especially if the light bulb were to be depicted in a pencil drawing, or in a cartoon context, it is very questionable whether the original light spreading capacities of the object is still thought of by the onlookers. In this sense, the ❤ and the light bulb may have become dead metaphors. These examples suggest that there is enough reason to believe that visual metaphors can differ from each other on conventionality.

In this paper, we set out to test Conceptual Metaphor Theory's claim that the processing of metaphors needs to be situated at the conceptual level and is largely independent of the mode in which these metaphors are expressed. Most research on metaphor processing has focused on the processing of verbal metaphors. The insights acquired through these studies provide an interesting framework to test CMT's claims.

CMT would predict that a visual metaphor is processed in a way similar to a verbal metaphor. Given that conventionality plays an important role in the processing of verbal metaphors, it should play a similar role in the processing of visual metaphors as well if CMT is correct. That is, conventional visual metaphors would be processed through categorization, novel visual metaphors through comparison. To set up such a study, it is pivotal to establish whether visual metaphors vary with respect to their perceived conventionality. This leads to the first research question:

To what extent do visual metaphors vary with respect to their perceived conventionality?

In addition, it would be interesting to assess whether the perceived conventionality of a metaphor depends on whether it is presented in a verbal or a visual format. The difference in processing verbal and visual metaphors may lie in the identification of the concepts involved in the metaphor. Neuro-scientific evidence suggests that regardless of whether a concept is verbally presented or visually depicted the same neuron sets are activated (Quiroga et al. 2005). If Holyoak and Thagard (1995) are

correct in claiming that metaphors are about the relation between concepts rather than about the relation between words, then the perceived conventionality of a metaphor should not depend on its presentation mode. This leads to the second research question:

To what extent is the perceived conventionality of a verbally expressed metaphor similar to that of its visually expressed counterpart?

If no correlation is found between the verbal and visual versions of a metaphor, this would raise serious doubts about the claim that metaphors should be situated at the conceptual level.

7.4.2 Method

To address the research questions, an experiment was conducted in which participants rated the conventionality of either verbally presented metaphors or their visually presented equivalents. The participants were told that they had to rate ideas for print advertisements. Given that metaphors are used frequently in advertisements, the task could be considered a realistic one. Participants received either the verbal expression, e.g., "this USB stick is a sponge" or a leaflet containing a picture of a the two concepts involved, e.g. USB-stick and next to it a picture of a sponge.

Pretest
For the visual representations of the metaphor, it was essential that the images were regarded as referring to the same concepts as the words used in the verbal equivalent. That is, the image of a sponge should evoke the concept SPONGE just as the word 'sponge'. To assess whether that was the case, a pretest was carried out among 27 undergraduate students. Participants were asked to name the object shown in the image. If fewer than 5 out of 6 participants used the exact same word to describe the image, the image was adapted or replaced and presented to new participants. The pretest resulted in the selection of images for the main experiment that were described with the exact intended word by at least five out of six participants.

Participants
Eighty Dutch undergraduate students of the Radboud University in Nijmegen, aged between 17 and 29 years (M = 20.43, SD = 2.63) served as participants. For all participants, Dutch was their mother tongue and 48.5% of them were female. None of the participants was entered in a language or communication related program. All participants were paid four euros for their participation.

Materials

The material consisted of 17 metaphors, which were all represented in both a visual and a verbal format. All metaphors contained a concrete source domain (e.g., sponge, battery) being compared to a concrete product representing the target domain (e.g., USB-stick, muesli bar). The metaphors were created by the researchers but most of them were based on metaphors used in real advertisements. However, none of these metaphors had been used in recent national campaigns.

All visual metaphors consisted of two images. To limit the influence of perceptual features, the images were stripped down (employing Photoshop software) to greyscale visual representations without any context or logo's. The order in which the images were presented could influence the evaluation of the metaphor. Forceville (2007) contends that in most metaphorical television commercials the source is introduced before the target, which may have participants expect the source to be presented left and the target on the right. To control for the potential effect of (left-right) order of the images on the evaluation, two versions of each visual metaphor were created that only differed in the order of presentation of the images. In Figure 7.3, two examples of the visual metaphors are presented, the first with the source domain on the left side, the second with the source domain on the right side.

Figure 7.3 Two examples of stimuli in the visual condition

For each of the visual metaphors, two verbal equivalents were developed: in one version the relation was expressed as "This X is a Y" (e.g., "This mueslibar is a battery" or "This USB-stick is a sponge"), in the other version the metaphor was expressed as a simile (e.g., "This mueslibar is like a battery" or "This USB-stick is like a sponge". Bowdle and Gentner (2005) reported that participants preferred the simile version when the metaphorical relation was considered novel, whereas they preferred verbal metaphor expressions when the relation was perceived as conventional. Bowdle and Gentner therefore concluded that the verbal metaphor realization ("This X is a Y") induces the categorisation process, whereas the simile version ("This X is like a Y") induces comparison.

As a consequence of this, control factors were different for both modalities. In the visual mode, we wanted to control for target presentation on the left or right: a factor that could not be introduced for the verbal mode without changing the meaning of the metaphor (e.g., compare 'this muesli bar is a battery' to 'this battery is a muesli bar'). On the other hand, only verbal metaphors were controlled for type: whether they were presented as metaphor or simile. Note that with regard to the visual structure of the stimuli (the spatial layout of the pictorial elements in the picture frame), all visual metaphors had the structure of a juxtaposition (according to the terminology of Phillips and McQuarrie 2004) or a simile (according to the terminology of Forceville 1996). We deliberately opted for this design choice, because we were interested in the conventionality of metaphoric comparisons of two domains. Whether the comparisons that are presented as pictorial juxtapositions relate to verbal similes, just as pictorial fusions or hybrids (where the source and target are fused together into a single image) relate to verbal metaphors (see, e.g., Van Mulken et al. 2010) has yet to be explored in more detail, particularly with regard to processing. Besides that, the creation of pictorial hybrids would differ from item to item, and this would have influenced the ease with which participants could have identified the individual domains, and consequently the perceived conventionality of the metaphor. Therefore, in the pictorial domain, visual structure was held constant.

Design

Half of the participants were presented with the visual metaphors, the other half with the verbal metaphors. For those who had to rate the visual metaphors, half saw the metaphors with the source domain on the left side whereas the other half saw the metaphors in the reversed order. For those rating the verbal metaphors, half received the "is" version, the other half was exposed to the "is like" version. The design resulted in four different conditions. For each of these conditions, a second version was developed in which the order in which the metaphors were presented, was reversed.

Procedure

Participants were told that they would be presented with concept versions of print advertisements, which they had to evaluate using a print questionnaire followed by an interview. After reading the instruction containing an example of a concept version of a well-known advertisement (either in the verbal or visual mode), the questionnaire started with two examples of concept versions. Participants were then given the opportunity to ask for further explanation if needed. Subsequently, the questionnaire with the seventeen stimuli had to be filled out individually.

The level of conventionality was measured by three seven point semantic differentials with the ends of the poles being 'new – standard', 'obvious – far-fetched' and 'unusual – usual'.[2] The reliability of the scale was good (a = .82), both for the evaluation of visual metaphors (a = .80) and verbal metaphors (a = .84).

