Anna Wilson*

It's Time to Do News Again

https://doi.org/10.1515/zaa-2020-2016

Abstract: The study analyses the RT USA show "The News with Rick Sanchez" to identify persuasive and manipulative communication strategies realised by the host's integration of speech and co-speech gestures.

What cognitive systems, conceptual processes and schemas does the speaker employ to construct and communicate viewpoint? What linguistic and gestural units enable certain viewpoint construction? What role does social context play in the interpretation of these units?

After a quantitative overview of the use of body-directed and related outward-directed gestures in 180 minutes of RT recordings, the study offers a fine-grained qualitative analysis of viewpoint construction behind four situations from these RT data. It develops an advanced cognitive-linguistic approach anchored in conceptual blending (Fauconnier and Turner 2002) and the cognitive system of force dynamics (Talmy 1988, 2000). The study argues that this approach is successful in revealing speech-gesture integrations as triggering viewpoint blending for the purpose of manipulation and is useful for transferring manual qualitative analysis to analysis at scale.

Keywords: multimodal communication, media, RT USA show host, co-speech gesture, body-directed gesture, viewpoint blending, force dynamics, multimodal viewpoint blending triggers

1 Introduction

RT (formerly Russia Today) is widely viewed as communicating pro-Russian state viewpoints to international TV audiences while trying to maintain the appearance of a channel of information independent of any influence. To achieve that, RT show hosts employ various multimodal strategies to persuade and manipulate. They work to make the communication process engaging, and to communicate information memorable to the viewer and easy to take in. They also rely on the integration of speech and co-speech hand gestures as one cognitive and communicative process.

***Corresponding author: Anna Wilson**, Oxford School of Global and Area Studies, University of Oxford, Oxford, UK, E-mail: anna.wilson@area.ox.ac.uk

This study explores this cognitive and communicative process to engage with some of the central questions posed by scholars working on viewpoint analysis in multimodal communication: What motivates the speaker to engage in viewpoint construction in a certain way, to shift or to displace viewpoint, to integrate and communicate two or more viewpoints simultaneously? How does the speaker's motivation for viewpoint construction relate to social context? What cognitive systems, conceptual processes and schemas does the speaker employ to construct and communicate viewpoint? What linguistic and gestural units enable certain viewpoint construction? What role does social context play in the interpretation of these units?

The study analyses the RT USA show "The News with Rick Sanchez" to identify persuasive and manipulative communication strategies realised by the host's integration of speech and body-directed gestures (BDG). The study sees these speech–BDG integrations as multimodal units that trigger and support viewpoint blending processes underlying the host's application of the multimodal strategies. The study thus finds it useful to introduce and apply the notions of the RT host as a viewpoint blend as well as of *multimodal viewpoint blending triggers* and *cues* (from now on: 'triggers' and 'cues'). The RT host is defined as a blend in the sense that he routinely embodies conceptual blending of multiple selves – his own self, others' real and imagined selves, and entities as selves. Others' real and imagined selves may include the selves of the RT TV team, the show's guests, and the audience's/viewers' selves. The entities' selves may be those of countries (Russia, other countries) or groups/categories of people or companies, times, locations, or products that the host associates himself with. He blends all these selves to communicate pro-Russia state viewpoints. Multimodal viewpoint blending triggers and cues are defined as multimodal units – speech–gesture units in this study – which emerging as blends themselves, in turn trigger the construction of new blends in the mental viewpoint web of unfolding communication. Media communication carries multiple viewpoints and viewpoint dimensions. These viewpoints are often constructed through more than one modality (Chilton 2014; Dancygier and Sweetser 2012; Dancygier and Vandelanotte 2017b; Dancygier et al. 2016; Vandelanotte 2017; Vandelanotte and Dancygier 2017). The proposed notions allow us to focus on three dimensions of viewpoint construction in their interrelation: (1) the motivation behind viewpoint construction and its relation to social context; (2) specific local stages of triggering and supporting viewpoint construction; (3) the relation of local viewpoint to the global viewpoint blending construction underlying an RT show as a whole or its topic segments (news reports on certain topics within the show). Considered together in a more systematic way, these three dimensions of viewpoint blending help us to appreciate how the construction of viewpoint at all local levels amounts to viewpoint at a global level,

and how the host as a blend orchestrates this process of accumulating and structuring a global viewpoint. Furthermore, the introduction of the notions of multimodal 'triggers' and 'cues' takes us a step further in pursuit of enabling the large-scale analysis of viewpoint blending. It does so through researching and identifying linguistic-gestural units more suitable for coding in working with large messy media multimodal datasets in a semi-automatic way.

In analysing 114 instances of integrations between speech and accompanying body-directed gestures made by the show's host, the study identifies three main and inter-linked strategies of multimodal communication: establishing authority; maintaining a neutral tone; and dealing with the alternative voice. The study offers the analysis of viewpoint blending networks and the corresponding multimodal triggers and cues which enable the host's application of these strategies, while revealing linguistic-gestural-conceptual viewpoint configurations typical of the RT show under consideration. The study argues that conceptual blending provides useful tools for unpacking these viewpoint configurations. Analytical tools offered by the blending framework (Fauconnier and Turner 2002) allow us to examine speech–gesture units as triggers for the construction of meaning without overlooking their integral nature. Viewpoint construction is inherent to meaning construction. The study's analysis demonstrates how force-dynamics (Talmy 1988, 2000) helps us to reveal viewpoint construction in multimodal blending behind the identified communication strategies.

In the remaining sections, this paper presents the theoretical background and methodological approach to the investigation of speech–BDG integrations; the data; the method used for annotation; analysis of the data; and a concluding discussion of the findings of the study, suggesting future directions for research.

2 Theoretical Background and Methodological Approach

2.1 Co-Speech Gesture in Communication

The part co-speech gesture plays in persuasion and manipulation in media communication is not to be underestimated. Co-speech gesture has been argued to be *verbal* in the sense that it is inseparable from speech (Kendon 2004; McNeill 1992). Crucially, there is evidence that the relation between speech and accompanying gesture is a two-way road. Language and thought influence co-speech gesture, and co-speech gesture influences speech and thought (Kita and Özyürek 2003; Kita et al. 2017). As they interact and integrate with speech, gestures communicate information

to the viewer and can do that very effectively (Hostetter 2011; Kendon 1994). Co-speech gestures benefit the viewer when dealing with more complex communicative tasks.[1] Gestures have been found to boost memory of events communicated to the viewer linguistically (Aussems and Kita 2019). There is evidence that gestures are produced more frequently when the speaker is highly keen to communicate information to the viewer (Kelly et al. 2011). This can be linked to the speaker's intention to make the viewer's comprehension task easier, especially when the information communicated is conceptually complex. The viewer benefits from the speaker's increased gesticulation, as gestures lower the high (complex) conceptual load on the viewer through the load's distribution between modalities (Melinger and Kita 2007). Co-speech gestures are found to play a key role in linguistic metaphor interpretation (Lewis and Stickles 2017), with metaphor seen by many cognitive linguists a complex conceptual process enabling persuasion and manipulation in media.[2] Beattie and Shovelton's (2005) study of such persuasive and manipulative data as TV advertisements demonstrates that a combination of speech and gesture makes semantic communication better, compared to speech alone. Co-speech gestures also play a large part in maintaining interaction (Bavelas et al. 1992) and managing and controlling it (Jacquin 2017; Wehling 2017).

2.2 Co-Speech Body-Directed Gesture and Viewpoint

Co-speech body-directed gestures (BDGs) constitute a phenomenon much richer than just pointing. Despite a valuable contribution made by Cooperrider (2014), this area remains largely understudied.[3] Cooperrider conducted an empirical analysis of 40 interviews hosted by the American TV show *The Tavis Smiley Show*, and provides BDG classification and a discussion of the/possible/certain conceptual processes underlying the use of BDGs. His study focuses on producing an all-encompassing typology based on a description of speech–BDGs that seems to be driven by formal rather than functional characteristics. It provides some insights into conceptual processes behind co-occurrences of BDG with discourse units, including observing their relation to viewpoint. For example, Cooperrider relates the use of BDGs to radical shifts in viewpoint – from pointing to Self as Self to pointing to Self as some

1 E.g. under noise, see Rogers, William T (1978). "The contribution of kinesic illustrators toward the comprehension of verbal behavior within utterances." *Human communication research* 5.1, 54–62.

2 See overview in Pleshakova, Anna (2018). "Cognitive approaches: media, mind, and culture." Colleen Cotter & Daniel Perrin eds. *The Routledge handbook of language and media*, 77–93. London and New York: Routledge.

3 For an overview of previous research in relation to BDG see Cooperrider 2014.

Other's Self – and explains them through the wide-spread phenomenon of body blending. That is when the speaker conceptualizes her own body as both the speaking-gesturing body and the body of another referent (Cooperrider 2014, 13). This phenomenon observed by Cooperrider for BDGs relates to a bigger phenomenon of viewpoint multiplicity, multidimensionality and creativity in speech and co-speech gesture explored by scholars working on gesture and viewpoint.[4] One dimension of such creativity is gesture's ability simultaneously to account for various perspectives (Taylor and Tversky 1996), character and observer viewpoints (McNeil 1992; Parrill 2012; Turner et al. 2019), immersed and non-immersed experiential viewpoints (Mittelberg 2017), dual viewpoints (Parrill 2009), and to deviate from speech in the viewpoint communicated (Guilbeault 2017).

