NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Olga Rubtsova
Elena Gorbunova

CATEGORIZATION OF ILLUSTRATED EMOTIONS IN VISUAL STORYTELLING CONTEXT

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: PSYCHOLOGY

WP BRP 134/PSY/2022

Olga Rubtsova[1], *Elena Gorbunova*[1,2]

CATEGORIZATION OF ILLUSTRATED EMOTIONS IN VISUAL STORYTELLING CONTEXT

There are a few factors related to the categorization of illustrated emotions. The most notable ones are suggested to be the following: facial expression, body posture, and emanata. The current study focuses on revealing the roles of these three factors, as well as their combinations with the addition of the narrative context that currently seems to be of very high practical significance. Illustrated emotional images were used in the experiment, and the presence of different pictorial factors was varied. The accuracy of categorization and reaction time of the participants were analyzed for different conditions. The results suggest that emanata and body posture play the least significant role in emotional categorization, especially when the contextual information is absent. At the same time, the addition of all three pictorial elements facilitated the emotional categorization. However, the roles of the specific combinations of the introduced factors were not revealed, which requires further clarification.

Keywords: illustrated emotions, emanata, emotional categorization, visual storytelling.

JEL Classification: Z.

[1]National Research University Higher School of Economics, Laboratory for Cognitive Psychology of Digital Interfaces Users
[2]National Research University Higher School of Economics, School of Psychology

Introduction

Recognition and categorization of emotions

Emotions are available for perception mainly through the physical features of the face, and the process of emotion recognition requires the work of the visual system (Calvo & Nummenmaa, 2015). The facial features form specific combinations that are recognizable to a wide specter of people, thus allowing the process of emotion recognition and categorization. The categorization in this case could be described as a cognitive process of attributing such combinations of facial elements to the same or different categories (Calvo & Nummenmaa, 2015). Emotions can be recognized based on the specific features (e.g., Beaudry et al., 2013). Feature based approach argues that separate facial elements, like mouth or eyes, are enough for the emotion recognition. Holistic approach argues the opposite: only the whole face can transmit emotional information (Piepers & Robbins, 2012). As well as that, holistic approach leaves room to other factors that can influence emotion recognition, in addition to facial features. There is data supporting both feature based (e.g., Chen & Chen, 2010; Lipp et al., 2009) and holistic (e.g., Calder & Jansen, 2005) approaches. It is possible that in reality there are both holistic and feature based mechanisms when it comes to emotional perception and categorization. In particular, there are other factors related to emotional categorization that should be considered together with facial elements in order to better understand the cognitive mechanisms of this process.

The second most important factor after facial expression when it comes to emotional categorization, at least in holistic view, seems to be body posture. Similarly to the facial expressions of basic emotions, body postures associated with basic emotions are recognized even when the facial expression is not presented (De Silva & Bianchi-Berthouze, 2004). Some studies have shown that the accuracy of receiving emotional signals from the body is comparable to facial emotion recognition, and sometimes it can be even higher (Atkinson et al., 2007; de Gelder, 2009). At the same time, it was pointed out that bodies and faces are rather processed as a whole, constituting one salient unit (Aviezer et al., 2012). The research by Schouwstra and Hoogstraten (1995) showed that the evaluation of spinal positions of the drawn figures did not lead to a strict correlation between body postures and specific emotions. Rather, one spinal position was related to a whole specter of emotional states. This finding suggests that the factor of body posture, despite being relevant to emotional categorization, is not as significant as the factor of facial expression. In one study the task for the participants was to guess the emotion for the stimuli consisted of facial and body images (Meeren et al., 2005). The images were either

congruent (the face and the body represented the same emotion) or incongruent (the face and the body represented different emotions). The results showed that the accuracy of the judgment increased for the congruent stimuli, while the reaction for them was significantly faster in comparison to incongruent images. Thus, the combination of facial expression and body posture that both reflect some specific emotional state leads to better and faster processes of categorization. Another study by Buisine et al. (2013) tested whether the body postures stimulate the emotional guidance for animated characters. This study is particularly interesting since it follows a holistic approach to emotional categorization. The authors emphasize that the emotions were studied in contextually rich conditions. The findings of the study showed that scenarios rich in contextual details are very advantageous for research of emotional recognition. At the same time, the presence of the body improved the accuracy of judgment, even when the body postures were neutral. Overall, the contextual information plays a significant role in identifying emotions from the body, as it was presented in the previous research (e.g., Lankes & Bernhaupt, 2011; Volkova et al., 2014).

Specifics of illustrated emotions

Illustrated emotional states differ from real ones in a few aspects. Cartoon faces and bodies are typically drawn with the use of hyperbolization and deformation of specific features. For example, the eyes of the characters are usually unproportionally large, making them more expressive than the eyes of a real human being (Liu et al., 2019). Overall, exaggeration is very widely used in illustration and animation industries, primarily due to its effects on enhancing character's emotional states and intentions (Hyde et al., 2016). It has been reported that emotions conveyed by illustrated faces are perceived as more intense than photorealistic ones. In particular, it was shown for the emotion of sadness (Zhang et al., 2021). It is important to add that not only the emotional aspect is specifically emphasized by the artists. For instance, in order to make the scene more dynamic, motion lines are commonly used. It is not just the artist's preference, though, because it has been shown that such techniques improve the viewers' understanding of the image. In case of motion lines, they enhance the perception of movement of objects and persons for the audience (Ito et al., 2010; Kawabe & Miura, 2006). The purpose of an artist is to convey the message in the most efficient way, hence, all types of proportional variations or additional pictorial elements can be used in order to transmit the message quickly and comprehensively.

