



## UvA-DARE (Digital Academic Repository)

### Multimodal news framing effects

Powell, T.E.

**Publication date**

2017

**Document Version**

Other version

**License**

Other

[Link to publication](#)

**Citation for published version (APA):**

Powell, T. E. (2017). *Multimodal news framing effects*.

**General rights**

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

**Disclaimer/Complaints regulations**

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

# Chapter 1

## Introduction: Framing politics visually and verbally

---

“Between [image and text] the blur occurs. Each is altered by the other again and again, back and forth.” Taryn Simon, photographer and artist.

The interplay between image and text is central in today’s media experience. Improvements in image processing in the latter half of the twentieth century hailed a “visual turn” in news media (Magin, 2015; Mitchell, 1994). Nowadays almost every news article, whether online or in print, contains a picture or video to draw us in to a story, and TV, with its rich audio-visuals, remains the most popular source for news audiences (Fletcher et al., 2015). Thus, whenever we read or watch the news a seemingly effortless fusion of the visual and verbal modalities occurs. But how do audiences deal with “the blur” of multimodality and extract meaning from a news story?

The centrality of images in this process can be explained by human’s evolutionary past. The vertebrate eye evolved some 500 million years ago, compared to 50,000 years ago for human speech; whilst the first cave drawings are dated at 30,000 years, compared to the earliest writing system approximately five thousand years ago (Parker, 2003; see also Grabe & Bucy, 2009). This comparatively more ancient visual perceptual and communicative apparatus enables us to quickly extract meaning from eye-catching news images. Moreover, the elevation of particularly powerful images to “iconic” status is evidence that visuals are not merely the “window-dressing” of news media (Bucy & Grabe, 2007, pp.670). One example comes from the Vietnam War in 1972. Several Asian children run toward us along a road, screaming, with soldiers and a dark screen of smoke in the background. The primary subject of the photo, a nine-year-old girl Phan Thi Kim Phuc, had stripped herself of her clothes after being splashed with napalm. Captured by a *Life* magazine photographer, the ‘Napalm Girl’ image has been described as a “gut-wrenching photo which seemed to encapsulate the moral horror and military futility of war,” (Judson, 1995, pp. B1; Hariman & Lucaites, 2007).

This “savage image” was flashed across the front pages of nearly every newspaper across the globe that year. It was not only emotionally evocative, but also a gift to U.S. anti-war propaganda. According to the *Los Angeles Times* the picture, “came to symbolize, more than any other photograph taken in Vietnam, the atrocity of the war,”

(Coburn, 1989, pp.8) and helped to move the U.S. to a peace agreement seven months later (Perlmutter, 1998). Ascribing an anti-war message to the photograph was an easy task. However this was not everybody's reading of the image. Echoing suspicions of then President Nixon, General William Westmoreland, America's chief commander in Vietnam from 1964-1968, claimed that it was a fake. He asserted that the girl had actually been burned in a hibachi accident at a family barbeque (Shannon, 1986). On this occasion the unambiguous nature of the image helped to quash such competing explanations. Nevertheless, it reveals how different words can turn the same graphic image into a tool for contradictory political purposes – in this case to promote or undermine the anti-war movement.

The interaction between words and images in delivering meaning to news stories is at the heart of this dissertation. It is also of vital importance to citizens' political decision-making and, in turn, broader public opinion. In order to make sense of the complexity of political events we all, more or less consciously, engage in a process in which we construct and continually update simplified mental models of the world around us based on prior experience, existing knowledge and incoming information (Friston, 2005). These models, or "pictures in our heads" (Lippmann, 1922), are invaluable in helping us to navigate all aspects of our political lives, from forming opinions about an issue and discussing it with friends, to casting a vote in a general election. Since most citizens have minimal direct experience and knowledge about many political issues (Wanta & Hu, 1993) these inevitably incomplete mental models are therefore highly susceptible to influence by news media. Even more so for geographically and culturally distant foreign affairs issues. Indeed, verbal rhetoric as well as graphic images of war, conflict and crisis have been attributed a leading role in driving public opinion and, in turn, policy making (Livingston, 1997; Perlmutter, 1998). In addition to Napalm Girl in Vietnam, U.S. humanitarian interventions in Somalia and the Balkans in the 1990s, recent conflicts in Libya, Syria and Iraq, and the image of the drowned Syrian boy Alan Kurdi during the European refugee crisis of 2015, attest to this (Entman, 2003; Robinson, 2002).

This dissertation aims to establish how and why citizens' political opinions and behaviours are influenced by visual and verbal modalities in news media. To address these fundamental questions, the empirical chapters comprise experiments in which the visual and verbal streams in actual news media content are manipulated. In doing so I aim to bring scholarly understanding of the meaning-making of news up to date with today's multimodal media reality.

