Uncanny resemblance. Words, pictures, and conceptual representations in the field of metaphor

Uncanny Resemblance. Words, pictures, and conceptual representations in the field of metaphor

Abstract: What is the relation between the three following elements: words, pictures, and conceptual representations? And how do these three elements work, in defining and explaining *metaphors*? These are the questions that we tackle in our interdisciplinary contribution, which moves across cognitive linguistics, cognitive sciences, philosophy and semiotics. Within the cognitive linguistic tradition, scholars have assumed that there are equivalent and comparable structures characterizing the way in which metaphor works in language and in pictures. In this paper we analyze contextual visual metaphors, which are considered to be the most complex ones, and we compare them to those that in language are called indirect metaphors. Our proposal is that a syllogistic mechanism of comprehension permeates both metaphors expressed in the verbal modality as well as metaphors expressed in the pictorial modality. While in the verbal modality the metaphoric syllogism is solved by inference, we argue that in the pictorial modality the role of inference is performed through mental imagery.

Introduction

The study of metaphor in language within the cognitive linguistic tradition, precedes by a couple of decades the study of metaphor in images. As a consequence, the latter field of research is often influenced by methods and theories borrowed from the former field of research. Such tools are adjusted and applied to the analysis of multimodal contexts, under the assumption that there are equivalent and comparable structures characterizing the way in which metaphor works in language and in images (equipollence hypothesis, Mairal and Ruiz de Mendoza, 2009: 154). This assumption, however, has to be taken with caution, especially when compared findings reported in various cognitive and neuro-scientific studies, on how humans process images and language. In particular, while some authors claim that the semantic information derived from pictorial and verbal stimuli is processed in much the same way (e.g., Caramazza, 1996), others argue that the semantic information encoded in pictures and words is processed differently and along two functionally independent but still interconnected cognitive systems. This allows the creation of multiple modality-specific semantic representations (e.g., Paivio, 1971; 2010; Glaser, 1992). This hypothesis is based on empirical studies showing that, for example, phonological/orthographic information and lexical variables such as word frequency play a little role in image processing (Taikh et al., 2015). Moreover, pictures seem to trigger faster and deeper emotional response compared to words (DeHouwer and Hermans, 1994; Hinojosa et al., 2009). In addition to the similarities and differences observed in how images and words are processed, these two semiotic systems (the pictorial and the linguistic one) appear to categorize experience in different ways, on the basis of different features. Each of these two modes appears to have specific strengths and specific weaknesses in the type of information that can be encoded and represented (Author 2016; 2017; 2018). For example, while entity-related perceptual features like colors and shapes are more easily, straightforwardly, and accurately represented within the pictorial system, temporal dynamics and contingencies are better represented within the linguistic semiotic system.

In this contribution we elaborate a series of theoretical observations aimed at illustrating similarities and differences between these two modes of representation (words and images) in relation to their ability to construct and represent metaphor, a cognitive phenomenon that bridges between modes of expression (i.e., pictorial, linguistic, etc.) and conceptual structures (Lakoff and Johnson, 1980), as well as between semiotic structure (how the metaphor can be represented) and cognitive processing (what type of operations are involved during its comprehension and production).

Visual metaphors as elliptic syllogisms

Let us try to roughly define what a visual[1] metaphor is. The scientific literature proposes a variety of definitions (e.g., Forceville, Urios-Aparisi 2009; Perez Sobrino 2017; Steen 2018), from which the following formulation can be derived: visual metaphors are highly constructed images that present perceptual incongruities; such unexpected elements stimulate the viewer to attempt an alternative (figurative) reading, and construct a cross-domain comparisons to make sense of the image. As the reader might appreciate, such definition encompasses elements related to the structure of typical visual metaphors (i.e., the presence of perceptual incongruities) as well as to the processing of visual metaphors (i.e., the fact that the viewer is stimulated to construct cross-domain comparisons).

[1] Within the cognitive linguistic tradition, metaphors expressed within the pictorial semiotic system can be broadly defined as *visual* metaphors, while within the semiotic tradition metaphorical images are often defined as *pictorial* metaphors (e.g., Author et al. 2018). Moreover, while within the cognitive linguistic discipline the various ways in which metaphors are manifested are defined as *modalities* of expression (visual, verbal, etc.), within the semiotic view these are often defined as *modes*. The terminology used to define these images is therefore very much tied and dependent on the discipline in which such phenomena are discussed. Nonetheless, recent attempts have been made to clarify and distinguish various aspects involved in the semantics of such terminology. For example, Forceville and Urios-Aparisi (2009) proposed to distinguish between metaphors in images that encompass only one semiotic system (the pictorial one, like in many artworks), and metaphors in images that encompass more than one semiotic system (for example the pictorial and the verbal ones, like in many advertisements). While the first type of tropes is defined monomodal, the latter type is defined multimodal. For a recent discussion on this matter, refer to Author (accepted).

While the body of literature on visual metaphor structure is well developed, we here focus on the latter aspect, which is, theorizing the mental operations that underlie the construction of a cross-domain comparison, starting from visual metaphors.

In doing so, we aim at highlighting the mechanism by means of what an interpreter of visual metaphors grasps the message conveyed by the pictures. According to Lakoff and Johnson, if a metaphor is not a matter of language but a matter of thought (Lakoff & Johnson 1980), then the mental process behind metaphors should be (at least partially) the same for the interpretation of both, verbal and pictorial metaphoric expressions. What kind of mechanism are we talking about? Here we assume, with Arthur C. Danto, that metaphors functioning is similar to that of a syllogism (Danto 1981: 169-171). In a syllogism, indeed, we have two premises (P1 and P2), through which is possible to infer a conclusion (C).

The most famous example of syllogism is the following:

(P1) If all men are mortal;
(P2) and Socrates is a man;
(C) then Socrates is mortal.