It has been argued that simplified, or iconic, images convey emotional information better than realistic images. In a study by Kendall et al. (2016) the experimenters used several

conditions of images in order to test facial recognition. The conditions varied from photos of real faces to rotoscope images to cartoon images. The results showed the increase of accuracy along with the cartoonization of the images. Similar results were obtained for emotional recognition in cartoon faces (Wessler & Hansen, 2021). The theory behind such findings suggests that the processing of iconic and realistic objects communicating emotional information is different. The underlying mechanisms suggest that symbolic images are more effective in their communicative function, being present in all sorts of advertisements, social media and other crucial industries. As well as that, such images are not as complicated as realistic ones, having more low-level features. Low- level features, for instance, contrast, have been shown to be very effective in facilitating emotion in faces and improving facial recognition (Gray et al., 2013; Halit et al., 2006).

When it comes to the question of categorization of illustrated emotions, it is not clear whether the categorical or dimensional approach is more appropriate. On the one hand, a lot of artists use the categorical approach when it comes to expressing the character's emotions. There are whole guides for the illustrators based on the most prominent categorical theories, mostly Ekman's theory of basic emotions (e.g., McCloud, 2006). Most of the time, it is necessary to make sure the emotion is interpreted only one specific way. Otherwise, the whole meaning of the image would be lost, and the goal of the artist would not be achieved. It becomes even more critical for the visual stories, for instance, in the form of comics or manga, where several images are used in order to convey a certain plot. In this case, misinterpretation of character's emotions, feelings or states could lead to the difficulties in following the storyline. Thus, using commonly recognized facial features associated with specific emotions is quite beneficial for the artist. Furthermore, data suggests that schematic faces are processed in a feature-based manner, in contrast to real faces that are processed more holistically (Prazak & Burgund, 2014). At the same time, emotions on illustrated faces can be quite ambiguous for the viewers. Research by Stamenković et al. (2018) tested 24 illustrated faces expressing different emotional states. The authors found out that there is little consistency in recognition among the participants, even for highly intense illustrations. They conclude that in most cases illustrated faces are accompanied by corresponding body postures or gestures that significantly increase the comprehension of the characters' internal states, feelings, and emotions. Thus, a dimensional approach can be successfully applied in the form of adding additional contextual perceptual information to the image.

Emanata

One of the elements that helps artists to hyperbolize the emotional aspect of their images specifically is called emanata. It is sometimes referred to as pictorial runes (e.g., Kennedy, 1982). These are the specific pictorial elements surrounding the character's heads or bodies or being placed on top of them that code a specific emotional state (Ojha et al., 2021). They are typically non-mimetic and facilitate the transmission of narrative information (Forceville, 2011). For example, a few straight lines around a character can convey the emotion of surprise, while many short lines could reflect the character is deep in thought (Itou et al., 2019). Emanata is a specific pictorial factor when it comes to emotional recognition, since different emanata do not correspond to specific emotions. In fact, one symbol could mean a whole range of things, including emotional states, physical materials, or environment conditions. For instance, the symbol of drops could be used to express sweat, tears, or water (Akai et al., 2015). Previous research on emanata also suggests that it can be perceived differently by the viewers. A study by Ojha et al. (2021) used droplets, spikes, spirals and twirls presented in a comic panels format to clarify the roles of these pictorial elements in conveying emotional information. They found out that the symbols did not correlate with emotions in one-to-one manner, rather, some symbols were more associated with particular states (e.g., twirls were more associated with confusion). However, it was concluded that emanata promotes the awareness of emotional information in the image. Overall, emanata seems to lack the systematicity when it comes to emotional categorization. At the same time, research showed that some particular emanata were in fact associated with specific emotional states. Seno and colleagues (2013) conducted an experiment putting vertical or horizontal lines over illustrated faces and showed that vertical, but not horizontal lines increased the perception of sadness.

In some cases emanata can be metaphoric, and this property makes it ambiguous without context from other emotional pictorial elements or surroundings. For example, gears around a character's head that can be additionally animated in complicated movements signal the viewer that the character is in the middle of thinking or solving an intricate problem. However, just relying on emanata will not be enough to correctly identify the character's emotion in this case. Thus, emanata can be described as an additional factor hyperbolizing, or intensifying an emotional state of a character, while not being sufficient for emotion recognition. In support of this claim, research by Kendall et al. (2020) involved studying face-upfix dyads that were either cartoons or photo images and showed that cartoon images are easier to recognize. The face and upfix could be congruent or incongruent, and the task was to judge the congruency of the paired

images. The results suggest that less resources of sustained attention are needed for information processing of cartoon images compared to paired photographs of real faces and objects. In particular, the gaze fixations on the cartoon pairs were significantly shorter, further implying the greater role of schematic or iconic images in efficient information processing.

The mechanisms of emanata could be explained through the Visual Language Theory (VLT) introduced by Cohn (2013). In this paradigm visual images have similar elements to language, for instance, affixes (suffixes, prefixes, circumfixes, etc.) can be represented by pictorial elements surrounding character's heads or bodies (Cohn, 2018). The lines of movement relate to such language elements as well. Thus, artists use graphical techniques in order to convey some story - they can express the character's movements in one static image, show the character in the middle of experiencing some emotional state, etc. However, separate elements of the image are not enough on their own to complete the meaningful story. It has already been mentioned that emanata (as well as some body postures) can be very ambiguous in isolation and have potential to be interpreted differently in combination with other visual elements of an image. Cohn (2016) suggests that the visual language exists only when grammar is present, that is when images create meaningful sequences. Such grammatical structures form parallel architecture that is perceived holistically. Thus, perception of illustrated images relies largely on the context surrounding the images.

