

University of Groningen

## The clarity and correctness of visualized thrust actions

Van der sluis, Ielka; Matoušková, Gabriela; Niemeier, Hannah; Popp, Sophia; Carstens, Josephine

*Published in:*  
Visual Communication

*DOI:*  
[10.1177/14703572221135240](https://doi.org/10.1177/14703572221135240)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Version created as part of publication process; publisher's layout; not normally made publicly available

*Publication date:*  
2022

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Van der sluis, I., Matoušková, G., Niemeier, H., Popp, S., & Carstens, J. (2022). The clarity and correctness of visualized thrust actions: a description and insights from users and experts. *Visual Communication*. Advance online publication. <https://doi.org/10.1177/14703572221135240>

### Copyright

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

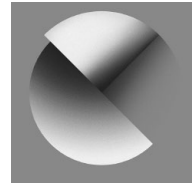
The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# The clarity and correctness of visualized thrust actions: a description and insights from users and experts



IELKA VAN DER SLUIS , GABRIELA MATOUŠKOVÁ,  
HANNAH NIEMEIER, SOPHIA POPP  
and JOSEPHINE CARSTENS  
University of Groningen, The Netherlands

## ABSTRACT

This article presents three studies that evaluate the effectiveness of instructional pictures that visualize Heimlich maneuver thrusts. Firstly, a corpus study is used to describe a collection of 30 pictures employing a model in which the angle, perspective, body and hand positions of the first-aid helper and the victim, the thrust action and its depicted results are analysed. Results show a large variation in the visualizations of the same action. Secondly, 56 potential users of the pictures filled out a questionnaire to evaluate the clarity and correctness with which the action is displayed in five representative pictures from the corpus with Heimlich's description of the thrust action as a reference. Results showed that the pictures were considered far from perfect. In particular, the clarity of the visualization of the helper applying pressure on the abdomen of the victim was considered the least clear and incorrect. Thirdly, six certified first-aid instructors gave their expert opinions on the five pictures. Their comments concentrate on the location of the hands which are depicted dangerously high on the front of the victim. These studies advance the annotation and evaluation methodologies for the investigation of visualized actions. Results serve authoring and analysis of multimodal instructions.

## KEYWORDS

annotation model • expert views • health communication • multimodal instructions • user study • visualized actions

Visual Communication 2022

Vol. 0(0) 1–26

© The Author(s) 2022



Article reuse guidelines: [sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI 10.1177/14703572221135240

 SAGE

The best rescue technique in a situation in which someone is choking is the Heimlich maneuver (HM). Heimlich (1975) presents the instruction for the HM procedure using the five actions reformatted on separate lines below:

Stand behind the victim and

wrap your arms around his [sic, the victim's] waist.

Grasp your fist with your other hand and

place the thumb side of your fist against the victim's abdomen, slightly above the navel and below the rib cage.

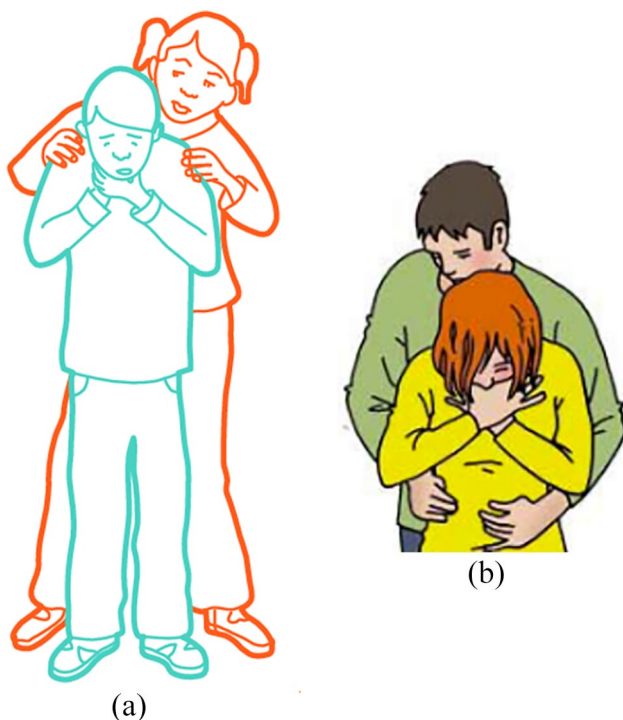
Press your fist into the victim's abdomen with a quick upward thrust.

An instructive text about the HM, such as first-aid instructions on posters, verbalizes a sequence of actions that should be carried out to perform a procedural task. Apart from the actions, verbal and pictorial presentations of the procedure may include control information that specifies, for example, the manner in which a particular action should be executed (with a quick upward thrust), the conditions that should apply (a person is choking and their life is threatened by a windpipe obstruction), the situational context (choking is signalled by an inability to speak, cough or breathe), advice and warnings to avoid undesirable effects (in the case of obesity or late pregnancy, apply chest thrusts instead of abdomen thrusts) (Van der Sluis et al., 2016a, 2017).

A quick search on the internet shows that HM instructions appear in many forms and very often contain more text than what is included in Heimlich's description presented above. Discovery of the action structure in these 'extended' procedural instructions requires proper processing of mainly the verbs in the instructions (Steehouder and Van der Meij, 2005; Van der Sluis et al., 2018, 2022; Zhang et al., 2012;). Actions in instructions most often occur in imperatives (Piwek, 2001[1998]; Steehouder et al., 2000), but can also include gerunds (e.g. to specify manner, as in 'standing behind the victim'). Alternative actions and contingencies are verbalized using modal verbs (e.g. 'Small children can just be turned on their side facing downward'), negations (e.g. 'There should be no pressure on the chest that can make breathing difficult'), or conditionals (e.g. 'If the victim is pregnant').

## **Visualized instructions**

In general, as also discussed in Clark (2016) and Pustejovsky (2018), depiction aids verbal communication. Similarly, although a verbalized instruction may be perceived as comprehensive, many studies have shown that users of instructive texts benefit from accompanying pictures that visualize the procedure (Houts et al., 2006; Katz et al., 2006). For instance, in the HM example, a verbalized instruction may tell the reader to stand behind the victim, but



**Figure 1.** (a) Stand behind the victim; (b) Wrap your arms around his (the victim's) waist.

Sources: (a) Het Oranje Kruis Jeugdboekje (tweede druk, 2017); (b) <http://leerjaar2-ehbo.jouw-web.nl/> Reproduced with permission where needed.

an accompanying picture could show the reader exactly how to do this in terms of, for example, proximity and orientation. The affordance offered via the visualization of the action adds to the comprehensibility of the instruction (Glenberg and Robertson, 1999). Comprehension in the case of HM is of great importance because an incorrectly performed thrust can cause severe injuries, such as breaking the victim's breastbone or piercing the victim's lungs (Koop, 1985).

Illustrated instructional texts that present the HM contain one or more pictures. Each picture visualizes an individual action from the sequence of HM actions in progress. These visualizations very often also show the results of the actions that were performed earlier in the sequence. For instance, Figures 1a and 1b both show the result of the Stand action in which the helper has positioned himself behind the victim. In addition, Figure 1b shows the process in which the helper wraps his arms around the victim.