The present study differs from Cooperrider (2014) in that it goes beyond the focus on the description of BDGs and their co-occurrences with speech units per se. It makes viewpoint blending rooted in BDG its main concern, and explores BDGs as triggering and supporting viewpoint blending in persuasive and manipulative communication.

2.3 Viewpoint and Viewpoint Dimensions, Conceptual Blending, and Schematic Systems

Conceptual blending presupposes that viewpoint resides in one or more mental spaces in the conceptual blending network, with the viewpoint space responsible for structuring or setting up of other spaces (Fauconnier 1997; Fauconnier and Turner 2002; Parrill 2012). The current study defines viewpoint as a phenomenon of human cognition which is inherent in speech and gesture, and thus inseparable from the emergence of meaning in conceptual blends. The conceptual processes upon which blending and integration rely, from activating elements of long-term knowledge structures in input mental spaces to the outer space mappings between these elements, and to selective projection and compression of vital relations in a blend, all, in this study's view, presuppose and support the emergence of meaning infused with certain viewpoint(s) in the blend. The conceptual integration networks can have a viewpoint space as an input mental space that viewpoint construction in the blend uses while running the network as a whole. This understanding relies on the multiplicity of voices and their dialogic inter-relation described by Bakhtin (1984, 1981).

4 See e.g. the overview in Mittelberg, Irene (2017). "Experiencing and construing spatial artifacts from within: Simulated artifact immersion as a multimodal viewpoint strategy." *Cognitive Linguistics* 28.3, 381.

The construction of blending networks can be triggered by linguistic and gestural cues (among others in various cognitive systems).[5] Both types of cues working interactively can activate elements of frames (long-term knowledge structures) of different kinds – thematic and structural – in input mental spaces, which then get mapped onto each other. In doing so linguistic and gestural cues influence the viewpoint arrangement as emerging in the blend. Taken together as instruments available to the conceptualizer, such cues are labeled as linguistic viewpoint and gestural viewpoint. These manifest all different ways in which content may be construed either linguistically or via gesticulation (Sweetser 2012). Such content is often constructed inter-subjectively (Mittelberg 2018; Sweetser 2012). While considering viewpoint in its relation to communicative behavior, Parrill (2012, 98) links linguistic and gestural viewpoints to the location of the conceptualizer and introduces the following concepts as dimensions of viewpoint construction: (1) conceptual – the real or imagined location of the conceptualizer; (2) linguistic – the linguistic manifestations of that location; (3) gestural – spatial cues to that location as seen in gesture. In its analysis the current study routinely considers all three dimensions. It sees speech–BDG interrelation and integration as the conceptual integration of two or more viewpoints. Speech–gesture constitutes a viewpoint blend in its own right, which becomes a multimodal trigger and cue in further viewpoint blending construction as communication unfolds.[6] The current study is interested in how spatial and temporal (physical) viewpoint cognitive viewpoint — temporal, spatial, epistemic, ideological (Dancygier 2011; Guilbeault 2017; Lou 2017). It is interested in how gestures working together with language add to information (Enfield 2009; Enfield et al. 2007), and package and schematise it (Alibali et al. 2000; Kita 2000; Kita et al. 2017), ultimately view-pointing this information to pursue various rhetorical goals. To this end, the study employs the concept of image schemas (Johnson 2008) and schematic systems (Ibarretxe-Antuñano 2006; Talmy 1988, 2000). These are seen as means for conceptual scaffolding serving and enabling complex conceptual blending processes in viewpoint construction (Dancygier and Vandelanotte 2017a).

5 See Mark Turner's (this volume) discussion of non-linguistic forms as 'suggesting' meaning construction via blending.

6 See Turner, M. (2014). *The Origin of Ideas: Blending, Creativity, and the Human Spark*. OUP USA. for discussion of complex mental webs routinely created by news show hosts through blending.

3 Data and Annotation Method

The study analyses seven video recordings for the RT (US) show "The News with Rick Sanchez" (NRS), totalling approximately 180 min, which aired between 5th August and 1st September 2020. The host of the RT show is Rick Sanchez (RS), and each show is *c*. 25–30 min long. The shows are available via RT TV, the RT websites and the RT YouTube channels. Video recordings of the shows were annotated using the ELAN 5.9 video annotation software (https://archive.mpi.nl/tla/elan). To ensure consistency and minimize annotation errors, each show was annotated by two coders.[7]

Table 1: RT Shows Analysed for This Study.

N	Broadcast	Date and time	URL
1	Fast cash online causing US debt explosion — borrower beware!	September 1, 2020 20:06	https://www.rt.com/shows/news-with-rick-sanchez/499641-fast-cash-online-causing-us/
2	'COVID-19' war imminent as China leads while US lags?	August 31, 2020 21:50	https://www.rt.com/shows/news-with-rick-sanchez/499559-covid-19-war-imminent-as/
3	Pompeo angry with Russia & EU over UN rebuke (full show)	August 21, 2020 19:32	https://www.rt.com/shows/news-with-rick-sanchez/498702-pompeo-angry-with-russia-eu/
4	New: Trump, Pompeo apply full-court press for US control of Iraq!	August 20, 2020 20:21	https://www.rt.com/shows/news-with-rick-sanchez/498594-new-trump-pompeo-apply-full/
5	Google 'illegally' sharing info with police, hack reveals	August 18, 2020 22:52	https://www.rt.com/shows/news-with-rick-sanchez/498369-google-illegally-sharing-info-with/
6	Dire warning: US-China war 'within 3 months'	August 6, 2020 19:29	https://www.rt.com/shows/news-with-rick-sanchez/497307-dire-warning-us-china-war/
7	Beirut explosion: missile strike, negligence, or sabotage	August 5, 2020 19:56	https://www.rt.com/shows/news-with-rick-sanchez/497210-beirut-explosion-missile-strike-negligence/

7 All annotations were verified by the author of this article. Agreement was 72%, and clashes in annotation were discussed and resolved.

3.1 Annotation

At the first stage of analysis all video recordings of NRS shows were annotated for:

- Identification of BDGs made by the RT show host, Rick Sanchez, as visually representing speech–gesture integrations central to communicative moves (in Enfield 2009: composite utterances as designed for anticipated interpreters). Time intervals for annotation were selected in accordance with interpretation needs for a speech–BDG integration under consideration. The main criterion here was linked to perception of a speech–BDG as a meaningful communicable unit (Guilbeault 2017), sufficient for its interpretation from the perspective of epistemic or ideological (Self–Other) viewpoint.
- Type of gesture.
  The typology offered in Cooperrider (2014) was used: (1) self-point, when "a speaker points to some part of the body to evoke the entire body"; (2) body-point, directed at a particular part or region of the body to narrowly evoke only that part; (3) body-anchor, which "involves indexing experiential notions—whether mental, emotional, or physiological—to a certain part or region of the body" (Cooperrider 2014, 5, 8, 9).
- Gesture morphology description. BDGs were annotated using the following characteristics: the spatial gesture chain of which the BDG is a part; gesture trajectory; part of body to which the BDG is directed; handedness; hand shape (fist, closed palm, open palm, bunch, pinch, thumb position, finger positions and direction, and index finger position); hand position in relation to the body; hand closeness to the body; BDG amplitude; number of BDG repetitions.
- Class of BDG gesture, recorded using the classification by Chu et al. (2014).
  Chu et al. (2014) distinguish between representational, conduit, and palm-revealing gestures, which may serve different functions and may be related to different social and cognitive skills. Chu et al. (2014, 2) define representational gestures as depicting "a concrete or abstract concept with the shape or motion of the hands (iconic gestures and metaphoric gestures in McNeill 1992), or point to a referent in the physical or imaginary space (concrete or abstract deictic gestures in McNeill 1992)". Conduit gestures are used to present the speaker's idea to the interlocutor/the viewer – "the palm of the hand faces upward and moves toward the listener as if to present a clearly formulated idea on the palm to the listener (conduit metaphor gesture in McNeill 1992; a type of interactive gesture in Bavelas et al. 1992)". Conduit gestures combine representational and interactive functions, by simultaneously presenting the speaker's idea or concept to the viewer and bringing the interlocutor into the

conversation. A palm-revealing gesture is seen a purely interactive gesture, but being an outward-directed gesture, it was not coded for this study's analysis.
– Speech transcripts were produced, with the word(s) on which BDG falls capitalised.