Visual storytelling

Contextual factors like background scenes, social situations and language play a significant role in emotion perception (Barrett et al., 2011). People typically judge individual emotional states based on the emotional states of the surrounding crowd (Griffiths et al., 2018). Moreover, the factors of facial expression, body posture and background scene have been reported to have independent and combinatorial roles in categorization of emotional information (Reschke & Walle, 2021). As well as perceptual contextual factors, illustrated emotions are almost always introduced in narrative contexts (Kukkonen, 2013). Such narrative contexts enable the viewer to appropriately comprehend the perceptual elements related to the emotional states of the character. The Kuleshov effect is interesting in this respect, since it illustrates the importance of the context surrounding an actor with regard to their emotional states' perception. Lev Kuleshov was a director who conducted a very well-known experiment concerning editing in cinematography (Prince & Hensley, 1992). He edited a close-up shot with a silent actor conveying a neutral facial expression with different shots: a shot of a bowl of soup, a shot of a dead person and a shot with a young girl. Even though in all the shots the actor did not express

any particular emotion, and the shot itself was exactly the same, the viewers of the film attributed different emotional states to the actor, based on the paired shot. They perceived a hungry man when the shot was accompanied by the image of soup, a grieving man when the shot was accompanied by the image of a dead woman and a happy man when it was accompanied by a little girl. Thus, context on its own already has a potential to induce the perception of particular emotions. Of course, when the narrative context is combined with the image of an emotional person or a character, these factors in combination facilitate the expressed emotions even further.

When it comes to illustrated emotions in visual narrative, a field of interactive storytelling games is particularly interesting. Such games allow the players to interact with the characters introduced in the game, make choices, and receive certain feedback depending on the chosen options. Such games are sometimes referred to as visual novels (Camingue et al., 2021). These games can be widely used in educational purposes (Camingue et al., 2020; Sullivan & Critten, 2014). As well as visual novels, illustrated and animated characters are used in the fields of artificial intelligence (Jaiswal et al., 2020). Virtual agents with natural emotional expression are trusted by people, and therefore, they can be successfully applied for different practical purposes. As well as that, illustrated and animated characters can be used in communicative interfaces in the form of avatars. Avatars usually represent the emotions and intentions of the users, helping them to better engage in online communication (Itou et al., 2019). Overall, there are a lot of practical applications for illustrated emotions, however, it is beneficial to study them not in isolation but embedded into narrative contexts.

Problem statement

There are several factors associated with emotional categorization for illustrated images, the three main ones being facial expression, body posture and emanata. It has been reported that those factors play significant but different roles in emotional categorization. However, the specific roles of these factors in combination with each other have not been experimentally clarified. In order to fully understand the mechanisms of illustrated emotion processing, it is necessary to study the role of all these factors separately as well as in combination.

This research is aimed to differentiate all the factors mentioned above and verify their combined roles in experimental conditions, which has not been done before. As well as that, studying emotional categorization is suggested in visual storytelling context, making it more natural and relevant for many practical fields, including artificial intelligence, gaming and communicative interfaces.

The primary experimental hypotheses of this study are:

1. Accuracy of emotional categorization should increase with the increase of perceptual factors conveying emotional information. Thus, when all the factors are combined, the categorization accuracy should be the highest. Emanata, being presented as the most ambiguous of all the three factors, is expected to cause more categorization errors.

2. Narrative context improves emotional categorization. When contextual information is present, the emotional categorization is expected to be better than when it is absent.

Method

Participants

There were 31 volunteers who participated in the experiment. There were 22 females and 9 males, their age varied from 18 to 32 (M = 21.32, SD = 2.47). Some of the volunteers were students of HSE University and received bonus points for the "Ergonomics and Usability" course or "Psychology" minor for participation in the study. Another part of volunteers was recruited via the public page "Cognitive Partymaker" in the social network VK (https://vk.com/). The participants automatically participated in a raffle of a gift electronic certificate for 1000 rubles. All the participants were required to be 18 years old or older, have normal or corrected to normal vision. They confirmed that they did not have neurological or psychiatric conditions. All of them gave their informed consent before the experiment.

Stimuli and design

Seven emotions were chosen to be used in the experiment, based on the results of the pilot study of the stimuli material. The illustrated images of ten emotional states (anger, confusion, surprise, irritation, relief, fear, happiness, embarrassment, sadness, and nausea) were presented to the participants (31 females and 3 males; age ranged from 18 to 32 (M = 20.03, SD = 2.59)). They were instructed to choose the appropriate emotion from a given list for each illustrated image. The illustrations of nausea, confusion, and irritation were revealed to be confusing to the participants, and they were excluded from the further study. Hence, the emotions used in the experiment were: anger, surprise, fear, happiness, sadness, relief, and embarrassment.

Stimuli material in the form of illustrations was created for each emotion according to the seven experimental conditions: facial expression, emanata, body posture, facial expression +

emanata, facial expression + body posture, body posture + emanata and facial expression + body posture + emanata. The images were created in Procreate v.526 software. They were 16.6°×14.49° in size for images with the head only and 19.35°×19.94° for images with the body. The examples of the images for each condition are presented in Figure 1.

