Multimodal news framing effects

Powell, T.E.

[Link to publication](#)

Chapter 5

Conclusion:
To multimodality and beyond

Summary of findings

How do visual and verbal modalities in news media combine to frame our understanding of political issues? In addressing this question this dissertation responds to long-standing claims that visuals play a powerful role in framing citizens' political perceptions, juxtaposed with an imbalance of evidence in favour of the verbal modality. As such, this dissertation contributes to the sub-field of political communication, as well as the broader communication science literature on media effects, by setting visual and verbal media modalities on an equal footing.

The relevance of multimodal news media to citizens' political opinions and behaviours was introduced in *Chapter 1*. A theoretical framework for multimodal framing effects was articulated and three hypothetical models were outlined. These models will help extract from the empirical findings the key implications for democratic processes that reflect public opinion. Four research questions were then formulated which defined the scope of this dissertation.

In *Chapter 2*, I addressed the fundamental question motivating this thesis: what is the contribution of the visual and verbal modality to multimodal framing effects? Using images and text from news articles about a little-known foreign conflict I showed that, when presented alone, visuals delivered stronger framing effects than text (RQ1). Moreover, when presented together, as in a typical news article, the contribution of each modality depended on the outcome variable: visuals drove framing effects for participants' behavioural intentions (e.g., to donate money and sign a petition), whereas text determined participants' opinions about the conflict (RQ2). I also took a first look at the mechanisms of multimodal effects by showing that emotional responses mediated framing effects driven by visuals and were less influential in effects that were guided by text. These findings provide nuance by showing that although visuals can be emotionally charged, they do not always play the leading role in multimodal effects.

I delved deeper into the underlying mechanisms of multimodal news framing in *Chapter 3* (RQ3). Using news articles drawn from two different political issues, experimental manipulations showed that when forced into a heuristic or systematic processing style, participants' opinions towards the issue and donation behaviour were influenced by the visual and textual modalities, respectively. However, individual

differences measures showed that those with a predisposition for heuristic or systematic processing were not always biased towards the visual or textual modality. For the little-known conflict issue, participants' predisposition for heuristic and systematic processing did bias their opinions and behaviours in line with the visual and text, respectively. However, for the highly salient issue of the European refugee crisis, these prior processing tendencies were not predictive of visual and textual effects. In line with prevailing theory, heuristic and systematic processing pathways are a good characterisation of how visuals and text are typically processed. However, characteristics of the individual and particularities of the political issue can interrupt these mechanisms and lead to less predictable interactive effects.

In *Chapter 4*, I compared the processing and effects of multimodal frames in different media formats (RQ4). Using coverage of the European refugee crisis, findings showed that news articles were processed more deeply than news videos and, in turn, delivered stronger framing effects on intentions to help refugees. No framing advantage was found for ostensibly psychologically-activating and vivid news videos. Moreover, for news articles compared to videos, the textual modality played a leading role in multimodal framing effects. These findings suggest that despite the increasing prominence and intuitively impactful qualities of news videos, they do not provide a particularly powerful platform for multimodal frames.

Taken together, these findings reveal that news framing effects of visual and verbal input occur through distinct processing pathways to produce nuanced effects on political opinions and behaviours. These effects depend on the medium via which they are delivered as well as citizens' individual characteristics and the outcome variable. Therefore, modalities, mechanisms and media formats matter in news framing research.

Overall Discussion

A move towards multimodality

When images are presented in isolation (i.e., uni-modally), they trigger an emotional response that can deliver stronger framing effects than text-alone. Thus, images alone can provide an unambiguous stimulus to opinion and behavioural change, even without the presence of a text to guide their interpretation (Müller et al., 2012). The same cannot always be said, however, for text alone, where the level of abstraction required meant that only those with sufficient knowledge were influenced by its frame. This moderation reinforces individual differences in the susceptibility to media effects (Nabi & Oliver, 2009; Valkenberg & Peter, 2013) and supports much past research showing increased textual framing effects with higher issue knowledge (Druckman & Nelson, 2003; Lecheler & de Vreese, 2012; Nelson et al., 1997; but not all, Haider-Markel & Joslyn,

2001; Kinder & Sanders, 1990; Schuck & de Vreese, 2006). Specifically, it is consistent with the idea that text is effective at conveying meaning but can fail if knowledge of its code and context is lacking (Geise & Baden, 2014). Taken together, in response to RQ1, I *conclude that visuals can produce stronger framing effects that text*. Furthermore, these findings provide clues to the emotional mechanisms of these effects which are explored in more depth in response to RQ3.