After the questionnaire was filled out, participants were interviewed. In this part of the experiment, all of the stimuli were presented again, one at a time and each one on a single page (and in the same order as in the questionnaire). While pretesting the experiment, we found that participants particularly liked to elaborate about their appreciation of the stimuli and whether they found it apt for advertising. Therefore the interview started with two filler questions: for each stimulus participants were asked to indicate to what extent they appreciated the comparison (*I like this comparison*) and considered it to be apt for advertising (*This is an apt comparison*). Both appreciation and aptness had to be rated on a single seven point scale (1 = disagree, 7 = agree) which was presented in print in front of the participants to help them verbalize their level of agreement in a grade.

To control whether participants interpreted the stimulus as a metaphor, they were asked for their interpretation of each stimulus. If they found it difficult to verbalize their thoughts, additional questions were asked ("What do you think is the ad maker's message?" or "Could you name any shared characteristics?"). It was checked that only initial metaphoric interpretations were gathered. If participants acknowledged that they perceived the metaphoric nature only at this secondary phase (during the interviews), their ratings for that particular item were removed from the dataset (n = 3). Likewise, cases were removed if more than one of the scores on the three seven point scales for conventionality was missing or not filled out correctly (n = 9). Finally, it was checked whether participants in the visual condition actually identified both target and source correctly. Most participants spontaneously mentioned the depicted items during the interview, but if not they were asked to name them (another 9 cases were deleted because of incorrect naming).

---

2. The 'conventional' poles were respectively on the right, left and again on the right side. This was done to prevent participants giving all dimensions the same score, but let them consider each decision separately.

Statistical and control analyses

To assess whether the (verbal and visual) stimuli were interpreted as metaphors, the participants' initial interpretations during the interview were checked by three independent raters (after they had had a training session in metaphors). If positive characteristics of the source domain where transferred onto the target, the interpretation was considered metaphoric, such as, when the 'lipstick is magnet' metaphor was interpreted as "this lipstick attracts men" or "it sticks perfectly to your lips". Interpretations that were considered as not metaphoric were interpretations that did not clearly involve metaphoric thinking (like "both belong to the same brand"). Besides that, interpretations that were negative about the product (like "everything will keep sticking to it") were removed from the dataset as well. The interrater reliability of the three raters was high (Fleiss' generalised Kappa = .83; Landis & Koch 1977).

Interpretations that were considered to be not metaphorical by at least two of the raters, were excluded for further analyses (211 cases, so that finally 84.04% of all cases remained for analyses). Before that, we checked whether the incidence of a metaphorical interpretation was related to whether the metaphors were presented in a verbal or visual format. A t-test showed that this was not the case (t (1327.11) = 1.44, p = .15). An additional t-test (t (672.01) = .51, p = .61) shows that it did not matter for the verbal items whether they were presented as a simile (14.1% incorrect interpretations) or a metaphor (12.6% incorrect interpretations). In the visual mode however, the t-test (t (638.19) = 2.37, p < .05) shows that participants who received the items with the target depicted on the right side made more errors (19.6%) when interpreting the visual metaphor then participants who received the items with the target depicted on the left side (12.8%). Although this is a significant difference (that will be handled in more detail in the discussion section), this showed to have no effect on the ratings on conventionality, since all the errors were deleted and only items that were interpreted metaphorically were included for our analyses on conventionality.

7.4.3 Results

Table 7.1 contains the average scores for both the visual and the verbal metaphors.

The first research question was about the extent to which visual metaphors differ with respect to their perceived conventionality. To address that question, a oneway anova with stimulus as factor was conducted to assess the differences in conventionality scores for the items in the visual mode. This revealed a highly significant effect of stimulus (F (16, 541) = 11.79, p < .001). Post hoc pairwise comparisons (Tukey's HSD) revealed that of the 136 potential significant differences, 43 were indeed significant.

Table 7.1 Mean conventionality scores for each item in both modalities, ranging from 1 (very unconventional) to 7 (very conventional), standard deviation between brackets (* = significant difference between the visual and verbal modality)

stimulus	mean conventionality	verbal conventionality	visual conventionality
target, source	M (SD)	M (SD)	M (SD)
car, rocket	5.62 (1.08)	5.59 (1.06)	5.65 (1.11)
watch, diamond	5.08 (1.17)	5.06 (1.08)	5.10 (1.27)
muesli bar, battery	4.90 (1.48)	4.94 (1.22)	4.86 (1.72)
gin, cannon*	4.34 (1.38)	3.94 (1.36)	4.75 (1.30)
lawn mower, electric shaver	4.13 (1.44)	4.32 (1.51)	3.89 (1.33)
lipstick, magnet*	4.01 (1.36)	4.61 (1.19)	3.35 (1.24)
piece of candy, ice cube*	3.97 (1.55)	3.24 (1.23)	4.96 (1.41)
mobile phone, tool case	3.96 (1.09)	3.96 (1.04)	3.95 (1.16)
chocolate, pearl	3.91 (1.36)	4.18 (1.37)	3.61 (1.30)
chain saw, tiger	3.81 (1.47)	3.84 (1.40)	3.77 (1.57)
microwave, robot	3.74 (1.39)	3.80 (1.42)	3.65 (1.37)
condom, umbrella	3.72 (1.39)	3.64 (1.35)	3.80 (1.45)
helmet, sea shell	3.48 (1.14)	3.63 (1.22)	3.30 (1.03)
beer, ventilator*	3.43 (1.56)	2.46 (1.13)	4.00 (1.51)
detergent, boxing glove*	3.41 (1.63)	2.66 (1.09)	4.20 (1.73)
USB-stick, sponge	3.03 (1.01)	3.24 (1.11)	2.81 (0.85)
laptop, dolphin	2.43 (1.26)	2.09 (0.99)	2.70 (1.40)

The second research question was about the extent to which the conventionality scores of the visual metaphors were similar to those of their verbal equivalents. A straightforward way to address this question is to compute the correlation between the conventionality scores for the verbal and visual metaphors. Indeed, the average verbal conventionality scores correlate significantly with the average visual conventionality scores ($r = .59$, $p < .05$). Although this correlation coefficient is considerable, it is far from perfect.

The question can also be addressed in a more complex way through a mixed model analysis. Our model included conventionality as dependent variable, modality as fixed factor and items and participants as covariates. To control for bias effects, several other factors (e.g., age, education, appreciation, aptness) were placed into the model, but for none of them did inclusion improve the model. This also holds for the fact whether a verbal item was presented as a metaphor or simile, and whether a visual item was presented with the target on the left or on the right side. This justifies that all verbal items were considered as one homogeneous group, and all visual items as another.

In the final model we found no overall significant main effect of the modality being verbal or visual (F (1, 22.47) < 1, p = .48) on the conventionality scores, implying that the average conventionality for verbal metaphors was equal to the visual metaphors. There were significant random effects of participants (b = . 16, SE = .04, p < .001) and of items (b = .42, SE = .21, p < .05) implying that variation in scores could be attributed to both differences between participants and differences between metaphors. These differences between metaphors were found when presented visually (b = .60, SE = .23, p < .05; similarly to the one way anova) as well as when they were presented verbally (b = .78, SE = .29, p < .01). However, a significant interaction of modality and item was found in the covariance structure (b = .27, SE = .11, p < .05) implying that for some of the metaphors, their conventionality scores differed depending on whether they were presented in a verbal or in a visual mode.

To assess for which metaphors presentation mode mattered, separate t-tests were conducted for each pair of verbal and visual metaphors. For five pairs, the difference proved significant (in *Table 1*, these items are marked with an asterisk). For 'lipstick is magnet', the verbal metaphor was perceived as more conventional than its visual counterpart (t (72) = 4.43, p < .001). For the other four metaphors, the visual version was experienced as more conventional than the verbal one ('detergent is boxing glove': t (64) = 4.37, p < .001; 'gin is cannon': t (71) = 2.61, p < .05; 'piece of candy is ice cube': t (57) = 5.02, p < 001; and 'beer is ventilator': t (55) = 4.05, p < .001). When correlating the conventionality scores for the verbal and visual metaphors without these five items, the correlation coefficient was almost perfect (r = .92, p < .001).

