Is there anything like visual argumentation?
A short exercise in methodical doubt[1]

In 2014 journal *Argumentation and Advocacy* was celebrating

> the groundbreaking work on visual argument that appeared in the journal's 1996 (double) issue on visual argument. Since that time, visual argument has become a central topic in argumentation theory and been featured in presented papers and published articles that explore case studies and investigate the possibility of a theory of visual argumentation (published on *Argthry*, 28th August 2014).

As an interested bystander who was not a partisan of visual argumentation (VA) nor an active participant in more or less heated debates around VA, I would like to start with a very short overview of these passed twenty years. Then—extensively commenting on Leo Groarke's paper 'Six Steps to a Thick Theory'—I will concentrate on some basic concepts AV is, in my view, lacking, but should be incorporated in their conceptual framework in order to better explain the following rhetorical problems: how visuals function, that is, how they get or catch the viewers, how the viewers break down the presented visuals, and how they reconstruct their meaning.

1 First version of this chapter (titled 'Against visual argumentation: multimodality as composite meaning and composite utterances') was published in Dima Mohammed and Marcin Lewinski, eds., *Argumentation and reasoned action, vol. I: Proceedings of the 1st European conference on argumentation*. Studies in Logic, vol. 62. (London: College Publications, 2016), 829–852.

Since *knowing is seeing* and *seeing is knowing* are deeply rooted and widely used metaphors in (not just) Western culture, such a rhetorical analysis, borrowing its tools from multimodal analysis, anthropological linguistics and (critical) discourse analysis, may importantly contribute to the thriving methodological discussion on how knowledge is extracted from the visuals, and how visuals generate knowledge.

Twenty years in a short overview

The way I say these twenty years of development of visual argumentation could be expressed contrastively, almost like an antithesis. On the one hand, the introduction to this double issue of A&A on VA, written by D. Birdsell and L. Groarke twenty years ago, is (understandably) still pretty cautious as to what visuals can do (all emphases are mine):

- '[...] the first step toward a theory of visual argument must be a better appreciation of both the *possibility* [!] of visual meaning and the *limits* of verbal meaning' (Birdsell, Groarke 1996: 2);
- '[...] we often *clarify* the latter (i.e., spoken or written words) with visual *cues* [...]' (Ibid.);
- '*Words* can establish a context of meaning into which *images* can enter with a high degree of specificity while achieving a meaning different from the words alone' (Ibid.: 6);
- '[...] diagrams can *forward* arguments' (Ibid.);
- 'The implicit verbal backdrop that allows us to *derive* arguments from images is clearly different from the immediate context created by the placement of a caption beside an image.' (Ibid.)

If we sum up: visuals may have some argumentative or persuasive potential (there is a *possibility* of visual meaning, visuals can *forward* arguments, and arguments can be *derived* from visuals) but they are usually (always?) still coupled with the verbal, and can achieve these argumentative effects only (?) in combination with the verbal. And the *pièce de resistance* Birdsell and Groarke are offering to illustrate the claims above (i.e., the possibility of visual argumentation) is an anti-smoking poster, published by the U.S. Department of Health, Education and Welfare in 1976 (I'll be commenting on it later on). Here it is:
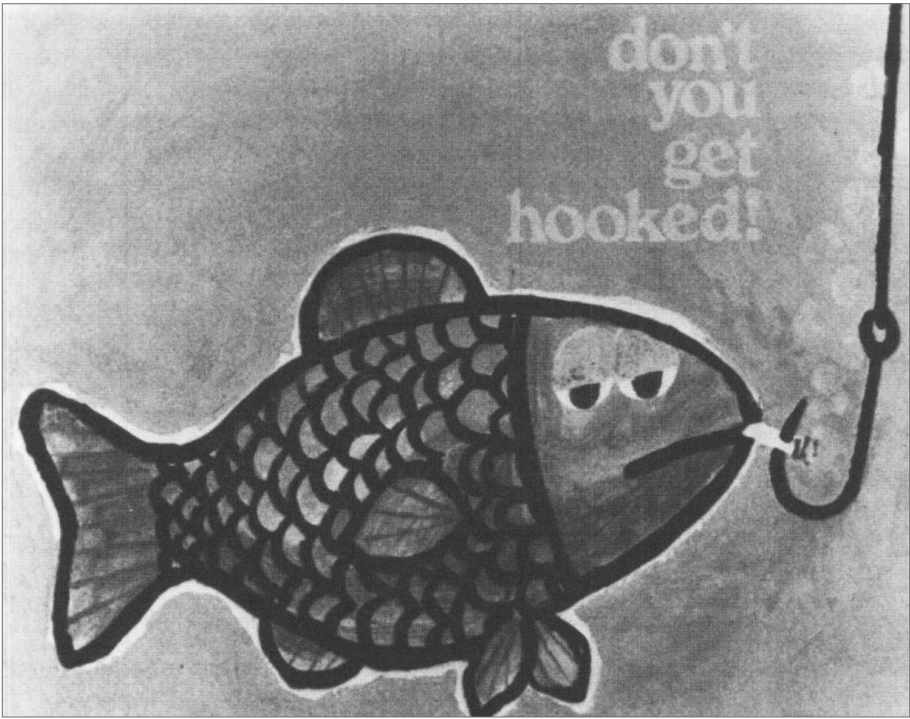
Figure 1. Smoking fish.