According to Danto a metaphor is like a particular type of syllogism – an elliptic one – where either one premise *or* the conclusion is omitted. Referring to Aristotle, Danto claims the metaphor to act like an enthymeme. What kind of features should an enthymeme have? In Aristotle's *Rhetoric*, the enthymeme is said to be the syllogism par excellence (*Rh.*, 1356b 5). It works just like any other syllogism but, if one of the premises is well known, the speaker omits it, leaving the completion task to the hearer (*Rh.*, 1357a 15-25). Following this description, Danto argues (Author 2017) that an enthymeme is a truncated syllogism lacking a premise or the conclusion, and «it yields a valid syllogism when, in addition to meeting the usual conditions of syllogistic validity, the missing line is an obvious truth […]: a banality» (Danto 1981: 170).
In the case of an enthymeme like "if you are a man, then you are mortal", we have the following scheme:

[(P1) Considering that all men are mortal],
(P2) If you are a man,
(C) then you are mortal.

As we can see, the implied term is P1, according to which all men are mortal.
So, if a metaphor works like an enthymeme and the blank line of an enthymeme is filled by the hearer with a banality, then also a metaphor should lie on trivial premises, (or trivial conclusions).
Adapting the scheme of the syllogism to the case of a metaphor like "Paul is a fox", the speaker says that Paul is metaphorically a fox, but she means that he is literally a smart person. The scheme is the following:

(P1) Paul is a fox,
[(P2) Foxes are sly],
[(C) Paul is sly].

As we can notice, we have two missing lines – P2 and C – that is to say one premise and the conclusion. The speaker utters only the first premise (P1) – which contains the target (Paul) and the source (a fox) of the metaphor – implying the other two lines of the argument[2]. Contrary to what Danto states, we believe that a metaphoric expression is like an elliptic syllogism with two missing terms – not only one – where a premise *and* the conclusion are omitted. We shall come back to this point at the end of the section.

[2] It should be further noted that P2 is not omitted in the same way as C. The hearer infers C through the line of reasoning summed up by the syllogism, while P2 is suggested by the source "fox". Indeed, the most peculiar feature that «the man in the street» usually attributes to the foxes is slyness. Slyness is more salient for foxes than for men, so we have a salience imbalance that determines the direction of the metaphorical expression. Ortony defines salience as the weight of the distinctive and prototypical features of the source that are not really distinctive and prototypical even of the target. According to Ortony, a good metaphor is the result of this imbalance. The properties of one of the two objects that match with those of the other object have to be more salient in the source than in the target (Ortony 1979). The salience imbalance, then, leads to C. In fact, the interpreter projects the prototypical feature (slyness) from the source domain to the target domain, deducing the conclusion of the metaphoric syllogism. It is remarkable that, starting from a false premise (P1), the hearer achieves a plausible conclusion.

However, the high or low level of knowledge whereby the hearer fills the blanks depends on the hearer herself: the interpretation stops when the interpreter's knowledge stops (Danto 1981: 127). Finally, if the syllogistic structure is the underlying mechanism of metaphors, both for verbal expression and for pictorial expression, then what really changes is the knowledge applied by the interpreter to solve the puzzle. It could be useful to make a distinction between two types of visual metaphors: those that require a low or medium level of knowledge, and those that need a massive interpretation. Representative of the first category are advertising pictures and, more in general, all the pictures that, in order to be quickly understood by the interpreter, leverage the knowledge shared by a large and well-defined community of people. In the second category we can include those pictures whose deep appreciation depends on a very specific knowledge, like many ancient paintings.

Considering the first category, the message conveyed by an advertising billboard must be grasped through a glance, so it has to be based on what Max Black calls *the system of associated commonplaces* that reflects «what the man in the street thinks about the matter» (Black 1955: 287).

Fig. 1

Take for example fig. 1, the two elements involved in the picture are a cat and a white soap bar. There is a clear incongruity coming from the fact that everyone knows that kittens usually do not lick soap but milk, so the viewer is called to mentally replace the soap bar with a milk bowl. The picture relies on a commonplace which is only suggested. While the soap bar is fully depicted, the milk bowl is only evoked by the shape and color of the soap bar, and by the fact that a cat is licking it. The billboard does not explain explicitly the comparison between the soap bar and the milk bowl, but by integrating contextual knowledge and genre-related knowledge the viewer has to fill these gaps, understand that this is an advertisement for a brand of soap bars, and deduce that these products are supposed to contain nourishing and natural ingredients that hydrate our skin like milk feeds a kitten.

A picture like this is considered a visual metaphor of the contextual type. According to Forceville, we can distinguish three types of visual metaphors, based on the ways in which source (S) and target (T) are conveyed within the pictorial semiotic system (i.e., based on the structure of the metaphoric expression): i) similes, where source and target are both fully depicted in the picture; ii) hybrid metaphors, where source and target are merged into a single entity; iii) contextual metaphors, where the source or the target is absent and has to be recovered through the pictorial context (Forceville 1996; 2002). Comparing a metaphor to an elliptic syllogism, it appears that in a contextual visual metaphor we have three ellipses, or missing terms: i) the target *or* the source; ii) the implicated message; iii) a missing premise that leads to the implicated message.

Van Mulken et al. have shown that the level of complexity of the pictures affects the degree of appreciation (Van Mulken, Le Pair, Forceville 2010). Contextual visual metaphors are perceived to be more difficult and they are on average less appreciated by viewers than hybrid visual metaphors (Van Mulken, Hooft, Nederstigt 2014).

In this paper we analyze precisely the contextual type, which are considered to be the most complex ones. We chose this specific type of visual metaphor because it seems to correspond to those that in language are called *indirect* metaphors, which are also the

most frequently used configuration within the verbal system (Steen et al 2010). An example of indirect metaphor in language is the following: "John *devoured* the new Rowling book", where *devoured* is used metaphorically. Within the context provided by the sentence, devoured stands for *read quickly and avidly*, or similar formulations. The context in both, this verbal example, as well as in the visual metaphor described above, relates to one of the two metaphor domains, while the metaphorical word/entity relates to the other metaphor domain. In the verbal example, the context of the sentence relates to T (the domain of reading), while the basic meaning of the metaphorical word, *devoured*, relates to S (eating). In the visual example, the context of the image relates to S (the domain of kittens), while the metaphorical entity relates to T (the domain of soap bars). In both cases is the context that (to different extents[3]) constructs a conflict with the metaphorically marked entity/word, and invites the reader/viewer to construct a figurative interpretation of the conveyed message.