The rest of this chapter is devoted to unpacking the relevance, rationale, theoretical framework and existing empirical knowledge required to address this overarching research question. Then the methodological approach and research design underpinning the empirical chapters is outlined. First, however, a fundamental question is addressed in more detail: why should we care about multimodal political communication at all?

## The democratic potential of multimodal news media

The idea that vivid visual images can influence public opinion and political policy – the concept of visual determinism (Perlmutter, 1998) – has been acknowledged as early as Plato's *Republic*. Then the philosopher was concerned that painters and poets would stir the emotions and fool the senses, conflating reality with falsehood. In his allegory of the cave, the prisoners condemned to darkness staring at images manipulated by puppeteers represented the lowest form of consciousness. By contrast, verbal dialogue that stimulates education, knowledge and ideas was the route to “good” citizenship and society (386a, 595a; Plato, Grube & Reeve, 1992; Goldstein, 2014).

These ideals were solidified during the Enlightenment, when increasing verbal literacy was regarded as the arbiter of logic and reason and considered as the building blocks of society. But is such a conception of good citizenship actually possible, or relevant to society today? In such a model, the average citizen would be expected to possess the willingness and ability to be informed about every policy issue, in writing or heard speeches, and the press should supply this information (Strömbäck, 2005). However, scholars have long acknowledged this as an unsustainable belief – there is simply too much to know that requires complex political expertise (Graber, 2003; Zaller, 2003). But is there an alternative way to stimulate political literacy and engage citizens in public life?

Visuals were used for informational purposes during the reformation (Bagchi, 2016) as well as in early news in materials aimed at the less literate in forms such as pamphlets and chapbooks (Zaret, 2000). Doris Graber (1990, 1996; see also Hariman & Lucaites, 2007) harnessed this principle to argue that visuals can promote the higher democratic ideals of an engaged and informed citizenry. She pointed out that human physiology ordains that most people prefer political learning involving visuals. In contrast to the processing of words through the brain's cortical pathways, visual input is initially processed through an evolutionarily old and fast route, directly from the thalamus to limbic system – the emotional centre of the brain. As such, compared to words, visuals are processed faster (Barry, 2005), requiring less conscious effort (Gazzaniga, 2004) and readily yield an immediate emotional response to guide the processing of news information (Bargh, 1988; Damasio, 2006).

Communication scholars have since proved that the combination of verbal *and* visual cues are essential for processing news messages (e.g., Graber, 1990; Bucy, 2000). For instance, Graber (1990) showed that recall of television news stories was enhanced by visuals, and others have since highlighted that memory for visuals often trumps verbal content (e.g., Drew & Grimes, 1987; Grimes, 1991; Lang, 1995; Reese, 1984). Grabe and colleagues (Bas & Grabe, 2015; Grabe, Bas & van Driel, 2015) have further broken from the “Gutenberg Legacy” that canonizes verbal content in political communication (Graber, 2001). They highlight that personalized emotional content,

and particularly visuals, show promise in closing the political knowledge gap between citizens of different education levels.

It seems, then, that the visual modality in vivid mediums such as TV – the very formats that advocates of Enlightenment ideals would most disdain – can promote political literacy in line with more lofty democratic objectives. However, learning is just one aspect of the meaning-making process which enables us to form opinions about political issues. As well as information, influence plays a role. Testament to this are political visuals in early modern placards and handbills with cartoonlike images ridiculing leaders (Burke, 1994), to the proliferation of twentieth century wartime propaganda. The question is, then: how does news media's "visual turn" (Mitchell, 1994) influence the "pictures in our heads" that guide political decision-making?

### Framing in multimodal news media

Until the 1970s the paradigm of rational choice dominated psychological models of social and economic decision-making. This tradition considered human behaviour, even when applied to the political and sociological realms (Lohmann, 2008), as an individualistic exercise in interest calculation and excluded emotion-related variables (Grabe & Bucy, 2009). That was until the prevalence of 'irrationality', or, more accurately, decision bias and heuristics (mental rules-of-thumb) was highlighted by the Nobel Prize-winning work of Daniel Kahneman and Amos Tversky (e.g., 1979). One aspect of this work demonstrated a "framing effect" (Tversky & Kahneman, 1985): that small differences in presentation of otherwise equivalent risky choices (e.g., the proposed number of lives saved or lost by different government policies) dramatically influenced decisions of participants in their experiments.

This framing effect was later shown to prevail in non-risky political decisions (Levin, Schneider & Gaeth 1998). Moreover, it sparked an abundance of dual processing theories of reasoning and persuasion (Chaiken, 1980; Epstein, 1994; Evans & Stanovich, 2013; Kahneman 2011; Loewenstein, Weber, Hsee & Welch, 2001; Petty & Cacioppo, 1986; Zajonc, 1980). These models contrasted automatic, unconscious and heuristic decisions with deeper processing that is systematic, conscious and deliberative. In this framework, visuals – especially personalized and affectively charged images of suffering children – are thought to increase the likelihood of heuristically-driven decisions (Slovic, Finucane, Peters & MacGregor, 2007). Initially these findings were negatively construed as imperfections in human decision-making. However, in *Descartes Error*, Damasio (2006) showed that affective cues are an indispensable element of human brain functioning – providing physiological "markers" to guide the conscious and unconscious mental processes underpinning behaviour. Such findings helped to break the prevailing wisdom of emotion and rationality as synonymous with 'good' and 'bad' decision-making, respectively (see also Lau & Redlawsk, 2001).

