

Testing Three Measures of Verbal–Visual Frame Interplay in German News Coverage of Refugees and Asylum Seekers

VIORELA DAN¹
LMU Munich, Germany

MARIA E. GRABE
BRENT J. HALE
Indiana University, Bloomington, USA

Drawing from framing theory, this article operationalizes and tests three ways to measure how verbal and visual modalities interplay in audiovisual messages to produce meaning. The measures include (a) a ratio of verbal to visual frames; (b) an association rules learning (ARL) procedure; and (c) in-depth analysis of the full audiovisual material. As a step toward validating the measures, they were applied to a sample of German television news stories ($n = 98$) about refugees and asylum seekers. Though the three measures produced varied results, verbal–visual frame redundancy and congruence were consistently more common than mismatches. Measures differed in the level of effort required to implement them, sample sizes they could handle, and the informative value of results. Future studies are advised to combine the ARL procedure with an in-depth analysis.

Keywords: multimodality, framing, audiovisual redundancy, television

Gripes about the lack of visual focus in media research have grown steadily in volume (both bulk and decibel) over the past 20 years (Coleman, 2010; Dan, 2018; Graber, 2001; Messaris & Abraham, 2001; Powell, Boomgaarden, De Swert, & de Vreese, 2019). The substance of these complaints typically falls within the realm of the Gutenberg legacy (Graber, 2001), pointing to a long-standing cultural tradition that glorifies the written word as the conduit of serious knowledge while treating images as feeble sources of information. Yet a number of scholars have explored images for information value, persuasiveness, and mobilization

Viorela Dan: viorela.dan@ifkw.lmu.de

Maria E. Grabe: mgrabe@indiana.edu

Brent J. Hale: brjhale@indiana.edu

Date submitted: 2019–09–14

¹ This work was supported with 17,000 EUR by the Indiana University and Freie Universität Berlin under the Strategic Partnership Initiative. The authors thank Eva Reitmeier, Jakob Huber, Antonia Mann, Carolin Csakli, and Lea Königer, for their assistance in collecting the data; Juliana Raupp, for input on a research instrument; and Martin Borkovec, for his advice on the machine learning procedure.

Copyright © 2020 (Viorela Dan, Maria E. Grabe, and Brent J. Hale). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at <http://ijoc.org>.

potential (Coleman, 2010; Grabe & Bucy, 2009; Prior, 2014) that compelled a recalibration of our discipline's ambit. Indeed, the progression toward taking visuals seriously is notable—to the point where the old research tug between images and words is transcended by fully deployed multimodality.

The study reported follows that lead. We devote our attention to the assessment of modality interplay in audiovisual messages—by which we mean the way in which verbal and visual components interact with one another. Specifically, we are interested in the extent to which modalities support or counter each other in producing meaning. Full support is generally known as redundancy and partial support as congruence, while mismatch refers to instances where verbal and visual modalities convey meanings that are incompatible. Despite wide recognition of the differential impact of redundancy, congruence, and mismatch on attention and memory (Dan, 2018; Lang, 1995), current scholarship has yet to develop fine-grain measures for detecting these interplay scenarios in media content.

In response, we developed and tested three interplay measures, drawing conceptual and operational insights from framing theory (Dan, 2018; D'Angelo et al., 2019; Entman, 1993; Reese, 2007). Unlike most existing work that assesses the semantic relatedness between audio and visual modalities, we measured the extent to which frame interplay emerged—as redundant, congruent, or mismatched. Focusing on frame interplay, including the gradual uncoiling of meaning over the course of a message, rather than semantic (word-to-image) relatedness arguably offers a more nuanced approach to understanding audiovisual meaning-making.

Given the methodological pluralism of our field, various frame interplay measures might be needed. Developing and comparing three measures offer a good start to that end. We report here on two quantitative measures, including one that determines the ratio of modality interplay and another based on machine-learning procedures, and a third involving a qualitative–quantitative mixed-method approach. A sample of German television news stories about refugees and asylum seekers (RAS) was subjected to analysis to compare the three measures we designed. As such, this project served as a means of developing audiovisual analysis tools while also offering insights into the framing of a contemporary news topic that has been described by the German Interior Minister, Horst Seehofer, as “the mother of all problems” (“Migration,” 2018).

Multimodality in Audiovisual News

Two bodies of research, one related to audiovisual production and the other related to memory formation, point to the merits of analyzing the interplay between verbal and visual components of audiovisuals to understand the packaging of news information and user comprehension thereof. Video editing practices commonly used in the journalism profession are consequential to the construction of meaning. Audiovisual journalists typically take a word- or visual-centered approach to storytelling, each with its strengths and drawbacks. In Germany and Great Britain, journalists often start from a visual basis (Silcock, 2007), laying down video and adding verbal narration to the visual track as a way of explaining or extending the visual content. This visually driven reporting practice limits reporters to convey information that can be visualized. It also stands in stark contrast to the so-called wallpaper video (Shook, 1994) approach used predominantly in U.S. journalism that employs visuals as a secondary decorative layer to the verbal narrative. In other words, journalists operating from this verbal-centered approach roll video out like

wallpaper over the verbal narration that contains the gist of the story—often not taking advantage of how powerful images can be in conveying ideas (see Shook, 1994; Silcock, 2007).

As a result of editing practices, three potential visual–word relationships could occur: redundancy, mismatch, and congruence. Audiovisual redundancy is known in the industry as say-dog-see-dog storytelling (Stewart & Alexander, 2016). Put simply, audio and video components echo each other. For example, the voice-over accompanying footage of a politician surrounded by adoring crowds describes her as a candidate with high mass appeal. Audiovisual mismatch comes about when there is no crossmodal correspondence. Some news topics are notoriously hard to visualize, and visual material is often unavailable, prompting journalists to turn to stock footage. One of the best-known cases of this dates back to the Reagan presidency, when Lesley Stahl, of CBS News, aired what she thought was a critical report on Reagan. Though the report’s narration was sharply critical, the video track was highly flattering of Reagan. Expecting criticism from the White House after the report aired, she instead received word that the audiovisual mismatch was seen as beneficial to the president. Audiovisual congruence falls somewhere between redundancy and a mismatch. Known in the industry as touch-and-go linkage (or cross-scripting), this production feature is marked by dynamic sequences that allow the two modalities to converge for a few seconds (touch) and diverge again (go; Shook, 1994). For instance, a news story about a political candidate might describe him as a tireless campaigner, with images of him working the rope lines (touch), but transition to scenes inside his campaign office (go) when the voice-over addresses the lack of support for the candidate within his political party. Previous studies suggested that about two-thirds of TV news stories exhibit characteristics more clearly aligned with say-dog-see-dog and touch-and-go linkage than with crossmodal mismatch (Graber, 1990; Walma van der Molen, 2001).