On the other hand, in the last ten years or so, visuals are more and more often presented by the proponents of VA as directly and unambiguously offering arguments by themselves, without any intervention or help from the verbal (or any other code), and not being conditioned or in any other way dependent on the verbal at all. Here are two reconstructed examples (I say reconstructed because I was unable to get the original materials from the authors).

The first one is a square ball, used as an example by one of the presenters at the 2014 ISSA conference. It was a small drawing of a square ball (unfortunately, the presenter wouldn't send me the exact drawing) with 'China' written on it, obviously cut from some newspaper or magazine, but presented without any immediate context: it wasn't made obvious to which section of the newspaper the visual belonged to (and the presenter would not explain it), nor could we see the neighbouring articles (and the presenter wouldn't explain that either). But he was very explicit in claiming that the argument offered by the visual itself was more than obvious: 'The Chinese football sucks!'

The counter-argument came up in the discussion. A colleague in the audience understood the square ball with the 'China' inscription on it as a metaphor of corruption in the PRC. Another colleague understood it as a metaphor of a hybrid socio-political system: turbo-capitalism under the leadership of the Central Committee of the CPC.

To sum up, the discussion showed very clearly that the argument was not evident from the drawing itself, otherwise so different interpretations could not have been possible. But, if the drawing would have been framed appropriately (so that we were able to see *where* in the paper the drawing was published, in *which* section, or *what* were the neighbouring articles), such an appropriate and sufficient framing would disambiguate the interpretation(s).

Here is another example of insufficient framing:

Figure 2. Notre-Dame Gargoyles I.

Figure 3. Notre-Dame Gargoyles II.

Figure 4. Notre-Dame Gargoyles III.

A photo resembling the three above (unfortunately, this presenter wouldn't send me the exact photo either) likewise was presented at the IPrA conference in New Delhi in 2013, with almost the same words as the square ball at the ISSA 2014 conference: 'What the argument is, is obvious from the photo itself.'

Framing in visual argumentation

But, are possible or potential arguments supposedly contained in the visuals really so obvious? We should recall what already Ch. S. Peirce had pointed out more than hundred years ago (Peirce 1931-58: 2.172): 'Nothing is a sign unless it is interpreted as a sign.' In other words, nothing is interpreted as a sign (i.e., representing or referring to something else) unless there is *intention* to see it and to understand it as a sign.

And these signs (consider figures 2, 3, and 4) can have many different interpretations (if not framed appropriately and sufficiently):

-   view of Paris (or one of the views of Paris);
-   view of Paris from Notre-Dame;
-   Notre-Dame on the background of Paris;
-   Postcard greetings from Paris;
-   some memorial photos from/of Paris;
-   details of Notre-Dame architecture;
-   examples of sacral architecture;
-   motives from the Notre-Dame outer walls;
-   mythological motives from the Notre-Dame architecture;
    even
-   excerpt from a book on plumbing (these Gargoyles were often used as gutters).

What is my point in enumerating all these? Simply, that *we should first know what the (immediate) context of a visual is*, and *only then proceed with the interpretation and meaning construction*. Or, in Wittgenstein's words (Wittgenstein 1953/1986: I-#663): 'Only when one knows the *story* does one know the *significance* of picture.' Which is, if we ponder a bit about this problem, just a corollary of a much more famous 7th thesis from his *Tractatus Logico-Philosophicus*: 'Whereof one cannot speak, thereof one must be silent.' Applied to visuals, we could paraphrase it as: until we know what the visual is (all) about, we cannot talk about it.

Or put it in the terms of what I will be proposing: we have to frame the visual (or the verbal, for that matter), and perform a frame analysis first (i.e. before proceeding to any kind of meaning construction).

Goffman's frames

Frames I will be concentrating on in this paper are not semantic frames as developed and defined by Charles Fillmore in 1977 (though even semantic frames (may) have a role in potentially argumentative interpretation of visuals as I will try to point out at least fragmentary), but frames that help us organize our everyday experience, frames as developed by sociologist Erving Goffman in his influential book *Frame Analysis: An Essay on the Organization of Experience* (1974).

What are Goffman's frames? In his own words:

When the individual in our Western society recognizes a particular event, he tends, whatever else he does, to imply in this response (and in effect employ) one or more frameworks or schemata of interpretation of a kind that can be called primary. I say primary because application of such a framework or perspective is seen by those who apply it as not depending on or harking back to some prior or 'original' interpretation; indeed *a primary framework is one that is seen as rendering what would otherwise be a meaningless aspect of the scene into something that is meaningful.* (Goffman 1974: 21)

Goffman distinguishes between natural and social frameworks. Natural frameworks 'identify occurrences seen as undirected, unoriented, unanimated, unguided, purely physical'. (Ibid.: 22) Social frameworks, on the other hand,

provide background understanding for events that incorporate the will, aim, and controlling effort of an intelligence. [...] *Motive and intent are involved, and their imputation helps select which of the various social frameworks of understandings is to be applied.* (Ibid.: 24)