More relevant to everyday media practice is how visual and verbal framing devices *combine* to influence political opinions and behaviours. Based on the unique communication-related potentials of images and text, one might expect an interaction between their salience-enhancing and meaning-making qualities – with image-text congruence producing strongest effects (Geise & Baden, 2014). However, main effects of images over behaviours and of text over opinions point to outcome-specific multimodal effects. The visual modality and its emotional consequences provide an "emotivational" stimulus to action (Frijda, 1988; Valentino et al., 2011), for instance by evoking sympathy for suffering victims which triggers helping behaviour, such as donations or petition-signing. In contrast, text seems better suited for relaying arguments and causal statements about who did what to whom and why which, in turn, change political opinions (Entman, 2003; Messaris & Abraham, 2001). Of note is that this textual impact was not present when viewed in isolation, which suggests that visuals play an additional role in capturing attention and drawing the reader into a story (Zillmann et al., 2001). In response to RQ2, then, *I conclude that the contribution of visual and verbal modalities to framing is nuanced and dependent on the outcome variable*.

The conclusions thus far are broadly in line with theoretical predictions about multimodal frames that, "depending on the message type, the availability of intuitive representations, and the knowledge brought to the task, different modalities may be better suited to convey specific meaning" (Geise & Baden, 2014, p.17). Indeed, the nuanced findings suggest that multimodal framing effects depends not on the modality per se, but on the degree to which framed cues are present, stimulate attention and are relevant to the political decision at hand. This chimes with proposals from NGO research about poverty aid appeals: Emotive negative visuals can be effective at producing monetary donations in the short-term, however audiences may experience "aid-fatigue" with no real engagement in an issue without more substantive textual arguments (Hudson et al., 2016b). A comparison of visual and verbal contributions to multimodal effects over time would shed light on this contention. Going further, employing a competitive framing approach (Chong & Druckman, 2007; Lecheler & de Vreese, 2013) in longitudinal studies would maximise external validity and better reflect today's diverse news media environment.

5

Inside the 'black box'

How multimodal news frames are processed to deliver the effects described above was the focus of RQ3. In addressing RQ1 and RQ2 we peeked inside the 'black box' to show that visuals play a leading role in evoking emotionally mediated effects. To explicitly test the processing pathways to multimodal effects I drew on well-established and operationalizable concepts from persuasion research: Chaiken's (1980) *heuristic-systematic theory*. Experimental manipulations showed that relatively more heuristic and systematic processing does underpin visual and verbal effects, respectively. This is in line with the unique properties of visuals and text (Barry, 2005), their theorized role in framing (Coleman, 2010; Geise & Baden, 2014) and media effects more broadly (Graber, 1980; Lang et al., 1999). However, importantly, individual differences data emphasised that these processing pathways also depend on the particular political context at hand, the individual characteristics of the viewer and particularities of the news article. Indeed, striking and emotionally-charged visuals might trigger some viewers to engage in a more detailed and systematic scrutiny of a news story (Russell, Marcus & MacKuen, 2000). Furthermore, a news text can contain cues to heuristic decision-making, such as party positions, candidate endorsements or emotional language (Lau & Redlawsk, 2001; Nabi, 2003).

From these findings I *conclude that there exist qualitatively distinct but interactive processing pathways to multimodal framing effects*. What does this mean in practice? From neuroscience we know that brain anatomy is wired for fast and automatic processing of visuals (Gazzaniga, 2004; Lang et al., 1999; Pessoa & Adolphs, 2010). Thus, as citizens go about their daily, habitual and minimally-effortful news routines (e.g., checking a news site on a smart-phone, skimming a free newspaper on the train, glancing at the evening news broadcast), visual content is easily processed without triggering engagement of systematic processing systems. Given the majority of citizens' low interest in political affairs, this can be described as a "default" processing mode for news media (Zaller, 1992). The empirical findings show that in this minimally effortful mode, subsequent decisions are more likely to be driven by easily-processed visual cues. However, such automatic routines are often interrupted and systematic processing is engaged – for instance by characteristics of the stimulus or issue (e.g., an especially salient topic), individual motivations (e.g., prior interest in an issue), or environmental cues (e.g., a friend offering their opinion). In such instances that trigger more effortful processing, results suggest that decision making of the viewer will be more readily influenced by textual cues in news frames.