Experimental work, both in communication (Crigler, Just, & Neuman, 1994; Zhou, 2004) and in neuro and cognitive science (De Gelder & Bertelson, 2003; Fujisaki, Shimojo, Kashino, & Nishida, 2004; Spence, 2014), tested the impact of modality interplay on audiences and produced two findings that inform our work. First, information processing varies across modalities—with regard to brain regions, pathways, speed, and priority. Second, audiovisual messages yield different effects based on the extent to which what is heard is matched by what is shown. In short, redundancy tends to advance memory and comprehension (Son, Reese, & Davie, 1987; Walma van der Molen & Klijn, 2004). By contrast, modality mismatch leads audiences to cognitively disentangle multimodal content, favoring visual over audio modalities in memory formation (Grimes, 1991; Hsia, 1968). In summary, redundancy appears to be most conducive to memory formation, followed by congruence and mismatch.

Framing Theory and the Assessment of Modality Interplay

Framing theory recognizes that storytelling constructs reality by selecting and emphasizing certain aspects while disregarding others (Entman, 1993; Reese, 2007). Framing scholars note that frames generate a range of meanings—from imprinting definitions of social issues, like RAS, to assigning blame and offering solutions and moral evaluations (Entman, 1993; see D’Angelo et al., 2019). Soon after framing entered the area of journalism studies in the 1990s, the idea emerged that visuals contain frames, much in the way that words do (Coleman, 2010; Messaris & Abraham, 2001). After a period of exclusive focus on the verbal component of messages (Matthes, 2009), scholars have shifted interest into visual framing. When applied to the study of audiovisual modality interplay, framing theory suggests that assessing the

relationship between verbal and visual frames may be more telling (Dan, 2018) than analyzing semantic relatedness among these modalities—as work in the audiovisual redundancy tradition has done.

These two approaches can be set apart from each other in the following ways. Frame interplay analyses have a deductive trajectory—starting with frames that were identified in existing literature pertaining to the topic. In the case of our study, this meant drawing from existing bodies of research that have shown that a number of negative and positive frames are persistently used in news stories about RAS around the world. Frame interplay analyses assess how identified frames unfold in audio and visual modalities across the story and to what degree frames in both modalities echo or contradict each other. The redundancy tradition is based in largely inductive procedures that focus on single sentences (or portions of sentences) and how these are matched or countered by visuals. For example, the semantic relatedness perspective would categorize a news story as redundant when visuals shown on-screen illustrate what is heard (i.e., the persons, locations, and events mentioned in the narration are displayed when mentioned) without consideration of potential frames contained in the story. By contrast, a perspective informed by framing theory would first detect the presence of a frame in either modality and then track how it unfolds verbally and visually to assess its frame interplay.

The two approaches also vary in focus. Frame interplay analyses account for verbal and visual cues that may be temporally scattered across a message, whereas semantic relatedness studies are focused on the simultaneous occurrence of verbal and visual content. Thus, frame interplay studies are mostly concerned with how meaning is constructed temporally over the course of an audiovisual message (in which the audio and visual modalities may interact asynchronously), making the analysis a holistic exercise. In contrast, the aim of semantic relatedness is to determine if the visuals shown on-screen illustrate what is heard at the same time (i.e., persons, locations, and events mentioned in the narration, as they are mentioned). Accordingly, a message would have to fulfil different criteria to be categorized as redundant, congruent, or mismatched through each tradition. What a semantic relatedness perspective might identify as a story with mismatched audio and visual channels, a frame interplay analysis might categorize as frame redundant. Both approaches have strengths in application, serving different goals in the studies that they serve. While the semantic relatedness perspective has a long tradition that also informed experimental work on audiovisual redundancy, the frame interplay concept that we explicate is in an early developmental stage. Content analytical studies employing the frame interplay measures we propose here will be well positioned to contribute to experimental work on media cognition, including information processing of stereotypes and memory formation related to different conditions of frame interplay.

Developing Three Measures of Audiovisual Frame Interplay

Our aim is to develop and test three fine-grain measures of audiovisual frame interplay in media content. Measure 1 represents an adaptation of an existing ratio for the computation of frame interplay from two sets of mono-modal data, collected separately from verbal and visual modalities (Dan, 2018). The ratio is computed in SPSS for each news story in sample by dividing the number of redundant frames in the two modalities (verbal and visual) by the highest number of possible frame pairs in that story. The decision to truncate audiovisual material into its verbal and visual component before data collection is rooted in previous research suggesting the human brain's information processing bias for images. Indeed, attempting to identify

verbal and visual frames straight from audiovisual material is likely to be difficult. Specifically, frames in the verbal modality may be missed when they differ from those in the visual modality (Dan, 2018). Nonetheless, coding mono-modal material, as done in Measures 1 and 2, might subvert the very phenomenon we hope to measure—audiovisual frame interplay. For this reason, Measure 3 also includes an in-depth analysis of full audiovisual material—in addition to knowledge of data collected from mono-modal material.

Measure 2 is our own, and constitutes a suggestion for determining frame interplay using association rules learning (ARL), a rule-based machine-learning method (Agrawal, Imieli, & Swami, 1993; Kotsiantis & Kanellopoulos, 2006) implemented in R. Though still new in media studies, this method is widely used in other disciplines. In economics for instance, ARL is used to identify regularities in consumer behavior. In e-commerce, it enables purchasing recommendations based on the items already located in virtual baskets (e.g., customers who bought razors are made aware that aftershave is often bought with it). We use ARL to identify regularities in reporting practices in much the same way. Specifically, if a journalist uses Frame X, they are likely to also use Frame Y. Thus, Measure 2 assesses the joint distribution of verbal frames and visual frames—that is, frame pairs consisting each of one verbal frame and one visual frame that occur together in the material. Moreover, the procedure identifies those frame pairs that stand good chances to be generalizable beyond the material analyzed.

Measure 3 sets criteria for assessing two types of modality interplay (congruence, mismatch) rather than frame prevalence. This analysis is limited to those audiovisual messages identified for featuring at least one instance of nonredundant frames across modalities. Thus, messages with frame redundancy across modalities and those featuring frames in only one modality are disregarded. Measure 3 requires that coders conducting the in-depth analysis are provided with the coding decisions of mono-modal coders. Specifically, they should know which verbal and visual frames were conveyed in each audiovisual message, whether they were countered or not, and when the frames and/or the countering occurred. They record if the different frames in the two modalities are congruent or mismatched, and provide justifications for each coding decision in an open field. Thus, coders are trained to recognize modality interplay—congruence or mismatch, which are operationally defined—based on the interrelationship between frames and intrarelation of counterframing. Congruence and mismatch are not conceptualized as mutually exclusive categories. Rather, Measure 3 allows for the possibility that verbal and visual modalities interact in different ways within the same news story.