There are different frames one can apply to a single event/entity, as in our two reconstructed examples with a square ball and the Notre-Dame Gargoyles, but 'we tend to perceive events in terms of primary frameworks, and *the type of framework we employ provides a way of describing the event to which it is applied*'. (Ibid.: 24)

For a contextualized illustration, let us go back to the smoking fish advertisement (Figure 1). The authors (Birdsell and Groarke) first admit that 'visual images can, of course, be vague and ambiguous. But this alone does

not distinguish them from words and sentences, which can also be vague and ambiguous'. (Birdsell, Groarke 1996: 2) Than they qualify this poster as 'an amalgam of the verbal and the visual' (ibid.), which, again, sounds quite acceptable. But then they conclude: 'Here the argument that you should be wary of cigarettes because they can hook you and endanger your health *is forwarded by means of visual images ...*' (Ibid.: 3) Which is obviously not the case. Without the verbal part, 'don't you get hooked!', the poster could be understood (framed) as a joke, as a cartoon, where, for example, smoking is presented as such an ubiquitous activity that even anglers use cigarettes to catch fish. *Only when we add the verbal part*, 'don't you get hooked!'—where 'hooked' activates a (this time semantic) frame of (semantic) knowledge relating to this specific concept (Fillmore 1977: 76–138),[2] which includes 'get addicted', and is, at the same time, coupled with a visual representation of a hook with a cigarette on it—is the appropriate (intended) frame set: the poster is now understood as an anti-smoking ad, belonging to an anti-smoking campaign.

Mental spaces

Equally problematic and ambiguous is the UvA poster Leo Groarke is using in his 'Logic, Art and Arguing' (1996: 112):

---

2   Probably the most widely known slogan describing the basic feature of frame semantics is due to Ch. Fillmore: 'Meanings are relativized to scenes.' And an often cited example by Fillmore (1997) demonstrating the above slogan is the difference in meaning between the following two sentences:
(1)   I spent three hours *on land* this afternoon.
(2)   I spent three hours *on the ground* this afternoon.
      The 'background' scene refered to in (1) would be a sea voyage while (2) refers to an interruption of an air travel.

Figure 5. UvA chief administrators.

Groarke's argument goes as follows:

The black and white photograph [...] presents the university's three chief administrators in front of the official entrance to the university. Especially in poster size, the photograph makes a stark impression, placing all this confident maleness in front of (visually blocking) the university's main entrance. *According to the committee,* which commissioned the poster, it is a 'statement' which effectively makes the point that 'we want more women at our university' and 'still have a long way to go in this regard. (Groarke 1996 ibid.)

But, if we are not acquainted with the committee's 'statement' that they want more women at their university (as, I guess, an 'average' Amsterdamer is not), and we just, walking the streets of Amsterdam, bump into this poster with three corpulent males, 'stating' 'UvA for Women', it is not at all clear how the poster was intended to be framed (by its authors). Is it (simply) a bad joke? Should it be taken ironically, maybe cynically, as a meta-statement from somebody who knows and objects the fact that UvA is all male? There is even a (at least implicitly) sexist interpretation that all these males at UvA need more women.

In other words, because of the insufficiently unambiguous framing it is not at all clear that we (the observers) can (and even should) reconstruct the argument(ation) in question the way Groarke does:

The poster thus presents the argument:

$$P$$
$$\downarrow$$
$$C$$

where the premise P is *the* (visual) statement that 'The University of Amsterdam's three chief administrators are all men' and C is the conclusion that 'The University needs more women' (Groarke 1996: 111).

Even if we take P as rather unambiguous (which it is not; for one thing, the fact that the University of Amsterdam's three chief administrators are all men is not a matter of general knowledge), the arrow, leading to C, can in no way be so linear, unidirectional, or monotonic (if you want) as to lead directly and exclusively to C, interpreted as 'The University

needs more women'. C could have had many other interpretations (and P many other formulations, for that matter), for example: 'UvA doesn't need women!', 'UvA is a sexist institution', 'UvA needs some women to change appearances'.

Much more appropriate representation of how we can read the UvA poster, and how we should interpret it, could be formulated in terms of mental spaces (nowadays more popularly called blending theory). Like this:
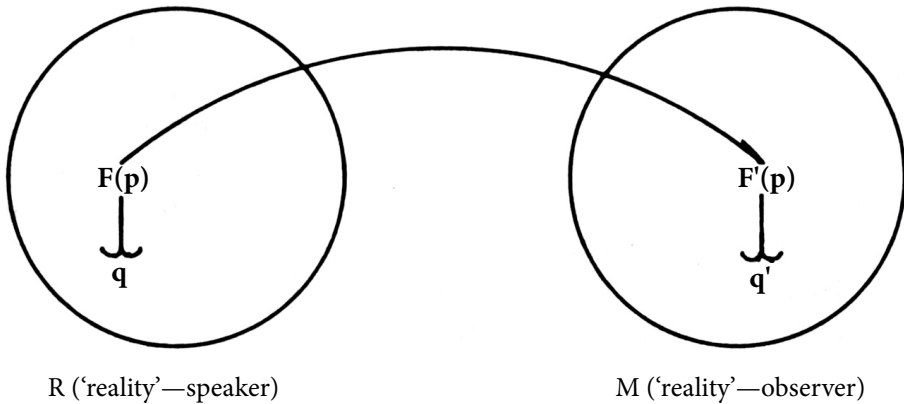

Figure 6. Construction of meaning in mental spaces.