In understanding metaphors expressed in the verbal mode, the hearer has to fill all the gaps left by the speaker, that is to say the missing terms (one of the premises and the conclusion of the reasoning implied by the metaphor). The speaker suggests something to the hearer by means of a «covert form of information transmission» (Sperber, Wilson 1986: 30). Making a metaphor, the speaker may want to send information to the hearer not by explicitly saying something but through some clues that trigger a set of inferential processes. In doing so, the speaker reveals her intention to send information avoiding a wholly overt communication.

In this paper we are going to explore what mechanisms are used to cue a comparison within the pictorial mode, and how these transform, when we move from the pictorial expression to the conceptual dimension.

[3] An in-depth overview of the different degrees of conflict involved in the processing of such expressions lies beyond the scope of our current study. Various theories have been proposed to explain when and how the metaphorical and the literal meaning are activated during the comprehension of metaphorical statements; a recent review and discussion of the various models proposed in the literature can be found in Cuccio (2018).

Our conviction is that, when understanding a visual metaphor, mental imagery plays the role that the inference plays in comprehending metaphors expressed in the verbal mode. Uttering a metaphorical statement, indeed, the speaker says that:

(P1) Peter is a lion,
[(P2) Lions are brave],
Which implies that:
[(C) Peter is brave].

The hearer has to infer C starting from P1, which contains both the target (Peter) and the source (a lion) of the metaphoric expression. The source is the key for P2. Through the source, indeed, the hearer gains the second premise, and finally, after the mapping from the source domain to the target domain, she achieves the conclusion.
As we have seen in the case of indirect metaphors, it is not always so simple to detect the target and the source. Let us consider "John devoured the new Rowling book" again. One needs a further effort to grasp the target domain from the context and then solve the related syllogism:

(P1) Reading is eating,
[(P2) Eating is putting something in the body],
[(C) Reading is putting something in the body].

Something similar happens for visual metaphors such as the image of the kitten (fig. 1) and fig. 2.

Fig. 2

Here we see alcoholic drinks (T, the target) taking the place of bullets (S, the source) in the gun cylinder and we deduce that drinking kills (C, the conclusion). How do we deduce S and C in the Figure 2?

Let us try to adapt the scheme of a syllogism to the visual metaphor.

(P1) Alcoholic drinks are bullets
(P2) Bullets kill

(C) Alcoholic drinks kill

Interestingly, none of the three lines is really depicted. P1 is incomplete, since bullets are omitted, while the achievement of P2 and C is left to the viewer's ability. The viewer has not only to infer the conclusion, but also to recover P2 and part of P1, like the hearer does in the case of an indirect metaphor. Our claim is that she performs the task by means of visual mental imagery.

Before investigating the relation between the pictorial and the verbal expression and the role of mental imagery in the comprehension of a visual metaphor, we would like to say something on another kind of connection, which is extremely important when talking about pictures: the relation between the visual metaphor-maker (or picture-maker) and the viewer.

The relationship between the picture-maker and the viewer

There must be a mutual agreement between the picture maker and the viewer in order to establish the presence or the absence of visual incongruences in a visual metaphor (Carroll 1994). Just an example to clarify the matter: when observing an ancient Greek krater with a depicted centaur, the viewer can't consider the mythological creature as an incongruence made up of a human part and a horse part, because the purpose of the picture maker was not to depict an incoherent subject, but to present the two elements – the human one and the feral one – as physically co-existing in a single organism. This is similar to what happens for chimeras, werewolves, the basilisk or the sphinx: in the case of the centaur, the cohabitation of inconsistent parts does not suggest the existence of an additional semantic dimension that can be accessed by mapping prototypical horse features onto the human element.

First of all, the picture-maker and the viewer negotiate that a picture is a visual metaphor if:

a) the picture-maker believes that the juxtaposition of different elements in her own picture depicts an incongruence and does not represent a religious or fictional character. Moreover, she has to believe that the viewer has the tools to grasp the depicted incongruence as an incongruence;

b) at the same time, the viewer believes that the picture-maker deliberately puts an incongruence in the picture[4]. She has not to interpret that juxtaposition of different and incongruous elements as belonging to a religious or fictional context, but she has to intend that incongruence as a spy of a further level of meaning.

Between the picture-maker and the viewer there has to be a relationship ruled by what Donald Davidson used to call *principle of charity*[5]. According to the philosopher, the principle of charity shapes the verbal transaction between the speaker and the interpreter during the communication. Paraphrasing Davidson, therefore, we could think of the principle as an *agreement in principle* between the picture-maker and the viewer, where the latter: i) assumes to share with the picture-maker a huge base of common beliefs; ii) holds the beliefs she ascribes to the picture-maker to be true. There could be no perfect correspondence between speaker's and interpreter's beliefs, but the important thing is that they agree with most of them, for «without a vast common ground there is no place for disputants to have their quarrel» (Davidson 1977: 244-245).

[4] In the scientific literature on metaphors in language, thought, and communication, a deliberate metaphor is defined as a metaphor that is constructed with the communicative goal of changing the reader's/listener's standpoint on a given topic (e.g., Steen 2008; 2011). While the literature on visual metaphors and deliberateness is still virtually absent, Author (under review) and Author (accepted) suggest that metaphors expressed within the visual (or verbo-pictorial) semiotic system typically display visual incongruities, and that such incongruities constitute the first step needed to attract the viewer's attention and stimulate her to stop in her track and start constructing a figurative reading of the picture.

[5] Davidson has never really defined the so-called *principle of charity*, although in various articles he provides some general indications, claiming that: «What matters is this: if all we know is what sentences a speaker holds true, and we cannot assume that his language is our own, then we cannot take even a first step towards interpretation without knowing or assuming a great deal about the speaker's beliefs. […] the only possibility at the start is to assume general agreement on beliefs. […]. The method is not designed to eliminate disagreement, nor can it: its purpose is to make meaningful disagreement possible, and this depends entirely on a foundation – *some* foundation – in agreement. […] charity is not an option, but a condition of having a workable theory […]. Charity is forced on us; - whether we like it or not, if we want to understand others, we must count them right in most matters» (Davidson 1974: 18-19). And again: «Charity in interpreting the words and thoughts of others is unavoidable […]: just as we must maximize agreement, or risk not making sense of what the alien is talking about, so we must maximize the self-consistency we attribute to him, on pain of not understanding him» (Davidson 1967: 313).