The three measures proposed here differ along at least four dimensions (see Table 1), including the methodological skills required to assess frame interplay, the type of data constituting the basis for the assessment, the time when this assessment is carried out, and the kind of results produced by the analysis. First, Measures 1 and 2 require quantitative methodological skills, including the computation of ratios and the use of machine-learning techniques in R, whereas Measure 3 uses a qualitative–quantitative mix. Second, Measures 1 and 2 rely on two sets of data collected from mono-modal material—one on the occurrence of verbal frames, the other on visual frames—which become linked during the data analysis process. In contrast, Measure 3 requires in-depth analysis of multimodal material, specifically cases that were identified during mono-modal data analysis as congruent or mismatched (as opposed to redundant). Third, stemming from their interdependency, Measures 1 and 2 assess frame interplay computationally based on mono-modal data collection, whereas Measure 3 requires in-depth analysis of multimodal material after mono-modal data analyses identified relevant cases.

Table 1. Three Measures of Audiovisual Frame Interplay.

Measures	Skills	Analysis base	Analysis timing	Results
Measure 1: Ratio	Quantitative (SPSS)	Mono-modal data (verbal, visual)	After mono-modal data collection	Continuous: 1.00 to 0.00 range
Measure 2: Association rules learning	Quantitative (R)	Mono-modal data (verbal, visual)	After mono-modal data collection	Association rules and plausibility scores
Measure 3: In- depth analysis	Qualitative- quantitative mix	Mono-modal (verbal, visual) results followed by multimodal analysis	During the multimodal data collection, itself scheduled after the mono-modal data collection	Insights on frame congruence and mismatch

Fourth, the three measures vary in the results they produce. Measure 1 produces a ratio with values ranging from 1.00 (redundancy in frames across modalities) to 0.00 (mismatch between the frames), where values between these poles indicate partial overlap between verbal and visual frames (i.e., congruence). Measure 3, the in-depth analysis, yields data on the nature of frame congruence and mismatch, and the frequency to which each of these two forms of multimodal interplay occurred.

Measure 2 returns a list of association rules and three values indicating the plausibility of each: support, confidence, and lift. "Support" denotes the frequency with which a set of items occurred in the data set, "confidence" indicates the likelihood of two items appearing together, and "lift" is the probability that an item's inclusion will prompt the presence (rather than absence) of another item (i.e., the two items constituting an item set). Support and confidence can range from zero to 100, but there is no fixed range of values for lift. Higher values indicate greater plausibility of the rule at hand. Lift values above 1 denote an item set, whereas values below 1 suggest the opposite (i.e., that an item's presence likely prompts another item's absence). These values are then used to identify plausible rules and compare the prevalence of rules as indicators of verbal-visual frame redundancy, congruence, and mismatch.

For example, assume the following rule: "If journalists use a verbal victim frame, they also commonly use a visual invasion frame." Assume further that the rule produced the following results: support = 20; confidence = 0.82, 0.79 (bidirectional); and lift = 2. This means that the verbal victim frame and the visual invasion frame co-occurred in 20% of stories (support). Also, 82% of the time journalists used a verbal victim frame, they also used a visual invasion frame, and 79% of the time journalists used a visual invasion frame they featured a verbal victim frame (confidence, bidirectional). Finally, the verbal victim frame and the visual invasion frame are likely to occur as an item set (lift value surpasses the threshold of 1). Put differently, the lift value suggests that the presence of the verbal victim frame in a news story prompts the appearance of the visual invasion frame.

Because the three proposed measures vary considerably, the first research question is extended to examine their differences in execution and the findings produced:

RQ1: How do the three measures vary in their assessment of modality interplay?

Framing Refugees and Asylum Seekers (RAS) in Audiovisual News

To answer RQ1, the three measures were implemented on a sample of German television news stories about refugees and asylum seekers (RAS). News about RAS was selected for the application of our proposed measures for a number of reasons. With more than 1 million new asylum applications since 2016, Germany has become a prominent destination for RAS (BAMF, 2017). In response, fiery debates about humanitarianism, security, the national economy, and cultural (in)compatibility have emerged. This topic receives frequent news coverage, drives polarization in German society, and generates interest in the scholarly community. Frames typically emerge in news that draw controversy and large volumes of coverage, evident from the 11 frames that have been identified in RAS news coverage. Though there is variance in explicating and naming these frames, there is consistency in motif. Seven frames are negatively valenced and revolve around the high number of people crossing borders (invasion), demands on governance structures to manage this influx (burden), political conflict and debate (political transformation), and threats to the safety (security threat), well-being (health threat), prosperity (economic threat), and identity (cultural threat) of the population receiving RAS. More positive frames address human rights (humanitarian) and suffering (victim) of RAS, and invoke the benefits of cultural cross-fertilization (multiculturalism) and financial gains (economic contribution) that newcomers afford. Studies that have implemented these frames showed that German news on RAS and migrants tends to be negative (e.g., Sommer & Ruhrmann, 2010).

Anti-RAS Frames

When the invasion frame is employed in news, RAS are presented as amorphous masses heading to the destination country, willing to breach borders at any cost, and threatening to outnumber the local population (Greenberg & Hier, 2001; Thiele, 2005). The negative tone is detectable in references to RAS numbers as overwhelming and migration across borders is reported in war and disaster mode (e.g., "fortress," "breach," and "flood"). The invasion frame is invoked visually through faceless masses of people; illegal or violent entry (e.g., heat images or breaking through fences); military-style battle animations (e.g., routes taken by RAS); and graphs that depict soaring numbers of RAS in red (Thiele, 2005).

The burden frame presents RAS as obstructing otherwise smooth administrative processes or creating challenges for bureaucracy (Estrada, Ebert, & Lore, 2016). For instance, the judicial system may be described as overwhelmed and unusually slow or police stations imposing vacation bans because of staff shortage. Existing studies have explored only the verbal extension of this frame.

The political transformation frame points to RAS as the impetus for unwelcome changes to the political process, including unpleasant debates, political extremism, protests, strained diplomatic relations, and the implementation of law and policy that infringes on personal freedom (Estrada et al., 2016;

Figenschou & Thorbjørnsrud, 2015). Like the burden frame, existing studies have not offered a visual explication of the political transformation frame.

The security threat frame presents RAS as having a propensity for violence, inclined to commit crime or engage in terrorism (Greenberg & Hier, 2001; Thiele, 2005; Van Gorp, 2005). Various transgressive behaviors are collapsed into this frame, including lying to authorities, drug dealing, unprovoked aggression, sexual abuse, murder, and terrorism. Visual manifestations of this frame include graphics of crime statistics and RAS shown in police custody (e.g., handcuffed or behind bars), in terrorist training, and armed or using weapons (Thiele, 2005; Van Gorp, 2005).