At the second stage, the study continued working with the Elan files created at the first stage of annotation and engaged in conceptual blending analysis of identified time intervals, while recording the results in the following Elan tiers:
– Input mental spaces and vital relations employed by the blend under consideration.The following outer and inner relations were considered where appropriate: change, identity, time, space, cause-effect, part-whole; representation, role, analogy, disanalogy, property, similarity, category, intentionality, uniqueness (Fauconnier and Turner 2002).
– Conceptual processes and schemas at work.
The study considered whether speech–BDG integration was rooted in or constituted a key part of: metaphor; metonymy; counterfactuals; irony; expository question (as described by Xiang and Pascual 2016); parody/dramatization/theatricality/re-enactment/epistemic stance; force-dynamics (Mittelberg 2018; Oakley 2005; Talmy 2000).
– Self–Other relation.
The study recorded whether the BDGs were used to signify Self's Self or Other's Self and what was the conceptual relation to the Other in each instance;
– Viewpoint strategy.
The study recorded viewpoint strategy in relation to the viewpoint constructed earlier in communication (whether by the host or his interlocutor) and activated by the host for the construction of the local viewpoint during the time interval under consideration as a mental input. The study considered such viewpoint strategies as: set-up, support, extension, reinforcement, summarising, shift, displacement, dismissal, buffering.
– Inter-modality relation (IMR), i.e. the interaction between speech and gesture in providing semantic information.
The study used the typology offered in Caldognetto and Poggi 26–27 September 1997, and differentiated between IMR of: repetition (both modalities convey the same meaning); addition (one or both modalities add complementary congruent information); substitution (giving information not provided in the other modality; contradiction (providing contrasting information); independent (with signals taking part in different communicative plans).

At the third stage of analysis, all resulting annotations were downloaded from Elan as tab-delimited files and transformed into Excel spreadsheets. All annotations were checked again for possible errors and to ensure that labels/results were presented in a unified manner. The resulting list contained 114 instances of speech–BDG integrations made by the host, for which qualitative analysis aiming to identify multimodal communication strategies, and some quantitative analysis was performed.

4 Analysis and Results

4.1 Gesture Formulas

The annotation process and analysis quickly made it evident that for the purposes of viewpoint blending analysis in persuasive and manipulative communication, BDG cannot be considered without simultaneously considering the outward-directed gestures (ODG) closely linked to them. Therefore the study annotated for gesture configurations that incorporated BDG alongside outward-directed gestures, and identified a list of the gesture formulas. These gesture formulae co-occur with discourse units of varied length. The analysis here considers those gesture formulae or chains which co-occur with speech units of sufficient length to act as multimodal triggers or cues for persuasive and manipulative viewpoint construction.

Table 2 summarises the types and frequencies of the body-directed gestures and related outward-directed gestures made by the talk show's host. BGD stands for body-directed gestures, and ODG stands for outward-directed gestures. BDG.E stands for a BDG of small amplitude as compared to the ODG it is closely linked to. An underscore '_' marks a tight relation between BDG and ODG signifying that it is not possible to interpret the gestural viewpoint here from the perspective of viewpoint blending in unfolding discourse and communication without considering BDG_ODG as one functional unit. The relation marked with 'and' signifies a less tight relation and a possibility for the interpretation of a BDG in its own right, but at the same time signalling that it is only in its relation to an ODG that the BDG can be interpreted as part of a speech–gesture trigger for further viewpoint blending in discourse and communication. A hyphen '-', as in BDG–ODG, marks the instances when we observe two or more BDGs appearing as the host's hand makes an outward gesture. In other words, as the host's hand/arm moves towards the interlocutor or the camera, he makes BDGs while pointing to himself with his fingers (index finger, pinch, or finger bunch). The occurrence count shows how many times a certain formula appeared in our resulting dataset. As the BDG can be repeated more than once in a formula the total count of BDGs for a particular

formula is provided in the 'BDG Repetition Count' to give the reader an idea of how frequently BDG appeared in our data. This frequency – 239 BDGs for 114 gesture formula occurrences, made by the show's host Rick Sanchez in 180 min of air time – demonstrates the overall importance of BDGs in multimodal viewpoint blending construction behind communication strategies in our data.

The list of gesture formulas presented in Table 2 illustrates three key findings:

(1) When considered as triggers or cues for viewpoint blending construction in persuasive and manipulative discourse, BGDs need to be analysed in their relation to ODGs to which they are closely related;
(2) BDG and BDG.E formulas – in which BDGs act as meaningful components of multimodal triggers and cues for viewpoint blending - constitute only 19.3% of all gesture formulas and 17.6% of all BDG repetitions. All other formulas manifest gesture configurations involving an inseparable relation between

Table 2: Types and Frequency of Body-Directed Gestures (BDG) and Related Outward-Directed Gestures Made by the Host in the 180 min of RT Shows Analysed in this Study.

Gesture formulas	Occurrence count	BDG repetition count
BDG	21	38
BDG.E_ODG	11	29
BDG_ODG	39	92
BDG.E_ODG and BGD	1	5
BDG_ODG and BDG_ODG	1	1
ODG_BDG.E	1	4
BDG and BDG.E_ODG	1	3
ODG and BDG and ODG	4	6
ODG and BDG_ODG	2	7
ODG_BDG_ODG	2	2
ODG_BDG	3	5
ODG and BDG	2	6
BDG.E-ODG	10	18
BDG and ODG	6	6
ODG and BDG and BDG	1	1
BDG-ODG	1	3
BDG_ODG and BDG and BDG_ODG	1	3
BDG and ODG_BDG	1	1
BDG and BDG	2	2
BDG and BDG-ODG	1	1
BDG and ODG and BDG	1	1
BDG and BDG_ODG	1	1
BDG.E	1	4
Totals	**114**	**239**

BDGs and ODGs if assessed from the perspective of viewpoint construction underlying persuasive and manipulative RT messages;

(3) Although a great variety of gesture formulas is observed as underlying viewpoint blending construction in persuasive and manipulative discourse, three types of gesture formulas appear to form the core for the data analysed. These three gesture formula types act as building blocks for the longer gesture formulas appearing in Table 2.

The most frequent gesture formulas observed for this study's data are:

(i) BDG_ODG (39 occurrences with 92 repetitions – constituting 38.5% of all gesture repetitions);
(ii) BDG (21 occurrences with 38 repetitions);
(iii) BDG.E_ODG (11 occurrences with 29 repetitions);
(iv) BDG.E-ODG (10 occurrences with 18 repetitions);

The BDG_ODG formula differs from the BDG.E_ODG formula only in how pronounced the BDG is; hence both formulas can be considered as belonging to one type. When considered as such, together BDG_ODG and BDG.E_ODG constitute 43.8% of total gesture formula occurrences and 50.6% of all BDG repetitions.

The second most frequent formulas are BDGs as contributing on their own to construction of multimodal triggers and cues for viewpoint blending. As mentioned above in (2), together with BDG.E they constitute 19.3% of all gesture formulas and 17.6% of all BDG repetitions.

The third place in terms of frequency is occupied by the BDG.E-ODG formula where BDGs occur within *one* outward-directed gesture, i.e. while the hand or arm is making an outward-directed gesture, a hand or fingers are making BDGs at the same time. Considered together with BDG-ODG, BDG.E-ODG constitute 9.6% of all gesture formulas and 8.8% of all BDG repetitions.

Overall, the table demonstrates that if we consider BDG as part of multimodal triggers and cues for viewpoint blending underlying persuasion and manipulation in this study's data, then they make a meaningful contribution to the underlying speech–gesture occurrences only in less than one-fifth of all given instances. In more than four-fifths of the instances, the relation of BDG to ODG or other BDG needs to be considered as part of the formation of such multimodal triggers or cues.

The further interpretation of the frequency of these formulas and detailed analysis including how they correlate with corresponding linguistic inputs and other characteristics of speech–gesture integrations for which we annotated is beyond the scope of this paper. However, the examples analysed below illustrating multimodal strategies at work should give the reader an insight into the workings of viewpoint blending that some of these formulas, including the most frequent BDG_ODG, represent.

4.2 Viewpoint Blending behind Multimodal Strategies

This study identifies three multimodal strategies applied by the RT NRS show host Rick Sanchez: establishing authority; maintaining a neutral tone; and dealing with the alternative voice. Three types of situations behind these strategies are distinguished: the host relying on a report by an RT correspondent, inserted as a video clip, to support the host's statements and arguments offered before or after the report; the host's interaction with one or more experts invited to the RT studio via video-conferencing (rather than in person, because of restrictions imposed by the coronavirus pandemic); and the host referring to previous events and situations as experienced and perceived by him personally, the RT team as a whole, or the show's expert guests.

4.2.1 Establishing Authority

Consider Example 1, in which the host relies on his interaction with an expert on explosions to establish both the credibility of information and his own authority as source of information (linguistic units on which the gestures under consideration fall are marked by capital letters italics). The host addresses the expert:

(1) *Listen, you do this for a living, you're called into court rooms TO BE ASKED QUESTIONS. I JUST called you, so I thank you.* (Show 6, 00:21:42:00-00:21:48:80; click to view or scan QR code)

The 'linguistic' input space represents a conceptual blend between two scenes – of a typical court room and of the RT studio. It does so through mapping the actions of calling to question, with their sets of participants (linguistically marked by *you* and *I*) and other space elements. The 'linguistic' input space for a court scene has more elements than the 'linguistic' input for the RT scene, including the markers for the RT show's guest's main occupation as a witness expert in courts, multiplicity of the 'calling into court rooms' events, and the main goal of the witness being called to court rooms – 'to be asked questions'. *JUST* provides a necessary linguistic trigger for activating the conceptual scene of the host questioning the expert minutes ago. This scene also feeds into the blend construction through TIME[8] and SPACE being mapped across all the mental inputs and scaled. The resulting 'linguistic' blend has the reliability of an expert's testimony projected from the 'court' scene, and thus signals a certain degree of credibility of the information communicated by RT. In the 'linguistic' blending network here, we also observe a viewpoint shift. *You* [the expert] is first fore-grounded – *Listen, you do this for a living*, and, then

8 Vital relations between mental spaces are given in small capitals.

subsequently, backgrounded via the use of the passive voice in *you're called* …. At the same time *I* [the host] is foregrounded through the use of the active voice.