7.4.4 Conclusion

Our research questions concerned whether visual metaphors vary with respect to their perceived conventionality and whether the perceived conventionality of a verbally presented metaphor is similar to that of its visually presented counterpart. With regard to the first question, we can conclude that indeed, visual metaphors can be placed on a continuum, ranging from novel to highly conventional metaphors. Whereas the juxtaposition of a laptop computer to a dolphin was considered to be the most novel comparison relative to the other juxtapositions in our experiment, the comparison of a car to a rocket was seen as fairly conventional. With regard to the second question, we found that the verbal and visual modalities indeed correlate substantially with regard to perceived conventionality. We have therefore reason to assume that metaphors are represented at a deeper conceptual level. The experienced conventionality of a metaphor seems to hold primarily for a specific comparison of concepts, rather than the verbal expressions that represent these concepts.

However, for five stimuli in our experiment, there were significant differences between the perceived conventionality of the verbal representation and the visual representation. Given that there was no main effect of modality, and the overall correlation of the items in the verbal and visual mode, these five metaphors are no reason to reject the hypothesis that the experienced conventionality of visual representations of metaphoric concepts is comparable to verbal metaphors that are based on the same conceptual comparison. This quintet of 'dissident' metaphors does raise the question whether these metaphors do not share a metaphoric relation at the conceptual level or whether there are other reasons why these metaphors are perceived differently depending on their mode of presentation.

An important assumption for the experiment was that the words and images in our materials would evoke the same underlying concepts. Through extensive pretesting, images were selected that would evoke the words used in the verbal version of the metaphor. The results of the main experiment revealed that in the vast majority of cases, participants indeed used these words when describing their interpretations. However, it was not possible to assess whether the words evoked the kind of image that was used in the visual versions. It is not unlikely that participants differed in the images that came to mind when processing the word 'candy', or 'cannon', or 'ventilator'. High conventionality ratings for some of the visual metaphors may have been the result of the fact that the words evoked slightly different concepts than the visual image did, concepts that might be considered less easy to map onto the target domain.

In spite of the fact that we tried to select visual items that represent superordinate classes, there are more than one representative or prototypical images for these concepts. In addition, prototypical representations of these concepts may vary from person to person. In case of candy, someone can think of bonbons, sweets, liquorice or other confections. The candy in our experiment was the typical sourball in wrapping paper, but some participants may have thought of bulk candy, such as fudge, or 'tumtum' (a typical Dutch piece of candy). The same might be true for the ventilator, which can be represented mentally in several forms, such as a desk air blower, a propeller with blades hanging from the ceiling, air conditioners or even the computer fan. If the word 'ventilator' activated a concept other than the desk air blower, participants may have thought the relation between this type of ventilator and beer to be more unusual and far-fetched than when they were thinking about a desk air blower. It would be interesting to see what kind of image participants would draw when being presented with the word 'candy' or, when presented with images of different kinds of candy, to see whether they consider the image used in the experiment as being the most typical representation of the concept.

The concept of something seemingly simple as a car can be communicated in numerous ways, not only visually, but verbally as well (consider e.g. automobile, bolide,

convertible, motorcar, roadster or wagon). However, in the visual domain, visual grammar plays a considerable role: camera angle, point of view, framing, colours, contrast, focus, depth of field all matter when choosing a style of representation. This is much less a problem in the case of verbal representation (where only font and spacing are involved).

What is more, the informativeness of the visual items in our experiment was probably higher than that of the corresponding words. The inherent rich character of visuals may help to stimulate metaphorical thought. For instance, the picture of a piece of candy gave much more precise information about the particular kind of candy (the sourball in plastic wrap) than the verbal counterpart ('piece of candy') did, which may have made attribution mapping easier in the visual mode. The same may hold for the 'detergent is boxing glove', which was also perceived as more conventional in the visual mode. The image of a boxing glove appearing to be in upward movement may have evoked a stronger sense of "fighting dirt" compared to people who "saw" a boxing glove lying on the table or hanging on the wall when processing the word "boxing glove". Besides that, the image of detergent may be considered as a prototypical product in advertisements containing images, which might have biased the scores on conventionality for this particular comparison. Phillips (1997) has shown that familiarity with the target can influence the interpretation of metaphors.

Only one metaphor was experienced as more conventional in the verbal than in the visual mode: 'lipstick is magnet'. In Dutch, besides the literal meaning of 'magnet', also the figurative meaning has a lemma in the dictionary. Probably, the figurative meaning of 'magnet' is a dead metaphor in Dutch. In the context of a comparison or simile, the verbal meaning of magnet then no longer evokes its original meaning and the category 'attraction' is called for instantly, whereas a visual magnet still refers to its 'literal' self initially. The fact that the image still evokes the concept of magnet whereas the word no longer does, may be due to the inherent iconic character of visual representations of concepts. Besides that, a large part of conventionalization may occur at a conceptual level, while metaphors can have different careers depending on the frequency of use within a particular modality. The categorization process – the process of referring to categories instead of palpable objects – may evolve at a different pace in the two modalities, as a result of the extent to which they are perceived as conventional. Future research should investigate whether modality differences are indeed characteristic of these so called 'dead' or highly conventional metaphors.

Thibodeau and Durgin (2011) have already shown that conventionality is often hard to measure and therefore often confused with aptness by participants. We controlled for this effect and indeed we found a significant, but moderate correlation between conventionality and aptness ($r = .55$, $p < .001$). However, it is hard to

imagine that laymen judgments on conventionality would *not* extend along aptness. We believe that our operationalization of conventionality is fairly reliable, due to the fact that the three different scales ('new – standard', 'obvious – far-fetched' and 'unusual – usual') to measure conventionality were not only highly reliable (see Cronbach's alpha scores in the procedure section), but also quite comprehensible for laymen and therefore not likely to be mistaken for aptness (which was scored on a secluded scale). We controlled for aptness and appreciation in the mixed model analysis, but factors such as informativeness, iconicity and arousal could as well be considered to have an influence on conventionality judgments.

7.5 General conclusion

Although there are many factors that could be considered as significant factors on metaphor processing, verbal and visual representations of metaphors seem to be modality independent with regard to the evaluation of conventionality. Because conventionality is such an important factor in metaphor processing, this suggests that the way a metaphoric representation is processed rather depends on the conventionality of the conceptual idea, than on its mode of representation. Now that we know that visual metaphors can differ with respect to their perceived conventionality, the next question is whether conventionality influences the processing of visual metaphors in a similar way as it influences the processing of verbal metaphors. Future research should investigate whether novel visual metaphors are processed by making a comparison and more conventional ones by categorization. This would mean that a picture of a car juxtaposed to a rocket (a conventional metaphor) is processed relatively fast, whereas a helmet juxtaposed to a sea shell most likely requires more processing time.