The cultural threat frame presents RAS as observing values, norms, and customs that are incompatible with the destination community's way of life (Balabanova & Balch, 2010; Thiele, 2005; Van Gorp, 2005). Islam is often described as an undesirable influence, leading to an abatement of secularism or an abandonment of gender equality ideals. Visually, this frame is conveyed through emphasis on differences in dress, grooming, and habit (Thiele, 2005; Van Gorp, 2005). Examples include RAS wearing burkas, praying in gender-segregated mosques, harassing women on the street, or eating delicacies that would make the destination community recoil in distaste.

The economic threat frame presents RAS as financially draining destination countries (Balabanova & Balch, 2010; Estrada et al., 2016; Lawlor & Tolley, 2017; Thiele, 2005; Van Gorp, 2005). Paradoxically, RAS are sometimes described as unmotivated or unskilled and other times as competitors for jobs and other resources. The economic threat frame emerges in images of idle people lining up for social assistance or graphic depictions of economic downturn and soaring expenses (Thiele, 2005; Van Gorp, 2005).

The health threat frame constructs RAS as diseased and potentially contagious (Greenberg & Hier, 2001; Thiele, 2005). This includes suggestions that their arrival may reactivate diseases long eradicated in destination communities. Connected to the preservation of public health, there might be innuendo about avoiding physical contact with RAS. Visuals showing RAS handled by medical staff wearing gloves and surgical masks, or emphasis on physical examinations, treatment, or signs of illness align with this frame (Thiele, 2005).

Pro-RAS Frames

The humanitarian frame emerges from conceptualizations of human rights, by the Geneva Conventions, national constitutions and treaties, as well as societal norms and values. For instance, offering shelter, nutrition, and safety to RAS is presented as a matter of course, whereas volunteering and the work of NGOs are praised (Figenschou & Thorbjørnsrud, 2015; Lawlor & Tolley, 2017). Previous studies have not addressed how the humanitarian frame would emerge visually.

The victim frame presents RAS as suffering in their countries of origin, leading to their decision to flee. Victimization is also referenced in reportage on the adversity of the journey to the destination country, the border crossing, and the arrival in the country of destination (Figenschou & Thorbjørnsrud, 2015; Thiele, 2005; Van Gorp, 2005). This hardship may entail political persecution, violence, hostility, racism,

exploitation, and anxiety. When the victim frame is conveyed visually, it includes images of people who look exhausted, in physical pain, or in despair (Thiele, 2005; Van Gorp, 2005).

The multiculturalism frame presents RAS as a source of enrichment to the cultural life of the destination country (Balabanova & Balch, 2010; Thiele, 2005). Differences in values, norms, and religious views are constructed as valuable cultural additions and local populations welcoming these differences. Images of mixed-race groups enjoying each other's company and juxtapositions of RAS with symbols such as the Statue of Liberty have been argued to elicit this frame visually (see Johnson, 2003).

Finally, the economic contribution frame presents RAS as contributing to the national economy of the destination country—and by extension—to the welfare of its citizens (Balabanova & Balch, 2010; Thiele, 2005). This frame finds expression in journalistic accounts of RAS as either skilled or eager and able to receive training. The visual dimension of this frame has not been explicated in existing research.

Methodology

Sample

Television news remains the most used source of information in Germany and worldwide (Engel & Breunig, 2015; Pew Research Center, 2007), driving the decision to focus on this platform for this study. News stories were selected from the most popular newscasts in Germany aired on two public and two commercial channels. All newscasts that aired during four constructed weeks from September 2016 to September 2017 (112 newscasts in total) were considered for inclusion in the analysis. The sample was built by watching each newscast in full—first to identify stories that contained specified keywords and then to determine if these stories focused centrally on RAS. A coder was hired and trained for this two-stage selection procedure. One author double-coded approximately 20% of the newscasts ($n = 22$) to assess the reliability of story selection. Intercoder reliability was acceptable for keyword selection (Krippendorff's $\alpha = .95$) and story-focus selection ($\alpha = .96$). Of the 115 news items that were originally selected, 98 were coded as centrally focused on RAS and included in the analysis. Public TV news stories were slightly overrepresented: News items from ARD Tagesschau ($n = 30$) and ZDF heute-journal ($n = 29$) made up 60.2% of the sample, with commercial channels yielding fewer items: Sat.1 Nachrichten ($n = 22$) and RTL aktuell ($n = 17$).

Data Collection

Two sets of coders worked on collecting mono-modal data: One set coded the prevalence of frames in the verbal modality of stories, and the other team collected data from the visual modality. A subsequent third set of coders collected data from both modalities to produce the multimodal data set. Accordingly, three different codebooks were employed by three different sets of coders. Table 2 summarizes the differences between the three sets of coders, the correspondence between coders and codebooks, and the modality that was analyzed.


Table 2. Coders, Codebooks, and Coded Material.

Coders	Instrument focus	Coded material
Set 1: Mono-modal (verbal)	Codebook 1: Verbal frames	Audio track of all news stories
Set 2: Mono-modal (visual)	Codebook 2: Visual frames	Muted video track of all news stories
Set 3: Multimodal (audiovisual)	Codebook 3: Audiovisual frames	Full audiovisual versions of sample subset

Coding Instruments

Three codebooks were designed, one each for the audio and visual tracks and a third for multimodal coding of the full audiovisual versions of stories. The mono-modal codebooks contained variables that operationalized the 11 RAS frames of interest. The operational definitions were developed in close alignment with existing studies, as reviewed earlier.

Four visual frames, not coded before, were operationalized. The burden frame was defined by visuals suggesting excessive demands on bureaucracy, such as folders piling up on desks or graphs showing increases in processing times—followed/preceded by images of RAS. The political transformation frame was recorded for visual material of protests/riots or RAS followed/preceded by scenes suggesting conflict (e.g., graphics that depict conflict through fracture lines among headshots of politicians). The humanitarian frame was explicated as images of NGOs assisting RAS and of banners demanding that RAS be treated humanely. The economic contribution frame was signaled in scenes of RAS in a work setting or by graphs showing positive economic developments related to their work efforts. Figure 1 shows example images. Some visual frames were conceptualized as potentially cued by a single type of image, whereas others (i.e., burden, political transformation, and economic threat) require a combination of cues, as suggested by the brackets and the plus signs in Figure 1. Multiple cues work in concert through succession or juxtaposition of RAS images that invoke frames of excessive demands (burden), conflict (political transformation), or expense (economic threat).