To further support the users of these instructions, the included pictures may be enhanced with schematic forms such as lines or arrows with

context-dependent meanings (Tversky, 2014). For instance, lines drawn close to a depicted person may indicate that the body is moving, arrows may point at a location or an object, or indicate a direction (Heiser and Tversky, 2006), and colour blobs may attract attention to a particular location and offer a symbolic meaning to the situation portrayed (Tversky et al., 2000). The comprehensibility of schematic forms, however, seems much dependent on demographic factors. For instance, for poorly educated South Africans, the use of motion indicating arrows and lines was found to be less effective to convey motion than the use of hands and body gestures (Carstens et al., 2006; Hoogwegt et al., 2009).

The diversity of HM instructions accessible via the internet also displays a variety of pictures that use the extensive potential of visual language to depict how the first-aid technique should be applied. Some examples included in the study presented in this article are displayed in Figure 2. This variety, especially in situations of life and death, raises questions like: How should authors visualize the actions? and Which visualizations are most effective?



**Figure 2.** Illustrations from left to right: (1) Mayo Clinic; (2) leereanimeren.nl; (3) Het Oranje Kruisboekje, 2011; (4) Het Oranje Kruisboekje, 2016; (5) Radiolab<sup>1</sup>.

### **Action identification, description and evaluation**

Arguably, the evaluation of singular visualizations only makes sense when it is known how the content is depicted exactly. In our pilot studies (Van der Sluis et al., 2016b) we found that participants interpret and value pictures that accompany instructional texts very differently. In order to understand the meaning of effectiveness in the evaluation of pictures (i.e. to comprehend why particular visualizations are more effective than others), we advocate conducting corpus studies to describe and analyse the concept of investigation. Conceivably, the description of the various aspects used to visualize a particular action will allow the specification of the attributes that make a picture effective. There are numerous theories and methods proposed to describe visual content on its own as well as in multimodal contexts (Bateman, 2008; Bateman

et al., 2017; Kress and Van Leeuwen, 2020; Rose, 2016; Smith et al., 2004). Various reviews and more concrete studies have also explored the schemes chosen by authors to present actions in static pictures (De Souza, 2008; De Souza and Dyson, 2008; Parkinson et al., 2020; Thomson, 2021; Tufte, 1997). In order to approach animated demonstration of actions in static visualizations, authors often use sequences of pictures to show the results of the various actions involved in the procedure. Other approaches, as discussed in Doran (2019), visualize movement in one picture by depicting an object in a number of stages in a sequence where the stages are connected with arrows. By analysing multiple pictures that visualize the HM thrust action as well as the result of some of the actions performed earlier in the sequence leading up to the thrust action, this article seeks to complement current methods to investigate visual content.

In addition, by specifying the aspects important to the visualization of actions, the proposed analysis may also be beneficial for current efforts to automatically caption pictures. Nowadays, it is possible to recognize visualized objects automatically. In the case of the HM pictures exemplified in Figure 1, the two depicted persons can probably be identified due to rapidly advancing techniques that employ (pretrained) CNN models (Lu et al., 2019, 2020; Tan and Bansal, 2019) for learning on large collections of pictures (Lin et al., 2014; Sharma et al., 2018). The identification of depicted actions – and thereby the inferences about the contextual roles of the objects or persons depicted – is still challenging (Birmingham et al., 2018; Jensen and Lulla, 2008; Karpathy and Fei-Fei, 2015; Socher et al., 2014; Stanfield and Zwaan, 2001). However, fusing techniques that combine vision and language modes seem promising (Bisk et al., 2020; Chen et al., 2020; Li et al., 2019; Su et al., 2020). Potentially, identification of the actions in the text can inform the identification of these actions (and/or their results) in the accompanying pictures (Ghanimifard and Dobnik, 2017; Hürlimann and Bos, 2016; Parcalabescu et al., 2021; Vedantam et al., 2015). For instance, in Silberer and Pinkal (2018), it is shown that spatial relations of the objects in a visual scene can be used together with syntax-based techniques to develop accurate models for image description generation. By investigating the characteristics of the visualization of actions in multimodal documents, such as the instructions from which the HM thrust visualizations used in this study are derived, we hope to contribute to the automatic identification of visualized actions.

Finally, the work presented in this article shares insights in evaluation methodologies set up to investigate the perception of visualized actions. As stated previously, visualizations can be described in many ways (e.g. perspective, drawing style, colouring, graphical enhancements), but the effectiveness of these aspects is likely to depend on the characteristics of the viewer. In the domain of first-aid communication, authors of multimodal instructions are relating to first-aid instructors as well as the general public. These audiences use the instructions for different purposes, for example,

teaching how to perform the HM and using the HM to save a victim's life, respectively. Therefore, perception studies that address the characteristics of the viewers (see Wilson and Landon-Hays, 2016; Zoss et al., 2010) are likely to result in a variety of pictorial aspects that add to the effectiveness of a visualized action.

### **Studying thrust visualizations**

The three studies presented in this article serve to illustrate the iterative development of a model to annotate visualized HM thrusts and to evaluate their effectiveness. Different evaluation cycles are used to investigate the effectiveness and completeness of the descriptive categories. The studies aim to answer the research question: *How should the thrust action of the HM be visualized in terms of clarity and correctness?*

Firstly, a corpus study was conducted to describe pictures taken from illustrated verbal instructions that were written in Dutch or English. The studied corpus only includes pictures that visualize the thrust action in progress. As a starting point, the corpus was analysed employing a model with attributes to describe the composition of the pictures as well as content-related aspects that describe the visualized people. The corpus study was used to determine a balanced subset of pictures from the corpus for evaluation purposes.

Subsequently, two evaluation studies were conducted to investigate the perceived quality of five HM thrust visualizations as well as to further improve the annotation model. The studies target different audiences. The evaluation studies offer insight into aspects that the participants noticed and considered beneficial or harmful to the clarity and correctness of the HM thrust visualizations and that are therefore worthwhile analysing. The studies also show considerable differences between the two different user groups in terms of the visual aspects that they consider to be important.

### **CORPUS STUDY**

The corpus study presented below was set up to describe the way in which the thrust action is visualized. An annotation model was developed to find out more about the visual features used to depict the HM thrust action. The pictures are annotated and analysed accordingly.

#### **Data set**

The corpus study uses a subset of the PAT corpus<sup>2</sup>, a collection of over 300 first-aid instructions that include pictures and text (Van der Sluis et al., 2016a), complemented with new instances retrieved from the internet. The final corpus of the project presented in this article comprises 30 pictures that

**Table 1.** Annotation model for the HM thrust action.

Attribute	Values
Perspective	Middle shot, Long shot
Angle	Display from the Front, Left or Right
Victim's body	Standing, Sitting
Victim's hands	Straight down, One hand on throat, Both hands on throat
Helper's body	Bent forward, Straight
Helper's hands	Wrist grab, On top of each other, Not visible
Thrust action	Lines around the body, Shirt wrinkles, Arrow, Colour, None
Result	Object, Arrow, Line, None

stem from different instructions that are meant to instruct the general public on how to execute the HM thrust action with the ultimate goal of saving life in a life-threatening choking situation. Given the variety of thrust visualizations discussed in the introduction of this article, the corpus is balanced in that it contains 15 drawings and 15 photographs. The drawings as well as the photographs differ in many ways, for example, drawing styles, colouring and backgrounds. As a result, the annotation model resulting from this corpus study will be applicable to describe the HM thrust action independent of the visualization. The corpus selection criteria were as follows:

- Victim and helper are both visible.
- Helper is standing behind the victim facing the victim's back.
- Victim is standing or sitting in front of the helper facing away from the helper.
- Helper is performing the thrust action with arms placed around the victim's waist and hands touching in front of the victim's waist.