The processing states described above are, of course, idealised and, in reality, attention can be briefly piqued and processing capacity temporarily increased, whether to a high or medium capacity. In such conditions, textual and visual cues are likely to interact in political decision-making. As part of this process, a long list of factors may

determine the reliance on text and visuals in frame processing. These include: interest and knowledge of the topic; content and complexity of the news item; its specific composition; an individuals' trait personality characteristics and state motivations and mood; their prior attitudes towards the issue or actors; the specific political decision at hand; the media format and the viewing experience it engenders; and one's immediate or anticipated environment. These are all prime variables for future studies investigating the link between visuals and text and heuristic and systematic frame processing. In the *Theoretical Implications* section that follows I offer concrete hypotheses on several of these factors.

Multimodal media formats

The use of videos in online news is on the rise, for mainstream news sites and social media platforms alike (Kalogeropoulos et al., 2016). Despite their ostensibly psychologically-activating properties, multimodal frames delivered in video form were *less* effective in stimulating helping behaviour than news articles. The self-paced reading of a news article led to deeper processing of the news story (e.g., Sundar, 2000; Tukachinsky et al., 2011) which, in turn, increased participants' intentions to help refugees. Furthermore, this more elaborative processing of news articles compared to videos also increased the influence of textual content in political decision-making. By comparison, news videos – a highly visual medium with moving images that resemble real-world visual experience – did not stimulate strong emotional responses nor influence opinions or behavioural intentions.

These findings offer answers to RQ4: *multimodal news videos do not deliver stronger framing effects than articles. In fact, by virtue of being processed more deeply, articles deliver stronger effects. Moreover, this deeper processing causes the verbal component of a multimodal news frame to exert stronger effects in article compared to video format.* The absence of effects for vivid news videos runs counter to theoretical expectations of media effects researchers (Green et al., 2004 Taylor & Thompson, 1982) and framing theorists more specifically who argue that frames operate through increasing the salience and accessibility of considerations in the mind of viewers (Entman, 2003). Should not videos, with their rich and salience-enhancing qualities, exert stronger framing effects? Contemporary conceptualizations of framing effects have argued to the contrary: viewers also assess a frame's *applicability* (Scheufele & Tewksbury, 2007) to their own "constructed-reality" of political issues (McQuail, 2005). If deemed applicable in light of a citizen's prior knowledge and attitudes towards the issue, then a frame will exert an effect (Scheufele & Tewksbury, 2007). I argue that news articles allow the reader to reflect, re-read and introspectively imagine the story at their own pace, and this facilitates both accessibility *and* applicability processes. The time afforded for associations related to the issue to be brought forward in the mind of the reader

5

enables accessibility, whilst the opportunity to actively and systematically reflect and disambiguate parts of the frame facilitates applicability considerations.

Therefore the qualities of the media format influence multimodal framing effects. Framing research has, aside from the focus on the emotional content of news frames (e.g., Lecheler, Schuck & de Vreese, 2013; Nabi, 2003), been largely agnostic to fundamental characteristics of the stimulus – for instance medium, complexity and emotional arousal. Instead, the field has deployed numerous issue-specific frames, studied characteristics of the receiver as moderators of framing effects (Slothuus, 2008), and a variety of factors (including emotions) as mediators (e.g., Lecheler, Bos & Vliegenthart, 2015). Throughout this dissertation stimuli were matched stimuli on several fundamental factors pertaining to media complexity and emotionality. Future framing research should respond to critics of the field (Cacciatore et al., 2016; Scheufele, 2004) in a similar manner by taking such stimulus factors into account, or even systematically manipulating them. Indeed, much could be gleaned from the approach of the broader mass communication literature which has sought to chart stimulus characteristics that increase or decrease media learning and persuasion (e.g., Lang, 2000). Investigating whether factors important for media learning are also relevant in framing is a prime avenue for further research. Doing so would shed much needed light on the extent to which these separate fields of media effects share underlying mechanisms (e.g., Baden & Lecheler, 2012), and, importantly, would elucidate framing as a paradigm described as fractured and fading (Cacciatore et al., 2016; Entman, 1993).