News framing differs from Kahneman and Tversky's equivalence framing, although they do rely on related principles of individuals using presentation cues to simplify complex decisions via their emotional and cognitive faculties (Druckman, 2004; Nabi, 2003). Frames in news occur when journalists take necessary steps to simplify reporting about complicated issues by emphasising the most salient aspects at the expense of others (de Vreese, 2005). Frames help citizens engage with news by providing "interpretative packages" with a "central organising idea" that gives "meaning [...] to an unfolding strip of events" (Gamson & Modigliani, 1989; Goffman, 1974). This simplification process necessarily introduces bias in the presentation of an issue, and even more so for particularly complex and distant foreign affairs issues. As such, concerns abound that (especially traditional) media outlets readily adopt frames professionally produced by politicians, organisations and their spin doctors with the purpose of manipulating public opinion in their favour (Entman, 2003). Indeed, the express aim of public diplomacy is to exercise "soft power" through the media (Nye, 2008).

Whether and how frames in news influence citizens' perceptions and political opinions is the study of framing effects research (Chong & Druckman, 2007; de Vreese, 2005). For instance, coverage of a Klu Klux Klan rally framed either as a free speech issue or a disruption of public order can influence subsequent perceptions, opinions and behaviours towards the group (Nelson, Oxley & Clawson, 1997). Although the concepts underpinning the mechanisms of framing effects have been described as vague (amongst other terms; Scheufele & Iyengar, 2012; Scheufele, 2004), they are thought to involve two processes: accessibility and applicability. Frames make certain associations and ideas more *accessible* in mind when considering a political issue – also a central concept in agenda setting and priming research (Scheufele & Tewksbury, 2007). Importantly, whether a news frame exerts an influence is determined by how *applicable* it is to an individual, which depends on a host of factors such as one's interests, knowledge and pre-existing attitudes about the issue (Scheufele & Scheufele, 2010). As such, framing effects emphasise a "constructed reality" approach (McQuail, 2010) to making sense of political news.

Visuals operate as framing devices by visualizing and emphasizing a particular aspect of an issue (de Vreese, 2005; Entman, 1993; Tankard, 2001). The way in which visuals are manifest in news frames is determined by a multi-level "winnowing process" (Schwalbe, 2006, p. 269); including the photo journalist's selection of what to capture and how (e.g., angle, perspective, cropping) and editorial decisions about what images to publish beside an article's text, and in what size and position (Kress & van Leeuwen, 1996). Content analyses of resultant visual frames have focused on sources of powerful visuals – war and conflict (Fahmy & Neumann, 2011; Griffin, 2004; Parry, 2010), protest (Corrigan-Brown & Wilkes, 2012; Gitlin, 1980), natural disasters (Fahmy, Kelly & Kim, 2007) and election campaigns (Coleman & Banning, 2006; Grabe & Bucy, 2009).

A select few have gone further to analyse visual and verbal content simultaneously (Huang & Fahmy, 2013; Ojala, Pantti & Kangas, 2017).

Importantly for this dissertation, and despite the purported power of visuals, framing effects research is only beginning to catch up with the multimodal reality of news media (Coleman, 2010). Indeed, despite scholars having bestowed visuals with the potential to “revitalise” framing theory (Coleman, 2010; Scheufele & Iyengar, 2012), our understanding of visual effects still lags behind the verbal modality. Knowledge about the interactive effects of multimodal (visual and verbal) frames is scarcer still.

## Visual and verbal framing effects

The contribution of visuals and verbals (i.e., the written and heard word) to framing effects depends on their unique characteristics (Geise & Baden, 2014; Kress & van Leeuwen, 2001). Visuals, especially those containing graphic content (Scharrer & Blackburn, 2015) or social cues (Findlay & Gilchrist, 2003), tend to be the first thing that capture readers’ attention (Garcia & Stark, 1991). Compared to text which is less salient, news visuals are psychologically activating – increasing heart rate (Lang, Newhagen, & Reeves, 1996) and appreciation (Reeves, Newhagen, Maibach, Basil, & Kurz, 1991) – and are particularly effective at evoking an emotional response (Iyer & Oldmeadow, 2006). In the domain of memory and learning, the picture superiority effect refers to the way named images are better recalled than named words due to the concrete imagery they generate (Nelson, Reed, & Walling, 1976; Paivio, 1991). In news media, memory for the visual modality typically exceeds the verbal modality (e.g., Lang, 1995), especially for compelling negative images (Newhagen & Reeves, 1992).