If a metaphor's evaluation of conventionality is independent of modality, then the process of conventionalisation might be independent of modality as well. Many verbal metaphors seem to be derived from visual experiences. Lakoff & Johnson (1980, pp. 14–21) show e.g. how verbal expressions that are based on the conceptual idea that MORE IS UP (like 'prices are going through the roof' or 'temperature is rising') are derived from the visual experience of liquid levels rising upward when something is filled with more liquid. Experimental research could reveal whether the repeated activation of a metaphoric concept in one modality can cause a conventionality effect of the same metaphor in another modality. For instance, when one actively processes a series of verbal (or visual) metaphors containing the same novel source domain (like Bowdle and Gentner did in their 2005 *in vitro* experiment), this should lead to conventionalisation and hence a shift in processing from comparison to categorisation. A new realization of the conventionalised metaphoric

concept should then be processed via direct alignment with an abstract category, regardless whether the metaphor is presented visually (or verbally). Future research should investigate whether these expectations are true.

Future research could also focus on a surprising finding that was beyond the primary scope of this study: participants who received visual metaphors with the target depicted on the right side where less likely to come up with a metaphoric interpretation than participants who received visual metaphors with the target depicted on the left side. This is in the opposite direction of what can be expected on behalf of the findings of Forceville (2007), who found that in most tv-commercials the source is introduced before the target. Perhaps creators of advertisements have a tendency to 'start' an advertisement with an eye-catching element: a source domain that deviates from our expectations. Besides that, positioning of the product on the right side might be obvious, because in most cultures this side is associated with positive attributes (see, e.g., Wilson & Nisbett 1978; Lakoff & Johnson 1980; Sutton-Spence & Kaneko 2007). Our findings show that for an onlooker, target placement on the right side might be problematic for interpretation. In verbal language the typical 'X is Y' structure of metaphors follows the subject-verb-object of subject-object-verb order that is common in most languages. In the visual mode one would expect the subject to be introduced before the object as well. Based on the left to right reading direction in most cultures, this would be on the left side. Like in verbal language, one has to form an idea of the product before processing additional information that is taken from the source, and before interpreting the advertisement as a whole. This is a process that might have consequences for creators of pictorial advertisements. A cross-cultural study including cultures with contrasting reading directions and mirrored advertisements could give more insight into the function of source and target order in pictorial juxtapositions. In addition, eye movement patterns when viewing visual metaphors could be measured by making use of an eye tracker.

Finally, the question whether visual metaphors can 'die' like verbal metaphors still remains. Is it possible for an onlooker to overlook or forget the iconic character of an image? Although the light bulb is a highly conventional metaphor for a bright idea, it is difficult to ignore the representation of a physical object and what it stands for. However, now that governments around the world have passed measures to phase out incandescent light bulbs, it is not totally inconceivable that new generations will refer to brilliant ideas by using bulbs without knowing what the original function of this object was. The career of visual metaphor might also be a matter of time.

Acknowledgements

NWO for funding the project, Ellen Webbink for her help with statistical analyses, Merel Balfoort, Britt de Lange, Charissa Koetsier and Funda Roes for assisting as experiment leader and rater, and the members of the "VisMet" group for their useful comments and feedback.

References

Bowdle, B., & Gentner, D. (1999). Metaphor comprehension: From comparison to categorization. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the twenty-first annual conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Bowdle, B., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, 112, 193–215. https://doi.org/10.1037/0033-295X.112.1.193

Forceville, C. (1996). *Pictorial metaphor in advertising*. London: Routledge. https://doi.org/10.4324/9780203272305

Forceville, C. (2005). Visual representations of the idealized cognitive model of anger in the Asterix album La Zizanie. *Journal of Pragmatics*, 37, 69–88. https://doi.org/10.1016/j.pragma.2003.10.002

Forceville, C. (2007). Multimodal metaphor in ten Dutch TV commercials. *The Public Journal of Semiotics*, 1, 19–51.

Forceville, C. (2009). Non-verbal and multimodal metaphor in a cognitivist framework: Agendas for research. In C. J. Forceville & E. Urios-Aparisi (Eds.), *Multimodal Metaphor*. Berlin, New York: Mouton de Gruyter.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive linguistics*, 8, 183–206. https://doi.org/10.1515/cogl.1997.8.3.183

Hoeken, H., Swanepoel, P., Saal, E., & Jansen, C. (2009). Using message form to stimulate conversations: The case of tropes. *Journal of Communication*, 19, 49–65.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

Kövecses, Z. (2005). *Metaphor in culture: Universality and variation*. Cambridge: Cambridge Universtity Press. https://doi.org/10.1017/CBO9780511614408

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9781139173865.013

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago [etc.]: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Pilosophy in the flesh*. New York: Basic Books.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. https://doi.org/10.2307/2529310

Maes, A., & Schilperoord, J. (2008). Classifying visual rhetoric: Conceptual and structural heuristics in interpreting and classifying visual rhetoric. In B. J. Phillips & E. F. McQuarrie (Eds.), *Go figure: New directions in advertising rhetoric*. Armonk, NY: Sharpe.

McGlone, M. S. (2007). What is the explanatory value of a conceptual metaphor? *Language and Communication*, 27, 109–126. https://doi.org/10.1016/j.langcom.2006.02.016

Müller, C., & Cienki, A. (2009). Words, gestures, and beyond: Forms of multimodal metaphor in the use of spoken language. In C. J. Forceville & E. Urios-Aparisi (Eds.), *Multimodal Metaphor*. Berlin, New York: Mouton de Gruyter.

Murphy, G. L. (1997). Reasons to doubt the present evidence for metaphoric representations. *Cognition*, 62, 99–108. https://doi.org/10.1016/S0010-0277(96)00725-1

Ortony, A. (1979). *Metaphor and thought: second edition*. Cambridge: Cambridge University Press.

Phillips, B. J., & McQuarrie, E. F. (2004). Beyond visual metaphor: A new typology of visual rhetoric in advertising. *Marketing Theory*, 4, 113–136. https://doi.org/10.1177/1470593104044089

Phillips, B. J. (1997). Thinking into it: Consumer interpretation of complex advertisement images. *Journal of advertising*, 26, 77–87. https://doi.org/10.1080/00913367.1997.10673524

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435, 1102–1107. https://doi.org/10.1038/nature03687

Steen, G. J. (2007). *Finding metaphor in grammar and usage*. Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/celcr.10

Sutton-Spence, R., & Kaneko, M. (2007). Symmetry in sign language poetry. *Sign Language Studies*, 7, 284–318. https://doi.org/10.1353/sls.2007.0020

Thibodeau, P., & Durgin, F. H. (2011). Metaphor aptness and conventionality: A processing fluency account. *Metaphor and Symbol*, 26, 206–226. https://doi.org/10.1080/10926488.2011.583196

Van Enschot, R., Hoeken, H., & Van Mulken, M. (2008). Rhetoric in advertising: Attitudes towards verbo-pictorial rhetorical figures. *Information Design Journal*, 16, 35–45. https://doi.org/10.1075/idj.16.1.05ens

Van Enschot, R., & Hoeken, H. (2015). The occurrence and effects of verbal and visual anchoring of tropes on the perceived comprehensibility and liking of TV commercials. *Journal of Advertising*, 44, 25–36. https://doi.org/10.1080/00913367.2014.933688

Van Mulken, M., Le Pair, R., & Forceville, C. (2010). The impact of perceived complexity, deviation and comprehension on the appreciation of visual metaphor in advertising across three European countries. *Journal of Pragmatics*, 42, 3418–3430. https://doi.org/10.1016/j.pragma.2010.04.030

Wilson, T. D., & Nisbett, R. E. (1978). The accuracy of verbal reports about the effects of stimuli on evaluations and behavior. *Social Psychology*, 41, 118–131. https://doi.org/10.2307/3033572

CHAPTER 8

Conclusion

Gerard J. Steen
Universiteit van Amsterdam

The goal of this book was to report on the outcomes of a four-year research program entitled 'Visual metaphor: A psycholinguistic approach' (2009–2012). The program comprised three projects that led to various studies and findings. In this chapter I will discuss these findings as reported in the various chapters of this book and suggest some possible avenues for future research that may be derived from our work in relation to new developments in metaphor research. The overall framework of this discussion concerns the productivity of adopting a psycholinguistic approach to the study of visual metaphor.