Let us consider fig. 1 again and fig. 3.

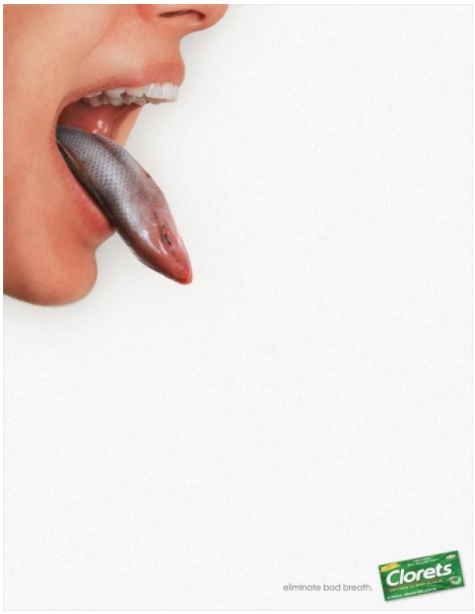

Fig. 3

The viewer has to think that: i) she shares some beliefs with the picture-maker, and ii) the beliefs she ascribes to the picture-maker are not so different from her own. From this point of view, the most relevant common belief when we talk about visual metaphors lies in the fact that two elements, usually not connected in that way, can acquire a new semantic value if they are juxtaposed within the same picture. Without this common conviction, there wouldn't be the right premise for talking about visual metaphors. Perhaps, the viewer and the picture-maker may not share the belief involving the soap to be like milk, or the tongue to smell like damaged fish, but to disagree with these beliefs they must share a lot of other beliefs. For instance, the viewer has to believe in the existence of cats and soaps, in the fact that a cat assumes that position when licking milk, which is a white drink produced by mammals, and so on, endlessly. Essentially, the agreement on a specific belief doesn't make it true, but «much of what is agreed must be true if some of what is agreed is false» (Davidson 1977: 245).

So, to grasp the meaning conveyed by a visual metaphor there is no need for the viewer to share every single belief with the picture-maker. Moreover, the fact that the viewer does not believe that a car could be wild and powerful like a horse (fig. 4) should not lead to the conclusion that the picture-maker and the viewer organize the world using different and incommensurable conceptual schemes: «Different points of view make sense, but only if there is a common coordinate system on which to plot them; yet the existence of a common system belies the claim of dramatic incomparability» (Davidson 1974: 6).

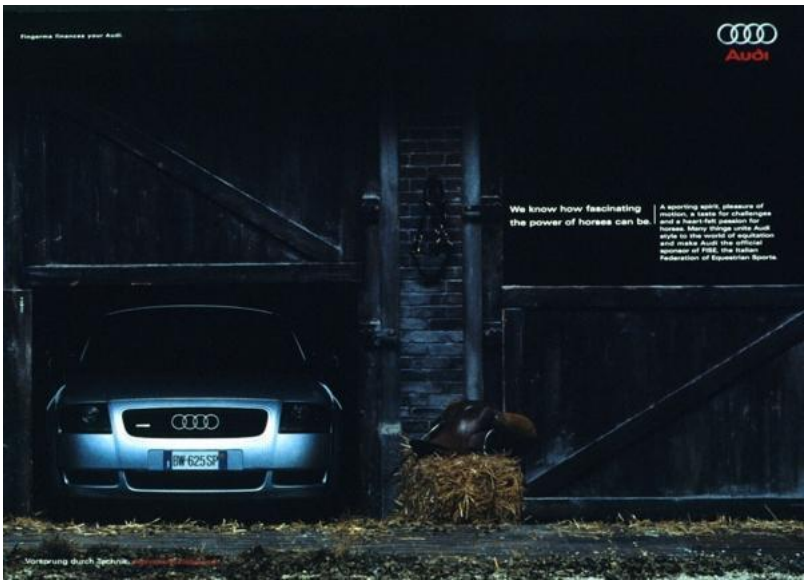

Fig. 4

A good model to describe the relationship between the picture-maker and the viewer is the one sketched by Davidson in *A Nice Derangement of Epitaphs* (1986), where he argues that, at any moment of a speech transaction, both the speaker and the interpreter have in mind what might be called a theory. From the interpreter's side, the theory is adjusted to the evidence available to her: gestures, dress, character, and whatever else can be gathered observing the speaker's behavior. When the speaker utters something, the interpreter modifies her theory, reconsidering past interpretations of particular utterances in the light of new evidences. According to Davidson, before starting the discussion the theory is called *prior theory*, because it concerns the way in which the hearer is prepared to interpret the speaker's utterances.

During the discussion, the theory is called *passing theory*, because it concerns the way in which utterances are *de facto* interpreted (Davidson 1986: 168-169).

Now, after having transferred this model to the pictures, let's try to think of the same process from the picture-maker's side. What is the picture-maker's prior theory? It is what she believes the viewer's prior theory to be. First of all, the picture-maker should take into account the representational code the viewer refers to. For example, if a painter intentionally depicts the figures on the background bigger than those on the foreground, ignoring perspective ground rules, it can mean that this painter: i) knows that the viewer will consider the deformation as a marker of something that has to be interpreted in a certain way; ii) has ignored the viewer's prior theory, which provides, among other things, that a painter complies with specific depiction rules. With regard to i), the picture-maker has to hypothesize that the viewer will interpret every deviation from the standard as a signal of the existence of a further level of meaning. The juxtaposition of incongruous elements is one of that deviation.

Interestingly, unlike the relationship between the speaker and the hearer, the relation between the picture-maker and the viewer is an unbalanced one. The picture-maker's passing theory can't be recorded, because what she offers is an immutable picture which will not be adjusted on the basis of the communication with an interlocutor (a viewer, in this case). The viewer's passing theory, on the contrary, can be adjusted and recalibrated only according to picture-maker's beliefs, which do not come directly from her, but from the picture itself. Thus, the picture provides a starting point for the viewer, and moving backwards from it a viewer can trace the intentions of the picture-maker. Nevertheless, the picture-maker is required to guide the heuristic operation, trying to conduct the viewer through a proper exploration of the picture.