The meaning of visuals is accessed faster than words (Smith & Magee, 1980). This is because images (both still and moving) are analogies of physical objects and events in the non-mediated environment, whereas words are abstract symbols that bear no physical resemblance to their referents (Messaris & Abraham, 2001; Sontag, 1977). However, since many camera and editing operations – such as zooming, panning and tilting – resemble human perceptual experiences, and because news visuals are often stereotyped, conventionalised and stylistically repetitive (Gartner, 2011), viewers are often not explicitly aware that media portrayals are constructed or artificial (Grabe & Bucy, 2009). In contrast, the written or heard word possesses an explicit propositional syntax for relaying meaning which is especially good for clearly conveying who did what to whom and why (Entman, 1993; Messaris & Abraham, 2001). Therefore, whilst decoding text does require a level of knowledge and abstraction not necessary for visuals (Messaris, 1994), verbal content is typically a more explicit purveyor of meaning (Geise & Baden, 2014).

A fledgling body of literature have shown that these qualities influence framing effects. Grabe and Bucy (2009) systematically related image- and sound-bite framing

in American presidential campaigns to electoral outcomes – ranging from candidate image management and issue position, to editing techniques such as camera angles and zoom (see also Boomgaarden, Boukes & Iorgoveanu, 2016; Gadarian, 2014; Graber, 2001; Keeter, 1987; Rosenberg, Bohan, McCafferty & Harris, 1986; Sullivan & Masters, 1988; Todorov, Mandisodza, Goren, & Hall, 2005). Fewer studies have focused on how the visual framing of issues themselves can deliver effects. Insights from exemplification theory have shown that manipulation of an image whilst keeping an article's text constant can affect perceptions of an issue (Gibson & Zillmann, 2000; Zillmann, Gibson & Sargent, 1999). Other studies have shown that graphic images of conflict can influence support for war (Gartner, 2011; Scharrer & Blackburn, 2015) and that visual effects can depend on participants' interest in an issue (Arpan et al., 2006; Soroka, Loewen, Fournier & Rubenson, 2016). Pfau et al., (2006; see also 2008) showed that an image plus caption can be particularly effective at eliciting an emotional response and opinion change. However, in contrast to studies implying the superiority of images, Domke, Perlmutter and Spratt (2002; see also Perlmutter, 1998) criticised the widely-held notion that vivid visuals drive public opinion. They argued that image content and accompanying narratives interact with individuals' considerations to shape their affective and cognitive reactions. On the weight of evidence, then, the individual role of visuals and text in framing effects seems broadly in line with their unique characteristics and psychological consequences. However, evidence of their integrative influence is scarce.

### Moving towards multimodality

In their recent synthesis of this body of work, Geise and Baden (2014) proposed a *multimodal* theory of framing effects, and helped set the stage for the empirical investigations in this dissertation. At the heart of this work was the premise that visual and linguistic frame processing should not be understood as fundamentally distinct logics (Coleman, 2010; Mueller, 2007). What shapes frame processing is not necessarily the modality per se, but the degree to which the respective properties of visual, textual, or other framing devices are actualized (Geise & Baden, 2014). Specifically, they formulated several theoretical propositions about the contribution of images and text to multimodal frame processing. First, visuals are superior to text in their salience attribution (proposition 1a *picture superiority*), but texts have a more constrained and conventional structure (proposition 1b *open/conventional structure*). Second, visual representations permit multiple (iconic, indexical, symbolic/conventional) strategies for decoding meaning depending on the viewers' available knowledge, whereas text promotes a more direct matching of signifier and signified (proposition 2 *polysemy/strong codes*). Third, textual frames specify the nature of and relations between decoded elements, whilst visuals merely suggest associations between elements and provide

less guidance about their specific nature (proposition 3 *associative/propositional*). Fourth, the richness and ambiguity in visuals compared to texts results in a surplus of information available for integration, and thus produces higher variability in the formulation of a central organizing idea (proposition 4 *iterative/focused integration*). Thus, in Geise and Baden's model, frames emerge from the interaction of modality-specific potentials and the inter-semiosis of multimodal meaning (Martinec & Salway, 2005; van Leeuwen, 2012). However, as yet, no studies have empirically tested these theoretical expectations to examine the relative power of visuals and text when presented alone, and, importantly, their integrative contribution to multimodal framing effects.