With regard to media format, future research can usefully draw on the work of Annie Lang (e.g., Lang, 2000; 1995). She showed that, although numerous studies argue for the persuasive effects of vivid audio-visual news, structural characteristics of the medium may qualify these effects. For instance, arousing content, production pacing (Lang et al., 1999) and related and unrelated camera cuts (Lang et al., 1993) were observed to influence memory of TV content, especially the verbal stream. Such factors could also help account for why framing effects are relatively weaker in news videos since their inherent qualities may interrupt frame processing, as evidenced by their more shallow processing. This dissertation represents a first step in bringing the careful consideration of stimulus characteristics in different media formats to the study of framing effects. Future steps should investigate how viewing these formats on media platforms like Facebook that contain highly integrated multimodal content might produce different effects to the more structured offerings of mainstream online news sites.

The democratic reality of multimodal news frames

What are the implications of these findings for democratic processes that reflect public opinion? In this section, I return to the hypothetical models of multimodal framing effects outlined in *Chapter 1*. They relate to the relative power of visuals versus text in

the framing process and structure the discussion that follows about how multimodal news frames contribute to democracy. The models are: *1) Strong visual effects; 2) Weak visual effects; 3) Moderate and conditional visual effects.* In light of the conclusions of the three empirical chapters I consider which model is best supported. Note again that these perspectives conform to an ideal-type where stylized models have been substituted for the complexity of reality and some overlap necessarily exists between them. The perspectives can be most usefully deployed as they are here: as a tool to aid investigation of multimodal framing effects, as well as media effects more broadly – across different topics, contexts, populations and message-types.

Depending on the political decision at hand – either forming an opinion or deciding how to behave – visuals or text exert a relatively stronger influence, respectively (RQ2). This is firmly in line with the conditionality of Model 3. However, the comparatively stronger influence of visuals compared to text when they are presented in isolation (RQ1), as well as the mediation of visual effects by emotions, also speaks to Model 1. The emotional outpouring of aid in response to the image of Alan Kurdi in 2015, as well as natural disasters such as the 2004 Indian Ocean tsunami and 2010 Haiti earthquake coverage, are testament to potential power of images. However, counter to theories promoting powerful visual media effects, for instance the CNN-effect (Livingston, 1997), the findings show that verbal content can play the leading role.

More support for Model 3 was provided by the investigation into the mechanisms of multimodal framing (RQ3). Although heuristic and systematic pathways characterise visual and verbal frame processing, this did not hold for all individuals and issues. In a default low-effort mode, visuals can effectively impart information and influence decisions with minimal processing demands (Lang et al., 1999). However, depending on a number of factors – including the motivations of the viewer and the topic at hand – citizens may be inclined to systematically process an accompanying text.

Compared to news videos, the deeper processing and stronger effects produced by news articles (RQ4), especially for the textual modality, is in line with minimal visual effects proposed by Model 2. In contrast with scholars advocating the power of dynamic TV visuals (Graber, 2001), the perception of emotionally-evocative visual frames might be prone to a "first-person effect" (Perlmutter, 1998). This involves journalists or politicians projecting their own emotional response to visuals, which is magnified by their proximity to the issue, onto the population at large – to whom they have access due to their profession. Thus, the real-world impact of visuals may be overstated, and particularly so for news videos. Instead, by providing the opportunity for deliberative processing, news articles and the text therein can be effective framing devices.

Taken together, the weight of empirical evidence points towards Model 3: *moderate and conditional influence of visual and verbal modalities in framing effects*. This is further supported by the relatively small effects throughout in all empirical chapters. With this

in mind I now offer a normative assessment of this dissertations' conclusions in light of the aspirations of liberal participatory democratic values (Strömbäck, 2005; Althaus, 2012).

Contrary to the Enlightenment ideals of literacy and reason as the vessel for political decision-making, I argue that the influence of vivid visuals need not be a cause for democratic concern. Visual effects on political opinions and behaviours are contingent on the outcome variable, context, and medium and thus are not large and across-the-board. Moreover, visual effects can sometimes be overshadowed by verbal input. That is not to say that visual political effects are non-existent. On the contrary, research has emphasised their impact, particularly when judging candidates (Grabe & Bucy, 2009; Todorov et al., 2005), and many instances – such as the image of Alan Kurdi in 2015, the beheading of western journalists by ISIS in 2014, and Sarin gas attacks in Syria in 2013, to name a few recent examples – illustrate that an emotional response to images trigger (at the least) a discussion about international affairs issues and raise them up the public agenda (Wanta, 1988). Stimulating public deliberation about politics is certainly a positive contribution to democracy. However, deciding complex matters of foreign policy solely on an emotional response is not uniformly desirable. Fear in response to images of massed migrants at border fences, for instance, should not be the single determinant of immigration policy. Fortunately, my findings show that such reactions are probably not universal enough to engender an enduring change in public opinion. Moreover, practicalities of the situation and the historical and unfolding political dynamics are likely to outweigh visually-evoked popular sentiment in policy-making (Perlmutter, 1998; Robinson, 2002).