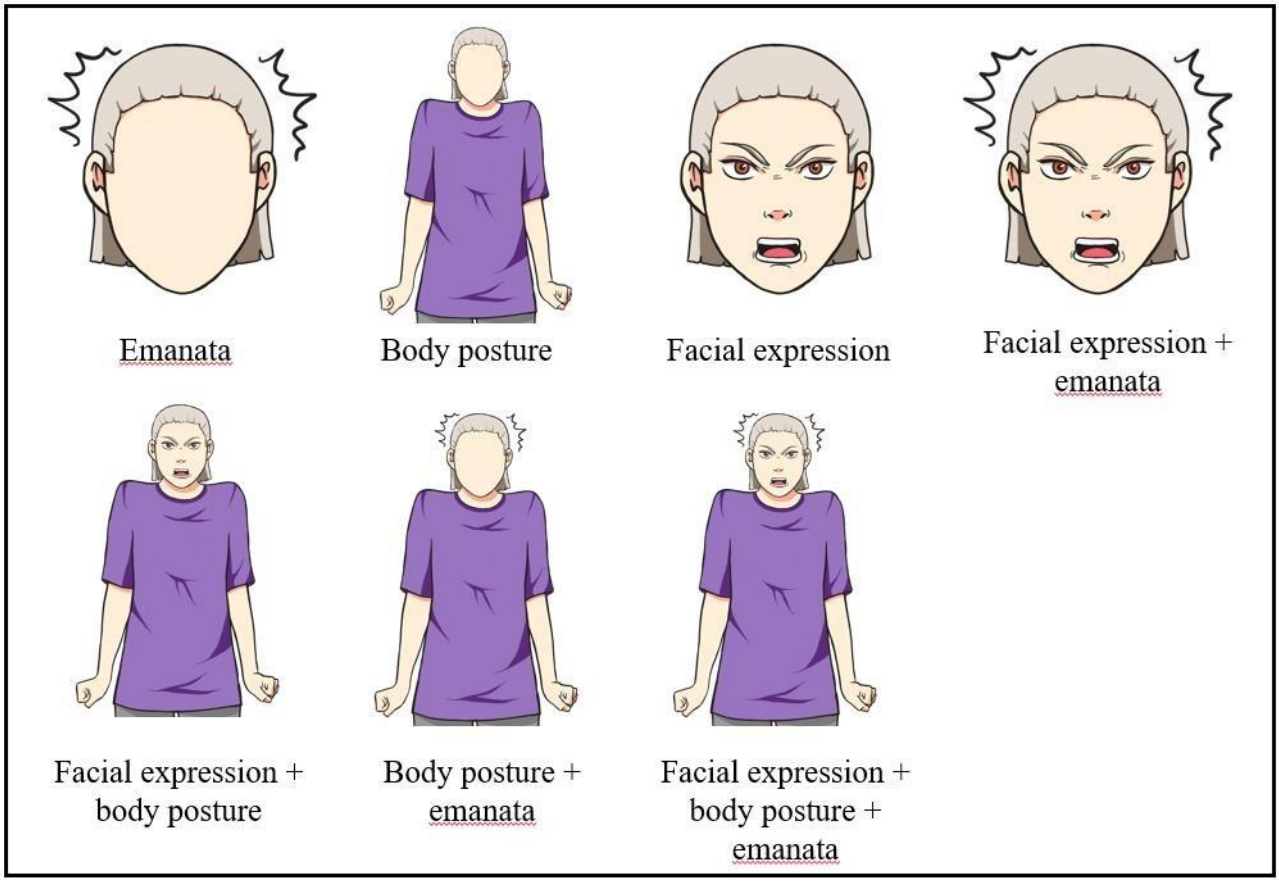

Figure 1. Example of emotional stimuli used in the experiment. All the conditions are presented

For each of the image type conditions, as well as for each emotion, contextual stories were created in order to introduce the narrative factor. The stories were written in a form of short descriptions of situations that happened to the character illustrated in the image and provided foundation for the further emotional reaction of the character. The emotion itself was not mentioned in the description. The text was put in the middle of the rectangle that was 16.67°×10.71° in size. An example of a contextual situation description is presented in Figure 2.

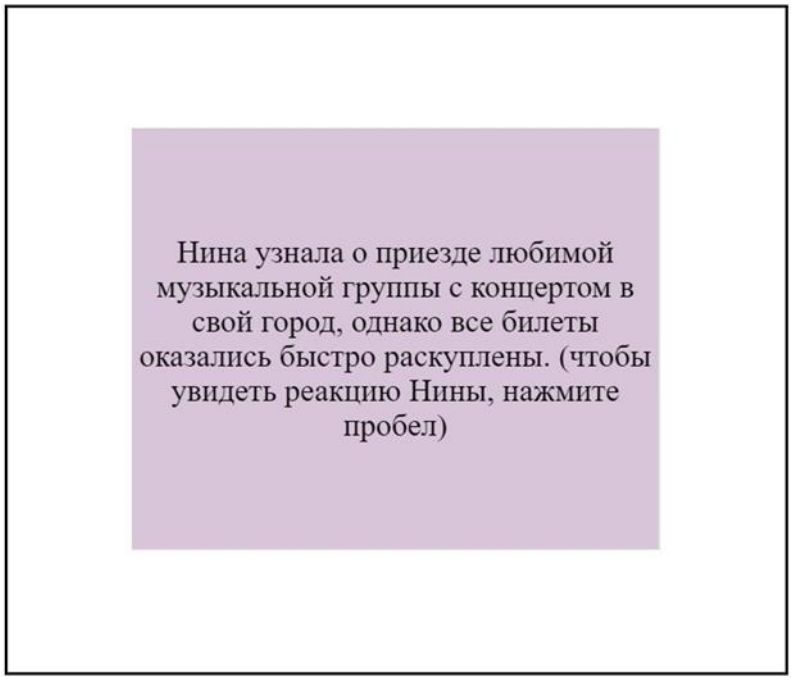

Figure 2. Example of contextual situation description used in the experiment. The situation refers to the emotion of sadness. The text reads: “Nina found out about the arrival of her favorite musical group with a concert to her city, but all the tickets were quickly sold out. (in order to see Nina’s reaction, press spacebar)”

Overall, there were 14 main conditions in the experiment, according to the factors of narrative context (absent or present) and type of emotional image (facial expression, body posture, emanata, facial expression + emanata, facial expression + body posture, body posture + emanata, facial expression + body posture + emanata). All seven emotions were used for each condition. All the conditions were presented to each participant (within-subject design).