These communication-related potentials of visuals and text suggests that they play unique but complementary roles in the mechanisms of multimodal framing effects (Geise & Baden, 2014). The salience-enhancing properties of visuals appear to make them particularly effective at the increasing the *accessibility* of considerations in the mind of the viewer. Indeed, inclusion of an image can increase the salience of one issue over another, prime that topic in the minds of citizens (Fahmy, Cho, Wanta & Song, 2006; Scheufele & Tewksbury, 2007) and, in turn, raise it up the public agenda (Wanta, 1988). By contrast, the formal syntax of a verbal narrative should more effectively disambiguate news frames and thus may play a stronger role in determining the *applicability* of a media frame to one's pre-existing knowledge and beliefs (Domke et al., 2002). These functions are, of course, are not mutually exclusive. For instance, visuals may capture one's attention and evoke emotions, but they can also trigger cognitive elaboration (Russell, Marcus & MacKuen, 2000). This idea connects to dual pathway theories of persuasion (e.g., Chaiken, 1980; Kahneman 2011; Petty & Cacioppo, 1986): Visuals may typically trigger processing that is relatively more heuristic, affective and based on source cues, whereas verbal content may rely on more detailed systematic and cognitive processing (Chaiken & Eagly, 1976; Epstein, 1994; Pfau, Holbert, Zubric, Pasha & Lin, 2000; Sparks, Areni & Cox, 1998). Despite these ideas, no studies have systematically examined the processing of visual and verbal input in multimodal framing effects.

Finally, although the reviewed literature focuses on newspaper articles and TV, there have been no direct comparisons of how multimodal framing effects differ in different media formats. This is becoming increasingly important since news producers and publishers are devoting more and more resources to news videos in place of (or in addition to) static articles (Kalogeropoulos, Cherubini & Newman, 2016). With the moving images of online video and TV providing a richer and more vivid depiction of reality (Pincus, Wojchieszak & Boomgaarden, 2016; Slater, Rouner & Long, 2006), one might expect the visual modality to dominate in these formats. In the domain of memory and learning, the interactive effect of dynamic visual and verbal streams in TV news have been subject of much attention. This has produced mixed findings, with

Annie Lang (1995; 2000) helping to clarify this body of work using a Limited Capacity Information Processing approach to argue for an overall memory benefit for visuals at the expense of verbal input. An important stipulation in this work was that structural characteristics inherent in multimodal TV and video content – including production techniques such as camera cuts and zoom – must be carefully conducted in order not to overload viewers' limited processing capacity (Lang, Geiger, Strickwerda & Sumner, 1993). Whether these factors apply to multimodal framing effects and contribute to meaningful differences between media formats is, as yet, unknown.

Despite the ever-accumulating body of literature, it is clear that multiple questions remain about how visuals and text interact to deliver multimodal framing effects. Before unpacking the main research questions of this dissertation, the next section first describes three hypothetical models of multimodal framing effects. These provide extra context for the research questions that follow and guide later discussions about the democratic implications of the empirical chapters.

## Competing perspectives on multimodal framing effects

In this section the potential scope of visual and verbal effects in multimodal framing are outlined with the help of three hypothetical models. These models are useful not only for considering the potential findings that this dissertation might deliver, but, importantly, to connect results of the empirical chapters with their implications for democratic processes. The three hypothetical models in this section synthesise findings from the reviewed literature along with aggregate-level theories of media effects on public opinion and policy making. It is important to keep in mind that this dissertation focuses on multimodal framing effects at the individual level. Also note also that these perspectives are stylized: they depict simplified models of reality and are not entirely distinct – some overlap exists in the extent to which the visual and verbal modalities exert an influence. For clarity and succinctness the models are named according to the strength of visual effects in relation to verbal effects in the framing process.

### *1. Strong visual effects*

This model predicts that visuals trigger a robust emotional response and have strong effects on political opinions and behaviours. Visuals index reality and forge a connection between the subject of a news issue and the public (Chouliaraki, 2006) to stimulate emotions – sympathy for the starving child, anger at perpetrators of gun crime, fear towards threatening militiamen – which, in turn, moves citizens to action. The outpouring of donations, petition-signing and demonstrations that followed the publication of the picture of the drowned Syrian boy Alan Kurdi provides a recent example which, some have claimed, (temporarily) influenced refugee policy across Europe (Vis & Goriunova, 2015). In this perspective, visual effects should be widespread

and uniform – akin to the hypodermic needle of media effects (McQuail, 2010) – in that they apply to the majority of citizens, irrespective of individual differences in, for example, knowledge, and political beliefs. Moreover, visual effects should be magnified by vivid visual dynamic media, such as TV and online video, as they overpower the medium’s verbal stream and one’s ability for reasoned thought (Morrow, 1993).

A strong media effects model emphasising the role of powerful visuals, particularly on TV, is the “CNN-effect” (e.g., Livingston, 1997; Robinson, 2002). Formulated in the 1990s, when foreign policy making, particularly US interventions in humanitarian crises in the Balkans and Somalia, were said to have been triggered by the broadcast of images of the ethnic cleansing of Bosniaks and emaciated Somali children on networks such as CNN. Proponents of the CNN-effect claimed that images contributed to “empathy framing” (Entman, 2003) which changed public opinion and subsequent policy towards war. Then, to the reverse effect, images from Somalia, and previously in Vietnam, of dead and mutilated US soldiers, were said to make the public war averse (but see Robinson, 2002).