## Annotation model

Table 1 presents the annotation model in terms of attributes and values. The corpus includes visualizations of victims that are either standing or sitting. Because the position of the victim is likely to affect the body position of the helper, some attributes were used to describe the pictures. For instance, the attribute Perspective specifies if the picture presents the scene using a Long shot in which helper and victim are fully visible, or a Middle shot where the helper and victim are visible from the waist up. The Angle, described from the perspective of the viewer, either shows a frontal view or a view from the side, where the helper and victim are both facing towards the Left or towards the



Right. The model includes a description of the body posture of the victim (i.e. Standing, Sitting) and of the helper (i.e. Bent forward, Straight).

Other attributes were included to describe the visualized performance of the thrust action and the scene in more detail. The model includes a description of the hand positions of the victim (i.e. Straight down, One hand on throat, Both hands on throat) and of the helper (i.e. Wrist grab, On top of each other, Not visible). The pictures in the corpus also include additional features to amplify the visualization of the thrust action, which are described as Lines around the body (e.g. lines aligned with the helper's back to indicate movement), Shirt wrinkles, Arrows and coloured areas (Colour) specifying the compression performed by the helper at the front of the victim's body. The Result of a successful thrust action, i.e. visualized as an obstacle ejected from the victim's mouth, is described in terms of Object, Arrow and Line.

The annotation model was developed in an iterative process. Five annotators (four of whom attended a first-aid course that included HM) prepared a preliminary model by studying and analysing a test subset of the corpus ( $n = 8$ ) together. Subsequently, the annotators individually applied the preliminary model to annotate the validation subset of the corpus ( $n = 22$ ) employing the VGG Image Annotator (VIA)<sup>3</sup>. The agreement between the five resulting silver standard annotations was calculated with Fleiss Kappa (Fleiss, 1971; Fleiss et al., 2003) and valued according to the directions by Landis and Koch (1977). The agreement scores, together with the notes of the annotators, were used to find and resolve errors in the annotations and to improve the preliminary annotation model. For instance, for the attribute Helper's hands, the value On top of each other was introduced because the visibility of the helper's fingers was not always sufficient to determine how the hands were connected (e.g. in some cases the fingers seemed interlaced). Some values of the Thrust action turned out to include multiple features and the values were specified in more detail accordingly. Finally, the whole corpus ( $N = 30$ ) was annotated with the improved annotation model (Table 1) and double-checked by two annotators.

## RESULTS

Table 2 presents the corpus in terms of the attributes specified in the annotation model for 15 photographs and 15 drawings separately. Most of the pictures present a Middle shot (83.3%;  $n = 25$ ). The action is usually depicted using a Left or Right Angle (86.7%;  $n = 26$ ) as opposed to a view from the Front (13.3%;  $n = 4$ ). In 90 percent ( $n = 27$ ) of the pictures, the victim is Standing instead of Sitting (10%;  $n = 3$ ). The position of the victim's hands varies: in 43.3 percent ( $n = 13$ ) of the pictures the victim's hands are hanging down, while in 56.7 percent ( $n = 17$ ) one or both hands are located at the victim's throat.

The helper's body is mostly visualized in an upright position (86.7%;  $n = 26$ ) instead of Bent forward (13.3%;  $n = 4$ ). The helper's hands are mostly

**Table 2.** Frequencies and percentages for the attributes that describe the Perspective, Angle, Victim's and Helper's bodies and hands (with percentages calculated per attribute), as well as the number of pictures that include additional features to visualize the Thrust action and the Result of this action (where the percentages are calculated for the whole corpus as the values are not mutually exclusive).<sup>a</sup>

Attribute	Value	Photo (n = 15) (%)	Drawing (n = 15) (%)	Total (N = 30) (%)
Perspective	Middle shot	13 (86.7)	12 (80)	25 (83.3)
	Long shot	2 (13.3)	3 (20.0)	5 (16.7)
Angle	Front	2 (13.3)	2 (13.3)	4 (13.3)
	Left	8 (53.3)	6 (40.0)	14 (46.7)
	Right	5 (33.3)	7 (46.7)	12 (40.0)
Victim's body	Standing	13 (86.7)	14 (93.3)	27 (90.0)
	Sitting	2 (13.3)	1 (6.7)	3 (10.0)
Victim's hands	Straight down	7 (46.7)	6 (40.0)	13 (43.3)
	One hand on throat	4 (26.7)	2 (13.3)	6 (20.0)
	Both hands on throat	4 (26.7)	7 (46.7)	11 (36.7)
Helper's body	Bent forward	2 (13.3)	2 (13.3)	4 (13.3)
	Straight	13 (86.7)	13 (86.7)	26 (86.7)
Helper's hands	Wrist grab	2 (13.3)	6 (40)	8 (26.7)
	On top of each other	10 (66.7)	8 (53.3)	18 (60.0)
	Not visible	3 (20.0)	1 (6.7)	4 (13.3)
Thrust action	Lines around the body	0 (0)	1 (6.7)	1 (3.3)
	Shirt wrinkles	11 (73.3)	8 (53.3)	19 (63.3)
	Arrow(s)	0 (0)	8 (53.3)	8 (26.7)
	Colour	1 (6.7)	1 (6.7)	2 (6.7)
	None	4 (26.7)	4 (26.7)	8 (26.7)
Result	Object	1 (6.7)	3 (20.0)	4 (13.3)
	Arrow(s)	0 (0)	8 (53.3)	8 (26.7)
	Line(s)	0 (0)	1 (6.7)	1 (3.3)
	None	14 (93.3)	10 (66.7)	24 (80.0)

<sup>a</sup>The ratio is offered via the frequencies and the total number of pictures in the set. The percentages, of course, give a coarse indication of the accuracy of the measurements, because they indicate parts of pictures instead of the pictures as a whole.

placed On top of each other (60%;  $n = 18$ ), or shown in a Wrist grab (26.7%;  $n = 8$ ); in 13.3 percent ( $n = 4$ ) of the pictures the position is Not visible. The Thrust action is amplified in some of the pictures in the corpus by using additional features ( $n = 30$ ). In the photographs, mainly Shirt wrinkles were present ( $n = 11$ ; 73.3%). In the drawings not only Shirt Wrinkles ( $n = 8$ ; 53.3%) but also Arrows ( $n = 8$ ; 53.3%) were frequently added. Often more than one additional feature was present in one picture. Apart from the Shirt

wrinkles, the other additional features (i.e. Lines around the body and Colour) were mainly used in the drawings. In eight pictures (26.7%; 4 photos and 4 drawings), there were no additional features included. The visualization of the result of the HM, where the item that caused the airway blockage comes out of the victim's mouth, is depicted using additional features in six pictures (20%; 1 photo and 5 drawings). In these cases, different features (i.e. Object, Arrow, Line) are often combined.

## **USER JUDGEMENTS**

The study presented below was set up to evaluate the clarity and correctness of five representative pictures from the corpus. As the pictures in this project stem from instructions that are targeting lay persons, the authors distributed a questionnaire via their LinkedIn networks.