8.1 Discussion of the research presented in this book

As noted throughout this book, the reconceptualization in the late 1970s of metaphor from a figure of speech to a figure of thought has been essential in the area of metaphor studies (cf. Gibbs 2008). Metaphor is now generally considered to be fundamental to the human conceptual system, language being merely one possible modality of expression of metaphorical cognition. Yet the wide range of nonverbal metaphor and its processing has neither been empirically investigated on the same scale nor with the same rigour as metaphor in language. The overarching goal of this book has therefore been to exploit the vast psycholinguistic expertise on metaphor in language for a new, behaviourally founded approach to the identification, description, and process modelling of metaphor in one of these nonverbal manifestations, static visuals.

For this purpose, we have employed the Career of Metaphor Theory (Bowdle and Gentner 2005) as our starting point. This is because it combines insights from the two leading current approaches to metaphor in language: Conceptual Metaphor Theory (Lakoff & Johnson 1980, 1999) and Class-Inclusion Theory (Glucksberg 2001). The relevance of this framework for visual metaphor may be illustrated by the following constructed variants of an advertisement for a mineral water (see Figure 8.1).

https://doi.org/10.1075/celcr.18.08ste

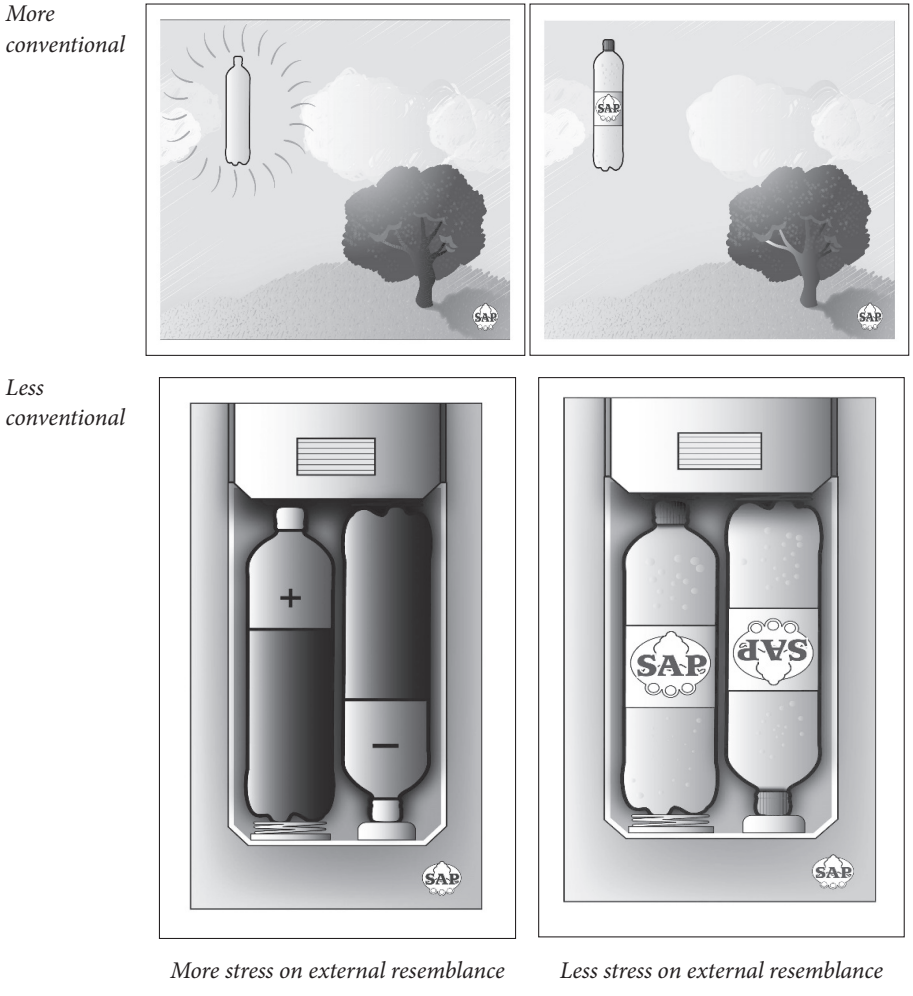


Figure 8.1 Visual metaphors: Water is energy

If one wants to communicate the message that this water provides energy, the designer may choose to compare the entity of water ("the target") to another entity ("the base") that is frequently used to represent energy, e.g., the sun, or to one that has not been used much to do so, e.g., a battery. A second choice concerns the external resemblance between the two entities, which may be stressed (the left-hand versions) or downplayed (the right-hand versions). Frequency and external resemblance are central aspects of visual metaphor.

Frequency and external resemblance contribute to general metaphor properties of conventionality and type, respectively. Metaphor conventionality and metaphor type play a fundamental role in the Career of Metaphor Theory (Bowdle & Gentner

2005). Since metaphor is now seen as a conceptual phenomenon that is relatively independent of the modality in which it is expressed, psycholinguistic insights into verbal metaphor processing provide new and motivated hypotheses for the processing of visual metaphors and the influence of conventionality (including frequency) and type (including external resemblance) in that modality.

For metaphor processing in language, the Career of Metaphor Theory has revealed interactions between a number of important factors, the three most important ones for this book being *metaphor conventionality* and *metaphor type* (part of metaphor structure) and their effects on *processing:*

1. **Metaphor conventionality** is the degree to which there is an established metaphorical connection between a particular conceptual base domain and target domain for a specific linguistic expression. Gentner and Bowdle propose that metaphors exhibit a career, from novel to conventional, and that different types of processes guide understanding novel and conventional metaphors. For *conventional metaphors*, the repeated use of a certain concept as a base (e.g., the sun) leads the base to get associated with a certain category that it typifies, i.e., 'energy-providing things'; this enables language users to infer that the metaphor communicates that a certain target (the product) provides energy (*categorization*). For *novel metaphors*, on the other hand, the reader first has to compare the two concepts (mineral water, battery) in order to infer that the product provides energy (*comparison*).
2. **Metaphor type** refers to the properties of the base concept that are relevant for the interpretation of a metaphor. For example, batteries are devices that provide electrical energy to machines. By analogy, Spa may be compared to batteries because it relates to people like batteries relate to machines: providing physical energy to the human body. By contrast, Spa may also be compared to a battery because the external appearance of a bottle of Spa can be mapped onto the external appearance of a battery. *Metaphor by analogy* leads to *relation mapping, metaphor by external resemblance* leads to *attribute mapping*. Combinations of both types of metaphor as well as mapping operations are also possible.

Metaphor conventionality and metaphor type and their effect on processing provide a set of fundamental parameters for comparing metaphor processing in language with metaphor processing in images, since there is reason to believe that conventionality and type have a different relation with processing in the visual modality:

1′. **Metaphor conventionality**: Bowdle and Gentner argue that as a metaphor becomes increasingly conventionalized in the verbal modality, the activation of the figurative category evoked by the base concept (energy) becomes dominant compared to the activation of the concept itself (sun) when a particular word

(*sun*) requires figurative interpretation. The question is whether the same holds for repeatedly used metaphors in the visual modality. In language, words typically have more than one conventionalized meaning, including metaphorical ones (Steen 2007). In visuals, the objects involved in the metaphor usually have one meaning: the batteries "mean" 'battery', and the Spa bottles "mean" 'Spa bottle'. Visuals are iconic whereas language is symbolic. This suggests that, in contrast to language, most metaphors in visuals require the identification of two distinct objects or domains of reference first *("iconic processing"*), which then need to be aligned and matched in subsequent conceptual mapping operations. The question arises which of these visuals can be called conventionally metaphorical.