The 'visual' input space includes the representation of two participants: the host (the speaker) who is gesticulating and the expert (the listener). The TV audience (from now on the viewer) sees both the host and the expert guest simultaneously on the screen as it is divided in two parts. On the utterance *Listen, you do this for a living* we observe an ODG, further foregrounding *you* here along with the fact that being an expert is his profession. Four repetitions of a 'body-directed gesture–outward-directed gesture' configuration (BDG_ODG) follow on *TO BE ASKED QUESTIONS I JUST*. This representational gesture is performed by the host's left hand. The closeness to the host's body is medium, the gesture amplitude is small. The palm of the host's hand is open, facing the host's chest, fingers slightly bent inwards, thumb extended. The motion trajectory can be described as 'swing'. The gestural input carries a viewpoint shift between the host and the interlocutor.

The BDG_ODG and the corresponding speech input *TO BE ASKED QUESTIONS I JUST* are integrated through TIME, SPACE, PART-WHOLE, IDENTITY, CHANGE and REPRESENTATION vital relations. Selective projection and compression of mappings between spaces result in the speech–BDG_ODG blend, which in turn acts as a multimodal trigger for further conceptual and viewpoint blending.

As mentioned before, the viewer sees the host and the expert guest simultaneously on the screen divided in two parts. This arrangement in itself serves as a visual trigger to conceptually map the host with the guest expert. The main trigger here though comes from the BDG_ODG ability to further window the viewer's attention.[9] BDG_ODG works to foreground the speech unit, which constitutes a more compressed blend, linguistically marked as the passive structure *TO BE ASKED QUESTIONS* (in the court scene input) and active *I JUST* (in the RT studio scene input). The Actor position is gapped (Talmy 2000, 2015) in the 'court' input space, but it is foregrounded through windowing the viewer's attention in the 'RT studio' input space. The resulting more compressed blend is framed by the event PARTICIPANT INTERACTION,[10] that both input scenes — 'court' and 'RT studio' — share. The blend is further structured by the force-dynamics schema of Shifting Impingement.[11] (Talmy 2000). In the

9 On the process of windowing attention see Talmy (2000; 2015).
It is important to note that introducing the concept of 'windowing attention' through gesture is different from linguistic approaches that look into gestures aligning with particular words. Hence the problem of gestures often preceding the speech units to which they may have semantic relation by up to 80 ms is less acute here (McNeill, David (2016). *Why we gesture: The surprising role of hand movements in communication*. Cambridge University Press., 19; cf. Hoffmann 2017, 4–5).

10 Names of frames are given in uppercase.

11 Force-dynamic patterns at work are signalled by initial capitals for each word.

blend, the on-set causation *I* [the host] is foregrounded as calling the expert and questioning him. This has a further consequence. The outer-space relations of ANALOGY/DISANALOGY and CHANGE between the judge in the 'court' input and the host in the 'RT studio' input are compressed into the resulting blend as the UNIQUE RT host with the PROPERTY of the court judge's authority.

Interestingly, at this stage of blending, the CYCLE event frame is projected from the gesture input. Although the CYCLE frame was activated by the grammatical and lexical means at the earlier stage of blending, at this stage the linguistic 'cycle' information is outside the window of the viewer's immediate attention. It is likely that the viewer retains some of this cycle information from earlier. Furthermore, now the activation of the CYCLE frame is reinforced by the gestural input. This distribution of 'cycle' information across time and modalities demonstrates the host's intention to distribute the conceptual load for the purposes of more efficient communication of information related to his establishing of authority.[12]

Crucially, there seems to be another stage of blending here, triggered by both the visual arrangement on the screen and the speech–BDG_ODG. The screen arrangement enables the construction of the blend in which the host and the viewer engage in fictive interaction. The BDG_ODG representational and conduit functions are blended through the combination of two schemas — self-orientated and interlocutor-orientated (as identified for gestures by Mittelberg 2018). The former supports the dominance of the host in the 'fictive interaction' blend and the latter supports the intersubjective 'force dynamic' understanding between the host and the viewer in this blend.[13] As noted by Parrill (2012, 99) the use of the second-person pronouns is likely to prompt the recipient to adopt the internal conceptual viewpoint. Thus, in our case *you* combined with the gesture carrying both representational and interactive 'intersubjective and control' functions may prompt the viewer to dramatize their internal position in the 'fictive interaction' blend, thereby experiencing the authority of the RT host with its all emotional dimensions and being prepared to accept it.[14]

---

12 On the distribution of conceptual load in communication see Melinger, Alissa & Sotaro Kita (2007). "Conceptualisation load triggers gesture production." *Language and Cognitive Processes* 22.4, 473–500.

13 Cf. interactive gestures as control gestures, Wehling, Elisabeth (2017). "Discourse management gestures." *Gesture* 16.2, 245–276.

14 On dramatization and theatricality in fictive interaction, see Brandt, Line (2008). "A semiotic approach to fictive interaction as a representational strategy in communicative meaning construction." *Mental spaces in discourse and interaction* 170, 109. and Pascual, Esther (2014). *Fictive Interaction: The conversation frame in thought, language, and discourse*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

4.2.2 Maintaining Neutral Tone

The next strategy under consideration here is maintaining a neutral tone. The strategy is understood as the host's intent to appear neutral and objective as far as the linguistic modality is concerned, while at the same time communicating stronger messages via gesture.

Consider Example 2:

(2) *We have team coverage, right here, on the News with Rick SanchEZ, WHERE WE BELIEVE it's time to do news again.* (Show 7, 00:02:41:49-00:02:49:19; click to view or scan QR code)

This utterance is repeated at the beginning of each RT NRS show, and thus gains the status of a slogan. The linguistic input here tells us that the RT show offers news coverage prepared by the RT team at the present space and location, and that they believe that it's time to do news again. The visual input has the host — Rick Sanchez — as embodying his self and the self of 'we' and the self of 'News with Rick Sanchez' and the associated selves' cognitive state and epistemic stance of believing. The host's fist (right hand) goes close to his head, and then the hand comes down. Both linguistic and gestural cues engage in over-explicit referencing to the RT news show. Such referencing occurs through the repetition of the pronoun *we*, as well as the lexical units indicating the show—*where, right here, coverage, news, Rick Sanchez*. The referencing here relies on the frame-metonymic

relation.[15] These lexical means signify RT's location, time, participants, action, event, cognitive state, and epistemic stance.

Chafe (1976, 1994) links the frequency and the choice of referring expressions to the part a referent plays in a discourse at a particular time. This is linked to the referent's importance, accessibility, and its cognitive status or givenness. The frequently mentioned referent is seen as view-pointed (Parrill 2012). The host's over-explicit referencing behaviour is a rhetorical strategy, aiming to foreground the RT show characteristics central to which is its new independent approach to news making. The input space for *it's time to do news again* on its own constructs the meaning of 'it's time to broadcast today's news show with Rick Sanchez' and is the mental space for 'real' with its epistemics of certainty. The high degree of certainty here is also supported by the visual input, as the viewer is already watching the RT NRS news show. This 'real' input space of *it's time to do news again* clashes with the input space for *WHERE WE BELIEVE*, upon which it is syntactically dependent in the linguistic input space. The space for *believe* is the 'unreal' space in a sense that "it is not epistemically endorsed by S [the Speaker/Subject] but may, privately as it were, be regarded by S as counterfactual, possible, probable, improbable, etc." (Chilton 2010, 197). These spaces clash as the viewer is not satisfied with the interpretation that the RT team broadcasting their news show is counterfactual. The resulting blend rooted in this clash generates a richer counterfactual meaning. This in turn enables the viewer to interpret the utterance as implying that there had been no real, independent, good news coverage for a long time up until RT came onto the scene, and the RT team believes that it is time to engage in producing such coverage. The first shape of BDG windows the viewer's attention and foregrounds *WHERE WE BELIEVE*. It also adds to the overall emergent semantic meaning. Calbris (2011) argues that the hand shape of a fist carries an idea of strength and force and, by extension, all possible semantic components of strength and force, physical or psychological. She also links it to necessity and gestural exclamation. A simple search on the internet gives multiple testimonies showing that people associate fist gesture as an emblem with pride and a raised fist with the idea of solidarity. In Example 2, the semantics carried by the BDG-fist raised to head is further reinforced by a head gesture which is made in the direction of the fist. Such a head gesture is interpreted as 'effort against force' in Гришина (2017).