## 2. *Weak visual effects*

This perspective considers visual effects as minimal, mainly due to their inherent ambiguity (Geise & Baden, 2014) and their conative qualities – that they trigger thoughts that are dependent on the experiences of the viewer (Zelizer, 2010). Instead, verbal narratives in news should play a larger role since they provide a clear meaning to stories which helps citizens determine whether a news story is applicable to their prior proclivities. Any effects that visuals might evoke – for example on emotions – are fleeting and fade quickly. As such, visuals in certain circumstances might lead to political behaviours such as donations to help identifiable individuals – in the case of Alan Kurdi, for example – but they have no major or lasting effect on public opinion at the aggregate level. Limited visual effects should not be magnified in richer media formats, but the qualities of news articles, allowing re-reading and more careful processing of textual content of news frames, might amplify the influence of the written word. As such, rational processing of news, and particularly the verbal modality therein should determine framing effects (Scheufele & Tewksbury, 2007).

Perlmutter (1998; 2005) and colleagues (Domke et al., 2002) argued that the magnitude of visual effects are limited and overstated. Contrary to widely-claimed power of visual images, Perlmutter claimed that a *first-person effect* is at play. Here, those who have a strong reaction to an emotionally charged news image assume that others uniformly feel the same way, talk about the image with others, and in turn propagate this belief. This phenomenon is particularly strong for political elites for whom political media and its implications are magnified and play an integral part in their daily lives. These individuals also possess a platform from which an image can be popularised by projecting their

perception of powerful visuals onto the public at large. As such, Perlmutter argued that visuals can have an individual-level effect, but their aggregate effects are overstated and outweighed by the history, context and real-world events surrounding a political issue. Citizen's political opinions may be shaped over time by repeated textual messages (Breckler & Wiggins, 1989) which have a greater potential for framing effects.

### *3. Moderate and conditional visual effects*

This model represents a middle ground, where visuals may exert a strong emotional response and influence behaviours and opinions, however this depends on the news issue at hand and the experiences and motivations of the individual (e.g., Nabi & Oliver, 2009). Visuals are attention grabbing and their connotative qualities can help define an issue. These effects could be stronger for those disinterested in politics since visuals require minimal processing effort, and for images of high profile news issues that more easily reach citizens via incidental news exposure. By contrast, for political sophisticates and obscure issues minimally covered by mainstream media, ostensibly powerful visuals may provide a stimulus to elaboration and discussion with peers which can merely reinforce existing beliefs. In such cases, a detailed message delivered via the verbal modality should prove more influential. Furthermore, this perspective predicts that effects of different media formats are not universal. For some, news articles provide a desired opportunity to elaborate on political issues. For others, the image- and sound-bites of audio-visual videos and TV are a good fit, especially for those minimally motivated to process the news. Furthermore, certain production techniques might also attract or distract citizens from the frames presented.

A key perspective here, albeit one less applicable to aggregate-level effects, is that of Lang's (2000) Limited Capacity Information Processing Model of media learning. Here, certain aspects of news media, including the motivational and personal relevance, medium, complexity and emotionality determines the information processing cascade and eventual outcomes for message storage, encoding and retrieval (e.g., Lang, 1995; 2006). Importantly, if the properties of a certain media message overload an individual's limited capacity then retention of it will be minimal. The extent to which Lang and colleagues' observations extend to framing effects and public opinion is yet to be established. However, NGO and international aid research provide a relevant extension to the aggregate level, where shocking negative visuals, for instance showing third-world poverty, are often employed to stimulate large-scale helping behaviours, such as donations. However, over time, routinized visuals may lead to desensitisation and "aid-fatigue" and weaken longer-term engagement (Hudson, Van Heerde-Hudson, Dasandi, & Gaines (2016b). Instead, repeated substantive verbal messages might be more effective at encouraging longer-term engagement with the issue. As such, this perspective emphasises text-image relations and more nuanced effects.

This dissertation will shed first empirical light on multimodal framing effects by revealing which of these three perspectives is most appropriate. Research reviewed thus far can help to narrow down expectations in advance of the empirical chapters. The weight of theoretical argument suggests that strong visual effects (model 1) are possible, due to the assumed power of news visuals. However, after factoring in relevant existing empirical evidence, especially from studies that have manipulated visuals whilst keeping text constant (e.g., Domke et al., 2002) as well as studies of media memory (e.g., Lang, 1995), moderate and conditional visual effects (model 3) seem most plausible.

### Focus of this dissertation

This dissertation focuses on addressing a few fundamental research questions as a first step in understanding the interactive contribution of visual and verbal input to multimodal framing effects. Below, highlights from the reviewed literature are summarised and the research questions are outlined.