Procedure

The experiment started with instructions being presented. The participants were explained the structure of the experiment and their task. As well as that, the main character involved in the experiment was introduced to them as Nina, a student who finds herself in various life situations. The participants were instructed that they would read the descriptions of these situations and then see Nina’s reactions to them that they would have to guess. They were also instructed that sometimes they would see just the reactions without the preceding story descriptions. Moreover, they were warned that in some cases Nina’s reactions would not be full, for instance, they would not see Nina’s facial expression. In these cases, the participants were instructed to guess Nina’s

emotion on the grounds of other presented features. If everything was clear to them, they proceeded by pressing the ‘spacebar’ on the keyboard. Each trial started with a rectangle being presented in the middle of the screen. The rectangle contained either a description of a situation with context for emotional reaction or an instruction to press ‘spacebar’. For the former option, the participants read the situation, after which the instruction to press ‘spacebar’ was presented. Next, the emotional reaction of the character was presented with the duration of one second. The time limit was introduced in order to increase the participants’ attention to the task. The picture of the character belonged to one of 7 emotions and one of seven conditions introduced in the experiment design. Subsequently, four rectangles were presented containing the names of the emotions that the participant would have to choose from. The participants had to choose the only option that corresponded to the previously presented emotion of the character and press the corresponding button on the keyboard. They had to press the ‘up arrow’ if they chose the upper option, the ‘down arrow’ if they chose the lower option, the ‘left arrow’ if they chose the left option and the ‘right arrow’ if they chose the right option. Then the cycle started again. The order of the trials was fully randomized for each participant. A schematic representation of the procedure is presented in Figure 3.

The experiment was created in PsychoPy v.2022.1.2 software and run via Pavlovia platform (https://pavlovia.org/) for online experiments.

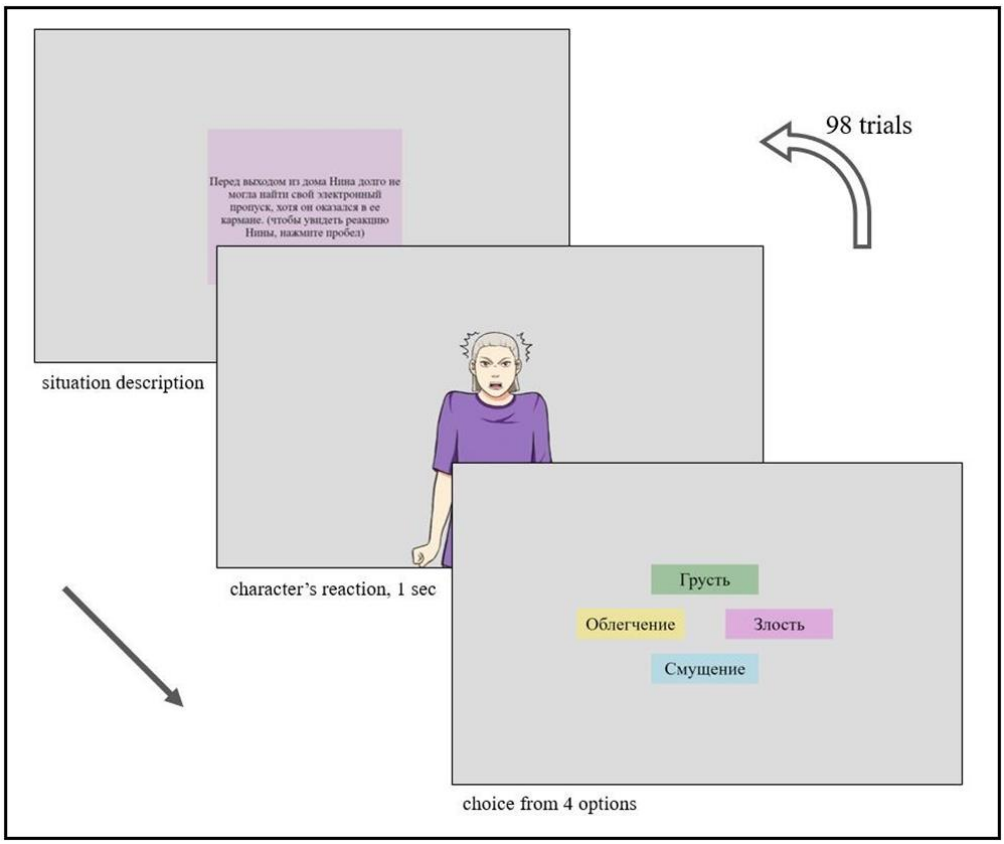

Figure 3. Procedure of the experiment. The trial represents the condition with a present contextual situation description, facial expression + body posture + emanata in the illustrated image, emotion of anger

Results

The percentages of correct answers were calculated for each condition, and the reaction times (RT) for each answer were measured. Two-way repeated measures ANOVA was used in order to reveal the role of factors "presence of contextual situation" (two levels - present or absent) and "image type" (seven levels - facial expression, body posture, emanata, facial expression + emanata, facial expression + body posture, body posture + emanata, facial expression + body posture + emanata), as well as their interaction. In order to further study the specifics of the interaction, additional rmANOVAs with pairwise comparisons using Bonferroni adjustment were conducted, as well as paired sample t-tests. The Greenhouse-Geisser corrections were always applied, when Mauchly's sphericity tests were significant. Accuracy and RT were analyzed separately. The analysis was made using SPSS v. 22.0.0.0.

Accuracy

Two way rmANOVA revealed a significant effect of presence of the contextual situation factor ($F(1,30) = 8.561$, $p = .006$, $\eta_p^2 = .222$), a significant effect of the image type factor ($F(3,104) = 7.229$, $p < .001$, $\eta_p^2 = .194$) and a significant interaction ($F(6,180) = 2.91$, $p = .01$, $\eta_p^2 = .088$). The results are presented in Figure 4. Since the factor interaction was significant, additional paired sample t-tests were conducted in order to compare conditions with the presence and absence of a contextual situation for the different emotional factors. T-tests showed the significant differences for the images with the body posture ($p = .004$) and for the images with the body posture + emanata ($p = .017$). The results of t-tests are presented in Table 1. In order to compare different types of images, additional rmANOVAs with pairwise comparisons were conducted separately for the conditions with and without the contextual story description. The results show that when contextual story was present, the factor of image type was significant ($F(4,111) = 5.525$, $p < .001$, $\eta_p^2 = .156$). Accuracy was significantly lower for identifying the images with emanata compared to the images with body posture + emanata ($p = .011$) and compared to the images with facial expression + body posture + emanata ($p = .015$). For the trials with no context preceding the image the factor of image type was also significant ($F(6,180) = 4.957$, $p < .001$, $\eta_p^2 = .142$). Accuracy was significantly lower regarding emanata condition compared to facial expression + body posture condition ($p = .015$) and compared to facial expression + body posture + emanata condition ($p = .025$), moreover the accuracy was significantly lower in body posture condition compared to facial expression condition ($p = .037$), compared to facial expression + body posture condition ($p = .039$) and compared to facial expression + body posture + emanata condition ($p = .032$). The results of pairwise comparisons are presented in Table 2 and Table 3.