As we have already seen in the first section, our claim is that mental imagery plays the role that the inference plays in understanding metaphors expressed in the verbal modality. Before going ahead, therefore, we need to define what we mean when we talk about mental imagery and why it is important for visual perception.

Visual perception and mental imagery

Mental imagery refers to «our ability to reactivate and manipulate visual representations in the absence of the corresponding visual stimuli […]. Visual mental imagery is employed when one answers questions, e.g., about the shape of the tail of one's dog, in the absence of the corresponding visual stimulus» (Ganis & Schendan 2011: 239). Kosslyn et al. had previously formulated a similar definition in the following terms: «Visual mental imagery is 'seeing' in the absence of the appropriate immediate sensory input […]. Imagery is distinct from perception, which is the registration of physically present stimuli» (Kosslyn, Behrmann & Jeannerod 1995: 1335).

Visual mental imagery is essential in visual perception. Bence Nanay expresses an interesting point about the relation between mental imagery and perception. Nanay argues that mental imagery is necessary for some forms of perception, like amodal perception, defined as the representation of the occluded parts of perceived objects. We perceive a part of an object *amodally* if we do not receive any sensory stimulation related to it. Nanay claims that if we are looking at a cat whose tail is not visible because it is occluded by a picket, we represent the tail by means of mental imagery. He explains that this could be argued on empirical basis. First, modal completion (the ability to perceive as distinct objects those areas that don't exhibit any difference of color and shape if compared to the surrounding areas, see Kellman & Shipley 1991: 143-144) and amodal completion depend on the same neural mechanisms. Secondly, neuroscientists noted that looking at the sides of the Kanizsa triangle (fig. 5) and visualizing objects in our mind activate the same patterns in the primary visual cortex (Lee & Nguyen 2001: 1907; Kosslyn, Thompson & Albert 1995: 496-498). Now, it is known that, following Michotte's account, "seeing" the Kanizsa triangle relies on modal perception, thus there is a very close link between modal perception and visual mental imagery. If modal and amodal perception depend on the same mechanisms, then also amodal perception and visual mental imagery share the same neural processing. Finally, according to Nanay, this means that representing the occluded parts of perceived objects is a form of mental imagery (Nanay 2010, 2016b). Since we move in a non-transparent world, we experience amodal perception continuously in our everyday life, therefore we represent those non-visible part of objects through visual mental imagery again and again: «[W]hat we take to be perception is really a

mixture of two things: sensory stimulation-driven perception and amodal completion» (Nanay 2018: 8).

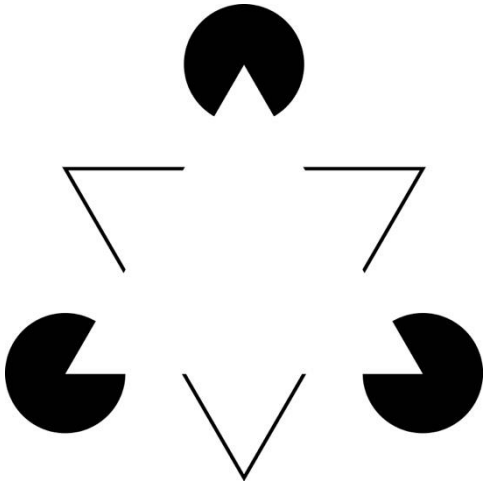

Fig. 5

So, if mental imagery is so pervasive, then it is also fundamental in perceiving, appreciating and interpreting pictures, as well as, for our scopes, visual metaphors. But, how?

In the next sections we will see which features of a visual metaphor trigger our mental imagery. They are: i) the sharing of the occlusion shape; ii) quasi-perceptually represented *aesthetically relevant properties*.

Invariance Principle, occlusion shape and occlusion size

George Lakoff argues that during the mapping from source to target, conceptual metaphors preserve a sort of structural scheme. Obviously, the source's structure is not wholly projected onto the target, otherwise we would have a full identification between the two domains. Lakoff calls this structural layout *image-schema structure*, where image schemas can be defined as recurring structures within our cognitive processes, formed from our bodily interactions, which can motivate patterns of our understanding and reasoning (e.g., Johnson 1987).

If we consider a verbal metaphoric expression such as "during the debate John attacked Paul, but Paul defended his position quite well", and the underlying conceptual metaphor ARGUMENT IS WAR, the image-schema structure of WAR involves force dynamics related to the defeat of an enemy. Such schemas are mapped without any changes onto the target domain if and only if there are correspondences between certain selected features of ARGUMENT domain and certain selected features of WAR domain. Lakoff calls this principle the *Invariance Principle*. It further establishes that, if the metaphorizing term is an object that can act as a container, its external shape will be conceptually mapped onto the target "exterior", while interiors will be mapped onto interiors, and so on (Lakoff 1993: 215).

While Lakoff attributes the Invariance principle to the structure of conceptual metaphors which he derived from the observation of verbal metaphoric expressions, we hereby explore to what extent is this principle applicable to metaphors expressed within the pictorial mode, commonly called visual metaphors. In fig. 1, the source-schema structure mapped onto the target includes what John Hyman would define the *occlusion shape* of the milk plate, which is the smallest mark one would need to make on a sheet of glass, placed between the observer and the plate, in order to completely hide the plate (Hyman 2006: 75-76). The sharing of the occlusion shape is a peculiar device of many contextual visual metaphors. In fig. 6 the hose shares its occlusion shape with a snake coiled around a limb, while in fig. 3 a fish shares the occlusion shape with the tongue. Adapting Lakoff's Invariance Principle to this latter picture, then, the external outlines of the fish will correspond to the external outlines of the tongue. In this case, therefore, what is maintained invariant between the source and the target is the occlusion shape, which can be compared to an image-schematic representation that characterizes a conceptual metaphor derived from a visual metaphoric expression, as opposed to a verbal metaphoric expression.