Frame	Example images
Anti RAS-frames	
Invasion	
Burden*	
Political transformation*	
Security threat	

Cultural threat	
Economic threat	
Health threat	
Pro-RAS frames	
Humanitarian*	
Victim	

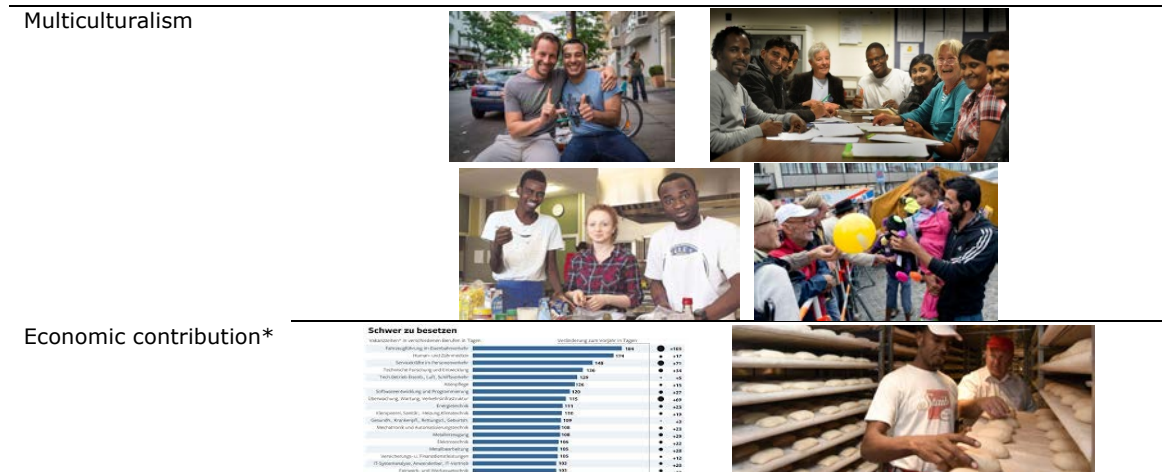


Figure 1. Images illustrating the 11 visual frames for refugees and asylum seekers (RAS).
 Note. The operational definitions of the visual frames marked with an asterisk were developed inductively because there are no known existing studies that employed visual measures for these frames.

Coders attended to a story at least three times for each round of data collection: first without pause, once with as many pauses as needed, and finally without pause. The individual news story served as the unit of analysis. Dichotomous measures were used throughout (yes = 1; no = 0), and frames were each coded holistically, as either present or absent. In addition to assessing if a frame appeared, coders documented whether a frame was countered. They used two fields to record the time when a frame appeared or was countered. Frame countering was operationalized as the appearance of a frame and information that stands in opposition to the frame, thereby undoing or undermining the frame. Coders were trained to identify both explicit and implicit instances of countering. Explicit countering included statements such as “This is utter nonsense,” or the placement of a question mark, an “X,” or a stop sign over an image. Subtle countering occurred through elaboration on wider contexts that undermines the frame. For example, a statement such as “Young male RAS engage in crime as much as German men in this age group” represents countering through contextual elaboration. Visually, this could come about through graphics that juxtapose a RAS-incriminating graph with one showing the prevalence of the same problem among home country nationals.

The third codebook set criteria for assessments of modality interplay rather than assessment of frame prevalence. Specifically, this instrument measured the extent to which verbal and visual modalities worked together or against each other in framing RAS. Coders collected data from stories that were identified for featuring at least one instance of unmatched frames across modalities (*n* = 47). Thus, stories with frame redundancy across modalities (*n* = 30) and stories that featured frames in only one modality (*n* = 21) were not included in this data collection round. Coders conducting the multimodal analysis were provided with the coding decisions of mono-modal coders. Specifically, they knew which verbal and visual frames were conveyed in each news story, whether they were countered or not, and when (time stamped) the frames and/or the countering occurred. They were instructed to watch the news stories multiple times, record if the different frames in the two modalities were congruent or mismatched, and provide justifications for each

coding decision in an open field. Coders were trained to recognize modality interplay—congruence or mismatch—based on the interrelationship between frames and intrarelationship of counterframing.

Congruence was defined by instances where (1) different yet compatible frames (e.g., humanitarian and victim; invasion and burden) were present in the two modalities; (2) different yet compatible frames were present in addition to redundant frame pairs (political transformation in both modalities); (3) a frame in one modality (e.g., victim) was supported by the countering of another frame (e.g., security threat) in the other modality. As noted above, the mismatch category was designed to uncover the audiovisual messages in which the two modalities conveyed contradictory interpretations. Mismatch was documented in stories that featured (1) incompatible frames such as security threat in the verbal and the humanitarian frame in the visual modality; (2) the same frame (e.g., security threat) in both modalities, but countered in one modality (e.g., a sex offense charge against an RAS is dismissed as Russian propaganda). Congruence and mismatch were not conceptualized as mutually exclusive categories. Rather, Measure 3 allowed for the possibility that verbal and visual modalities interacted in different ways within the same news story.

Coder Training and Reliability

Three sets of coders, each set consisting of two coders, were hired and trained separately to collect mono-modal (audio or visual) and multimodal data. One of the authors served as the reliability check on the multimodal coding. Training was conducted over the course of four sessions, lasting approximately five hours each. Reliability for mono-modal coding was assessed using Krippendorff's alpha. All reliability coefficients were above .80—specifically, invasion (verbal = .83, visual = .88), burden (verbal = .82, visual = .84), political transformation (verbal = .89, visual = .88), security threat (verbal = .80, visual = .85), cultural threat (verbal = 1, visual = .84), economic threat (verbal = 1, visual = .84), health threat (verbal = .82, visual = 1), humanitarian (verbal = .84, visual = .93), victim (verbal = .87, visual = 1), multiculturalism (verbal = 1, visual = .91), and economic contribution (verbal = 1, visual = 1). There was perfect agreement on the occurrence of countering in both modalities.

Multimodal data collection included consensus coding procedures. Material was coded independently, and justifications for coding decisions were recorded. During meetings, coding decisions were compared, and differences were discussed. Full agreement was reached on the congruence category, and coders conferred on four news items to make decisions about mismatched frames ($\alpha = .83$).

Findings

Descriptives on RAS Framing

Invasion, political transformation, victim, humanitarian, security threat, and burden frames were most prevalent across modalities. Together, these frames made up more than 85.25% of framing occurrences. Frames about cultural aspects, the economy, and health were rather seldom (see Table 3). Overall, few frame countering instances were recorded ($n = 31, 31.6\%$). Once articulated, frames are generally left undisputed.

Table 3. RAS Frames Frequencies.