Although the counterfactual construction as supported by linguistic viewpoint already relies on forces (Chilton 2010), the BDG brings the force-dynamic schema to

15 On frame metonymy in persuasive communication see Sweetser, Eve (2017). "Metaphor and metonymy in advertising: Building viewpoint in multimodal multi-space blends." *Journal of Pragmatics* 122, 65–76.

the blend which is anchored in visual representation of body and physical strength. The force-dynamic schema of Shifting Impingement here clashes with the schema of Cognitive State evoked by *WE BELIEVE*.[16] As Fauconnier and Turner (2002) point out, bigger clashes between input mental spaces generate the emergence of more imaginative meanings in blends. The emergent meaning in the blend under consideration is twofold. Firstly, it has the semantic components of 'we, here, now, believe, strongly', which in their total get much closer to the epistemic 'we know'. Secondly, the blending network runs to support the emergence of a counterfactual 'before us there was no news coverage'. That in turn is likely to support the emergence of the deontic meaning in the blend — 'it's necessary for us to engage/we must engage in reporting real independent news with all the effort and force we can invest'.[17] The resulting complex blend is embodied by the host, and creates a powerful confident 'self'. *WHERE WE BELIEVE*, integrated with the BDG, is likely to also trigger the viewer's willingness to conceptually include themselves in a scene, i.e. assume an internal conceptual viewpoint, which is associated here with strength, necessity, and pride. This study argues that although the viewer is likely to construct a conceptual viewpoint based on Example 2 as engaging and carrying deontic meaning, she is also likely to do this without registering it, as a lot of the conceptual load here is carried by gesture. Equally, it would be hard to hold the host or the RT accountable for communicating such strong messages as in Example 2, as on the purely linguistic level the example looks more neutral and innocent.

4.2.3 Dealing with the Alternative Voice

As mentioned earlier, one of RT's main goals is to maintain the appearance that their news coverage is objective and independent. Hence, from time to time shows invite guests as alternative voices, meaning that the guests express viewpoints conflicting with the pro-Russian state messages communicated by RT. In our data we identified three such instances supported by BDGs. One example included employing a complex counterfactual scenario in which Russia was in fictive dialogue with the West over Russia's rapidly approving a COVID-19 vaccine and the West's disapproving Russia's actions as over-hasty. In this scenario Russia was embodied by the RT host and the West by his expert guest, a professor of regulatory

16 On schematic systems of cognitive state, see Talmy 2000 and Ibarretxe-Antuñano, Iraide (2006). "Leonard Talmy. A windowing onto conceptual structure and language: Part 2: Language and cognition: Past and future." *Annual Review of Cognitive Linguistics* 4.1, 253–268.

17 On epistemic, deontic, and counterfactual meanings in viewpoint construction, see Chilton, Paul (2010). "The conceptual structure of deontic meaning: A model based on geometrical principles." *Language and Cognition* 2.2, 191–220.

and quality sciences at USC. BDGs played a key part in construction of pro-Russia ideological viewpoints in this 'fictive interaction' scenario, while helping the host to shift between his Russia's Self viewpoints and his Other's (US) Self viewpoints. Viewpoint shifts as embodied by the host worked to further pro-Russia's viewpoints.

Consider Example 3 (Show 7), four separate utterances by the host, all with BDGs:

(3) a. *What RUSSIA seems to be signalling is that it's COMFORTABLE enough* […] (00:15:57:70-00:16: 01:00; click to view or scan QR code below)
b. *THIS IS YOUR BIGGEST FEAR IS that* […] (00:21:00:00-00:21:01:46; click to view or scan QR code below)
c. *AND, WE'VE made a decision that we're READY to go* […] (00:21:37:00-00:21:39:77; click to view or scan QR code below)
d. *Hard to be able to look at it FROM SO FAR AWAY and make that decision* […] (00:21:39:80-00:21:43:27; click to view or scan QR code below)

(3a) (3b) (3c) (3d)

This example illustrates the viewpoint shifts underlying the host's manipulative technique supported by various BDGs made on words not obviously referring to the host himself, but to concepts with which he identifies.

The host's integration of speech units (foregrounded through windowed viewer's attention) and BDGs in (3a) construct a character viewpoint of Self through gesture and an observer viewpoint through the linguistic input. In (3b) again the linguistic input supports an observer viewpoint construction, whereas the BDG constructs a character viewpoint of Self for the professor – expert guest, while also labelling her opinion as her biggest fear. In (3c) the host's integration of speech units and BDGs support his dramatization of the conversation between Russia and the West. Here, both the linguistic and gestural input create a character (Russia) viewpoint embodied by the host. In (3d) the dramatization continues and the linguistic input marks the observer viewpoint. At the same time the host's making a BDG on *FROM SO FAR AWAY* clearly supports his identifying himself with this 'from so far away' – the USA/the West. These viewpoint shifts work to present the host to the viewer (US or Western TV audience) as one of them – Self, while also conceptually connecting them to Russia's Self through him. This is

likely to result in blending of the viewer's Self with the Russia's Self,[18] as part of viewer's conceptualisation process. The viewpoint shifts enabling this conceptual 'Self' blending can be interpreted as shifts in force vectors (Chilton 2010, 2014) in the argumentation and rhetoric under consideration. Furthermore, the viewpoint conceptualisation here can be described in terms of pointing and placing offered in Clark (2003). It becomes important not only whether the host points to himself to accompany certain lexical and grammatical units, but also how he 'places' BDGs as discourse unfolds to enable shifts in force vectors.[19] As Turner (this volume, p. 459) points out, "very little information needs to be communicated, even in cases where the communicator wishes the receiver to construct a full and complete meaning immediately". The analysis of our example shows that both body-directed gestures and their placings can function as triggers and cues for a certain kind of viewpoint blending. The latter enables full, complete, and immediate construction of persuasive and manipulative meaning in the viewer's mind. Obviously, gesture is not the only cognitive system upon which the host relies in his conversation with the professor to ensure that Russia's position on approving the vaccine is seen in the most favourable light. At the very least, he employs an epistemic stance of *seems to*, a counterfactual scenario of fictive interaction between Russia and the West, and dramatization of this scenario. However, it is evident that through placing BDG gestures at the right moments and employing viewpoint shifts through shifts in force vectors, the host blends the viewer's Self and Russia's Self through his own Self. This is intended to make the viewer assume Russia's ideological viewpoint.

Considering the linguistic input on its own in Example 3 does not reveal the full information communicated by the host. It only becomes visible if one looks at speech and BDG 'Self' gestures in combination. Linguistic input signals more objectivity and less certainty, compared to when it is integrated with gesture. Working together the speech and gesture support the displacement by the host's pro-Russia viewpoint of the pro-West viewpoint communicated by the guest expert.

The speech–BDG integration presented below in Example 4 is chosen for a reason that in contrast to the example of the fictive interaction re the vaccine just mentioned, the host struggled to displace the guest's alternative viewpoint. The guest was Larry King and the conversation was about the Democratic National Convention (DNC) which this year was held online because of the COVID-19 pandemic. The host started the conversation with an attempt to frame the DNC in a negative light, by repeatedly referring to DNC organisers failing to run the convention smoothly from a

18 Cf. Turner (2014, 102–103).

19 Cf. discussion of non-linguistic units as prompting blending in the scenes of suggestive blended classic joint attention, Turner this volume.

technical perspective, and referring to some speakers as lacking skills in dealing with video-conferencing software or other technical nuances.

Example 4 (Show 4):
The host:

a. *It's been everything from interesting TO NOVEL to A TECHNICAL DUD*. (00:21:54:59-00:21:59:67; click to view or scan QR code)

On the NOVEL the host makes a rotating gesture with both hands towards his torso, with hands taking turns in being close to the body. Open palms are facing the body, thumbs are extended. On the *TECHNICAL DUD* the host moves both hands to his body and crosses them on his chest (closed position).

Once Larry King is brought in the RT studio via video-conferencing, he is asked for his opinion on the DNC. Larry responds that it was a well-produced show considering that it was done without crowds. He hails Barack Obama's speech as brilliant, and remarks that in his opinion (and he notes that he goes back as far as Roosevelt), nobody is in Obama's league.

The host ignores Larry's viewpoint, and tries to re-direct the conversation back to the topic of technical failings by the DNC organisers. He actively searches for Larry's support with this re-framing, saying:

b. *GUYS LIKE YOU AND I – JUST A MOMENT AGO I had* […]. *I KNOW HOW TO DO THIS, I've been doing it for 30 years*. (00:23:23:79-00:23:33:79; click to view or scan QR code)

The *GUYS LIKE YOU AND I* is accompanied by his right-hand ODG_OBG with the hand going toward Larry and then to the host's chest. His right-hand ODG_BDG repeats on *JUST A MOMENT AGO*. Simultaneously the host's left hand goes towards the host's torso, the palm is half closed, index finger pointing inwards, thumb extended. On *I KNOW HOW TO DO THIS* the host's right hand goes close to his chest, palm is half closed with the index finger pointing to the chest.

The host is clearly keen to communicate the message about the DNC's technical failings in the foregrounded fashion. He tries to establish authority in judging on the matters in focus, by referring to his recent and past success in dealing with technical questions. Moreover, through constructing linguistic and gestural viewpoint in a certain way he attempts to tie Larry and himself together conceptually. The gestural configuration with the host's ODG_BDG performed by his right hand and the simultaneous BDG performed by the left hand seems to carry two slightly different functions in two interconnected mental spaces. The right hand performs a representational function extending the representation of 'guys like us' beyond its occurrence in time. This is to continue tying Larry and the host together semantically

and pragmatically. The host's right-hand gesture also fulfils an interactive function signalling to the viewer that the exchange of arguments continues and the host has evidence in support of his arguments to present. At the same time the left hand seems to be directing the viewer's attention to this evidence – the description of the event which has just taken place incorporating the Self (the host) as competent in technical matters. The gesture configuration windows the linguistic unit *JUST A MOMENT AGO I*, which immediately triggers the construction of the 'evidence' space. The use of the pronoun *I* accompanied by the BDG with half closed palm and the index finger pointing both foregrounds the 'evidence' space and the host's own role in the 'evidence' space.