Figure 6 should be read (interpreted) as follows. R stands for the 'reality' of the speaker (speaker's mental space), M for the 'reality' of the observer (observer's mental space). *p* represents the poster in question, *F(p)* its (intended) premise, and *q* its (intended) conclusion in R. In M, on the other hand, *p* still represents the same poster in question (hence the long arrow connecting the two spaces), but *F'(p)*, the observer's premise, and *q'*, the observer' s conclusion, may be quite different from speaker's premise and speaker's conclusion (depending on the context (time and place of encountering the poster), observers' experience, their social and cultural background, education, gender, and many other, even bio-neurological and cognitive factors). On top of that, M spaces may be multiplied in relation to R space, precisely because of different observers' different (social, cultural, etc.) background, education, gender, and many other factors.

Polyphony

A bit different mechanism seems to be at work in Marlboro advertisements Asimakis Tseronis used at the Brač Argumentation Conference in 2012. Actually, these were not advertisements but 'subvertisements', produced by a group called Adbusters (a name that is rather indicative as to what they are doing to advertisements).

Chronologically, the original Marlboro advertisements come first, of course. The background is always the American (Wild?) West, represented in warm, yellowish and brownish colors, and in the foreground there is always one or several cowboys. They may be smoking or not, but a pack of Marlboro cigarettes together with the company's logo is always highly visible and sets the frame (= we are talking cigarettes advertisement here, not, for example, westerns, or horse breeding).

What do Adbusters do to these original ads? They can't use the company's logo and packs of cigarettes, of course, so they use the standardized Marlboro background (warm, yellowish and brownish colors in the background, several cowboys in the foreground) to activate the appropriate frame with the observers (= this is (about) Marlboro). And the text within this familiar 'Marlboro country', implicitly and indirectly, alludes to the missing packs of cigarettes.

Like in Figure 7:

Figure 7.

Maybe even more efficient is the following parody. On the original advertisement we see cowboys on horses in a winter landscape, with Marlboro

packs in the lower right corner (for obvious reasons, we can't reproduce the photo). On the 'busted' version, we just see the horses in an empty grave-yard, covered with snow, while the tombstones symbolically replace packs of cigarettes.

Figure 8.

What is the mechanism at work here? It appears that a kind of 'gestalt' (warm, yellowish/brownish colors in the background, cowboys in the fore-ground ...) sets the frame (= Marlboro advertisements), while the text or the setting in the photo activates a (kind of) polyphonic reading: we can only make sense of and understand the busted advertisements if we con-nect them to the original advertisements, i.e. we can only understand them on the background of the original ads, i.e. as a kind of meta-ads.

When I am mentioning polyphony, I am referring to Bakhtin, of course, but even more explicitly to Ducrot's theory of polyphony, informed by Bakhtin, but much more elaborated. You may recall that Ducrot (2009: 32–44) is distinguishing between a producer, a locutor and several enuncia-tors/utterers or uttering positions. A producer is the person/organization ... that is the 'material' author of a given piece of text (or visual). In our case, the producer(s) would be the Adbusters (and their collaborators), the peo-ple who produced the anti-ads in question, those who had the idea, set the scenery, took the photo, developed it, and so on ...

A locutor is the entity (person, organizationa ...) that is (symbolically) responsible for the message of the ad. In our case the message could be reconstructed as something like: 'Smoking kills'. But this (meta)message is obviously only possible because there is an interplay of (at least two) enunciators or uttering positions within the locutor; the first one declaring that smoking is cool/attractive/adult (the original Marlboro ads) ..., and the second one subverting, criticizing such a position (the Adbuster ads). And the criticism as the main theme of the Adbuster ads prevails as the main message.

Rhizome and superdiversity in visual argumentation —a commentary

At this point, it may be worth briefly mentioning that in dealing with visuals, with construction of meaning and interpretation in visuals, we are necessarily dealing with the so-called rhizomatic structure and rhizomatic reading.

Rhizome is a (philosophical) concept developed in 1980 by two French philosophers, Gilles Deleuze and Félix Guattari (2005), and defined as theoretical approach that

> ceaselessly establishes connections between (different) semiotic chains, organizations of power, and circumstances relative to the arts, sciences, and social struggles. A semiotic chain is like a tuber agglomerating very diverse acts, not only linguistic, but also perceptive, mimetic, gestural and cognitive ...' (Deleuze, Guattari 2005: 28)