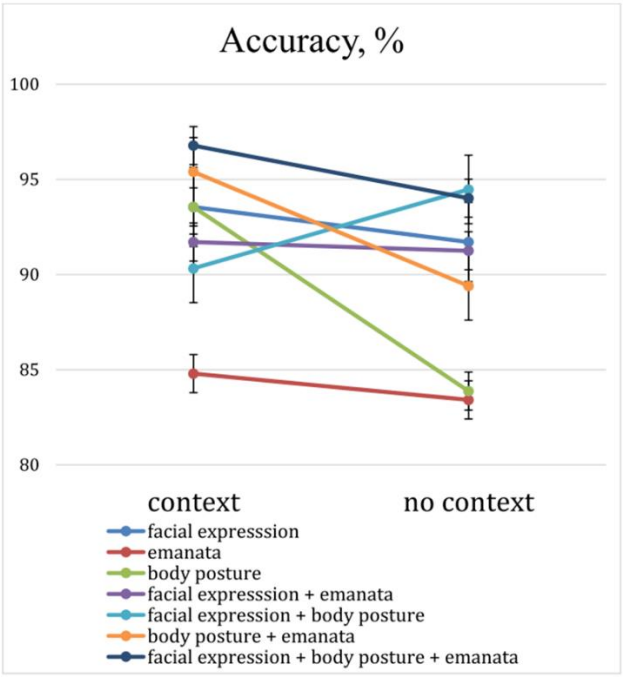

Figure 4. Results for accuracy. Error bars represent standard error means

Table 1. Results of t-tests (accuracy) for conditions with and without narrative context, the cells contain the descriptive statistics, p-values (p) and effect sizes (d) of t-tests.

	narrative context is present	narrative context is absent	p	d
facial expression	M = 93.55 SD = 11.57	M = 91.7 SD = 11.53	.38	.16
emanata	M = 84.79 SD = 16.88	M = 83.41 SD = 16.12	.63	.083
body posture	M = 93.55 SD = 13.72	M = 83.87 SD = 17.59	.004*	.613
facial expression + emanata	M = 91.7 SD = 16.4	M = 91.24 SD = 14.6	.845	.03
facial expression + body posture	M = 90.32 SD = 10	M = 94.47 SD = 15.94	.095	.311
body posture + emanata	M = 95.39 SD = 10	M = 89.4 SD = 13.79	.017*	.497
facial expression + body posture + emanata	M = 96.77 SD = 13.15	M = 94 SD = 7.17	.246	.261

* significant p-values

Table 2. Results of pairwise comparisons for different types of emotional stimuli in the condition with narrative context (accuracy), the cells contain p-values (p).

	facial expression	emanata	body posture	facial expression + emanata	facial expression + body posture	body posture + emanata	facial expression + body posture + emanata
facial expression	-						
emanata	.295	-					
body posture	> .99	.169	-				
facial expression + emanata	> .99	> .99	> .99	-			
facial expression + body posture	> .99	> .99	> .99	> .99	-		
body posture + emanata	> .99	.011*	> .99	> .99	.116	-	
facial expression + body posture + emanata	.68	.015*	> .99	.529	.056	> .99	-

* significant p-values

Table 3. Results of pairwise comparisons for different types of emotional stimuli in the condition without narrative context (accuracy), the cells contain p-values (p).

	facial expression	emanata	body posture	facial expression + emanata	facial expression + body posture	body posture + emanata	facial expression + body posture + emanata
facial expression	-						
emanata	.077	-					
body posture	.037*	> .99	-				
facial expression + emanata	> .99	.308	.383	-			
facial expression + body posture	> .99	.015*	.039*	> .99	-		
body posture + emanata	> .99	.632	> .99	> .99	> .99	-	
facial expression + body posture + emanata	> .99	.025*	.032*	> .99	> .99	> .99	-

* significant p-values

Reaction time

Two way rmANOVA revealed a significant effect of presence of the contextual situation factor ($F(1,30) = 19.809$, $p < .001$, $\eta_p^2 = .398$), a significant effect of the image type factor ($F(4,118) = 8.183$, $p < .001$, $\eta_p^2 = .214$) and a significant interaction ($F(4,124) = 4.371$, $p = .002$, $\eta_p^2 = .127$). The results are presented in Figure 5. Since the factor interaction was significant, additional paired sample t-tests were conducted in order to compare conditions with the presence and absence of a contextual situation for the different emotional factors. T-tests showed the significant differences for the images with facial expression ($p = .011$), emanata ($p = .006$) and

body posture (p = .003). The results of t-tests are presented in Table 4. In order to compare different types of images, additional rmANOVAs with pairwise comparisons were conducted separately for the conditions with and without the contextual story description. The results show that when contextual story was present, the factor of image type was significant (F(4,132) = 3.018, p = .017, η_p^2 = .091). RT was significantly higher in emanata condition compared to facial expression + body posture condition (p = .028). For the trials with no context preceding the image the factor of image type was also significant (F(4,113) = 8.36, p < .001, η_p^2 = .218). RT was significantly higher in emanata condition compared to facial expression + body posture condition (p = .001), body posture + emanata condition (p = .003) and facial expression + body posture + emanata condition (p < .001), moreover RT was significantly higher in body posture condition compared to facial expression + body posture + emanata condition (p = .022). The results of pairwise comparisons are presented in Table 5 and Table 6.