The shift toward formats in which eye-catching images are presented with accompanying text kept to a minimum means that an image's impact when viewed alone is no longer a trivial matter. Until now, no studies have compared the relative power of images and text in framing effects, which is also a fundamental prerequisite for understanding multimodal interactions. This motivates the first research question:

*RQ1: When viewed in isolation, are images or text a more powerful in delivering framing effects?*

Since news articles most often consist of one or more images accompanying a text, a more important question for media theory and practice concerns the unique contribution of the visual and verbal modalities to multimodal effects. Multimodal frames in which the different modalities reinforce each other should benefit both from the salience and vividness of visuals, and from the guided structuring of linguistic representations. Where different modalities conflict, it is still unclear whether salient but ambiguous visuals or clear but less salient news texts prevail (Geise & Baden, 2014). Thus, the second research question is:

*RQ2: What is the contribution of the visual and verbal modality to multimodal framing effects?*

Mediated visual and verbal input are processed differently. Visuals are initially processed quickly via emotional centres in the brain and, as such, their effects are thought to involve more emotional and heuristic processing pathways. In contrast, verbal input

is processed more slowly, requires more effort in the abstraction of meaning from its formal structure, and thus is associated with more systematic and cognitive processing. The third research question relates to whether these pathways characterise visual and verbal processing in multimodal framing effects.

*RQ3: What are the mechanisms of multimodal framing effects: are visual and verbal modalities processed via more heuristic and systematic pathways, respectively?*

The different qualities of static (i.e., news articles) and dynamic (i.e., news videos and TV) news media suggest that the influence of visual and verbal modalities in multimodal framing effects might depend on the communication medium. Dynamic audio-visual media are psychologically activating and may bias the visual stream. Static image-text articles, in contrast, are less vivid, and the more effortful reading required may bias framing effects toward the verbal modality. This begs the final research question:

*RQ4: How does the processing and effects of multimodal frames differ in dynamic (videos) and static (articles) news media?*

## Research design

Three studies comprise this dissertation, all of which make use of the web-based experiment. A web-based experiment provides many of the benefits of a typical lab experiment, but takes place in an online setting in which news is often encountered. This provides participants with the comfort of an environment in which they would normally browse the web for news (Horton, Rand & Zeckhauser, 2011). The studies make use of representative U.S. (chapters 2 & 3) and Dutch (chapters 3 & 4) samples. For all studies these samples are drawn from different participant pools. Thus, our findings represent populations from both an archetypal Liberal media system (the U.S., Hallin & Mancini, 2004) and a Democratic Corporatist media system (the Netherlands), providing a reasonably representative picture of Western media audiences (Brüggemann et al., 2014).

The experimental method is fitting since experiments possess a number of attributes ideally suited to studying multimodal media effects and their underlying mechanisms. Importantly, the causal direction is unambiguous since the viewing of a stimulus precedes an effect. Moreover, random allocation to experimental conditions ensures that any potentially confounding participant characteristics should be randomly, rather than systematically, distributed amongst the conditions of interest. Thus the experimental researcher can be sure that self-selection into media exposure – the endogeneity problem – is not at play (Jackson, 1992). Furthermore, stimulus exposure can be carefully controlled by the experimenter to ensure that an effect is due to the manipulation of interest. All studies in this dissertation involved rigorous and multiple

pre-tests to ensure that the visual and verbal stimuli did indeed depict the frames of interest. Importantly, through pre-testing, several potentially confounding factors were controlled for, such as stimulus emotionality and complexity, that have been shown to be important in media effects research and particularly so for the study of visuals (e.g., Lang, 1995). By doing so, the internal validity of the studies here are assured.

In all three empirical chapters that follow, a visual-verbal congruence (also known as redundancy) design is employed. This involves creating a multimodal stimulus in which the frames conveyed by the composite visual and verbal modalities are manipulated in a factorial design so that they are matching or non-matching. This design is informative for media effects researchers since it answers fundamental questions about which modality is more influential in determining outcomes on a dependent variable, or whether there is an interactive effect of visual and verbal input. The design was the driving force behind a body of literature about the contribution of visual and verbal modalities to media learning and memory (summarised in Lang, 1995). Furthermore, incongruence is relevant to media practice: it reflects the unintentional mismatches that sometimes occur in news, where visuals, often added as an afterthought and in haste from a limited pool of wire images, can be inappropriately paired with an article's text. The same problem can occur with TV, where stock images or visuals from the start of a story no longer match with coverage as it progresses. In some cases, incongruence can be attributed surprisingly large effects (Fahmy, Bock & Wanta, 2014). This design has, to the best of my knowledge, rarely and only recently been applied in the study of multimodal framing effects (see Boomgaarden et al., 2016; Powell, Boomgaarden, De Swert, de Vreese, 2015; Seo & Dillard, 2016).