The concept was borrowed from botany and dendrology, where *rhizome* is a modified subterranean stem of a plant that is usually found underground, often sending out roots and shoots from its node. The rhizome also retains the ability to allow new shoots to grow upwards. If a rhizome is separated into pieces, each piece may be able to give rise to a new plant.
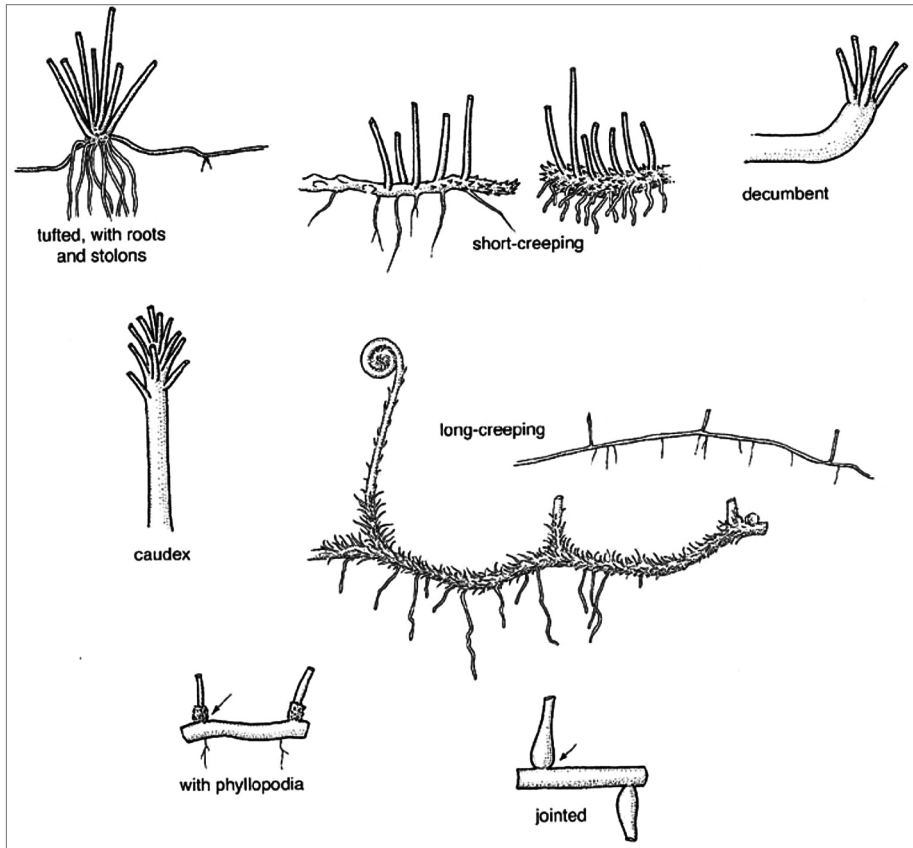
Figure 9. Rhizome, plant.

Or, in a more abstract and generalized form:
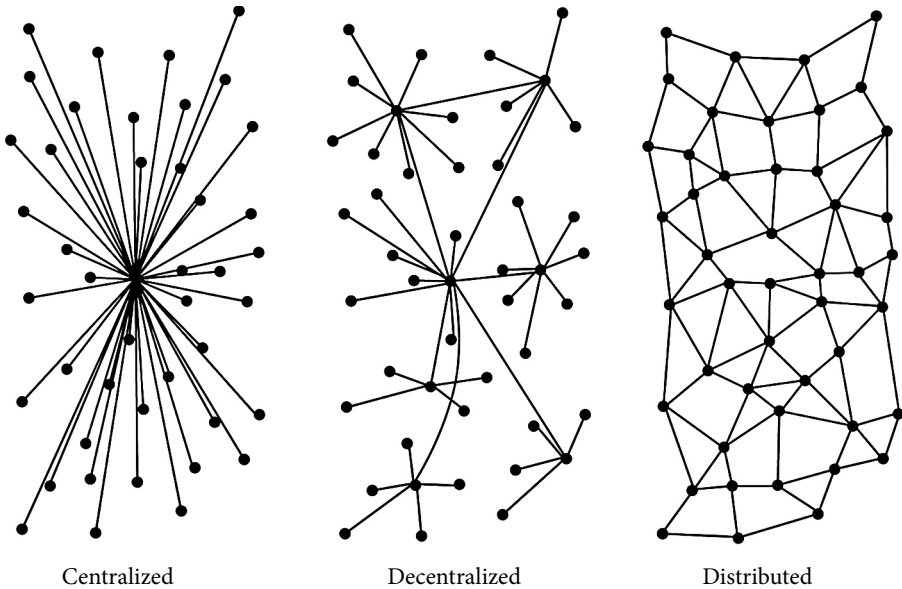
| Centralized | Decentralized | Distributed |

Figure 10. Rhizome, abstract.

Rhizome and rhizomatic structures become conceptually especially interesting if coupled and integrated with a (rather) new sociological concept that is rapidly gaining importance, the concept of superdiversity. Superdiversity is a concept coined by sociologist Steven Vertovec, and he describes it as:

> [...] a dynamic interplay of variables among an increased number of new, small and scattered, multiple-origin, transnationally connected, socio-economically differentiated and legally stratified immigrants who have arrived over the last decade (Vertovec 2007: 1025).