Fig. 6

As we just said, in a visual metaphor target and source share the occlusion shape. However, the occlusion shape of one of the two terms has to be often modified in size in order to recall the occlusion shape of the other term. Let's consider fig. 7.

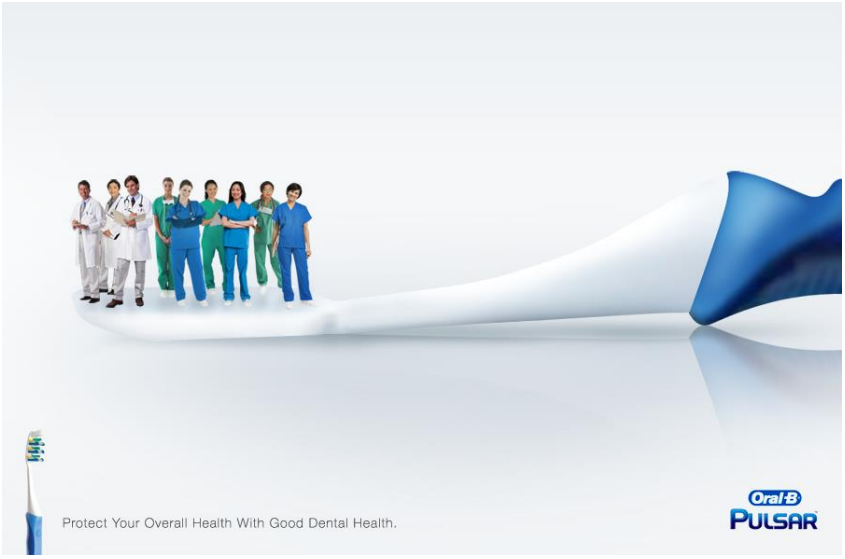

Fig. 7

The bristles of the toothbrush are much smaller than a group of people, so the occlusion shape of the group of people has to be caught from a different point of view to perfectly overlap the occlusion shape of the bristles. The size of the people depends on the position of the eye along the line of sight which ideally links the viewer and the object. Hyman calls this variable the *occlusion size* (Hyman 2006: 98). In a visual metaphor, one of the two terms usually modifies its occlusion size in order to fit the shape of the other term. However, bristles are not really depicted in the picture, they are only suggested by the people's silhouette. Whereas in the linguistic mode the two terms are not suggested by parts or details, because their role as target or source is established by the syntax, in the pictorial mode we have to define roles by means of visual equivalent of some figure of speech, like synecdoche or metonymy. Sometimes, pictures suggest that observers have to interpret the most complete depicted term as the metaphor target, other times happens the opposite and the metaphor target becomes the incomplete or absent element. Relying on Forceville's categories, as reported in the first section, we can say that there are visual metaphors that work *in praesentia*, showing both target and source, and visual metaphors that work *in absentia*, showing only one term. In this second case, there are no parts of the entity represented, the viewer has to infer the entity by collecting and integrating information retrieved from the context.

Aesthetically relevant properties represented by mental imagery

Nanay explains aesthetically relevant properties (ARP) in this way: «properties are aesthetically relevant if attending to them makes an aesthetic difference» (Nanay 2016a: 67). In relation to pictures, an ARP induces a new experience of the picture itself. Consider the *Allegory* painted by Agnolo Bronzino in 1546 for the king of France Francis I (fig. 8).

Fig. 8

At first glance the picture seems a beautiful scene of love, with Venus and Cupid in the foreground. Let's assume we don't know much about this artwork. If we look on the right side of the painting we see an apparently nice girl in a green dress, who however hides a reptile-like body, lion's claws and the tail of a serpent. She's offering a honeycomb with one hand while she is using the other to hide a look-alike scorpion tail. If we look carefully, we can see that her hands are inverted (fig. 9): attached to the right arm is the left hand, and attached to the left arm we find the right hand. With what would normally be the "good" hand, that is indeed the "evil" hand, she offers a sweet treat, while hiding some poison in what is really the "good" hand. In short, she's lying. After we discover the hand inversion trick, the painting immediately appears different: it no longer looks like a beautiful scene of love. So, the inversion of the hands is an aesthetically relevant property. In fact, according to the art historian Erwin Panofsky, who based his analysis on sixteenth-century emblem books, the figure of the young girl is a crasis of Deceit, Hypocrisy and Fraud allegories (Panofsky 1939: 86-91). This is a case of ARP depicted in the painting, but not all these properties are

actually visible, some must be represented by the viewer through mental imagery (Nanay 2016: 87).

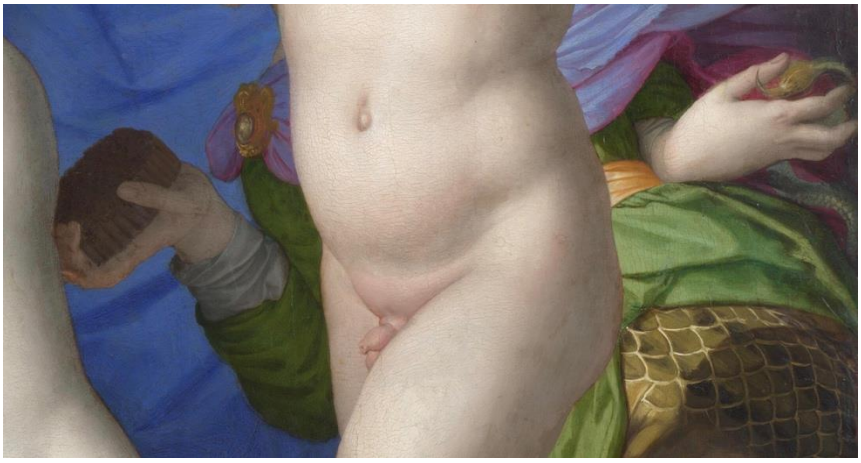

Fig. 9

Let's consider a painting like *A Bigger Splash*, by David Hockney (fig. 10). It depicts a splash in a Californian swimming pool. Let us stress that it is a "splash" and not a "dive". We do not see anyone diving in the swimming pool. We only see the final result, that is to say, splashing water. We have to represent by mental imagery that a few seconds before someone dove into the pool, which probably is, in the context of the painting, the most relevant moment.