2′. **Metaphor type**: in language, relational properties are more likely to get mapped than attributive properties. In fact, attribute mapping is regarded as only marginally important and even potentially disruptive. In processing visual materials, however, it may be the very external appearance (attributes) of the objects in an image which is given careful attention, as when the Spa bottles are made maximally similar in shape (e.g., size) to the batteries. Attributes are therefore hard to ignore by the viewer and *attribute mapping* may be expected to be relatively more important in visual metaphor processing than in verbal metaphor processing. How this affects relation mapping is an open question.

In sum, we have used current knowledge about metaphor in language processing to address how conventionality and metaphor type affect visual metaphor processing. In contrast with verbal metaphor processing, we hypothesized that iconic processing and attribute mapping are typical of visual representation. Chapters 5, 6 and 7 charted these processes in detail and answered the question how the roles of relational mappings and conventionality are affected.

In order to examine the validity of our visual metaphor identification method (Chapter 3) and relate it to visual metaphor processing by people simply viewing multimodal messages, we then continued with presenting a pilot study collecting think aloud data about metaphorical visuals (Van den Heerik et al. this vol.; cf. Šorm and Steen 2013). The results of this study yielded findings about the various levels of meaning at which people process and represent metaphorical visuals. These findings throw more light on the structures and functions of visuals in cognitive processing.

In Chapter 6, Van Weelden et al. (this vol.) presented one study that showed that similarity in shape functions as a bootstrap mechanism for the activation of metaphoric relations, between objects that have distinct functions and for which no ready-made common category is available. For these object combinations more correspondences were produced (both attributive and relational) and thinking time to do so decreased as a result of shape similarity. Overlap in objects' shape,

so structural descriptions, activates shared semantic (i.e., conceptual) knowledge. When semantic knowledge is already shared, however, additional overlap in shape does not particularly strengthen the semantic relation.

In their second study, participants were faster in finding and producing a correspondence when the objects had similar shapes, as compared to when the objects did not look alike. Similarity in shape resulted in the production of more correspondences between the objects as well. Regarding the nature of the produced correspondences, there were no differences between the proportions of both the attributive and relational correspondences for the similarly and dissimilarly shaped object pairs. Hence, participants found both more attributive and relational correspondences.

In Chapter 7, Hodiamont et al. (this vol.) charted the conventionality continuum. To identify visual metaphors that vary with respect to conventionality, participants were asked to provide interpretations and to rate the level of conventionality for a sample of metaphors that consists of metaphors that visualize verbally dead metaphors, ones that are frequently used, and ones that are rather unique and presumably novel. In order to assess whether the conventionality of the visual metaphor is related to its verbal counterpart, a verbal expression of each of these visual metaphors was rated for conventionality by different participants. It turned out that visual metaphors can be placed on a continuum, ranging from novel to highly conventional metaphors. Whereas the juxtaposition of a laptop computer to a dolphin was considered to be the most novel comparison relative to the other juxtapositions in our experiment, the comparison of a car to a rocket was seen as fairly conventional. The authors also found that there is no substantial difference between the two modalities. They conclude that metaphors are represented at a deeper conceptual level. The experienced conventionality of a metaphor seems to hold primarily for a specific comparison of concepts, rather than the verbal expressions that represent these concepts. This project has enabled us to position visual metaphors on a conventionality continuum and to assess the relation between the conventionality of metaphors in the visual and verbal code.

Chapters 5, 6 and 7 therefore provide evidence that supports our novel psycholinguistic approach to visual metaphor processing in comparison with verbal metaphor processing. Distinct steps and operations, types of mapping (relations and attributes) and their conventionality appear to have their own position in metaphor processing. Distinct steps each play an understandable role that can even be ordered for analytical purposes in the proposed procedure for metaphor identification called VISMIP (Chapter 3). For type of mapping, attributes play a more important role in visual metaphor processing than in verbal processing. And for conventionality, visual arrangement seems less important than underlying conceptual mapping, for these stimuli at least, which in language use shows a different interaction with an equal role for both factors (Bowdle & Gentner 2005).

Apart from these central questions about metaphor processing, we also wanted to develop an independent, psycholinguistically motivated grasp of the structure of visual metaphors. Our program therefore also aimed at the development of a method for visual metaphor identification and analysis in printed texts. It capitalized on the method developed for verbal metaphor identification by Gerard Steen and colleagues (Pragglejaz Group 2007; Steen 2007; Steen, Dorst, et al. 2010). This method has been successfully applied to two corpora and in an increasing amount of research elsewhere. The main goal of our program was to reconceptualise the method for visual metaphors, elaborate an appropriate operational definition, and attune it to the more specific concerns of the present project that have to do with metaphor conventionality and type.

Joost Schilperoord's chapter on visual incongruities laid the groundwork for this discussion. He explored various visual structures evoking incongruity and showed how these structures may be interpreted by different conceptual operations. His formal taxonomy for both may be of great help to many visual metaphor scholars. His main point was that visual incongruity is not a necessary and sufficient condition for metaphorical interpretation. Visual incongruity may be resolved in various ways, and Schilperoord's map of the field may be of eminent service for further explorations of the way in which visual incongruity in relation to message topic and genre can manifest intricate interactions in producing or not producing visual metaphor.

The next stage addressed the visual objects involved in the metaphorical relation (Šorm and Steen this vol.). It tackled the following issues, directly derived from the verbal metaphor identification method MIPVU:

1. What counts as a distinct visual form which can indicate a base and/or target concept?
2. What counts as an incongruous visual form (IVF)?
    a. What is the referent or function of the IVF in the picture?
    b. What is the usual referent of the IVF?
3. Is the usual referent of the IVF sufficiently distinct from its referent or function in the picture?
4. Is the referent or function of the IVF sufficiently comparable to its usual referent?

In addressing these questions, the following visual metaphor identification procedure was formulated (Table 8.2).

With the clarification of these conceptual and operational issues of visual metaphor identification and analysis in Chapter 2 and 3, it was possible to make a beginning with the construction of a visual metaphor database, called VisMet 1.0. A collection of images was put together across a range of genres, including cartoons, advertisements, posters for public campaigns, and so on. For each of these visual

documents, permission was obtained for inclusion on the website, which is now a public resource that presents all documents for access. Each of the documents displays additional information about expressive, conceptual and communicative aspects of the visual metaphor as well as about producer and distributor. It is our goal to scale up the database and add crowd sourced tags to the images in order to engage the public as well as study their interpretations.