### **Participants**

To evaluate five pictures in which the thrust action is visualized, 56 participants (40 females and 16 males), aged between 20 and 68 ( $M = 33.3$ ;  $SD = 12.5$ ) filled out a Qualtrics questionnaire. The participants mostly resided in the Netherlands ( $n = 39$ ) and Germany ( $n = 11$ ) which matches the origin of the pictures used in the corpus. Of the total number of participants, 43 said that they knew what the HM is, 6 were unfamiliar with the term and 7 were unsure about their knowledge of HM; 32 participants had no experience with executing HM, 24 had experience with HM and chose one of the following options: they had either executed HM themselves ( $n = 4$ ), seen somebody else perform HM ( $n = 5$ ), or both ( $n = 1$ ), or they had taken a first-aid course that included HM ( $n = 14$ ). Although participants' knowledge of the HM is lower than one would hope for in general, it was expected that the participants would not have any knowledge or experience with the HM at all and that all the participants would need time to familiarize themselves with the procedure to make proper judgements about the pictures to be evaluated in this study. Therefore, all participants were offered time to study the description and use of the HM before answering any of the evaluative questions.

### **Materials**

The five pictures evaluated in this study are presented in Figure 2. Table 3 presents a description of the pictures in terms of the corpus annotation. The pictures (2 photographs, 3 drawings) were chosen to reflect the variation in the corpus. Each of the pictures was presented, accompanied by the HM description (see Figure 3) and two questions about the visualization of the thrust action in the pictures.

Note that the HM description was presented as consisting of four steps, where the first step included the first two actions in the HM pro-

**Table 3.** Description of the figures used in the evaluation study according to the annotation model.

Attribute	Picture 1	Picture 2	Picture 3	Picture 4	Picture 5
Picture Type	Drawing	Drawing	Photo	Photo	Drawing
Perspective	Middle shot	Long shot	Middle shot	Middle shot	Middle shot
Angle	Right	Left	Left	Right	Right
Victim's body	Standing	Standing	Standing	Standing	Standing
Victim's hands	Both hands on throat	Straight down	Both hands on throat	Straight down	Straight down
Helper's body	Straight	Straight	Straight	Straight	Straight
Helper's hands	On top of each other	Not visible	On top of each other	On top of each other	Wrist grab
Thrust action	Shirt wrinkles, Arrow	Lines, Colour	Shirt wrinkles	None	Shirt wrinkles, Arrow
Result	None	None	None	None	Object, Arrow, Line

Please have a look at the presented picture and answer the questions below.

For your convenience, we make the Heimlich maneuver text instruction available to you again:

- *Stand behind the victim and wrap your arms around his (the victim's) waist.*
- *Grasp your fist with your other hand.*
- *Place the thumb side of your fist against the victim's abdomen, slightly above the navel and below the rib cage.*
- *Press your fist into the victim's abdomen with a quick upward **thrust**.*



**Figure 3.** Presentation of HM instruction text.

cedure: (1) as also presented in the introduction of this article (i.e. Stand behind the victim and Wrap your arms around his [the victim's] waist). As the pictures stem from a corpus that includes only pictures in which the helper performed the thrust action, the result of the actions in this

Indicate in how far you agree that the aspects of the Heimlich maneuver are **clearly visualized**:

	completely disagree	disagree	slightly disagree	neutral	slightly agree	agree	completely agree
1. The location where the helper's hands should be placed on the victim's body.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. The way in which the helper should connect his/her hands.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. The pressure with which the thrust should be performed.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. The direction in which the pressure should be conducted (in and upwards).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Figure 4.** Presentation of Clarity question.

first step were entailed. The sentence was therefore not split but kept as Heimlich has presented it. In contrast, the second sentence of the instruction was split into two steps separating the two actions in it (i.e. Grasp your fist with your other hand and Place the thumb side of your fist against the victim's abdomen, slightly above the navel and below the rib cage). Since the corpus analysis had shown that the description of the helper's hands was not straightforward and since the annotators had difficulty judging the location of the hands, it was decided to investigate the judgements of the participants in the user study on these actions separately.

Firstly, the participants were asked to indicate how far they agreed that particular aspects of the thrust action were clearly visualized (see Figure 4). In this question we concentrated on the helper as described in the corpus study: Location of the helper's hands, Connection of the helper's hands, Pressure indication and Direction in which the pressure is applied, where we split the Thrust action attribute in terms of the application of pressure and the direction of the applied pressure. The Clarity question was presented using a seven-point Likert scale (1 = completely disagree to 7 = completely agree).

Secondly, participants were asked to indicate if the sub-actions included in the HM description (i.e. Stand & Wrap, Grasp, Place, Press) were incorrectly visualized, or explicitly indicate that all sub-actions were correctly visualized. The Correctness question was presented using check boxes, one for each of the sub-actions that was incorrectly visualized and one to indicate that all sub-actions were correctly visualized (see Figure 5).

Indicate if you think that particular aspects are **not correctly visualized** in respect to the Heimlich maneuver text instruction.

- 1. Stand behind the victim and wrap your arms around his (the victim's) waist.
- 2. Grasp your fist with your other hand.
- 3. Place the thumb side of your fist against the victim's abdomen, slightly above the navel and below the rib cage.
- 4. Press your fist into the victim's abdomen with a quick upward thrust.
- 5. All the mentioned aspects (1-4) are correctly visualized.

**Figure 5.** Presentation of Correctness question.

## Procedure

Table 4 presents the results collected with the evaluation study. The results did not differ significantly between the participants with and without first-aid experience; in terms of clarity and correctness, the participants without experience were slightly more critical with respect to Pictures 2, 3 and 4.

*Clarity:* The Location of the helper's hands was generally perceived as clear. The Connection of the helper's hands was perceived as unclear in Picture 2 ( $M = 2.09$ ;  $SD = 1.20$ ), Pictures 1 and 4 scored best in terms of the clarity of the connection; the view of the helper's hands, in the drawing and photograph, respectively, is slightly more frontal compared to Picture 3, and not obstructed by arrows (cf. Picture 5) or by the victim's arms (cf. Picture 2). Overall, the Pressure conducted by the helper's hands was perceived as the least clear, where Pictures 2, 3 and 4 are unclear and Pictures 1 and 5, in which the pressure is indicated with additional arrows, scored almost neutral. The Direction of the thrust action was also perceived as unclear in Pictures 2, 3 and 4. Overall, in terms of clarity Pictures 1 and Picture 5 scored best.