Fig. 10

Imagery view of visual metaphors

The interpreting process of a visual metaphor could be summed up as follows. During the first phase, the interpreter looks at a picture like fig. 2. In doing so, on the basis of her beliefs, she has some expectations: in this case, she expects to see a gun with bullets. However, expectations are disappointed by the picture itself, which exhibits a clear incongruity, that is to say, alcoholic drinks in place of bullets. If the verbally encoded message of the visual metaphor is "drinking kills", what are the two elements which act as target (T) and source (S)?

They are, respectively, the alcoholic drinks (T) and the bullets (S). There are two clues that allow the viewer to recognize the bullet as the source of the visual metaphor: one visual clue and one contextual clue. The visual one is suggested by the sharing of the occlusion shape between the bottle and the bullet, while the contextual one is provided by the general pictorial context, i.e. bottles are not scattered on the ground, but they take the place of the bullets. Bullets are, in this case, an *aesthetically relevant property*, that the interpreter represents by means of visual mental imagery (VMI). VMI is triggered by the occlusion shape and the pictorial context. So, if the bottle is the target, the interpreter completes the visual metaphor providing the source thanks to VMI. The imbalance between the mental image of the bullet and the real picture with bottles triggers the interpreting process.

During the second phase, the viewer is involved in the genuine interpreting process. At this stage, she forms a new mental image, more complex than the previous one, that allows her to visualize the interactions between target and source, recognizing a lot of implicatures (not only "the alcoholic drink is a bullet", but also "drink kills like a gun does", and so on). This second order visual mental imagery provides some inspectable and re-interpretable mental images, which play a fundamental role in cognitive processes. According to Ronald Finke, visual mental images can be reinterpreted. More specifically, the content they express never corresponds to the one

used to create them. Reinterpretation adds more information that is the essence of their cognitive peculiarity (Finke , Pinker & Farah 1989). The process begins from interpreted, encoded, and memorized contents (the concepts of "smoke", "disease", "bullet", "gun", and so on). When these contents take form in the mental image, they then produce something new, which is similar to a percept and enables some of the visual system processes.

Several experiments have shown that there is an overlap between visual mental imagery and visual perception. If mental images and perceived stimuli are represented in the same way, then they can be processed in a similar manner (Borst & Kosslyn 2008). Probably this is the key for starting a discussion on the role of mental images in the interpretation of such complex pictures as visual metaphors.

Conclusions

Starting from the assumption that there are equivalent and comparable structures characterizing the way in which metaphor works in language and in images, we tried to find the middle term between words and pictures. Our proposal is that the middle term lies in a syllogistic mechanism of comprehension that permeates metaphors expressed in the verbal modality (for the sake of brevity called verbal metaphors) as well as metaphors expressed in the pictorial modality (for the sake of brevity called visual metaphors). Saying that A is metaphorically B, the speaker implies that B has the peculiar feature C, suggesting that, if A is metaphorically B, then A has literally the feature C, which fits as well for A as for B. To give a real example: saying that Paul is a fox, the speaker makes two operations: (1) she assumes that the most characteristic feature of a fox is (supposedly) slyness; (2) she suggests that, if Paul is metaphorically a fox, then he is literally sly, since slyness fits well for Paul as well as for foxes.
We argue that a verbal metaphor is like an elliptic syllogism with two missing terms: one premise and the conclusion. The three lines of the metaphoric syllogism are the following, where P1 and P2 are the premises and C the conclusion:
(P1) Paul is a fox,
[(P2) Foxes are sly],

[(C) Paul is sly].

As we can see, P2 and C are the two missing terms: the hearer has to provide P2 in order to achieve C.

Considering "Paul" as the target and "fox" as the source of the metaphor, it is not always so simple to recognize the two domains involved in a metaphoric expression. There are metaphors, called *indirect* metaphors, where either the target or the source has to be recovered through the context. Indirect metaphors are very similar to contextual visual metaphors, indeed the source or the target is absent, and the viewer has to gain them analyzing the pictorial context. We focused our analysis on this particular type of visual metaphors. Both indirect metaphors and contextual visual metaphors present this sort of syllogistic scheme. Considering T as the target and S as the source:

(P1) If T is S,

(P2) and S presents the feature F,

(C) then also T presents the feature F.

In both modalities, verbal and pictorial, only P1 (which lacks T or S) is presented to the interpreter. The two other lines, P2 and C, require some cognitive operations to be recovered, as well as T or S.

How do verbal and visual metaphors suggest the lacking domain T or S? A first and simple answer is that they do this, in different ways, through the context. In a metaphor like "John devoured the new Rowling book", we have seen that within the context provided by the sentence, "devoured" stands for *read quickly and eagerly*. A contextual visual metaphor suggests T or S by means of specific pictorial devices. One of these devices is the sharing of the so-called occlusion shape (which has several similarities with Lakoff's Invariance Principle). The occlusion shape is the smallest mark one would need to make on a sheet of glass, placed between the observer and the object X, in order to completely hide the object X. In fig. 1 the occlusion shape of the soap bar evokes that of a bowl of milk, in fig. 6 the occlusion shape of the hose evokes that of a snake, while in fig. 3 the occlusion shape of the fish recalls that of a tongue, and so on.

The missing term in a contextual visual metaphor acts like an aesthetically relevant property. This means that, if attending to it, the viewer changes the way in which she

looks at the picture. If we did not represent the bowl of milk, the snake, or the tongue, these pictures would only be weird photographs. On the contrary, they become visual metaphors (also) because the picture-maker knows that the viewer will represent them and attend to them as aesthetically relevant properties. Moreover, we have seen that there are aesthetically relevant properties depicted in the picture and aesthetically relevant properties only suggested by the picture-maker, that the viewer has to represent in other ways. The missing T or S in a contextual visual metaphor belongs to this second category. So, the sharing of the occlusion shape triggers the representation of the missing domain as aesthetically relevant property, but what is it that allows us to represent T or S? Referring to Nanay, we have said that the aesthetically relevant properties not perceptually represented in the pictures are presumably represented by mental imagery, which is essential in visual perception. So, the viewer will represent T or S by means of mental imagery.