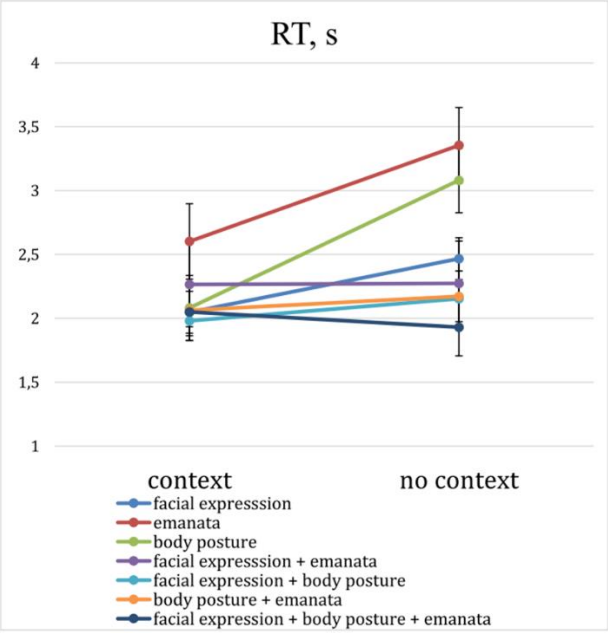

Figure 5. Results for RT. Error bars represent standard error means

Table 4. Results of t-tests (RT) for conditions with and without narrative context, the cells contain the descriptive statistics, p-values (p) and effect sizes (d) of t-tests.

	narrative context is present	narrative context is absent	p	d
facial expression	M = 2.05 SD = 0.91	M = 2.47 SD = 1.13	.011*	.408
emanata	M = 2.6 SD = 1.65	M = 3.35 SD = 1.57	.006*	.468
body posture	M = 2.08 SD = 1.41	M = 3.08 SD = 2.18	.003*	.544
facial expression + emanata	M = 2.27 SD = 1.84	M = 2.27 SD = 2.04	.945	.004
facial expression + body posture	M = 1.98 SD = 1.26	M = 2.15 SD = 1.12	.228	.145
body posture + emanata	M = 2.06 SD = 1.1	M = 2.17 SD = 1.26	.478	.092
facial expression + body posture + emanata	M = 2.05 SD = 1.25	M = 1.93 SD = 1.31	.514	.094

* significant p-values

Table 5. Results of pairwise comparisons for different types of emotional stimuli in the condition with narrative context (RT), the cells contain p-values (p).

	facial expression	emanata	body posture	facial expression + emanata	facial expression + body posture	body posture + emanata	facial expression + body posture + emanata
facial expression	-						
emanata	.27	-					
body posture	> .99	.117	-				
facial expression + emanata	> .99	> .99	> .99	-			
facial expression + body posture	> .99	.028*	> .99	> .99	-		
body posture + emanata	> .99	.138	> .99	> .99	> .99	-	
facial expression + body posture + emanata	> .99	.146	> .99	> .99	> .99	> .99	-

* significant p-values

Table 6. Results of pairwise comparisons for different types of emotional stimuli in the condition without narrative context (RT), the cells contain p-values (p).

	facial expression	emanata	body posture	facial expression + emanata	facial expression + body posture	body posture + emanata	facial expression + body posture + emanata
facial expression	-						
emanata	.057	-					
body posture	> .99	> .99	-				
facial expression + emanata	> .99	.055	.363	-			
facial expression + body posture	> .99	.001*	.07	> .99	-		
body posture + emanata	> .99	.003*	.103	> .99	> .99	-	
facial expression + body posture + emanata	.057	< .001*	.022*	> .99	> .99	> .99	-

* significant p-values

Discussion

Based on the results of the participant's accuracy there was a general trend of accuracy decrease when the narrative context was omitted. The significant differences for the context factor were obtained for the conditions of body posture and body posture + emanata. Hence, the factors of body posture and emanata seem to not be sufficient for accurate emotional categorization. Additional information, like the introduced contextual story, helps the participants' recognition. The comparison of different types of images shows that emanata is the

most ambiguous factor both when context is present and absent. It corresponds with the previous research that argues the lack of systematicity in emanata (e.g., Ojha et al., 2021). Emanata on its own does not really help the emotion recognition since it can be used in many various situations. However, the hypothesis that addition of emanata improves the emotional categorization did not find the confirmation. Conditions with combinations of different factors were generally similar in accuracy patterns. One explanation could be that the accuracy was quite high, and there was a significant overlap of different conditions. If accuracy is above 80% for the most ambiguous condition, then the task was pretty easy on its own, and so there was no place for significant improvements of participants' performance. At the same time, high accuracy could reflect the hypothesis that illustrated emotional states are generally better processed than realistic ones. The schematic appearance of the images, compared to how real people look, could facilitate the emotional signals conveyed by the illustrations. It is interesting that when there was no context, the body posture factor was not as helpful as other factors, the accuracy dropped to the same level as emanata. This result goes against the findings by Atkinson et al. (2007) and de Gelder (2009) who argued that body has potential to transmit emotional information with the same precision as face or even better. In this study emotional categorization by body posture was significantly less accurate than by facial expression. Therefore, facial expression seems to play a more significant role than body posture in illustrated emotions processing, at least in the absence of the narrative context.

Reaction time reflected the relevance of the narrative context for the conditions of facial expression, body posture and emanata. In case of context absence, it took the participants significantly more time to choose the emotional state of the character. Thus, it takes more cognitive resources to complete the categorization without any additional narrative elements. At the same time, since there were no significant differences for the other conditions, the combination of different perceptual factors facilitates the processing of emotional signals, increasing the speed of decision making. When the contextual information was present, RT was generally similar conditions, and the images with only emanata were categorized longer. It corresponds with the findings for accuracy, suggesting little role of emanata as a separate factor in illustrated emotions recognition. Furthermore, similarly to accuracy, when the context was absent, not only emanata but also the body posture factor increased the RT. Body posture and emanata were not as specific as other factors or their combinations, and the emotions could not be as easily identified.