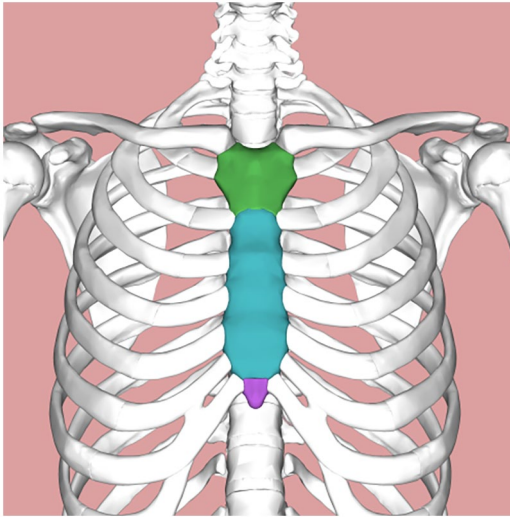
*Correctness:* In general, as we expected, most of the 56 participants who took part in our study found that the helper is Standing and Wrapping correctly in the five pictures. However, the helper Grasps the fist of his other hand incorrectly in Pictures 2 ( $n = 41$ ; 73.2%) and 3 ( $n = 21$ ; 37.2%). Less than 20 percent determined the helper's grasp to be incorrect in Pictures 1 ( $n = 5$ ; 8.9%), 4 ( $n = 6$ ; 10.7%) and 5 ( $n = 10$ ; 17.9%), where the view of the helper's hands is slightly more frontal compared to Pictures 2 and 3. More than half of the participants said that the helper does not correctly apply Pressure in Pictures 2 ( $n = 42$ ; 75%), 3 ( $n = 45$ ; 80.4%) and 4 ( $n = 45$ ; 80.4%), where no arrows are included. Overall, Picture 1 with an L-shaped arrow and the victim with his hands placed at his throat scored best, 25 participants (44.6%) said that the helper performed the actions correctly.

## Expert views

Six certified first-aid instructors (2 male, 4 female) volunteered to comment on the HM pictures presented in Figure 2. They were asked for their expert view; no further instructions were given. Results consist of 30 freely written

**Table 4.** Means and standard deviations for Clarity on a 7-point scale (1 = unclear to 7 = clear) plus frequencies and percentages for Correctness for the sub-actions included in the HM description.<sup>4</sup>

Question	Picture 1	Picture 2	Picture 3	Picture 4	Picture 5	
Clarity	Location	6.00 (0.79)	4.84 (1.68)	5.66 (1)	5.82 (0.97)	5.18 (1.42)
	Connection	5.66 (1.23)	2.09 (1.20)	3.96 (1.53)	5.36 (1.49)	4.59 (1.58)
	Pressure	3.88 (1.75)	2.71 (1.47)	2.91 (1.63)	2.54 (1.45)	3.88 (1.89)
	Direction	5.80 (1.20)	2.93 (1.63)	2.95 (1.62)	2.86 (1.67)	5.79 (1.16)
	Stand & Wrap	0 (0%)	0 (0%)	1 (1.79%)	1 (1.79%)	0 (0%)
Correctness	Grasp	5 (8.9%)	41 (73.2%)	21 (37.5%)	6 (10.7%)	10 (17.9%)
	Place	15 (26.9%)	42 (75%)	26 (46.4%)	18 (32.1%)	32 (57.1%)
	Press	17 (30.4%)	42 (75%)	45 (80.4%)	45 (80.4%)	19 (33.9%)
	All correct	25 (44.6%)	4 (7.1%)	5 (8.9%)	5 (8.9%)	14 (25%)



**Figure 6.** Parts of the sternum: manubrium (green), body (blue), xiphoid process (purple). Available at: <https://en.wikipedia.org/wiki/> licensed under CC BY-SA

text comments. In our analysis of these comments, a number of aspects of the visualized HM thrust action emerged and are addressed below.

*Helper's hands position:* In 21 cases, the experts said that the helper placed his hands too high on the victim's abdomen (Picture 1,  $n = 4$ ; Picture 2,  $n = 5$ ; Picture 3,  $n = 2$ ; Picture 4,  $n = 4$ ; Picture 5,  $n = 6$ ). Two experts said that the helper's hands are placed too high in all five pictures. In three cases (Picture 1,  $n = 2$ ; Picture 3,  $n = 1$ ) this aspect of the visualization was not commented on. Three experts find the helper's hands in Picture 3 correctly placed. One expert explicitly said that the little finger of the helper in Picture 3 is more or less located on the victim's belly button. The same expert remarked that maybe not on first sight, but when looking more closely at Picture 4, the helper's fist is placed in the correct location. This expert acknowledged that, in drawings, the helper's fist is often placed too high. One expert explained that, by placing the hands too high, the helper risks fracturing the bottom part (xiphoid process, Figure 6) of the sternum. Two experts said that, in general, if the helper placed his arms on the victim's hips, the placement of the hands would be correct and result in a proper oblique pull.

*Body positions:* Two experts said that the helper's position, a bit on the side at the back of the victim, is clearly visible in Pictures 2 and 4. One expert also appreciated the forward bend of the victim's body in Pictures 2 and 4, and also in Picture 5.

*Additional features:* The additional features in the pictures were valued by two of the six experts. Both said that the L-shaped pull is clearly marked by the arrow in Picture 1. One expert found the lines at the back of the helper in Picture 2 unclear. One expert found the lines drawn in Picture 5 unsuitable



and that they blurred the helper's hands, whereas one expert appreciated the visualized result in Picture 5.

*Realism:* One expert mentioned that she had never seen a victim grabbing his throat in the manner shown in Pictures 1 and 3. The expert hypothesized that this visualization may have been motivated by reasons of clarity. One expert found Pictures 4 and 5 the most realistic, although she acknowledged that the location of the helper's hands was incorrect in both pictures. Two experts mentioned the blue face of the victim in Picture 3 which, while indicating the danger of the situation, may be distracting, according to one expert. In contrast, another expert found that the victim in Picture 4 could have looked a bit more anxious. This expert also said that sunglasses as worn by the helper in Picture 5 are not recommended when providing first-aid because they may be considered intimidating.

*Clarity:* One expert said that Picture 5 was the most clear because it shows what the helper should do even though the location of the helper's hands is incorrect. Two experts found that Picture 3 presents the proper aid and two experts said the same of Picture 4.

Overall, the experts offered clearly motivated observations with which the attributes in the annotation model can be further developed, not only to describe the thrust action, but also to address the diversity of the corpus. In terms of correctness and clarity, it is key that the most important aspect of the thrust visualization (i.e. the location of the helper's hands) is not obscured by any graphical enhancement (e.g. arrows or lines) nor by the visualization of the helper and victim themselves (e.g. the victim's arms should not be obstructing the hands of the helper). In terms of completeness or coverage of the diversity of the corpus, various means can be used to emphasize the importance of particular aspects of the procedure (e.g. the direction of the thrust action) and the critical situation that potential viewers of the pictures should expect to encounter (e.g. the anxiety of the victim). The expert study was also invaluable for understanding the potential effects of details like the sunglasses, that initially were not taken into consideration in the corpus study.

## **CONCLUSION AND FUTURE WORK**

### **Discussion of findings**

In the corpus study, eight attributes were identified to describe the HM thrust action visualized in 30 different instructions. Results showed that the thrust action is most often visualized with a Middle shot from a Left or Right Angle. The victim's hands were either hanging Straight down, or the victim used one or both hands to grab his throat. It turned out that, when excluding the two pictures in which the victim was seated, the percentage of pictures in which one or both hands of the victim were located on the throat increased from 57 to 75 percent; no other differences were observed. The effect of the position of the victim's hands on the visibility of the thrust action that the helper is

performing needs further investigation. In some of the corpus pictures, the arms of the victim obstructed the view of the hands of the helper. The position of the victim's hands may contribute to the indicated danger. However, in the instructive context in which these pictures are used, the actual state of the victim may be less important than the manner in which the helper performs the action.