Frame	Verbal	Visual	Total		Countering	
			N	%	N	%
Invasion	47	30	77	18.38	8	8.2
Political transformation	45	27	72	17.18	0	0
Humanitarian	36	19	55	13.13	4	4.1
Security threat	29	21	50	11.93	2	2
Victim	28	29	57	13.60	3	3
Burden	28	18	46	10.98	5	5.1
Cultural threat	12	5	17	4.06	4	4.1
Economic threat	9	2	11	2.62	1	1.0
Multiculturalism	8	8	16	3.82	3	3.1
Health threat	2	9	11	2.62	0	0
Economic contribution	3	4	7	1.67	1	1
Total	247	172	419	100.00	31	31.6

Thus, in line with previous research, this study revealed a predilection of framing RAS in negative ways (Balabanova & Balch, 2010; Van Gorp, 2005). This can be explained by the fact that news is by definition negative (Bednarek & Caple 2017). It seems equally plausible that journalists, in response to public criticism against left-wing bias and fake news, overcompensate by reporting on RAS in negative ways. It could also be that news sources providing negative sound and image bites are more successful at making themselves heard by journalists than are sources conveying positive information.

Audiovisual Frame Interplay

Research Question 1 prompted a comparison of the three proposed measures. Table 4 comparatively shows the frequency of the three types of frame interplay across the three measures.

Table 4. RAS Frames Frequencies.

Frame interplay	Measure		
	1: Ratio	2: Association rules learning	3: In-depth analysis
Redundancy	39%	34.23%	—
Congruence	58.44%	14.82%	26.15%
Mismatch	2.6%	11.32%	5.39%
Congruence and mismatch	—	—	38.27%

Measure 1

Measure 1 produced a frame congruence ratio of .70 ($SD = .28$), indicating that more than half of the sample featured frame congruence. Frame mismatch occurred infrequently, whereas a third of the sample was categorized as audiovisually redundant.

Measure 2

Measure 2 identified 14 frame pairs with rule status (support $\geq .10$, confidence $\geq .25$, and lift ≥ 1), thus revealing patterns in frame co-occurrences. The rules can be categorized into three groups. Redundancy rules refer to cases in which a verbal frame is associated with its visual counterpart and vice versa. In congruence rules, frames in one modality are paired with different but compatible frames in the other modality. Mismatch rules refer to verbal frames co-occurring with different yet incompatible visual frames, and vice versa. More than one-third of the rules returned by ARL pointed to frame redundancy. Congruence and mismatch rules were less than half as prevalent (see Table 5).

Table 5. RAS Frames Frequencies.

Verbal frames		Visual frames	Support	Confidence	Lift	Count
Redundancy rules						
Political transformation	↔	Political transformation	0.27	0.58, 0.96	2.10	26
Invasion	↔	Invasion	0.26	0.53, 0.83	1.74	25
Victim	↔	Victim	0.23	0.82, 0.79	2.78	23
Security threat	↔	Security threat	0.20	0.69, 0.95	3.22	20
Burden	↔	Burden	0.17	0.61, 0.94	3.31	17
Humanitarian	↔	Humanitarian	0.16	0.44, 0.84	2.29	16
Congruence rules						
Invasion	↔	Political transformation	0.15	0.32, 0.56	1.16	15
Humanitarian	↔	Victim	0.15	0.42, 0.52	1.41	15
Political transformation	↔	Invasion	0.14	0.31, 0.47	1.02	14
Victim	↔	Humanitarian	0.11	0.39, 0.58	2.03	11
Mismatch rules						
Humanitarian	↔	Invasion	0.12	0.33, 0.40	1.09	12
Victim	↔	Burden	0.10	0.36, 0.56	1.94	10
Humanitarian	↔	Burden	0.10	0.28, 0.56	1.51	10
Burden	↔	Victim	0.10	0.36, 0.34	1.21	10

Note. Only frames occurring in at least 10% of the sample were considered for the identification of rules. The table includes merely the most plausible rules (i.e., those yielding the following values: support $\geq .10$, confidence $\geq .25$, and lift ≥ 1). The rules encompass 224 of the 371 verbal and visual frame manifestations recorded for the 11 different frames—this represents 60.38%; the remaining frame pairs did not meet the threshold for plausibility to be granted rule status.

Patterns of frame co-occurrences are summarized in Table 5 and plotted in Figure 2, with a different color assigned to each frame. The shape in which the frame names are given indicates the modality: bubbles were used for verbal expression of frames, whereas squares indicate visual expression. The size of bubbles and squares in Figure 2 reflects the prominence of frames in the sample—the larger, the more common the frame. Lift and confidence values serve as indicators of rule plausibility: The higher the lift of a rule, the darker the arrow; the thicker the arrow, the stronger the confidence. Arrows are bidirectional because the confidence of a rule is assessed both ways: from Frame X to Frame Y and from Frame Y to Frame X.

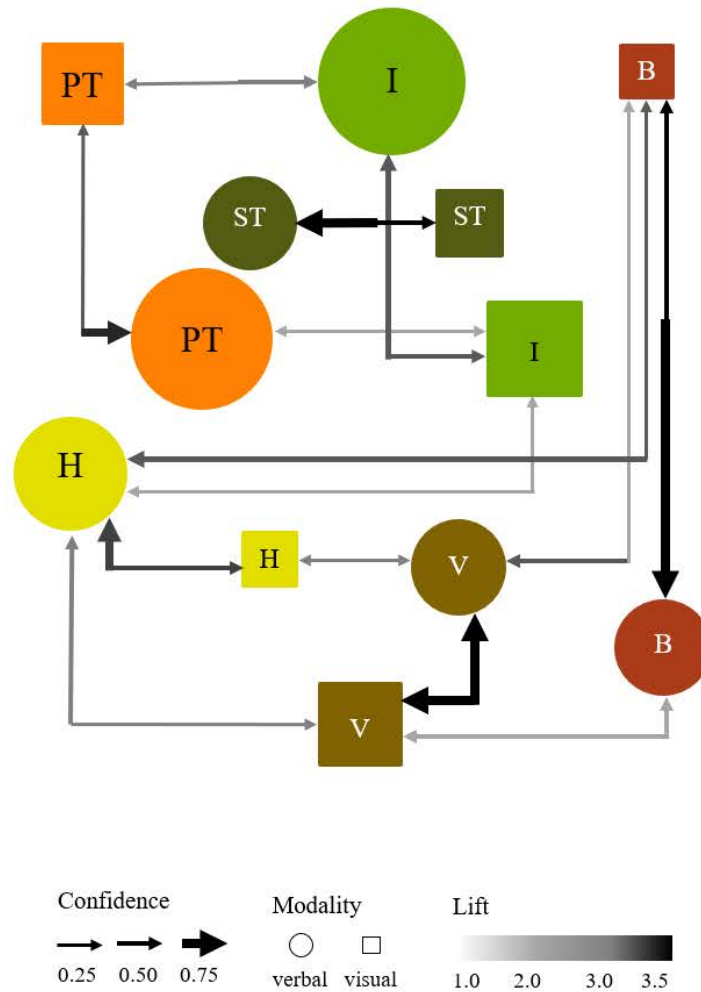


Figure 2. Joint distribution of verbal and visual frames.