Larry and the host continue interacting for some time in a similar fashion, with Larry hailing Obama, and wondering whether the host too found Obama sensational. The host responds by saying that many speeches were sensational if you watch them on their own on YouTube but that as a whole the DNC show was a disaster.

Larry King responds to that by saying:

c. *We are asking non-television people to put on a show. Not easy.* (00:24:30:27-00:24:37:36; click to view or scan QR code)

To that the host replies:

d. *That's exactly right. AND BY THE WAY WE'RE HYPERCRITICAL BECAUSE IT'S WHAT WE'VE BEEN DOING ALL OUR LIVES, so … guilty as charged, my friend Larry King.* (00:24:36:97-00:24:48:15; click to view or scan QR code)

It is clear that by now the host recognises his failure to frame the DNC in a negative light, so he clearly attempts to save his face by shifting the viewpoint yet again. His linguistic utterance is accompanied by two BDG and two ODG_BDG. On *AND BY THE WAY WE'RE* the host's right hand, open palm, thumb slightly extended, goes really close to his chest and stays there. On *HYPERCRITICAL* the hand touches the chest. The host is smiling. On *BECAUSE IT'S WHAT WE'VE BEEN DOING ALL OUR LIVES* the right hand moves to Larry and back to the host's chest two times. After that the host turns away from Larry King while looking at his papers and continuing with "so … guilty as charged, my friend Larry King." He is not gesticulating. His voice is quiet, his tone is unusually even. What is interesting in this example is the relation between the macrocosm of force-dynamics in argumentation and the microcosm of force-dynamics of its semantic and pragmatic construction (Oakley 2005). The BDG and ODG_BDG configuration provides a specific foregrounding configuration here. On *AND BY THE WAY WE'RE* the BDG foregrounds the host as part of the 'we' group, which is profiled by gesture's

configuration as really important and close to him on the mental and emotional side of things.

The host's BDG touches his chest on *HYPERCRITICAL*, thereby reinforcing the meaning of 'we' just created with its associated PROPERTY and INTENTIONALITY of *HYPERCRITICAL*. This BDG links 'we' to the action of 'criticizing' through the frame-metonymic relation. Speech–gesture integration here triggers the creation of the 'excessive' force schema. The utterance's grammatical and lexical triggers considered together draw on the CAUSAL CHAIN event frame to provide an excuse for applying excessive force — *BECAUSE IT'S WHAT WE'VE BEEN DOING ALL OUR LIVES*. Interestingly, the two ODG_BDG, co-occurring with this subordinate clause of reason, foreground the joint nature of the action of doing here. It is not entirely clear from the linguistic exchange exactly what unit in the previous discourse is targeted by the host's *IT'S WHAT*. The ODG_BDG gesture should help to disambiguate here in its function as a targeting gesture.[20] The ODG_BDG works to help the viewer find a target in Larry's *We are asking non-television people to put on a show*. It is the repetition of ODG_BDG that equips the viewer with the cue, enabling her to connect the trigger of *IT'S WHAT* to the target of *put on a show*. The representational ODG_BDG here also supports the construction of a proposition, which the host relies upon as a rhetorical technique for manipulation. The host constructs his utterance as propositional to include Larry in the 'applying excessive force' scenario, despite Larry not expressing any criticism about the DNC in his exchange with the host, but on the contrary praising it all the time. The 'excessive force' space created by the host at the local – micro – level, thus, clashes with the 'force-dynamic' space created by both interlocutors at the more global – macro – level. In the latter, the Balance of Strength is shifted to expose the host as a weaker party. The host's attempt to tie himself to the stronger party – Larry – through propositional construction triggered by speech–gesture integrations is not rhetorically successful. The host seems to fail to displace the conceptual 'inter-subjective' viewpoint structured by the force-dynamic scheme of argumentation. He is not successful in making the local 'inter-subjective' viewpoint of both him and Larry applying excessive force displace the viewpoint (of the DNC doing a great job in difficult circumstance) emerging on the global level during this debate. It seems that there is another reason why the BDG co-occurring with *HYPERCRITICAL* is further foregrounded in the gestural chain under consideration. The host's open palm touching his chest seems to be signalling the strong emotion linked to the embarrassment the host is experiencing.[21] Nijk (2019) notes that a displacement

20 On targeting gesture see Talmy, Leonard (2018). *The Targeting System of Language*. Cambridge, Massachusetts: MIT Press.

21 For BDG signalling feelings see the description of body-anchors in Cooperrider 2014.

scenario in viewpoint construction requires an internal arrangement of conceptual viewpoint. This is indeed the case in Example 4, where the inter-subjective internal viewpoint is repeatedly signified by the co-occurrence of pronoun *we* and ODG_BDGs. It is also the case in Example 3 where the host repeatedly communicates the internal viewpoint arrangement through both his BDG and the dramatization scenario he employs.

5 Discussion and Concluding Remarks

The approach employed by the current study demonstrated that viewpoint constructed locally cannot be fully interpreted without exploring its relation to global viewpoint, or in other words without considering viewpoint as discourse network (Dancygier 2011; Dancygier and Vandelanotte 2016). The importance of the relationship between local and global viewpoints becomes especially evident when we deal with integrations of linguistic and gestural viewpoints as multimodal triggers for viewpoint construction over the unfolding of the discourse. Recent empirical evidence suggests that a spatial schematic arrangement for a story conveyed by gesture stays with the recipient while she continues to take in story information now communicated without gesture (Sekine and Kita 2017). The examples analysed by this study show that it may be equally true that a schematic arrangement set up linguistically stays with the recipient and affects their interpretation of subsequently appearing gestures (Example 1, '*cycle*' information). This study links the observed phenomenon to specific configurations in conceptual load distribution across linguistic and visual modalities. It occurs when the speaker (here, the RT host) is keen to communicate a specific chunk of view-pointed information (the UNIQUE RT host with the PROPERTY of a court judge's authority). The conceptual characteristics of PROPERTY here are reinforced first linguistically and then gesturally, now coinciding with conceptual mappings between the RT host and the judge.

The study's findings also show that a conceptual blending framework can be successfully utilised to examine speech–gesture co-occurrences as multimodal constructions. Its toolbox allows us to unpack speech–gesture units without overlooking the integral nature of their interrelation (with integral and interrelated nature as understood e.g. by Enfield 2009 or Hoffmann 2017 [22]). Hoffmann (2017, 1) argues that for understanding such multimodal units as speech–gesture, one needs to appreciate the differentiation of "entrenched constructions that are stored

22 I am grateful to the editor of this volume, Peter Uhrig, for directing me to Thomas Hoffmann's work.

in the long-term memory from form-meaning pairings that are assembled in the working memory (online constructions)". Our study's analysis has shown that using blending as a method helps to shed some light on how speech–gesture units rely on evoking the elements of long-term knowledge in online meaning construction. For example, it demonstrates how the default function of body-directed gestures as self-points and body anchors (terms coined by Cooperrider 2014) feeds into more complex functioning of speech–gesture constructions as relying on viewpoint blending.

One of the strengths of the conceptual blending framework is that it offers a systematic set of precise tools for analysis, but at the same time it remains open in the sense that it can be developed further or combined with other tools, while maintaining the framework's main logic. This study has employed for viewpoint blending analysis the concept of windowing the receiver's attention, which was originally introduced by Talmy (2000, 2015) for purely 'linguistic' cognitive semantic analysis. It developed the concept further by successfully applying it to viewpoint analysis in multimodal communication, and arguing that the windowing of the viewer's attention forms part of a speech–gesture integration process. Furthermore, the windowing of the viewer's attention in mediated communication is underpinned by cognitive mechanisms of blended joint attention. Blended joint attention scenes are routinely created in mediated communication. In these scenes classic joint attention, which is prevalent in in-person physical communication, gets blended with the actual experience of mediated communication which is lacking such an in-person component.[23] Turner et al. (2019, 4) argue that mechanisms of blended joint attention may be used not just to address the viewer but also to persuade her. As this study's analysis has shown, BDGs support blended joint attention in the RT shows and may be used for persuasive and manipulative purposes. For example, blended joint attention may be achieved through both pointing to one's self (RT host's BDGs) and placing BDGs at the right moments as discourse unfolds. Both gesture and placing the gesture help to window the viewer's attention on particular speech units, thereby triggering and supporting a certain kind of viewpoint blending for persuasive and manipulative purposes. Thus, the study's findings add to our understanding of the intricacies of blended joint attention and its role in different kinds of mediated multimodal communication.

---

23 On blended joint attention see Thomas, Francis-Noël & Mark Turner (2011). *Clear and simple as the truth: Writing classic prose*. Princeton University Press, ibid.; Turner, Mark, Maíra Avelar & Milene Mendes de Oliveira (2019). "Blended Classic Joint Attention and Multimodal Deixis." *Signo, Santa Cruz do Sul* 44.79, 3–9., and Turner this volume.