That said, normative assertions about multimodal effects should also consider their influence over different sectors of society. More or less politically sophisticated individuals, for instance, process information differently which leads to different outcomes. Political sophisticates are typically good at using heuristic decision cues, such as a candidate's appearance, to make an informed voting decision, whereas non-sophisticates perform less well (Lau & Redlawsk, 2001). Conversely, the ease with which visuals are processed and impart information mean that they may help to close the knowledge gap between citizens possessing different levels of education (Grabe & Myrick, 2016). Thus, as well influencing visuals can inform, and future research aiming to deliver normative assessments about multimodal media should study how these outcomes interact in political decision-making.

To summarise: the evidence thus far points to citizens being only moderately influenced (see Model 3) and potentially better informed (e.g., Graber, 1996) through visuals in multimodal media. Ultimately, more focus is needed on the individual and stimulus-level differences that determine responses to multimodal news frames, which then should be linked to the aggregate level. As such, the jury is still out. However, when

considered in their proper multimodal context the democratic potential of visuals is promising.

Theoretical implications

The core theoretical contribution of this dissertation is to place visuals alongside the verbal modality to build a *multimodal* understanding of how news frames influence decision-making. This heeds the call of scholars like Doris Graber and Roger Masters to take visuals seriously in political communication, and builds upon the efforts of those who have focused more squarely on visual media effects (e.g., Grabe & Bucy, 2009). Indeed, multimodality has mainly been studied in the realm of learning and memory (Graber, 1990; Lang, 1995; Reese, 1984), much of which focused on non-political content (e.g., Lang, 2006; Unnava et al., 1994), where the visual and verbal modalities are considered in terms of their information value. Thus, the empirical chapters fill a considerable research gap in our understanding of meaning-making in multimodal political communication. Moreover, they contribute to broader media effects research by systematically studying the processing and effects of multimodal messages in different media formats.

The empirical chapters also build directly on Geise and Baden's (2014) theoretical formulations of multimodal framing introduced in *Chapter 1*. The findings provide much corroborating evidence for their propositions, but also introduce nuance (e.g., outcome-specific results) as well as counterintuitive findings (e.g., stronger effects in articles than videos and articles) which extend current theory. Thus I now spend some time comparing the empirical findings with Geise and Baden's (2014) theoretical assertions, addressing each of their propositions in turn. First, visuals are indeed superior to text in their salience attribution (proposition 1a *picture superiority* supported) and the open structure of visuals does allow interpretations and opinions to be guided by text. However, visuals can also convey a clear and powerful frame in their own right, which can evoke emotions and trigger political behaviours (proposition 1b *open/conventional structure* partially supported). Second, the conventional codes of text can be particularly reliant on prior knowledge to decode, whereas, as long as images present a clear frame they can be quickly and easily interpreted with minimal polysemy (proposition 2 *polysemy/strong codes* not supported). That said, results suggest that this is contingent on the context, the specific frame, and the media format. Third, the relatively more systematic processing of text does guide interpretation better than the heuristic cues of images (proposition 3 *associative/propositional configuration* supported). Fourth, rich visuals do seem to provide a surplus of information for integration. Particularly in video form where structural production features and forced-pacing can detract from the frame compared to the more focused integration of verbal content (proposition 4 *iterative/focused integration* supported).

5

This updated state-of-affairs not only highlights where theory and empirical evidence diverge, but helps to sketch a research agenda for multimodal framing theory. Future studies should identify the conditions under which visuals' relatively open and associative structure engenders polysemy and when it does not. One way to do this, in line with the earlier suggestion of increased focus on stimulus characteristics in framing research, is to systematically study ambiguity in news visuals. This would identify what level of visual clarity is required for visuals to drive interpretations, and when visual ambiguity would cause meaning to be determined by verbal content. This could be done with reference to specific political issues and contexts. For instance, by comparing the clear and readily-visualized contexts of war, conflict and protest, with issues that are more challenging and ambiguous in their visualisation, like climate change – where the onus is now on communication professionals and scholars to effectively convey the acute importance of this complex issue (O'Neill & Nicholson-Cole, 2009; Rebich-Hesphana et al., 2015). Of course, future research effort should also be devoted to cross-examining the other propositions supported by this dissertation.