Table 8.2 Set of instructions in the Visual Metaphor Identification Procedure (VisMIP). The instructions can be used to identify visual units that are related to metaphor

1. Look at the entire image, including visual and verbal elements, to establish a general understanding of the meaning.
   - 1a. Describe in just a few simple phrases the referential meaning of the image, i.e. what/who is being depicted here, what he is doing, where he is doing it, and so on. For example: 'Tall, blond man and spotted dog walk quickly in park. Smiling girl gives green apple to frowning boy under tree.' If the referential meaning is ambiguous and allows more than one interpretation, then give alternative descriptions.
   - 1b. Test whether there are any clues that tell you that more general and abstract meaning should be attached to what is described under step 1a. For example, if the exemplary image described under 1a is accompanied by the caption 'summer joy', we have a clue that the more abstract concept 'summer joy' should be attached to the referential meaning.
   - 1c. Reconstruct the point underlying the image.
   - 1d. Derive from step 1c the topic of the point, i.e. that about which the point is stated.
2. Structure the descriptive phrase(s) under step 1a. For example: 'Tall, blond man and spotted dog walk quickly in park. [Agent(man|tall, blond) Agent(dog|spotted) Action(walk|quickly) Setting(in park)] Smiling girl gives green apple to frowning boy under tree. [Agent(girl|smiling) Action(give) Object(apple|green) Recipient(boy|frowning) Setting(under tree)'.
3. Find incongruous visual units.
   - 3a. Decide for each unit under step 2 whether it is incongruous with the topic as formulated under step 1d ('topic-incongruous').
   - 3b. Decide for each topic-congruous unit under step 2 whether it shows properties that are incongruous with the properties that are typically true of that unit ('property-incongruous').
4. Test whether the incongruous units are to be integrated within the overall topical framework by means of some form of comparison.
   - 4a. For each incongruous unit under 3a, determine which replacing unit would be congruous with the topic AND would be coherent with the referential meaning of the image.
   - 4b. For each incongruous unit under 3b, determine which replacing unit would typically own the incongruous properties AND would be coherent with the referential meaning of the image.
5. Test whether the comparison(s) is/are cross-domain.
6. Test whether the comparison(s) can be seen as some form of indirect discourse about the topic as formulated under step 1d.
7. If the findings of tests 4, 5, 6 are positive, then a visual unit should be marked for metaphor.

The structure part of our study of visual metaphor therefore also led to new insights pushed by our psycholinguistic perspectives. Schilperoord's chapter showed how structure and process can be generally related to each other for the domain of visual incongruities, and how this map can be used for situating visual metaphor as one concrete visual rhetorical figure. An identification procedure was then presented in Chapter 3, by Šorm and Steen. And the experience with developing this tool was highly instrumental in collecting and analyzing the documents for the world's first public resource for visual metaphor, VisMet 1.0., with provisional structural descriptions for future discussion.

8.2 Further research

One consequence of our research is the development of a better view of where future research can go. What is important in this context is a new development in metaphor research. Ten years ago Bowdle and Gentner's (2005) Career of Metaphor Theory was probably the most complex theory of metaphor processing available, comprising insights from the two competing theories of Lakoff and Johnson (1980, 1999), Conceptual Metaphor Theory, and of Glucksberg (2001), Class Inclusion Theory. The appeal of the Career of Metaphor was that it could predict the incidence of two different metaphor processing strategies, comparison (related to Conceptual Metaphor Theory) and categorization (related to Class Inclusion Theory) on the basis of structural metaphor properties of conceptual conventionality versus novelty (examined in our book in structure in Chapter 4 and in process in Chapter 7) and linguistic form (examined in our book in structure in Chapters 2 and 4). What has happened since has been the development of a new theory, called Deliberate Metaphor Theory (Steen 2008, 2011, 2015, 2017), which takes these insights on board and extends them into new directions. These new directions also throw new light on the possibilities for future research on visual metaphor.

The most important proposal in Deliberate Metaphor Theory is that metaphors in language use are not just a matter of language and thought, as has been central in all three theories mentioned just now, but also a matter of communication. Metaphors occur in utterances that do not just display structural properties that have to do with linguistic form (e.g., simile versus metaphor) and conceptual structure (e.g., novelty versus conventionality, or relation-based mapping versus attribute mapping) but also display structural properties that have to do with their interactive function. The crucial issue in this regard is the possibility of using them *as* metaphors, as communicative devices that suggest to the addressee that the sender wants them to set up a cross-domain comparison as part of the referential meaning of the utterance. Examples include any explicit invitation on the part of a sender to

compare X to Y while X and Y come from different domains (*Shall I compare thee to a summer's day?*), any explicit statement on the part of a sender that X is like Y while X and Y come from different domains (*You are like a hurricane*). These are deliberate metaphors in that they are intentionally used as devices for people to do on line comparison between referents from two distinct domains, with the overt suggestion that such a comparison will trigger an 'alien' perspective on the target domain referent that the utterance is about. The utterance requires comparative processing, or cross-domain mapping, for its point to be understood: how can a lover be like a summer's day, or a hurricane?

This is not the case for non-deliberate metaphor use: in all examples coming from Conceptual Metaphor Theory, the main idea is that the source domain does not play a separate role in the referential meaning of the utterance, but that before then, source domain concepts are mapped on to relevant target domain entities that are the only ones to emerge in the referential meaning of the utterance. Thus, classic lexical illustrations of ARGUMENT IS WAR or LOVE IS A JOURNEY, two conventional conceptual metaphors described in Conceptual Metaphor Theory, all involve words presumably coming from the domains of 'war' or 'journey', but the referential meaning of the utterances they partake in is about 'argument' or 'love' only: people are not asked to think about war before they can understand *Lakoff attacked Glucksberg* or *We're at a crossroads*, at least not in the same way for setting up a referential representation of these utterances as with *Shall I compare thee to a summer's day* or *You are like a hurricane*.

Since the non-deliberate metaphors illustrated by Conceptual Metaphor Theory outnumber the deliberate ones as illustrated above by far, it is possible that most metaphor is processed without this type of referential online comparison. Most metaphor might even be processed with any comparison or cross domain mapping, instead being processed by lexical disambiguation (Giora 2003). Perhaps it is only deliberate metaphors in language use that require and exhibit processing by comparison across two distinct conceptual domains, or metaphorical processing.

Be that as it may, the relevance of these ideas for the structure and processing of visual metaphor is clear. Deliberate Metaphor Theory adds a third dimension to the two-dimensional model of the Career of Metaphor Theory and the prior models it comprises, and sees metaphor in utterances in discourse as a matter of language, thought and communication – or expression, conceptualization, and interaction. This raises the question how deliberateness as a communicative feature of metaphor may be conceptualized and researched in the structure and processes of visual metaphor, and how it relates to the conceptual and expressive properties of visual metaphor. It also raises the question how all of this may be compared with metaphor in language, and how this new psycholinguistic perspective on metaphor may be further developed and researched.

Deliberate metaphor in language use has been defined at the level of reference, where attention is required on the part of the addressee to the source domain as providing at least one or more referents from a different domain than the target. It is an intentional aspect of deliberate metaphors in language that they draw the addressee's attention to this different domain of reference. This does not happen for non-deliberate metaphors, where unconscious access to source domain word senses and concepts may be at play, but where these do not lead to setting up source domain referents in the state of affairs or situation model for the utterance, which is restricted to target domain referents only.

If non-deliberate metaphor defined in this way is completely predominant in language use, the opposite seems to hold for visual metaphor. Almost all visual metaphors discussed in this book have a referent from some source domain in two-model metaphors, or an incongruous referential property for one-model metaphors (Schilperoord this vol.; Šorm and Steen this vol.). The cartoon with Dutch politician Geert Wilders in the director's seat at a film set, for instance, sets up a situation with various incongruous referents coming from two domains or cognitive models, as was made explicit by VISMIP (Šorm and Steen this vol.). And the same holds for the advertisement for auto tyres that has an incongruous drain in the road, discussed by Schilperoord (this vol.). The one-model poster for the campaign reading discussed by Schilperoord also draws attention to the lack of an expected referent, the mouth, which has the metonymic function of evoking the topics someone can talk about. All of these are typical visual metaphors and they are all deliberate in that they draw attention to some incongruous referent that needs to be incorporated by on line comparison, which seems to be typical too.