Note. *I* = invasion; *B* = burden; *PT* = political transformation; *ST* = security threat; *H* = humanitarian; *V* = victim. Only frames occurring in at least 10% of the sample were considered for the identification of rules. The plot includes merely the most plausible rules (i.e., those yielding the following values: support $\geq .10$, confidence $\geq .25$, and lift ≥ 1).

It is reasonable to conclude that verbal occurrences of political transformation, invasion, victim, security threat, burden, and humanitarian frames were used together with their respective visual expressions in about 22% of the sample (redundancy rules). Moreover, we found indications of association between verbal and visual expressions of the frames' political transformation and invasion, and victim and humanitarian in about 15% of the sample (congruence rules). Finally, there was evidence of mismatch rules in about 10% of the sample. First, the visual expression of the victim frame co-occurred with the verbal expression of the burden frame. Second, the verbal humanitarian frame co-occurred with the visual invasion and burden frames.

Measure 3

Measure 3 involved the use of in-depth analysis of stories identified through Measure 1 as containing verbal and visual frames that were nonredundant ($n = 47$). The multimodal analysis revealed that more than two-thirds of stories contained frame congruence and mismatch over the course of a news story. About 5% of stories featured frame mismatch alone, whereas congruence alone was encountered in little more than a quarter of stories.

The two measures that assessed verbal–visual frame redundancy similarly revealed that this type of modality interplay occurred in about one-third of stories (Measures 1 and 2; see Table 3). Furthermore, all three measures consistently showed that frame mismatch was the least common type of modality interplay (see Table 3). Thus, all three measures are well positioned to assess these two ends of the spectrum. There were also large differences in results. Values for frame congruence ranged from about 15% (Measure 2), to little more than a quarter (Measure 3), and more than half of the news stories (Measure 1; see Table 3). Measure 3 revealed that only about a quarter of news stories were congruent because congruence and mismatch categories were not treated as mutually exclusive.

In light of our research question, a few differences between the three measures should be noted. First, the measures vary considerably in the level of effort associated with their use, with Measure 1 being the most straightforward, Measure 3 being the most complex, and Measure 2 striking middle ground (see Table 6). Measure 1 simply required the application of a formula in SPSS, whereas more sophisticated statistical procedures were required for Measure 2. In contrast, Measure 3 was the most time-consuming, necessitating greater coding effort for multimodal material. Therefore, concerning the effort required to code data, Measures 1 and 2 may be best suited for larger samples—with Measure 2 potentially requiring hundreds of instances for meaningful analysis, and Measure 3 for smaller samples. In trade-off, Measure 3 offered the most fine-grain assessment of interplay (see Table 6), whereas Measure 1 provided a comparatively crude indication of frame interplay, used optimally to produce a continuous variable for subsequent analyses (e.g., variations in modality interplay over time or across journalistic culture, medium, news topic). Measure 1 produced no information on frame pairs and did not account for the possibility that some heterogeneous frame pairs might be compatible (e.g., verbal burden frame and visual security threat frame), which could produce artificially low ratios. However, including if–then conditions to distinguish between compatible and incompatible frame pairs could enhance the precision of this measure. The advantage of Measure 2 is the identification of verbal and visual frame pairs, while factoring in the frequency with which such frame pairs occur in the data set and the likelihood of being

encountered in another sample of audiovisual messages on the same topic. Yet using association rules to reveal modality interplay merely shows co-occurrence, falling short of revealing if modalities worked together or against each other in audiovisual messages. In this regard, Measure 3 is best positioned to unveil the complexity of modality interplay in audiovisual messages because it allows assessment within the layers of a message.

Table 6. A Comparative Assessment of the Three Measures.

	Level of effort	Sample size	Granularity of insight
Measure 1: Ratio	Low	Large	Low
Measure 2: Association rules learning	Moderate	Large	Moderate
Measure 3: In-depth analysis	High	Small	High

Regarding these findings, we recommend that scholars interested in assessing modality interplay employ a combination of Measures 2 and 3. On applying Measure 2, we recommend reporting the frequency to which redundancy, congruence, and mismatch rules occurred next to the specific frame pairs for the congruence and mismatch categories and submit those to in-depth analysis through Measure 3. This procedure should reduce unnecessary effort and gain nuance in results.

Conclusion

The impunity of verbal-only studies is gradually collapsing under the growing evidence of visuals as conduits of meaning. Part of this movement is driven by conceptual and methodological curiosity—a line of work to which the present study belongs. We aimed to design and test three measures that could assess modality interplay in audiovisual media content (redundancy, congruence, and mismatch). In audiovisual messages, information and interpretative cues are temporally scattered across modalities. Accordingly, we argued that measures accounting for this dispersion of meaning deliver more nuanced data. Thus, we proposed focusing on verbal and visual frame pairs instead of semantic content combinations, arguing that measures informed by framing theory could breathe new life into audiovisual research by accommodating both modalities and their interplay (see Coleman, 2010).

Furthermore, our analysis suggested that the say-dog-see-dog approach to storytelling was quite common in German news coverage of RAS, and at the same time frames were rarely mismatched. This finding was expected and in line with previous studies that assessed microconcentrated semantic verbal-visual relationships (Graber, 1990; Walma van der Molen, 2001). This finding suggests that news messages are optimized for audience comprehension and memory. A finding of high frequency in mismatched frames would have signaled the potential for cognitive overload at the news user end—with the visual modality dominating as the driver in conveying information. That was not the case in our sample of RAS reporting. In fact, German television news about RAS is highly congruent and overwhelmingly negative in framing this highly charged contemporary social issue.

Taken together, these results suggest that communicators construct meaning by varying the degree of correspondence between verbal and visual frames. These findings also set the stage for future experimental work into how various frame pairs affect attention, comprehension, and memory for news (for a recent study along these lines, see, e.g., Powell et al., 2019).