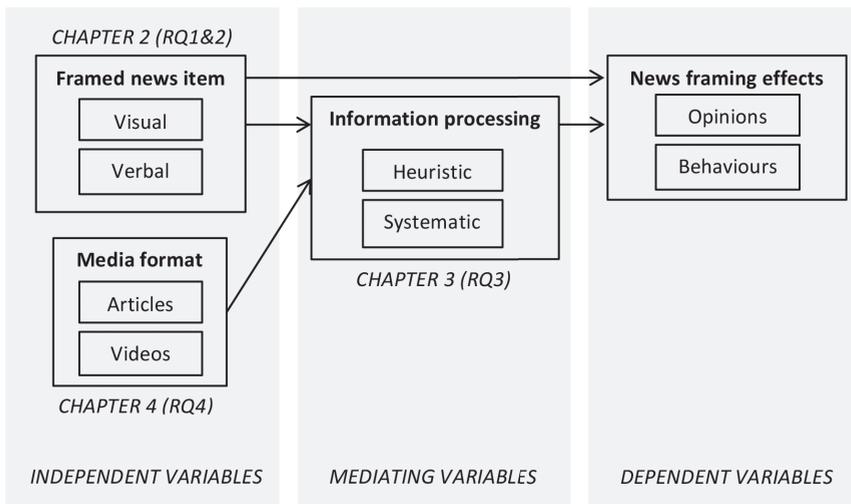
The political issues used to investigate multimodal framing effects come from war, conflict and crisis contexts – a ready source of vivid visuals. The first context involves intervention in a little-known foreign conflict (in the Central African Republic, CAR). The second context is the highly mediatized European refugee crisis. Therefore, both low (CAR conflict) and high (refugee crisis) salience issues are investigated. This multiple issue approach helps to make improve the generalizability of this dissertations' conclusions (Reeves, Yeykelis & Cummings, 2015).

The dependent variables measured are those regularly employed in framing effects research. The first key variable is opinions about the framed topic of interest – for instance whether a participant supports or opposes allowing Syrian refugees into Europe. The second key variable are behaviours, of which we measure two types. One is concerned with participants' intentions for political action – such as intention to donate money to a charity, sign a petition, or protest about an issue. In addition, actual monetary donations to charities relevant to the studied issues are taken. Arguably, donating behaviour should be less prone to social desirability effects and other confounds than self-reported opinion and intention measures.

## Outline of the dissertation

This dissertation consists of three separate studies. Each chapter documents one study and the chapters are self-contained. The chapters are empirical articles that contain abstracts, theoretical foundations, methods, results, and separate discussions and conclusions. *Figure 1* shows a conceptual schematic describing how the chapters are related.

*Figure 1.* Conceptual overview of the dissertation



*Chapter 2* proceeds by addressing the fundamental question of how the visual and verbal modalities in news articles (i.e., images and text) contribute to framing effects. This is first investigated for images and text in isolation (RQ1), followed by images and text combined as in a typical multimodal news report (RQ2). Due to the ability of visuals to evoke an emotional response, the mediation of these effects by discrete emotions is also measured as a first step in examining their mechanisms. The study is set in the US context, specifically concerning intervention in the conflict in the CAR, a context relevant to existing visual framing research, humanitarian intervention, and to recent conflicts in Libya and Syria.

The mechanisms of multimodal framing effects are systematically studied in *Chapter 3* (RQ3). Specifically, concepts from Chaiken's *heuristic-systematic theory* are used to assess the extent to which visual and verbal modalities exert framing effects via processing pathways that are more heuristic and systematic, respectively. These pathways were experimentally induced through manipulations of cognitive load and cognitive involvement, and dispositionally measured using the Need for Affect

(Sojka & Giese, 2006) and Need for Cognition (Cacioppo & Petty, 1982) scales. Two experiments set in two different contexts are presented. The first employs the same issue and frames as in chapter 2 but uses a different participant pool. The second draws on Dutch participants and their opinions and behaviours towards frames about the refugee crisis in Europe.

In *Chapter 4*, the contribution of the visual and verbal modalities to multimodal framing effects is compared across different media formats (RQ4). Using careful control of experimental stimuli, the same news stories were presented in the growing medium of news video (audio-visual dynamic media) and typical online news articles (image-text static media). The processing and effects of multimodal frames were compared, with the expectation that dynamic media would bias the visual modality and static media would enhance verbal effects. The European refugee crisis and Dutch context are used, but using a different participant pool to Chapter 3.

Finally, in *Chapter 5*, the findings from the three empirical studies are assimilated into broader conclusions, from which theoretical and practical implications are drawn. Using the models in this chapter as a guide, the democratic potential of multimodal news frames is re-assessed. The shortcomings of this dissertation are considered and future steps for developing multimodal framing theory are explored.

Taken together, this dissertation takes a first step towards studying framing effects in their proper multimodal context. The visual-verbal interactions investigated shed important new light on the democratic potential of multimodal news media by revealing their effects on political opinions and behaviours. At present, the impact of powerful news visuals alongside text are assumed but not quantified. By doing so, this dissertation brings framing theory up to date with today's multimodal media reality.