Especially when knowing that one can cause damage by thrusting one's thumbs into the victim's abdomen, the depiction of the helper's hands that perform the thrust action shows a surprising amount of variation. Given the description of the HM thrust action in the quote in Section 1 'Grasp your fist with your other hand' we would expect the helper's hands to be depicted in a Wrist grab, but this was the case in only eight pictures (26.7%). In 60.0 percent of the cases ( $n = 18$ ), the more general description On top of each other was chosen to describe the position of the helper's hands; no clear consensus could be reached amongst the five annotators on whether the helper's fingers were interlaced or not. In 13.3 percent of the cases ( $n = 4$ ) the helper's hands were not even visible. Curiously, none of the experts mentioned the connection of the helper's hands as a problem in the pictures. The helper's fist was mentioned explicitly by three experts. Additional features such as lines and arrows to specify the thrust action were used mainly in the drawings. The direction 'into the victim's abdomen with a quick upward thrust', i.e. in and upward, was specified with arrows in eight drawings.

The location of the helper's hands was not described in the corpus study. Although we are inclined to think that, in a number of cases, the location of the helper's hands is incorrect or indeterminable, we felt that experts were needed to ascertain if the pictures display the location correctly. This intuition was confirmed by the large differences between the users' judgements and the expert views on the location of the helper's hands. Given the results from the expert views, improvements in the annotation of pictures can be made by including the victim's body description (i.e. add Bent forward, Straight) and indications of urgency (e.g. facial expressions).

Interestingly, the participants who filled out the questionnaire considered other aspects in the visualization of the thrust action than the experts. The participants in the questionnaire seemed to concentrate on the visualization of the pressure applied by the helper, while the experts first checked the location of the helper's hands. Experts also checked the body positions of the victim and helper, and some appreciated indications of the urgency of the situation. Most experts did not comment on the additional features; perhaps these features were not noticed when included or were not missed when present. The experts' comments on the additional features that were collected rendered mixed results. In our corpus, only eight pictures include an arrow to visualize the thrust action; the quality of these arrows (e.g. L-shaped or other) needs further investigation. In our current studies on instruction comics presented on posters (Wildfeuer et al., 2022) that present the HM procedure, and in first-aid

instructions for children, these and other additional features appear often (cf. Bateman and Wildfeuer, 2014; Cohn and Maher, 2015; Forceville, 2011).

The questionnaire results showed that the drawing in Picture 1 is favoured in terms of clarity and correctness; in this picture, the arrow clearly indicates the pressure and the direction in which the pressure is applied. The added value of the arrow in Picture 1 is confirmed by two experts. Perhaps the non-experts are not so aware of the danger involved when placing the hands too high. The experts' views show mixed results, Pictures 3 and 4 (both photos) are most correct in terms of the placement of the helper's hands, while Picture 5 shows what the helper should do in the most realistic way. The preferred pictures (1, 3, 4 and 5) all use a Middle shot, but the Angle varies.

### **Guidelines for authors**

The work presented in this article concentrates on visualized HM thrust actions. To answer the question, 'How should the thrust action of the HM be visualized in terms of clarity and correctness?', we can conclude that a middle shot of the HM thrust action is most clear and correct, only if the helper's hands are placed at the correct location on the victim's abdomen, 'slightly above the navel and below the rib cage', to avoid piercing the victim's lungs with his own sternum. In addition, an L-shaped arrow, provided that it is not obstructing the view of the helper's hands, could help viewers to understand the direction of the thrust action. Obviously, for HM thrust actions and other lifesaving first-aid operations, such as placing a victim in a recovery position or resuscitating a victim, it is a matter of life and death that the pictures included in an instruction visualize the necessary actions clearly and correctly.

Authoring multimodal instructions, such as the ones from which the pictures studied in this article are derived, also includes combining text and pictures. While coherence relations between text and pictures have been categorized at various levels of granularity (Barthes, 1977; Martinec and Salway, 2005), the finer-grained coherence relations – i.e. elaborations, extensions and/or enhancements as proposed by Bateman and Wildfeuer (2014) and Halliday and Matthiessen (1985) – seem useful for application in guidelines to authors of HM and plausibly also in other instructions. For example, the text elaborates on the content of the picture if it describes the way in which the helper should hold his hands while the accompanying picture does not show the hand grasp in detail. Alternatively, a picture that shows how the windpipe obstruction is removed as a result of the thrust action extends a verbal instruction in which this result is not mentioned. And, by showing the position and stance of the helper behind a victim, a picture enhances the text in which the exact location and orientation of the helper are not specified. Corpus studies as presented in this article benefit the discovery of the aspects involved in composing multimodal instructions. Guiding authors in distributing the

information across modalities requires user studies to investigate to what extent visual and verbal content are best combined in terms of effectiveness (Van der Sluis et al., 2016b, 2017).

## **Visualization studies**

This article presented an approach for studying the visualization of a particular action. The study of the HM thrust action started with text (i.e. Heimlich's own description) and a variety of pictures that visualized the HM thrust action. Heimlich's description was analysed as containing five different actions (e.g. Stand, Wrap, Grasp, Place, Press). Pictures in procedural descriptions show results of actions that were performed earlier in the sequence of actions. Also, in the HM thrust visualizations, the results of the Stand and Wrap actions are visible. Because the thrust action was the focus of this study, the Grasp, Place and Press actions were considered most important. Differentiation between the analysis of the visualization of actions in progress and the visualization of the results of actions, especially when the actions are sequential and display temporal relations, are left for future work. Further directions for analysis include sequential actions in other multimodal documents such as instruction comics (Wildfeuer et al., 2022) instruction videos and animated pictures (e.g. static pictures with moving arrows). Currently, our studies in which we apply the same HM action structure to an illustrated instructive texts and instruction comics show that additional categories such as talking to the victim are required to address the displayed actions of the helper in comics.

The work presented in this article shows the added value of alternating corpus studies and evaluation studies. As envisioned, the annotation model turned out to be applicable to a variety of pictures (15 drawings and 15 photographs) and helped the selection of a balanced set of pictures for evaluation. The user studies that we conducted resulted in aspects of the visualizations that can be employed to further enhance the annotation model to include effective attributes for description. The participants' familiarity with and use of the visualized action clearly affected their perception and evaluation of the visualizations in terms of clarity and correctness. In future, versions of the annotation model attributes for addressing clarity, correctness, realism and graphical enhancement may turn out to be competing factors (e.g. a victim with his hands at his throat may not be realistic, but the stance offers a clear view of the helper's hands). Moreover, these factors may be perceived differently by other audiences than the mostly Dutch and German participants in our studies. Future work should therefore also include visualizations and users from other parts of the world.

## **ACKNOWLEDGEMENTS**


Our thanks to the anonymous reviewers for their very precise and useful comments on previous versions of this article. Thanks to Janina Wildfeuer and

Gisela Redeker for useful comments on the presentation of our work. Thanks to Charlotte Nolting-Hauff for her help with the corpus study. Thanks to Arie Wolters and Marianne Gruis at the K.N.V. EHBO afdeling Groningen for their help with contacting certified first-aid instructors to evaluate the pictures.

## FUNDING

The authors received no financial support for the research, authorship and publication of this article, and there is no conflict of interest.