Looking beyond Geise and Baden's model, more research is needed on how multimodal effects endure over time and in today's crowded, fragmented and competitive media environment. Indeed, future studies should focus on the extent to which motivated political reasoning, especially along partisan lines (Slothuus & de Vreese, 2010), applies to guiding our interpretation of visual as well as verbal news content. Another avenue to explore is the effects of the fleeting and fragmented viewing experiences promoted by scrolling through social media "timelines". Does scanning of multimodal content in this way promote a more heuristic information processing approach that biases visual content over text? Finally, comparative studies could illustrate how the present media landscape affects the emergence of recent "iconic" images (e.g., Alan Kurdi) compared to those from the twentieth century (e.g., the 'Napalm Girl' or 'Tank Man' images). The trajectory of images from different generations to icon status is likely to differ greatly. Whether their effects on public opinion also differ is an interesting empirical question.

Finally, this dissertation has empirically linked visual communication to frame processing research – identified as a lifeline for framing theory (Coleman, 2010; Cacciatore et al., 2016). Having taken this first step, I argue that a multimodal perspective can bring clarity to the underspecified mechanisms of accessibility and applicability (Scheufele & Iyengar, 2012). Specifically, I propose that the typically heuristic processing of salience-enhancing visuals have a particular ability to increase the accessibility of news frames in the minds of citizens. Visuals are both readily mediatized (e.g., the highly visual and personalized European refugee crisis) and quickly and easily processed by the viewer (Barry, 2005). Thus, highly visual issues should be both prominent in news and easy accessible in the mind of even minimally politically sophisticated citizens. In contrast, the systematic processing of a less

salient text should be an especially useful guide for determining the applicability of a frame. This may be particularly the case for complex issues, whose essence is less readily visualised (e.g., scientific and less temporally proximate issues such as climate change) and thus necessitate more systematic processing favouring textual content. So, for a highly salient issue that is subject to repeated news coverage, the visual content of multimodal news should increase its accessibility, but applicability may already be determined by one's prior attitudes. For a less salient issue, visuals can also engender accessibility as long as the topic is visualizable enough to stimulate interest in the issue, whilst a text would be important for an applicability judgment as long as the reader has sufficient knowledge to decode it. Such complex and non-linear relationships between contextual, individual and stimulus factors provides ample opportunity for hypotheses testing and specifying the concepts underpinning multimodal framing effects.

Practical implications

This dissertation offers up a number of concrete practical implications for producers as well as consumers of multimodal news media.

For journalists, the key take-away is that the choice of image to accompany a news text matters. Although a text is a stronger guide for the perceived meaning of the story and opinion formation (e.g., that measures to resettle refugees are failing and more resources should be devoted to help), the chosen image will determine viewers' emotional response and subsequent actions (e.g., sympathy for homeless refugees which triggers donating behaviour). Therefore an unintentional clash between visual and verbal elements can lead to unwanted consequences. Furthermore, even when images are displayed in isolation – such as in online photo galleries or albums on highly visual social media platforms like Facebook – viewers are still able to extract a visual's meaning and the resultant emotional response can influence political opinions and behaviours.

Journalists should also consider the viewing experience of their readers, as this can exert a predictable influence on multimodal media effects. Viewers who foresee debating a news issue are more likely to be influenced by a story's text. With this in mind, journalists could include a debate/chat function alongside news content to encourage deeper processing of textual arguments in their articles. By contrast, the perceptions and behaviours of a reader who is distracted (e.g., listening to music) will be swayed by news images. This is a useful insight for producers of charity material who could consider using emotional soundtracks to bolster the emotional and behavioural impact of their promotional images.

Political parties, charities, NGOs, and other publishers attempting to endorse a particular frame should consider the format of their media message. Videos may be rich and vivid, but articles are more likely to influence viewers' behaviours since they

5

are processed in greater depth. This is consistent with recent research showing that citizens still overwhelmingly opt for articles over videos when online (Kalogeropoulos et al., 2016), but is in contrast with increasing resources that have been devoted to creating video content in recent years (Bock, 2016).