And what could be the significance of this new concept, referring to immigrants (among others), for the analysis and interpretation of visuals? Exactly the possibility that increasingly different ethnic, cultural, educational, and ideological background of potential readers/interpreters (not necessarily immigrants, of course) in a more and more globalized, multi-ethnic and multicultural word, may imply even more different access points and interpretational paths in reading and interpreting visuals. In other words, the allegedly unidirectional and unproblematic arrow connecting P and C in Leo Groarke's interpretation of the UvA poster may not just be

multiplied in different ways, pointing in different directions, but may (and should) also change its shape, from straight to wavy or curved or even broken (indicating that the conclusion is not linear and uni-dimensional, but had to make many detours, stops, as well as several repeated (and reconstructed) starts; which makes it conceptually very close to enchrony, the concept we will introduce later in this chapter), depending on how complex the meaning and possibilities of its interpretation may be. There are many useful arrows in stock already, like:

∿

↪

↪

↯

↳

⇝

Which also implies that possible C's in this case (and many others, ambiguous or/and biased) may come not just in different forms and formulations, but also with different content and different versions and values attributed to this content.

This is the reason why the theory of visual argumentation would benefit from *concentrating more on different possible entry and exit points in representation of visuals and interpretation of hypothetical visual arguments.*

The reasoning is the seeing. Is it?

This is the reason why visual argumentation should concentrate more on *different possible entry and exit points in data representation* and *interpretation of hypothetical visual arguments.* As a kind of a case study—exposing possible *caveats* as well as *cul-de-sacs* of visual argumentation—we will concentrate on Leo Groarke's proposal of reconstructing visual arguments as presented and conceptualized in his 2013 article 'The Elements of Argument: Six Steps to a Thick Theory', published in the e-book *What do we know about the world?: Rhetorical and Argumentative perspectives.*

Here is the photo Groarke is taking as a starting point:

Figure 11. Fruit found on the Detroit River I.

If we just take the photo in Figure 11 *per se*, as it is (as we see it *prima facie*), without or before any verbal explanation, and not knowing anything about possible context(s), the photo could be framed in many ways. As, for example:

(1) introducing/showing a peculiarly looking fruit;

(2) preparing a snack (or some other kind of meal);

(3) showing/presenting a new knife;

(4) showing/presenting an efficient/robust/… knife;

(5) showing the protective gloves, or how do protective gloves look like/how we use them;

(6) warning that one should wear protective gloves when using a knife (demonstrating safety procedures), etc.

But Groarke does disambiguate the photo rather quickly with the following explanation (all emphases throughout the text that will follow are mine):

Consider a debate spurred by an unusual fruit I discovered during a kayak ride *on the Detroit River*. When my description ('nothing I recognize; a bumpy, yellow skin') initiated a debate

and competing hypotheses on the identity of the fruit, I went back and *took the photographs reproduced below. On the basis of these photographs, the fruit was quickly identified as breadfruit.*

So the frame in question is the first one mentioned: introducing/showing a peculiarly looking fruit. And here is how Groarke reconstructs the argument (actually the process of arriving from argument(s) to conclusion) in question:

> The argument that established this conclusion *compared my photographs to similar photographs found in encyclopaedia accounts of breadfruit*. One might summarize the reasoning as: 'The fruit is breadfruit, for these photographs are like standard photographs of breadfruit.' But this is just a verbal paraphrase. *The actual reasoning—what convinces one of the conclusion—is the seeing of the sets of photographs in question.* Using a variant of standard diagram techniques for argument analysis, we might map the structure of the argument as:

$$I^1 \qquad\qquad + \qquad\qquad I^2$$
$$\Downarrow$$
$$C$$

> where C is the conclusion that the fruit is a piece of breadfruit, I^1 is the set of photographs I took, and I^2 is the iconic photographs of breadfruit to which they were compared.

Comparing the visuals as argumentation

But should (and does) the reasoning really consist just of '*the seeing of the sets of photographs in question*'? Is just seeing and visually comparing photographs *from different sources* really enough for a reasoned, justified conclusion (in question)? And last but not least, let us not neglect Groarke's remark that 'on the basis of these photographs, the fruit was *quickly* identified as breadfruit'. Is the velocity of (visual?) reasoning to be considered a necessary and sufficient criterion for good argumentation?

In order to answer these questions, we will be replicating Groarke's procedure. Here are some photos of breadfruit found in different encyclopaedias:

Figure 12 Breadfruit at Tortuguero (Wikipedia, https://en.wikipedia.org/wiki/Breadfruit).

Figure 13. The fruit of the breadfruit tree—whole, sliced lengthwise and in cross-section (Wikipedia, https://en.wikipedia.org/wiki/Breadfruit).

Figure 14. Breadfruit (Healthy Benefits, http://healthybenefits.info/
the-health-benefits-of-consuming-bread-fruit%E2%80%8F/).

And here, again, are Groarke's two photos (from the point of view of
perception, processing and meaning construction, it is important for the
('argumentative') viewer that they are incorporated between new photos
(of breadfruit), and not just referred to by numbers (e.g. Figure 11)): the one
we have already seen:

Figure 11. Fruit found on the Detroit River I.

and the one we haven't seen yet:

Figure 15. Fruit found on the Detroit River II.

Please inspect these photos carefully. Is there really such a resemblance between the two represented fruits that we can *quickly* identify the fruit from the Detroit River as breadfruit? To put it in Groarke's words, I don't see that resemblance.