## ORCID ID

Ielka van der Sluis  <https://orcid.org/0000-0002-0151-5044>

## NOTES

1. Sources:
  1. <https://www.mayoclinic.org/first-aid/first-aid-choking/basics/art-20056637>
  2. <http://www.leerreanimeren.nl/luchtwegobstructie.html>
  3. Het Oranje Kruis (2011). *Het Oranje Kruisboekje, De officiële handleiding voor eerste hulp*. Thieme Meulenhoff, Amersfoort. ISBN 9789006921717
  4. Het Oranje Kruis (2016). *Het Oranje Kruisboekje, De officiële handleiding voor eerste hulp*. Thieme Meulenhoff, Amersfoort. ISBN13 9789006410341
  5. <https://www.wnycstudios.org/podcasts/radiolab/articles/273532-heimlich>All reproduced with permission where needed.
2. [https://www.rug.nl/research/clcg/research/dc2017\\_/dcresearch/projects/pat/](https://www.rug.nl/research/clcg/research/dc2017_/dcresearch/projects/pat/)
3. <https://www.robots.ox.ac.uk/~vgg/software/via/>
4. For correctness of the visualized subsections, participants could select multiple answers, unless they chose *All correct*. The percentages are based on the total number of participants ( $N = 56$ ) and are presented to provide a better picture of the proportion of the results, rounded to one decimal.

## REFERENCES

- Barthes R (1977) *Image–Music–Text*. Chichester: Macmillan.
- Bateman JA (2008) *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. Berlin: Springer.
- Bateman JA and Wildfeuer J (2014) A multimodal discourse theory of visual narrative. *Journal of Pragmatics* 74: 180–208.

- Bateman JA, Wildfeuer J and Hiippala T (2017) *Multimodality: Foundations, Research and Analysis: A Problem-Oriented Introduction*. Berlin: Walter de Gruyter.
- Birmingham B, Muscat A and Belz A (2018) Adding the third dimension to spatial relation detection in 2D images. In: Krahmer E et al. (eds) *Proceedings of the 11th International Conference on Natural Language Generation*, Tilburg University, The Netherlands, 5–8 November. Association for Computational Linguistics, 146–151. DOI:10.18653/v1/w18-6517.
- Bisk Y et al. (2020) Experience grounds language. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Online, 8718–8735. DOI: 10.18653/v1/2020.emnlp-main.703.
- Carstens A, Maes A and Gangla-Birir L (2006) Understanding visuals in HIV/AIDS education in South Africa: Differences between literate and low-literate audiences. *African Journal of AIDS Research* 5(3): 221–232.
- Chen Y-C et al. (2020) Uniter: Universal image–text representation learning. *Proceedings of European Conference on Computer Vision*. Berlin: Springer, 104–120.
- Clark HH (2016) Depicting as a method of communication. *Psychological Review* 123(3): 324.
- Cohn N and Maher S (2015) The notion of the motion: The neurocognition of motion lines in visual narratives. *Brain Research* 16(1): 73–84.
- De Souza J (2008) An illustrated review of how motion is represented in static instructional graphics. In: *Proceedings STC's 55th Conference*, 1–4.
- De Souza J and Dyson M (2008) Are animated demonstrations the clearest and most comfortable way to communicate on-screen instructions? *Information Design Journal* 16(2): 107–124.
- Doran YJ (2019) Building knowledge through images in physics. *Visual Communication* 18(2): 251–277.
- Fleiss JL (1971) Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5): 378.
- Fleiss JL, Levin B and Paik MC (2003) *Statistical Methods for Rates and Proportions*. Hoboken, NJ: John Wiley.
- Forceville C (2011) Pictorial runes in Tintin and the Picaros. *Journal of Pragmatics* 43(3): 875–890.
- Ghanimifard M and Dobnik S (2017) Learning to compose spatial relations with grounded neural language models. In: *IWCS 2017 – 12th International Conference on Computational Semantics – Long Papers*.
- Glenberg AM and Robertson DA (1999) Indexical understanding of instructions. *Discourse Processes* 28(1): 1–26.
- Halliday M and Matthiessen C (1985) An introduction to functional grammar. *Australian Review of Applied Linguistics* 10(2): 163–181.

- Heimlich HJ (1975) A life-saving maneuver to prevent food-choking. *Jama* 234(4): 398–401.
- Heiser J and Tversky B (2006) Arrows in comprehending and producing mechanical diagrams. *Cognitive Science* 30(3): 581–592.
- Hoogwegt H, Maes A and Van Wijk C (2009) Suggesting motion in health communication pictures: The effect of pictorial and extra-pictorial manipulation. *Communicatio: South African Journal for Communication Theory and Research* 35(2): 276–294.
- Houts PS et al. (2006) The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling* 61(2): 173–190.
- Hürlimann M and Bos J (2016) Combining lexical and spatial knowledge to predict spatial relations between objects in images. In: *Proceedings of the 5th Workshop on Vision and Language*, 10–18.
- Jensen J and Lulla K (2008) *Introductory Digital Image Processing: A Remote Sensing Perspective*. Englewood Cliffs, NJ: Prentice Hall.
- Karpathy A and Fei-Fei L (2015) Deep visual-semantic alignments for generating image descriptions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3128–3137.
- Katz MG, Kripalani S and Weiss BD (2006) Use of pictorial aids in medication instructions: A review of the literature. *American Journal of Health-System Pharmacy* 63(23): 2391–2397.
- Koop CE (1985) The Heimlich Maneuver. *Public Health Reports* 100(6): 557.
- Kress G and Van Leeuwen T (2020) *Reading Images: The Grammar of Visual Design*. London: Routledge.
- Landis JR and Koch GG (1977) An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 33(2): 363–374.
- Li LH et al. (2019) Visualbert: A simple and performant baseline for vision and language. *arXiv preprint arXiv:1908.03557*.
- Lin T-Y et al. (2014) Microsoft coco: Common objects in context. In: *Proceedings of European Conference on Computer Vision*. Berlin: Springer, 740–755.
- Lu J et al. (2019) Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. *arXiv preprint arXiv:1908.02265*.
- Lu J et al. (2020) 12-in-1: Multi-task vision and language representation learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10437–10446.
- Martinec R and Salway A (2005) A system for image–text relations in new (and old) media. *Visual Communication* 4(3): 337–371.
- Parcalabescu L et al. (2021) Seeing past words: Testing the cross-modal capabilities of pretrained VL models on counting tasks. 2021, 2012.12352.

- Parkinson J, Mackay J and Demecheleer M (2020) Putting yourself into your work: Expression of visual meaning in student technical writing. *Visual Communication* 19(2): 281–306.
- Piwek P (2001[1998]) Relating imperatives to action. In: Blunt H and Beun RJ (eds) *Cooperative Multimodal Communication: Second International Conference, CMC'98*, Tilburg, The Netherlands, January 28–30, Selected Papers, Vol. 2155. Berlin: Springer Science & Business Media.
- Pustejovsky J (2018) From actions to events: Communicating through language and gesture. *Interaction Studies* 19(1/2): 289–317.
- Rose G (2016) *Visual Methodologies: An Introduction to Researching with Visual Materials*. London: Sage.
- Sharma P et al. (2018) Conceptual captions: A cleaned, hypernymed, image alt-text dataset for automatic image captioning. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Vol. 1, Long Papers, 2556–2565.
- Silberer C and Pinkal M (2018) Grounding semantic roles in images. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2616–2626.
- Smith KL et al. (eds.) (2004) *Handbook of Visual Communication: Theory, Methods, and Media*. London: Routledge.
- Socher R et al. (2014) Grounded compositional semantics for finding and describing images with sentences. *Transactions of the Association for Computational Linguistics* 2: 207–218.
- Stanfield RA and Zwaan RA (2001) The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science* 12(2): 153–156.
- Steehouder M and Van der Meij H (2005) Designing and evaluating procedural instructions with the four components model. In: *Proceedings 2005 IEEE International Professional Communication Conference*, Limerick, Ireland, 10–13 July, 797–801.
- Steehouder M, Karreman J and Ummelen N (2000) Making sense of step-by-step procedures. In: *18th Annual Conference on Computer Documentation IPCC Sigdoc 2000. Technology and Teamwork. Proceedings. IEEE Professional Communication Society International Professional Communication Conference an. IEEE*, Cambridge, MA, 24–27 September, 463–475.
- Su W et al. (2020) VL-BERT: Pre-training of generic visual–linguistic representations. *International Conference on Learning Representations*. Available at: <https://openreview.net/forum?id=SygXPaEYvH> (accessed 31-10-2022).
- Tan H and Bansal M (2019) LXMERT: Learning cross-modality encoder representations from transformers. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*