The current study finds that the investigation of viewpoint through blending benefits from employing image schemas and schematic systems as an analytical tool. That allows for better understanding of how gesture schematises information and how speech–gesture integrations trigger the construction of meaning and associated viewpoint construction. The study argues that the force-dynamic schema in particular (Ibarretxe-Antuñano 2006; Talmy 1988, 2000) is useful for understanding the complex conceptual processes underlying RT's media communication. The study's analysis shows that the force-dynamic schema is pivotal in the construction of ideological viewpoints in persuasive and manipulative communication. It shows how speech–BDG and speech–BDG_ODG rely on force-dynamic conceptualisation when physical viewpoints inform cognitive and ideological viewpoint construction. It also demonstrates the link between the force-dynamics of argumentation and rhetoric at macro- and micro-levels of interaction (Example 4). Here, the study continues the discussion started by Oakley (2005), by bringing gesture into the exploration of force-dynamics in rhetoric. In so doing it also contributes to scholarship on multimodal argumentation and rhetoric in various media genres (Tseronis and Forceville 2017). Furthermore, this study's findings support the argument made by Dancygier and Vandelanotte (2017a) that force-dynamics schematise conceptualisation in blending, with the latter underlying multimodal constructions and units of various nature. It does so by showing examples of BDG and ODG further schematising (view-pointing) conceptualisations, which were first constructed linguistically, through force-dynamics. The current study claims that viewpoint shifts (character–observer) can be modelled as changes in force vectors,[24] applying this notion to the analysis of conceptual, linguistic and gestural viewpoint in their interelation and integration. This study also relates to observations and findings of Mittelberg (2018) in finding that BDG and BDG_ODG constructions play a crucial role in conceptualisation through image schemas because of the spatial and dynamic properties of gestures. Mittelberg (2018, 1) differentiates between three gestural patterns relying on image schemas and force-dynamics gestalts: "body-inherent/self-oriented (body as image-schematic structure; forces acting upon the body); environment-oriented (material culture including spatial structures), and interlocutor-oriented (intersubjective understanding)". The current study's findings show that as far as intersubjective understanding and other intersubjective conceptualisations are concerned, BDG need to be included in the interpretation process. The analysis of gesture formulas shows that in 93 instances out of the 114 in our data, BDG forms an inseparable component of force-dynamic schemas underlying intersubjective conceptualisation. The study argues that the pattern of the inter-relation of body-directed

24 On force vectors see Chilton (2010 and 2014).

gesture and outward-directed gesture needs to be added to Mittelberg's typology. This pattern relies on the event frame of PARTICIPANT INTERACTION and the relation of inter-subjectivity for viewpoint construction in communication.

Turner points out that news communication loaded with ideological viewpoints routinely relies on "cascades of nested selves in the mental webs" constructed by media, and on the implied and real selves of the hosts and the viewers (2014, 103). Our findings support this claim through an analysis of speech–gesture integrations. The current study argues that the RT host acts as an intermediate highly complex blend in his channelling of pro-Russian viewpoints. In doing so he relies on what Cooperrider (2014) defines as a microcosm of body-directed gestures, as anchored in the body — social, interactive, experiential, and biological. The current study finds that this microcosm is better understood when BDGs are explored in their interrelation with outward-directed gestures. Although valuable and fit for the purposes for which it was used, the approach employed by Cooperrider does not provide an opportunity to examine BDGs in their full capacity as multimodal triggers and cues for viewpoint blending. This is because Cooperrider's approach affords an examination of BDGs only, without taking into consideration outward-directed gestures as linked to body-directed gestures in their contribution to one viewpoint construction process. This study develops an analytical approach which is better suited for the purposes of viewpoint analysis. The approach allows us to investigate the role of body-directed gestures in interaction, and in creating an intersubjective viewpoint (cf. Mittelberg 2018; Sweetser 2012). It also adds to our understanding of representational and interactive gestures (Bavelas et al. 1992; Chu et al. 2014). As the study's examples demonstrate, body-directed gestures play a crucial part in the creation of viewpoint, with a BDG_ODG grouping structuring representational content and managing real and fictive interaction.

This study notes the ability of gesture to affect the overall speech–gesture viewpoint construction in that gestural viewpoint is communicated to the viewer without her registering it. There are empirical studies indicating that gestures visible to the viewer communicate information in a fashion somewhat invisible to them, i.e. under the viewer's conscious radar.[25] The viewer does not register what parts of information are conveyed by which mode. Once the information communicated by both language and co-speech gesture is processed by the viewer, its influence cannot be undone. For example, Kelly et al. (2010) showed that gestures cannot be ignored, even when people are asked to just make judgements on speech. Gesture-speech integration is automatic and tight. This is not surprising,

25 See overview in Guilbeault, Douglas (2017). "How politicians express different viewpoints in gesture and speech simultaneously." *Cognitive Linguistics* 28.3, 417, 435.

as empirical evidence indicates that the brain integrates information channelled through both modalities simultaneously (Özyürek et al. 2007). As this study's analysis has showed, such speech–gesture interrelation and interaction can enable the RT show host to communicate viewpoints to the viewer in a manipulative fashion, while on the surface remaining accountable only for the viewpoint communicated through the linguistic modality. Crucially, this study has introduced the notion of multimodal triggers and cues for viewpoint blending, which account for both linguistic viewpoint and gestural viewpoint in their interrelation and integration. The study argues that the notion of triggers and cues is a useful addition to the toolbox needed for the examination of viewpoint blending as a hugely complex network underlying such big multimodal discourses as TV news shows. This study argues that its analytical approach is successful in revealing the conceptual workings underlying RT's multimodal communication strategies of establishing authority, maintaining neutral tone, and dealing with the alternative voice. It is envisaged that the multimodal triggers and cues identified will enable the transfer from manual analysis of these multimodal communication strategies to analysis at scale, as supported by pipelines of computational tools for communication analysis in large multimodal datasets (Steen et al. 2018; Uhrig 2018). This constitutes the future direction for the research presented here. Another future direction would be to research how RT's multimodal strategies compare to strategies employed by other media, e.g. US English-language and Russian Russian-language domestic media.

References

Alibali, M. W., S. Kita, and A. J. Young. 2000. "Gesture and the Process of Speech Production: We Think, Therefore We Gesture." *Language & Cognitive Processes* 15 (6): 593–613.

Aussems, S., and S. Kita. 2019. "Seeing Iconic Gestures While Encoding Events Facilitates Children's Memory of these Events." *Child Development* 90 (4): 1123–37.

Bakhtin, M. M. 1981. *The Dialogic Imagination: Four Essays*. Austin: University of Texas Press.

Bakhtin, M. M. 1984. *Problems of Dostoevsky's Poetics*. Minneapolis: University of Minnesota Press.

Bavelas, J. B., N. Chovil, D. A. Lawrie, and A. Wade. 1992. "Interactive Gestures." *Discourse Processes* 15 (4): 469–89.

Beattie, G., and H. Shovelton. 2005. "Why the Spontaneous Images Created by the Hands During Talk Can Help Make TV Advertisements More Effective." *British Journal of Psychology* 96 (1): 21–37.

Brandt, L. 2008. "A Semiotic Approach to Fictive Interaction as a Representational Strategy in Communicative Meaning Construction." *Mental Spaces in Discourse and Interaction* 170: 109.

Calbris, G. 2011. *Elements of Meaning in Gesture*. Amsterdam; Philadelphia: John Benjamins.

Caldognetto, E. M, and I. Poggi. 1997. "Micro-and Macro-Bimodality." In *ESCA Workshop on Audio–Visual Speech Processing*. Rhodes, Greece.
Chafe, W. 1994. *Discourse, Consciousness, and Time: The flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.
Chafe, W. L. 1976. "Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View." In *Subject and Topic*, edited by C. N. Li, 25–55. New York: Academic Press.
Chilton, P. 2010. "The Conceptual Structure of Deontic Meaning: A Model Based on Geometrical Principles." *Language and Cognition* 2 (2): 191–220.
Chilton, P. 2014. *Language, Space and Mind: The Conceptual Geometry of Linguistic Meaning*. Cambridge: Cambridge University Press.
Chu, M., A. Meyer, L. Foulkes, and S. Kita. 2014. "Individual Differences in Frequency and Saliency of Speech–Accompanying Gestures: The Role of Cognitive Abilities and Empathy." *Journal of Experimental Psychology: General* 143 (2): 694–709.
Clark, H. H. 2003. "Pointing and Placing." In *Pointing: Where Language, Culture, and Cognition Meet*, edited by S. Kita, 243–68. Mahwah, N.J.: L. Erlbaum Associates.
Cooperrider, K. 2014. "Body-Directed Gestures: Pointing to the Self and Beyond." *Journal of Pragmatics* 71: 1–16.
Dancygier, B. 2011. *The Language of Stories: A Cognitive Approach*. Cambridge: Cambridge University Press.
Dancygier, B., and L. Vandelanotte. 2016. "Discourse Viewpoint as Network." In *Viewpoint and the Fabric of Meaning: Form and Use of Viewpoint Tools across Languages and Modalities*, edited by B. Dancygier, W. Lu, and A. Verhagen, 13–40. Berlin; Boston: De Gruyter Mouton.
Dancygier, B., W.-L. Lu, and A. Verhagen. 2016. *Viewpoint and the Fabric of Meaning: Form and Use of Viewpoint Tools across Languages and Modalities*. Berlin/Boston, Germany: De Gruyter, Inc.
Dancygier, B., and E. Sweetser, eds. 2012. *Viewpoint in Language: A Multimodal Perspective*. Cambridge: Cambridge University Press.
Dancygier, B., and L. Vandelanotte. 2017a. "Image-Schematic Scaffolding in Textual and Visual Artefacts." *Journal of Pragmatics* 122: 91–106.
Dancygier, B., and L. Vandelanotte. 2017b. "Viewpoint Phenomena in Multimodal Communication." *Cognitive Linguistics* 28 (3): 371–80.
Enfield, N. J. 2009. *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Cambridge: Cambridge University Press.
Enfield, N. J., S. Kita, and J. P. De Ruiter. 2007. "Primary and Secondary Pragmatic Functions of Pointing Gestures." *Journal of Pragmatics* 39 (10): 1722–41.
Fauconnier, G. 1997. *Mappings in Thought and Language*. Cambridge: Cambridge University Press.
Fauconnier, G., and M. Turner. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.
Guilbeault, D. 2017. "How Politicians Express Different Viewpoints in Gesture and Speech Simultaneously." *Cognitive Linguistics* 28 (3): 417.
Hoffmann, T. 2017. "Multimodal Constructs–Multimodal Constructions? The Role of Constructions in the Working Memory." *Linguistics Vanguard* 3: s1.
Hostetter, A. B. 2011. "When do Gestures Communicate? A Meta-Analysis." *Psychological Bulletin* 137 (2): 297.