All of this would fit squarely with Schilperoord's (this vol.) emphasis on intentional design of the visual metaphors discussed in this book. There is artful intention on the part of the designer, who deliberately constructs the kinds of visual incongruities that we have attempted to analyze here. This type of deliberate metaphor design is not what happens in most metaphor use in language, which is typically a matter of regular language use – only when explicit comparisons are used such as the ones mentioned above do we get a comparable experience of design involving two referential domains, as in poetic lines, headlines, or lines in advertising. Visual metaphor seems to be typically deliberate because of this genre background, whereas metaphor in language seems to be typically non-deliberate.

This genre aspect may also play a role in the tendency for visual metaphors to be rather innovative and creative – this is what designers are supposed to do when they spend time on creating their artwork. This may account for the fact that many visual metaphors are either novel and ad hoc or make use of conventional metaphors in innovative ways. Our corpus work on metaphor in language shows the opposite tendency: an estimated 99% of all metaphor in usage is conventional,

the required contextual meanings motivated by metaphor from some more basic sense (e.g. 'severely criticize' from 'attack') is part of the lexicon of a language and can be found in a modern users' dictionary. Visual metaphor seems to be typically novel whereas verbal metaphor is typically conventional and familiar.

All of this is also compatible with what Schilperoord (this vol.) and Hodiamont et al. (this vol.) call iconic processing. On the one hand, it seems inevitable in visuals that viewers engage in iconic processing, seeing the incongruous element for what it really is: we see a director's set, we see a drain, we see that a mouth is missing. At the level of reference, we see all of these source domain elements as separate, incongruous referents, perceiving them for what they are. On the other hand, these then need to be integrated into the overall picture by cross-domain mapping for visual metaphor because we attribute what Schilperoord calls an expressive function to them: they obviously are meant to 'mean' something else, and it is our job to work out what that is. This is why visual metaphors have a typical puzzle-like quality, which verbal metaphors typically do not have – unless they are deliberate and set up similarly puzzling cross-domain mappings between incongruous referents, which is not typical of metaphor in language use as it has been studied over the past four decades.

Both of these suggestions also accord with the difference in Deliberate Metaphor Theory between direct and indirect metaphor in language use (see Šorm and Steen this vol.). Utterances like *Shall I compare thee to a summer's day* and *You are like a hurricane* are metaphorical but feature a direct expression of the source domain: the lover is compared to a summer's day or a hurricane that is directly there as a referent in the utterance. This is to be contrasted with *Lakoff attacked Glucksberg*, where the source domain is indirectly present (if at all) in the utterance and needs to be reconstructed by additional analysis, which has been subjected to critical debate by various researchers (cf. Steen 2007 for an overview). It seems as if visual metaphors typically work like direct metaphors, their source domain concepts being directly expressed and therefore working as distinct referents in the picture. We do not reconstruct the director's set from a picture that is presumably about parliament and the Islam debate, we in fact see a director's set directly and need to resolve a puzzle that projects a coherent situation about Wilders in the Islam debate. Most visual metaphor may therefore not only be typically deliberate and novel but also typically direct, while most metaphor in language is typically non-deliberate, conventional and indirect.

In summary, the structure and process of visual metaphor may be approached with new vigor along the lines sketched out above. The research in this book has shown that adopting a psycholinguistic approach can be highly fruitful for both structure and process research. New developments in linguistic metaphor research that set up a three-dimensional model for metaphor (in expression,

conceptualization and communication) in connection with the genre events metaphors participate in are equally compatible with this psycholinguistic approach and raise new questions about visual metaphor. The comparison and contrast with verbal metaphor may only serve as a further platform on which to refine our research in attempting to answer these questions.

References

Bowdle, B. F. & Gentner, D. (2005). The career of metaphor. *Psychological review*, 112(1), 193–215. https://doi.org/10.1037/0033-295X.112.1.193

Gibbs, R. W. (Ed.) (2008). *Cambridge handbook of metaphor and thought*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511816802

Giora, R. (2003). *On our mind: Salience, context and figurative language*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195136166.001.0001

Glucksberg, S. (2001). *Understanding figurative language: From metaphors to idioms*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195111095.001.0001

Hodiamont, D., Hoeken, H., Van Mulken, M. (This vol.). Conventionality in visual metaphor.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago [etc.], University of Chicago Press.

Lakoff, G. & Johnson, M. (1999). *Pilosophy in the flesh*. New York: Basic Books.

Pragglejaz Group (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1), 1–39. https://doi.org/10.1080/10926480709336752

Schilperoord, J. (This vol.). Ways with pictures: Visual incongruities and metaphor.

Šorm, E. and Steen, G. J. (2013). Processing visual metaphor. *Metaphor and the Social World* 3(1), 1–34. https://doi.org/10.1075/msw.3.1.01sor

Šorm, E., and Steen, G. J. (This vol.). VisMIP: Towards a method for visual metaphor identification.

Steen, G. (2007). *Finding metaphor in grammar and usage: A methodological analysis of theory and research*. Amsterdam: John Benjamins. https://doi.org/10.1075/celcr.10

Steen, G. J. (2008). The paradox of metaphor: Why we need a three-dimensional model for metaphor. *Metaphor and Symbol*, 23(4), 213–241. https://doi.org/10.1080/10926480802426753

Steen, G. J. (2011). The contemporary theory of metaphor – now new and improved! *Review of Cognitive Linguistics* 9(1), 26–64. https://doi.org/10.1075/rcl.9.1.03ste

Steen, G. J. (2015). Developing, testing and interpreting Deliberate Metaphor Theory. *Journal of Pragmatics*. https://doi.org/S0378216615000946

Steen, G.J. (2017) Deliberate Metaphor Theory: Basic Assumptions, Core tenets, Remaining Issues. *Intercultural Pragmatics* 14(1), 1–24.

Steen, G. J., Dorst, L., Herrmann, B., Kaal, A., Krennmayr, T., & Pasma, T. (2010). *MIPVU: A method for linguistic metaphor identification in natural discourse*. Amsterdam etc.: Benjamins. https://doi.org/10.1075/celcr.14

Van den Heerik, R., Šorm, E., and Steen, G. J. (This vol.) Behavioral evidence for VisMIP.

Van Weelden, L., Maes, A. A., and Schilperoord, J. (This vol.). How visual form affects metaphorical conceptualization: The role of shape similarity.

Index

Metaphor has recently been reconceptualised as a fundamental part of the human conceptual system. It can hence be expressed in language but also in other modalities and media of communication, including gesture and body language, sound and music, and film and visuals. In spite of this theoretical landslide, however, the wide range of nonverbal metaphor and its processing has neither been empirically investigated on the same scale nor with the same rigour as metaphor in language. The overarching goal of this book is to report on the findings of a research program aimed at exploiting the vast cognitive linguistic and psycholinguistic expertise on metaphor in language for a new, behaviourally founded approach to the structure and processes of metaphor in one of these nonverbal manifestations, namely static visuals. The book presents concepts and methods for the identification and analysis of metaphor in document structure as well as new approaches to the study of visual metaphor processing. Its results are intended to further the development of an encompassing and robust cognitive-scientific theory of metaphor by including visual metaphor while also enriching our understanding of the communicative possibilities and effects of visual metaphor in multimodal discourse.

JOHN BENJAMINS PUBLISHING COMPANY