- (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, 5100–5111. DOI:10.18653/v1/D19-1514.
- Thomson EM (2021) The graphics of carving. *Visual Communication* 20(4): 552–576.
- Tufte ER (1997) *Visual Explanations*. Cheshire, CT: Graphics Press.
- Tversky B (2014) Visualizing thought. In: *Handbook of Human Centric Visualization*. New York, NY: Springer, 3–40.
- Tversky B et al. (2000) Lines, blobs, crosses and arrows: Diagrammatic communication with schematic figures. In: *International Conference on Theory and Application of Diagrams*, Berlin, Heidelberg, DE, September. Berlin: Springer, 221–230.
- Van der Sluis I, Eppinga AN and Redeker G (2017) Text–picture relations in multimodal instructions. In: *Proceedings of the IWCS workshop on Foundations of Situated and Multimodal Communication*.
- Van der Sluis I, Kloppenburg L and Redeker G (2016a) PAT Workbench: Annotation and evaluation of text and pictures in multimodal instructions. In: Hinrichs E et al. (eds) *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)* at COLING 2016, 131–139.
- Van der Sluis I, Leito S and Redeker G (2016b) Text–picture relations in cooking instructions. In: *Proceedings of LREC 2016, Tenth International Conference on Language Resources and Evaluation: Proceedings of the Twelfth Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation (ISA-12)*, 22–27.
- Van der Sluis I, Redeker G and Debreczeni S (2022) A text-based method to derive the main action structure in procedural instructions. In: *Proceedings of the Workshop on Annotation, Recognition and Evaluation of Actions (AREA-2)* to be held at (ESSLLI 2022), 1–10.
- Van der Sluis I, Vergeer R and Redeker G (2018) Action categorisation in multimodal instructions. In: *Proceedings of the Workshop on Annotation, Recognition and Evaluation of Actions (AREA) held at the 11th edition of the Language Resources and Evaluation Conference (LREC 2018)*, 22–27.
- Vedantam R, Lawrence Zitnick C and Parikh D (2015) Cider: Consensus-based image description evaluation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4566–4575.
- Wildfeuer J et al. (2022) No laughing matter!? Analyzing the page layout of instruction comics. Epub ahead of print 21 March. *Journal of Graphic Novels and Comics*. DOI: 10.1080/21504857.2022.2053559.
- Wilson AA and Landon-Hays M (2016) A social semiotic analysis of instructional images across academic disciplines. *Visual Communication* 15(1): 3–31.
- Zhang ZP et al. (2012) Automatically extracting procedural knowledge from instructional texts using natural language processing. In: *Proceedings*

*of the Eighth International Conference on Language Resources and Evaluation, Istanbul, Turkey, 21–27 May, 520–527.*

Zoss M, Siegesmund R and Patisaul SJ (2010) Seeing, writing, and drawing the intangible: Teaching with multiple literacies. In: Albers P and Sanders J (eds) *Literacies, the Arts, and Multimodality*. Urbana, IL: National Council of Teachers of English, 136–156.

## **BIOGRAPHICAL NOTES**

IELKA VAN DER SLUIS is a computational linguist, specialising in the areas of human communication, multimodal information presentation, human-computer interaction, natural language generation, affective computing and multimodal interaction. She researches the production and processing of the visual and verbal means that people employ to communicate as well as the contexts of use in which this happens through collection and analysis of corpora and empirical experimentation. The data collected in her studies offers new insights in the workings of human communication that can be used to guide authors of multimodal documents and to improve computational models for use in real-life applications. Dr. Ielka van der Sluis is a member of the Center for Language and Cognition Groningen (<http://www.rug.nl/research/clcg/>) and an Assistant Professor at the Department of Communication and Information Science at the University of Groningen. In the past she has worked as a Research Fellow at the Computational Linguistics Group at the Department of Computer Science at the Trinity College Dublin in Ireland, the Natural Language Generation Group at the University of Aberdeen, and the University of Tilburg. Detailed CV: <https://www.rug.nl/staff/i.f.van.der.sluis/>

*Address:* Center for Language and Cognition, University of Groningen, Oude Kijk in 't Jatstraat 26, Harmoniegebouw 1312-428, Groningen 9712 EK, The Netherlands. [email: [i.f.van.der.sluis@rug.nl](mailto:i.f.van.der.sluis@rug.nl)]

GABRIELA MATOUŠKOVÁ is a graduate of the study programme Communication and Information Studies at the Faculty of Arts, University of Groningen. She currently holds the position of event and community manager at the Aletta Jacobs School of Public Health in Groningen, connecting researchers from different disciplines to work together on projects addressing public health challenges and communicating science to the public through various public engagement programmes. Besides that, she is interested in health communication and effects of social media on health.

*Address:* as Ielka Van der Sluis. [email: [g.matouskova@student.rug.nl](mailto:g.matouskova@student.rug.nl)]

HANNAH NIEMEIER is currently completing her bachelor's in Communication and Information Sciences at the Faculty of Arts, University

of Groningen. She is currently doing a minor abroad at Queens University, Kingston, Canada focusing on Linguistic features in Science Fiction and Fantasy literature. Besides that, she is interested in environmental communication and health communication.

*Address:* as Ielka Van der Sluis. [email: h.niemeier@student.rug.nl]

SOPHIA POPP has completed here bachelors in Communication and Information Science at the University of Groningen and is currently pursuing her masters in Management of Information Systems and Digital Innovation at the London School of Economics and Political Science. She has conducted research in the fields of human-computer interaction, information communication technologies, digital networks, linguistics and multimodal information presentation. Moreover, she has gained insights into policy making research through an extracurricular program at the University of Groningen. Her most recent research interest is the field of data governance and generativity of digital platforms and infrastructures.

*Address:* as Ielka Van der Sluis. [email: l.s.popp@student.rug.nl]

JOSEPHINE CARSTENS founded her own company, after graduating from the bachelor programme Communication Information Studies at the University of Groningen. In her company she is running a career as a cinematographer and marketing consultant. At this moment she plans, consults and conducts film productions and social media campaigns, working for companies and private clients in Germany, Austria and the Netherlands. Her research interest surrounds the effect and meaning of multimodal information presentation, particularly, videography in multimedia communication.

*Address:* as Ielka Van der Sluis. [email: j.l.carstens@student.rug.nl]