Finally, citizens should be receptive to their own perceptual preferences when viewing news media. This dissertation showed that the opinions and behaviours of those who tend to rely on visual or verbal thinking in daily life will be more strongly influenced by images and text, respectively. Thus, to achieve a more gratifying news media experience, users would be advised to consider emerging online journalistic formats in which news content and modality can be tailored or personalized to their own preferences (Bas & Grabe, 2015; Nguyen et al., 2017; see also blendle.com). The qualities and effects of such personalized visual political communication is a ready avenue for future research.

Methodological reflections

This dissertation used simple manipulations of visual-verbal congruence to test fundamental questions in multimodal framing theory. By adopting this design from studies of media learning and memory, the empirical findings set the stage for further investigations of both multimodal framing and media effects more broadly. The design, for instance, can be used to answer questions such as how visual and verbal streams contribute to media effects over time, and the extent to which each modality contributes to news selection.

The degree of experimental control of potentially confounding factors is a particular strength of the empirical chapters. In all studies I matched the visual and verbal stimuli for several key variables in media effects research – including arousal, valence, salience, complexity, ambiguity and credibility. By doing so one can be highly certain that the findings are a result of the experimental manipulations and not an artefact. Maximising internal validity in this way is a novel approach to concept explication in visual communication and is a vital step towards heeding the call for increased rigour in framing research (Liu & Scheufele, 2016; Scheufele & Iyengar, 2012).

This strength, however, could also be considered a weakness. Careful pre-testing to select a small number of stimuli in specific political contexts has the drawback of limiting the generalizability of the conclusions. I attempted to mitigate this issue, especially in *Chapter 3*, by using multiple issues relevant to different countries. Moreover, the control of extraneous factors means that the stimuli in the empirical chapters can be considered as archetypal visual and verbal exemplars of the frames they depict. Nevertheless, to complement the approach taken in this dissertation and evaluate the veracity of its conclusions, I urge for a multiple stimulus and context approach in future experimental studies of multimodal framing effects (Reeves et al, 2015). Better still

would be to seek triangulation through alternative methodologies. For instance, by harvesting and analysing visual and textual content on social media and linking this to online or real-world behaviour such as protesting (Corrigall-Brown & Wilkes, 2011). Or through field experiments whereby naturally-occurring variation in the occurrence of visual and multimodal political media is exploited – such as estimating the political effects of the use of candidate photographs in voting booths in different countries.

The use of the experimental method necessarily limits external validity, although all stimuli in the empirical chapters did contain material drawn from actual news media. In *Chapter 4* this also included a professional broadcast journalist who reviewed the articles and provided voice-overs for the news videos. A related artificiality is that participants experienced forced exposure to one of the stimulus conditions and thus were not able to select content of their own choosing. For a first investigation into multimodal framing effects this was a necessity. Moreover, this issue might not be so great since particularly powerful visuals are typically offered by news wire companies which are then used by multiple outlets and shared widely on social media – thus increasing the potential for incidental exposure. Nevertheless, testing multimodal media effects in free-choice environment is a vital next step, especially since the diversity of media choice engenders (partisan) selective exposure which is reinforced by the filter bubbles of social media (Sunstein, 2009; Van Aelst et al., 2017). Estimating multimodal effects in conditions of selective exposure is a vital step in gaining a full picture of the contribution of visuals and text to media effects (Arcenaux, Johnson & Cryderman, 2011). Existing research suggests salient images could play a leading role (Zillmann et al, 2001). Moreover, this approach would help address the more fundamental question of the continued relevance of framing in today's highly fragmented media environment (Cacciatore et al., 2016).

Concluding remarks

People today are bombarded with more images than ever before in human history. How visuals interact with text in citizens' sense-making of political issues is an important and understudied phenomenon at the heart of this dissertation. Despite a prevailing wisdom that powerful news images are "mainlined directly into the democracy's emotional bloodstream," (Morrow, 1993, p. 36), this dissertation shows that the impact of visuals in the multimodal media environment is more nuanced. Visuals evoke an emotional reaction which can drive political behaviour, but their heuristic cues often play second fiddle to systematically processed verbal content, especially when presented in news videos compared to articles. As such, I argue that visuals play a positive role in connecting with an often politically-detached audience. Moreover, citizens' political preferences are not mindlessly given over to their gut-reaction to visual content. Rather, modalities, mechanisms and media formats matter when considering the effects of news

frames. By studying these factors this dissertation has taken a first step towards a fully multimodal empirical and theoretical explication of framing effects. Only once visuals are considered in their proper multimodal context do we achieve a clearer picture of their democratic potential.