Finally, our conviction is that, when understanding a visual metaphor, mental imagery plays the role that the inference plays in comprehending metaphors expressed in the verbal mode.

References

Aristotle. *Rhetoric*. Indianapolis-Cambridge: Hackett Publishing (2018).

Black, M. (1955). Metaphor. *Proceedings of the Aristotelian Society*, 55, 273-294.

Borst, G., & Kosslyn, S. M. (2008). Visual Mental Imagery and Visual Perception: Structural Equivalence Revealed by Scanning Processes. *Memory and Cognition*, 36 (4), 849-862.

Caramazza, A. (1996). Pictures, words and the brain. *Nature*, 383(6597), 216-217.

Carroll, N. (1994). Visual Metaphor. In J. Hintikka (Ed.), *Aspects of Metaphor* (pp. 189-218). Dordrecht: Kluwer Academic Publishers. Repr. in N. Carroll (2001), *Beyond Aesthetics: Philosophical Essays* (pp. 347-368). Cambridge: Cambridge University Press.

Cuccio, V. (2018). *Attention to metaphor*. Amsterdam: John Benjamins Publishing Company.

Danto, A. C. (1981). *The Transfiguration of the Commonplace. A Philosophy of Art*. Cambridge (Mass.)-London: Harvard University Press.

Davidson, D. (1967). Truth and Meaning. Synthese, 17 (3), 304-323.

Davidson, D. (1974). On the Very Idea of a Conceptual Scheme. *Proceedings and Addresses of the American Philosophical Association*, 47, 5-20.

Davidson, D. (1977). The Method of Truth in Metaphysics. *Midwest Studies in Philosophy*, 2, 244-254.

Davidson, D. (1986). A Nice Derangement of Epitaphs. In R. Grandy & R. Warner (Eds.), *Philosophical Grounds of Rationality* (pp. 156-174). Oxford: Oxford University Press.

De Houwer, J., & Hermans, D. (1994). Differences in the affective processing of words and pictures. *Cognition and Emotion*, 8(1), 1-20.

Finke, R. A., Pinker, S., Farah, M. J. (1989). Reinterpreting Visual Patterns in Mental Images. *Cognitive Science*, 13 (1), 51-78.

Forceville, C. (1996). *Pictorial Metaphor in Advertising*. London-New York: Routledge.

Forceville, C., & Urios-Aparisi E., (Eds.) (2009). *Multimodal Metaphor*. New York: Mouton de Gruyter.

Ganis, G., Schendan, H. E. (2011). Visual Imagery. *WIREs Cognitive Science*, 2, 239-252.

Glaser, W.R. (1992). Picture naming. *Cognition*, 42(1-3), 61-105.

Hinojosa J.A., Carretié L., Valcárcel M.A., Méndez-Bértolo C., & Pozo M.A. (2009). Electrophysiological differences in the processing of affective information in words and pictures. *Cognitive, Affective and Behavioral Neuroscience*, 9, 173–189.

Hyman, J. (2006). *The Objective Eye. Color, Form, and Reality in the Theory of Art*. Chicago: Chicago University Press.

Kellman, P.J., & Shipley, T.F. (1991). A Theory of Visual Interpolation in Object Perception. *Cognitive Psychology*, 23 (2), 141-221.

Kosslyn, S. M., Behrmann, M., Jeannerod, M. (1995). The Cognitive Neuroscience of Mental Imagery. *Neuropsychologia*, 33, 1335-1344.

Kosslyn, S.M., Thompson, W.L, Kim, I.J., & Alpert, N.M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, 378, 496-498.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago press.

Lakoff, G., Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, G. (1993). The Contemporary Theory of Metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 202-251). Cambridge: Cambridge University Press.

Lee, T.S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 98 (4), 1907-1911.

Mairal, R., & Ruiz de Mendoza, F. (2009). Levels of description and explanation in meaning construction. In C. Butler & J. Martin Arista (Eds.), *Deconstructing constructions* (153–198). Amsterdam/Philadelphia: John Benjamins.

Nanay, B. (2010). Perception and Imagination: Amodal Perception as Mental Imagery. *Philosophical Studies*, 150, 239-254.

Nanay, B. (2016a). *Aesthetics as Philosophy of Perception*. Oxford: Oxford University Press.

Nanay, B. (2016b). Imagination and Perception. in A. Kind (Ed.), *The Routledge Handbook of Philosophy of Imagination* (124-134). London and New York: Routledge.

Nanay, B. (2018). The Importance of Amodal Completion in Everyday Perception. *i-Perception*, 9, 1-16.

Ortony, A. (1979). Beyond Literal Similarity. *Psychological Review*, 86 (3), 161-180.

Paivio, A. (1971). Imagery and Verbal Processes. New York: Holt, Rinehart & Winston.

Paivio, A. (2010). Dual coding theory and the mental lexicon. *Mental Lexicon*, 5, 205-230.

Panofsky, E. (1939). *Studies in Iconology*. Boulder: Westview Press (1972).

Pérez Sobrino, P. (2017). *Multimodal Metaphor and Metonymy in Advertising*. Amsterdam: John Benjamins Publishing Company.

Sperber, D., Wilson, D. (1986). *Relevance. Communication and Cognition*. Oxford-Cambridge (Mass.): Blackwell.

Steen, G. (2008. The Paradox of Metaphor: Why We Need a Three-Dimensional Model of Metaphor. *Metaphor and Symbol*, 23 (4), 213-241.

Steen, G. (2011). The contemporary theory of metaphor -- now new and improved! *Review of Cognitive Linguistics*, 9 (1), 26-64.

Steen, G. (Ed.) (2018). *Visual Metaphor: Structure and Process*. Amsterdam: John Benjamins Publishing Company.

Taikh, A., Hargreaves, I.S., Yap, M.J. & Pexman, P.M. (2015). Semantic classification of pictures and words. *Quarterly Journal of Experimental Psychology*, 68, 8, 1502 -1518.

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.

Van Mulken, M., van Hooft, A., Nederstigt, U. (2014). Finding the Tipping Point: Visual Metaphor and Conceptual Complexity in Advertising. *Journal of Advertising*, 43 (4), 333–343.