Ibarretxe-Antuñano, I. 2006. "Leonard Talmy. A Windowing Onto Conceptual Structure and Language: Part 2: Language and Cognition: Past and Future." *Annual Review of Cognitive Linguistics* 4 (1): 253–68.

Jacquin, J. 2017. "Embodied Argumentation in Public Debates." In *Multimodal Argumentation and Rhetoric in Media Genres*, Vol. 14, 239.

Johnson, M. 2008. "The Philosophical Significance of Image Schemas." In *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, edited by H. Beate, and G. Josephe, 15–33. Berlin: De Gruyter.

Kelly, S., K. Byrne, and J. Holler. 2011. "Raising the Ante of Communication: Evidence for Enhanced Gesture Use in High Stakes Situations." *Information* 2 (4): 579–93.

Kendon, A. 1994. "Do Gestures Communicate? A Review." *Research on Language and Social Interaction* 27 (3): 175–200.

Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kita, S. 2000. "How Representational Gestures Help Speaking." In *Language and Gesture (Language Culture and Cognition Language Culture and Cognition)*, edited by M. David, 162–85. Cambridge: Cambridge University Press.

Kita, S., M. W. Alibali, and M. Chu. 2017. "How do Gestures Influence Thinking and Speaking? The Gesture-For-Conceptualization Hypothesis." *Psychological Review* 124 (3): 245.

Kita, S., and A. Özyürek. 2003. "What Does Cross-Linguistic Variation in Semantic Coordination of Speech and Gesture Reveal?: Evidence for an Interface Representation of Spatial Thinking and Speaking." *Journal of Memory and Language* 48 (1) 16–32.

Lewis, T. N., and E. Stickles. 2017. "Gestural Modality and Addressee Perspective Influence How We Reason About Time." *Cognitive Linguistics* 28 (1): 45–76.

Lou, A. 2017. "Viewpoint and Image Schemas in Multimodal Political Discourse." *Journal of Pragmatics* 122: 77–90.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal About Thought*. Chicago; London: University of Chicago press.

McNeill, D. 2016. *Why We Gesture: The Surprising Role of Hand Movements in Communication*. Cambridge: Cambridge University Press.

Melinger, A., and S. Kita. 2007. "Conceptualisation Load Triggers Gesture Production." *Language & Cognitive Processes* 22 (4): 473–500.

Mittelberg, I. 2017. "Experiencing and Construing Spatial Artifacts from Within: Simulated Artifact Immersion as a Multimodal Viewpoint Strategy." *Cognitive Linguistics* 28 (3): 381.

Mittelberg, I. 2018. "Gestures as Image Schemas and Force Gestalts: A Dynamic Systems Approach Augmented With Motion-Capture Data Analyses." *Cognitive Semiotics* 11 (1): 20180002.

Nijk, A. A. 2019. "Bridging the Gap Between the Near and the Far: Displacement and Representation." *Cognitive Linguistics* 30 (2): 327.

Oakley, T. 2005. "Force-Dynamic Dimensions of Rhetorical Effect." In *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, edited by B. Hampe, and J. E. Grady, 443–73. Berlin; New York: Mouton de Gruyter.

Özyürek, A., R. M. Willems, S. Kita, and H Peter. 2007. "On-line Integration of Semantic Information from Speech and Gesture: Insights from Event-Related Brain Potentials." *Journal of Cognitive Neuroscience* 19 (4): 605–16.

Parrill, F. 2009. "Dual Viewpoint Gestures." *Gesture* 9 (3): 271–89.

Parrill, F. 2012. "Interactions Between Discourse Status and Viewpoint in Co-Speech Gesture." In *Viewpoint in Language: A Multimodal Perspective*, 97–112.

Pascual, E. 2014. *Fictive Interaction: The Conversation Frame in Thought, Language, and Discourse*. Amsterdam; Philadelphia: John Benjamins.

Pleshakova, A. 2018. "Cognitive Approaches: Media, Mind, And Culture." In *The Routledge Handbook of Language and Media*, edited by C. Cotter, and D. Perrin, 77–93. London and New York: Routledge.

Rogers, W. T. 1978. 'The Contribution of Kinesic Illustrators Toward the Comprehension of Verbal Behavior Within Utterances." *Human Communication Research* 5 (1): 54–62.

Sekine, K., and S. Kita. 2017. "The Listener Automatically Uses Spatial Story Representations from the Speaker's Cohesive Gestures When Processing Subsequent Sentences Without Gestures." *Acta Psychologica* 179: 89–95.

Steen, F., C. Cristobal Pagán, H. Anders, J. Joo, I. Olza, P. Anna, S. Ray, U. Peter, J. Valenzuela, J. Woźny, and M. Turner. 2018. "Toward an Infrastructure for Data-driven Multimodal Communication Research." *Linguistics Vanguard: A Multimodal Journal for the Language Sciences* 1, article number 20170041, https://doi.org/10.1515/lingvan-2017-0041.

Sweetser, E. 2012. "Introduction: Viewpoint and Perspective in Language and Gesture, from the Ground Down." In *Viewpoint in Language: A Multimodal Perspective*, edited by. B. Dancygier, and E. Sweetser, 1–24. Cambridge: Cambridge University Press.

Sweetser, E. 2017. "Metaphor and Metonymy in Advertising: Building Viewpoint in Multimodal Multi-Space Blends." *Journal of Pragmatics* 122: 65–76.

Talmy, L. 1988. "Force Dynamics in Language and Cognition." *Cognitive Science* 12 (1): 49–100.

Talmy, L. 2000. *Toward a Cognitive Semantics*. Cambridge, Mass: MIT Press.

Talmy, L. 2018. *The Targeting System of Language*. Cambridge, Massachusetts: MIT Press.

Taylor, H. A., and B. Tversky. 1996. "Perspective in Spatial Descriptions." *Journal of Memory and Language* 35 (3): 371–91.

Thomas, F.-N., and M. Turner. 2011. *Clear and Simple as the Truth: Writing Classic Prose*. Princeton, NJ: Princeton University Press.

Tseronis, A., and C. Forceville, eds. 2017. *Multimodal Argumentation and Rhetoric in Media Genres*. Amsterdam; Philadelphia: John Benjamins.

Turner, M. 2014. *The Origin of Ideas: Blending, Creativity, and the Human Spark*. USA: OUP.

Turner, M., M. Avelar, and M. Mendes de Oliveira. 2019. "Blended Classic Joint Attention and Multimodal Deixis." *Signo, Santa Cruz do Sul* 44 (79): 3–9.

Uhrig, P. 2018. "NewsScape and the Distributed Little Red Hen Lab – A Digital Infrastructure for the Large-Scale Analysis of TV Broadcasts." In *Anglistentag 2017 in Regensburg: Proceedings. Proceedings of the Conference of the German Association of University Teachers of English*, edited by A.-J. Zwierlein, J. Petzold, K. Böhm, and M. Decker. Trier: Wissenschaftlicher Verlag Trier.

Vandelanotte, L. 2017. "Viewpoint." In *The Cambridge Handbook of Cognitive Linguistics*, edited by B. Dancygier, 157–71. Cambridge: Cambridge University Press.

Vandelanotte, L., and B. Dancygier. 2017. "Multimodal Artefacts and the Texture of Viewpoint." *Journal of Pragmatics* 122: 106.

Wehling, E. 2017. "Discourse Management Gestures." *Gesture* 16 (2): 245–76.

Xiang, M., and E. Pascual. 2016. "Debate with Zhuangzi: Expository Questions as Fictive Interaction Blends in Ancient Chinese Philosophy." *Pragmatics* 26.1: 137–62.

Гришина, Елена Александровна. 2017. *Русская жестикуляция с лингвистической точки зрения: корпусные исследования*. Москва: Языки славянской культуры.