References

- Agrawal, R., Imieli, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *SIGMOD Record*, 22(2), 207–216.
- Balabanova, E., & Balch, A. (2010). Sending and receiving: The ethical framing of intra-EU migration in the European press. *European Journal of Communication*, 25(4), 382–397.
- BAMF. (2017). *Aktuelle Zahlen zu Asyl* [Current numbers on asylum]. Retrieved from www.bamf.de/DE/Themen/Statistik/Asylzahlen/AktuelleZahlen/aktuellezahlen-node.html
- Bednarek, M., & Caple, H. (2017). *The discourse of news values*. Oxford, UK: Oxford University Press.
- Coleman, R. (2010). Framing the pictures in our heads: Exploring the framing and agenda-setting effects of visual images. In P. D'Angelo & J. A. Kuypers (Eds.), *Doing news framing analysis* (pp. 233–261). New York, NY: Routledge.
- Crigler, A. N., Just, M. R., & Neuman, R. W. (1994). Interpreting visual versus audio messages in television news. *Journal of Communication*, 44(4), 132–149.
- Dan, V. (2018). A methodological approach for integrative framing analysis of television news. In P. D'Angelo (Ed.), *Doing news framing analysis II* (pp. 191–220). New York, NY: Routledge.
- D'Angelo, P., Lule, J., Neuman, W. R., Rodriguez, L., Dimitrova, D. V., & Carragee, K. M. (2019). Beyond framing: A forum for framing researchers. *Journalism & Mass Communication Quarterly*, 96(1), 12–30.
- De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, 7(10), 460–467.
- Engel, B., & Breunig, C. (2015). Ergebnisse der ARD/ZDF-Langzeitstudie [Results of the ARD/ZDF longitudinal study]. *Media Perspektiven*, 7/8, 310–322.
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58.

- Estrada, E. P., Ebert, K., & Lore, M. H. (2016). Apathy and antipathy: Media coverage of restrictive immigration legislation and the maintenance of symbolic boundaries. *Sociological Forum, 31*(3), 555–576.
- Figenschou, T. U., & Thorbjørnsrud, K. (2015). Faces of an invisible population: Human interest framing of irregular immigration news in the United States, France, and Norway. *American Behavioral Scientist, 59*(7), 783–801.
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. Y. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience, 7*, 773.
- Grabe, M. E., & Bucy, E. P. (2009). *Image bite politics: News and the visual framing of elections*. Oxford, UK: Oxford University Press.
- Graber, D. A. (1990). Seeing is remembering: How visuals contribute to learning from television news. *Journal of Communication, 40*(3), 134–155.
- Graber, D. A. (2001). *Processing politics: Learning from television in the Internet age*. Chicago, IL: University of Chicago Press.
- Greenberg, J., & Hier, S. (2001). Crisis, mobilization and collective problematization: “Illegal” Chinese migrants and the Canadian news media. *Journalism Studies, 2*(4), 563–583.
- Grimes, T. (1991). Mild auditory-visual dissonance in television news may exceed viewer attentional capacity. *Human Communication Research, 18*(2), 268–298.
- Hsia, H. (1968). Output, error, equivocation and recalled information in auditory, visual and audiovisual information processing with constraint and noise. *Journal of Communication, 18*(4), 325–353.
- Johnson, M. (2003, May). *Immigrant images: U.S. network news coverage of Mexican immigration, 1971–2000*. Paper presented at the annual meeting of the International Communication Association (ICA), San Diego, CA.
- Kotsiantis, S., & Kanellopoulos, D. (2006). Association rules mining: A recent overview. *GESTS, 32*(1), 71–82.
- Lang, A. (1995). Defining audio/video redundancy from a limited-capacity information processing perspective. *Communication Research, 22*(1), 86–115.
- Lawlor, A., & Tolley, E. (2017). Deciding who’s legitimate: News media framing of immigrants and refugees. *International Journal of Communication, 11*(25), 967–991.
- Matthes, J. (2009). What’s in a frame? A content analysis of media framing studies in the world’s leading communication journals, 1990–2005. *Journalism & Mass Communication Quarterly, 86*(2), 349–367.

- Messariss, P., & Abraham, L. (2001). The role of images in framing news stories. In S. D. Reese, O. H. Gandy, & A. E. Grant (Eds.), *Framing public life: Perspectives on media and our understanding of the social world* (pp. 215–226). Mahwah, NJ: Erlbaum.
- Migration “mother of all political problems,” says German Interior Minister Horst Seehofer. (2018). *Deutsche Welle*. Retrieved from www.dw.com/en/migration-mother-of-all-political-problems-says-german-interior-minister-horst-seehofer/a-45378092
- Pew Research Center. (2007, October 4). *Where people get their news*. Retrieved from <http://www.pewglobal.org/2007/10/04/chapter-7-where-people-get-their-news/>
- Powell, T. E., Boomgaarden, H. G., De Swert, K., & de Vreese, C. H. (2019). Framing fast and slow: A dual processing account of multimodal framing effects. *Media Psychology, 2*(4), 572–600.
- Prior, M. (2014). Visual political knowledge: A different road to competence. *Journal of Politics, 76*(1), 41–57.
- Reese, S. D. (2007). The framing project: A bridging model for media research revisited. *Journal of Communication, 57*(1), 148–154.
- Shook, F. (1994). *Television newswriting: Captivating an audience*. New York, NY: Longman.
- Silcock, B. W. (2007). Every edit tells a story: Sound and framing routines of videotape editors in global news cultures. *Visual Communication Quarterly, 14*(1), 3–15.
- Sommer, D., & Ruhrmann, G. (2010). Oughts and ideals: Framing people with migration background in TV news. *Conflict & Communication Online, 9*(2), 1–15.
- Son, J., Reese, S. D., & Davie, W. R. (1987). Effects of visual-verbal redundancy and recaps on the TV news learning. *Journal of Broadcasting & Electronic Media, 31*, 207–216.
- Spence, C. (2014). Orienting attention: A crossmodal perspective. In A. C. Nobre & S. Kastner (Eds.), *The Oxford handbook of attention* (pp. 1–21, online source). Oxford, UK: Oxford University Press. doi:10.1093/oxfordhb/9780199675111.013.015
- Stewart, P., & Alexander, R. (2016). *Broadcast journalism: Techniques of radio and television news*. New York, NY: Routledge.
- Thiele, M. (2005). *Flucht, Asyl und Einwanderung im Fernsehen* [Flight, asylum and immigration on TV]. Constance, Germany: UVK.
- Van Gorp, B. (2005). Where is the frame? Victims and intruders in the Belgian press coverage of the asylum issue. *European Journal of Communication, 20*(4), 484–507.

Walma van der Molen, J. H. (2001). Assessing text-picture correspondence in television news: The development of a new coding scheme. *Journal of Broadcasting & Electronic Media*, 45(3), 483–498.

Walma van der Molen, J. H., & Klijn, M. E. (2004). Recall of television versus print news: Retesting the semantic overlap hypothesis. *Journal of Broadcasting & Electronic Media*, 48(1), 89–107.

Zhou, S. (2004). Effects of visual intensity and audiovisual redundancy in bad news. *Media Psychology*, 6(3), 237–256.