Breadfruit, as we have seen, has a kind of knobbly skin with spines or hard hairs, patterned with irregular, 4- to 6-sided face, while in the center there is a cylindrical core. On the other hand, the skin of the fruit found in the Detroit River seems smooth, without spines or hairs, covered with smooth irregular bumps, no 4- to 6-sided face, and there seems to be no cylindrical core in the centre (though that may be due to the lightning, the angle or some other disturbing factor).

Introducing the necessity of the verbal

In such a case (where some items/entities look alike, but don't quite the same), just 'seeing' is not enough, and it is wise if not necessary to consult other reliable sources, like verbal description.

Why verbal descriptions? Because in such a case there is not much else one can consult. On the other hand, language is still the only communicative 'medium' that is (rather) linear, straightforward, and unambiguous enough; in combination with pertinent visuals almost error-proof. And

if, when consulting encyclopaedias or other relevant sources, *we don't just check the photos, but the text as well*, we find the following description of breadfruit (please, pay special attention to emphases in italics):

> Breadfruit (*Artocarpus altilis)* is one of the highest-yielding food plants, with a single tree producing up to 200 or more fruits per season. *In the South Pacific*, the trees yield 50 to 150 fruits per year. *In southern India*, normal production is 150 to 200 fruits annually. Productivity varies between wet and dry areas. *In the Caribbean*, a conservative estimate is 25 fruits per tree. Studies *in Barbados* indicate a reasonable potential of 6.7 to 13.4 tons per acre (16-32 tons/ha).
>
> [...]
>
> Breadfruit, *an equatorial lowland species*, grows best below elevations of 650 metres (2,130 ft), but is found at elevations of 1,550 metres (5,090 ft). Its preferred rainfall is 1,500-3,000 millimetres (59-118 in) per year.
>
> [...]
>
> Breadfruit is a staple food *in many tropical regions*. The trees were propagated far outside their native range by Polynesian voyagers who transported root cuttings and air-layered plants over long ocean distances. (From Wikipedia: http://en.wikipedia.org/wiki/Breadfruit)

If we sum up, breadfruit is a tropical plant, usually found (and used) in tropical areas. It is, therefore, not very probable to find it in Ontario, in the Detroit River, though it is not completely impossible, of course, that a specimen of a breadfruit found its way into the Detroit River from one of the local Caribbean restaurants or stores.

But if relevant sources were indeed amply consulted (i.e. browsed through), and the point of departure in investigating the nature of the found fruit was not based on some kind of preconceived idea or a hunch that the Detroit River fruit looked very much like breadfruit, a neutral, objective and interested investigator should have easily found the following photos as well:

Figure 16. Maclura pomifera (Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Maclura_pomifera_Inermis_BotGardBln1105Fruits.jpg).

Figure 17. Maclura pomifera (Plants for a Future, http://www.pfaf.org/user/Plant.aspx?LatinName=Maclura+pomifera).

Figure 18. Maclura pomifera (Acta Plantarum, http://www.actaplantarum.org/acta/galleria1.php?aid=463).

Figure 19. Maclura pomifera (Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Maclura_pomifera_FrJPG.jpg).

And once more, here are the two photos of a fruit found in the Detroit River:

Figure 11. Fruit found on the Detroit River I.

Figure 15. Fruit found on the Detroit River II.

A close comparative observation between encyclopedic photos of this second fruit and the photos of breadfruit reveals that this second fruit looks much more like the fruit found in the Detroit River: its skin seems smooth, without spines or hairs, and it is covered with smooth irregular bumps, not 4- to 6-sided face as in the bread fruit.

And if we consult the verbal part of the encyclopaedia, connected to this fruit, we find the following (once more, please, pay attention to emphases in italics):

> *Macula pomifera*, commonly called *Osage orange, hedge apple, horse apple, bois d'arc, bodark, or bodock* is a small deciduous tree or large shrub, typically growing to 8-15 meters (26.49 ft) tall. It is dioecious, with male and female flowers on different plants. The fruit, a multiple fruit, is *roughly spherical, but bumpy*, and 7.6-15 centimetres (3–6 in) in diameter. It is filled with sticky white latex. *In fall, its color turns a bright yellow-green.*

> [...]

> Osage orange occurred historically in the Red River drainage of Oklahoma, Texas and Arkansas and in the Blackland Prairies, Post Oak Savannas, and Chisos Mountains of Texas. *It has been widely naturalized in the United States and Ontario.* (Wikipedia: http://en.wikipedia.org/wiki/Maclura_pomifera)

As you can see for yourself, the verbal description of *Macula pomifera* actually fits the Detroit River fruit much more accurately than the description of breadfruit. And since we learn that the Osage orange 'has been widely naturalized in the United States *and Ontario*' it is much more probable that it fell in the water someplace along the Ontario river than that it found its way into the river from one of the Caribbean facilities in Ontario.

Thousands of words and a single picture

What can we learn from this? Above all that sayings like: 'A picture tells a thousands words' should be indeed taken seriously. *But*, to be (absolutely) sure *which of these thousands words* refer to that particular picture we have in front of us in these particular circumstances, *we have to cut down (on) those words considerably.* On the other hand, *without any words at all, we can hardly identify the exact meaning of the picture*

In other words, there seem to be no pure visual arguments (as there are, probably, very few purely verbal arguments; if any at all), and instead of visual argumentation (or purely verbal argumentation, for that matter) we should (always) talk about multimodal argumentation and multimodal meaning (combining, in our case, at least visual and verbal, but other semiotic modes are usually involved as well, such as gesture and gaze). But

multimodal meaning and multimodal argumentation require different (expanded, at least) analytical framework, let us simply call it multimodal analysis. And in relation to that, I would like to emphasize a few points.

In cases where just 'seeing' is not enough, and we have to consult verbal (or other) sources (and incorporate other types of signs, like gestures, gazes ...), we should be talking of *enchronic analysis* (Enfield 2009). What is enchronic analysis?

> Enchronic analysis is concerned with *relations between data from neighbouring moments*, adjacent units of behaviour in locally coherent communicative sequences. (Enfield 2009: 10)

Enchronic analysis is therefore looking at *sequences of social interaction in which the moves that constitute social actions occur as responses to other such moves, and in turn these give rise to further moves*. The Detroit River fruit is exactly a case in point: from observation of the photos of the fruit taken on the river, we have to move to the observation of the photos in encyclopaedias. And to get more complete and accurate information we have to switch from photos to text, and incorporate the textual information as well. And to fine-tune our findings (understanding), we have to switch to yet other photos (if necessary), and from them to yet another text(s) (if necessary), and finally compare all these again with the initial photo (of the fruit taken on the river).

If, when consulting encyclopaedias, *we don't just check the photos, but the text as well*, and then go and (re)check other available texts and photos, and compare them with the initial photo(s), the final result we arrive at should be described as *composite meaning*, resulting in *composite utterances*, conceptualized as: '[...] a *communicative move that incorporates multiple signs of multiple types*'. (Enfield ibid.: 15)

Here is a visual example of a composite sign (with composite meaning), Enfield is using himself:

Figure 20. Willy Brandt in Warsaw Ghetto (published in Enfield 2009: 3).

And this is his analysis (Ibid.: 3-4):

While the kneeling posture may have an intrinsic, ethological basis for interpretation, this particular token of the behaviour has had a *deeply enriched meaning* for many who have seen it, because it was performed by this particular man, at this time and place. The man is Willy Brandt, chancellor of West Germany. *Once you know this, the act already begins to take on enriched meaning*. It is not just a man kneeling, but a man whose actions will be taken to stand for those of a nation's people. It is 7 December 1970, a state visit to Warsaw, Poland. These *new layers of information should yet further enrich your interpretation*. To add another layer: the occasion is a commemoration of Jewish victims of the Warsaw Ghetto uprising of 1943. [...] The body posture [...] is a *composite sign in so far as its meaning is partly a function of its co-occurrence with other signs*: in particular, the role being played by its producer, given the circumstances of its time and place of production. The behaviour *derives its meaning as much from its position on these coordinates as from its intrinsic significance*.

In place of conclusion

We are dealing with several layers of meaning here, resulting in complex amalgam of signs as a process and product of a sequence of meaning-making moves. First, there is a kneeling posture as such, with its prototypical meaning. Then there is the presence of Willy Brandt, at that time the chancellor of Germany, with a variety of different meanings being attached to him or his function. The chancellor of Germany taking this kneeling position creates the third (amalgam) layer of meaning. The fourth layer of meaning is provided by the information that this act of kneeling was part of Brandt's state visit to Warsaw, and the fifth layer is provided by the information that Brandt's kneeling act was part of the commemoration of Jewish victims of the Warsaw Ghetto.

Speaking of the photo as such, these five layers of meaning form an amalgam of signs. But even more layers of meaning may be added, depending on the background knowledge of the observer and interpreter, as well as the context in which the photo is interpreted.

In view of all that has been said, let us return to the fruit found in Detroit River. If after checking and re-checking different photos, different

texts, and the strange fruit that was found in Detroit River, we finally point (and probably gaze) at it, declaring: 'This fruit is not a bread fruit!', we have produced a composite utterance, enchronically embracing several, at least seven, layers of meaning (1) checking the photos of the Detroit River fruit, (2) checking the photos of breadfruit in different encyclopaedias, (3) checking the text that comments on these photos, (4) checking the Detroit River fruit again, (5) looking for photos of similar fruits, (6) checking the text that comments on these fruits, (7) rechecking the Detroit River fruit again), belonging to three types of signs (conventional signs: words/text; non-conventional signs: photos, gesture, gaze; symbolic indexicals: demonstrative pronoun 'this', linking the conventional and non-conventional signs).

Put in other words and more explicitly: reasoning is not and cannot be just seeing, and just seeing is not and cannot be reasoning. Consequently, there is no 'pure' visual, but only multimodal argumentation:[3] at least verbal and probably other codes should be taken into consideration in order to reach sufficient, satisfying and complete meaning interpretation. To gain analytic credibility and interpretive force, scholars working on visual argumentation should consider incorporating into their framework all these intermediate gradual steps, as well as all these mutually dependent concepts.

---

[3]   A large body of literature has already been published on multimodality. An excellent introductory study is the book by Gunther Kress, *Multimodality: A social semiotic approach to contemporary communication* (London: Routledge, 2010).