To take everything into consideration, the hypotheses of this experimental study were only partially confirmed. As can be seen from the graphs, the factor of emanata was the most

ambiguous, since the accuracy in this condition was the lowest, and the RT was the highest. Emanata does not correspond to specific emotions, therefore, in isolation from other information the judgment becomes more complicated. Similarly to emanata, body posture is not sufficient for the precise emotional categorization, but only when there is no narrative context. At the same time, the condition where all the factors were combined reflected the highest accuracy and the lowest RT. Thus, the combination of all the pictorial factors indeed leads to enhanced emotional recognition. The combinations of other factors did not show significant differences in the experimental conditions, so it is hard to make clear conclusions about the roles of separate combinations in emotional categorization. The narrative context played a significant role in the categorization of illustrated emotions, but not for all the factors. The results indicate that background narrative can improve the categorization when the present pictorial factors are ambiguous or when they are presented in isolation from other factors.

There are a few limitations in this experimental study. Firstly, too many factors were used at the same time, which probably caused the overlap of different conditions in the results. The further research in this particular area should focus on specific factors or specific combinations, rather than on them altogether. It could bring more light to the specifics of the underlying mechanisms of illustrated emotions perception. Moreover, since the task was rather easy for the participants, some modifications should be made for the further research. For example, the presentation time of the emotional stimuli could be decreased (in terms of milliseconds), so that the primary factors necessary for emotional categorization could be clarified. Finally, just behavioral experiments do not give the full information about how perceptual information is processed and how attentional mechanisms work. Accuracy and RT could be not enough to really separate different conditions in the experiment. It would be beneficial to involve the eye tracking method in order to identify the main areas of interest for the illustrated images. It could be the case that emanata are not paid attention to at all, and that is why adding this pictorial element to the image does not improve the performance, as it was reported in this study.

Conclusion

The research was focused on revealing the specifics of emotional categorization of illustrated images in the context of visual storytelling. Three primary factors associated with illustrated emotions were introduced: facial expression, body posture and emanata. As well as that, the role of contextual information, in particular, visual narrative context, was discussed in addition to its practical value.

In order to experimentally study the proposed topic, illustrated material was created with the separation and combination of all the three perceptual factors. The stimuli were first tested in the pilot study and then modified for the main experiment. Seven experimental conditions, represented by the types of illustrated images were varied. As well as that, for each emotion and for each condition contextual stories were created in order to introduce the narrative context factor.

The results revealed that while contextual factor was significant, it improved emotional recognition only for originally ambiguous factors. These factors were emanata and body posture. It took more time to identify images with only those elements presented, and the accuracy was significantly lower. At the same time, the combination of all the three perceptual factors seemingly improved the recognition of illustrated emotions. Thus, additional pictorial and contextual details can be successfully applied in various fields utilizing illustrated or animated characters in order to improve the experience of the users. However, the results of the experiment did not reveal the roles of combinations of perceptual factors, and this should be the focus of the follow-up research.

References

Akai, Y., Yamashita, R., & Matsushita, M. (2015, November). Giving emotions to characters using comic symbols. *Proceedings of the 12th International Conference on Advances in Computer Entertainment Technology*. https://doi.org/10.1145/2832932.2832979

Atkinson, A. P., Tunstall, M. L., & Dittrich, W. H. (2007). Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition*, *104*(1), 59–72.https://doi.org/10.1016/j.cognition.2006.05.005

Aviezer, H., Trope, Y., & Todorov, A. (2012). Holistic person processing: Faces withbodies tell the whole story. *Journal of Personality and Social Psychology*, *103*(1), 20–37. https://doi.org/10.1037/a0027411

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in Emotion Perception. *Current Directions in Psychological Science*, *20*(5), 286–290. https://doi.org/10.1177/0963721411422522

Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., & Tapp, R. (2013). Featural processing in recognition of emotional facial expressions. *Cognition and Emotion*, *28*(3), 416–432. https://doi.org/10.1080/02699931.2013.833500

Buisine, S., Courgeon, M., Charles, A., Clavel, C., Martin, J. C., Tan, N., & Grynszpan, O. (2013). The Role of Body Postures in the Recognition of Emotions in Contextually Rich Scenarios. *International Journal of Human- Computer Interaction, 30*(1), 52–62. https://doi.org/10.1080/10447318.2013.802200

Calder, A. J., & Jansen, J. (2005). Configural coding of facial expressions: The impact of inversion and photographic negative. *Visual Cognition, 12*(3), 495– 518. https://doi.org/10.1080/13506280444000418

Calvo, M. G., & Nummenmaa, L. (2015). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion*, *30*(6), 1081–1106. https://doi.org/10.1080/02699931.2015.1049124

Camingue, J., Carstensdottir, E., & Melcer, E. F. (2021). What is a Visual Novel? *Proceedings of the ACM on Human-Computer Interaction*, *5*(CHI PLAY), 1–18. https://doi.org/10.1145/3474712

Camingue, J., Melcer, E. F., & Carstensdottir, E. (2020). A (Visual) Novel Route to Learning: A Taxonomy of Teaching Strategies in Visual Novels. *International Conference on the Foundations of Digital Games*. https://doi.org/10.1145/3402942.3403004

Chen, M. Y., & Chen, C. C. (2010). The contribution of the upper and lower face in happy and sad facial expression classification. *Vision Research, 50*(18), 1814– 1823. https://doi.org/10.1016/j.visres.2010.06.002

Cohn, N. (2013). *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images.* London, UK: Bloomsbury.

Cohn, N. (2016). A multimodal parallel architecture: A cognitive framework for multimodal interactions. *Cognition, 146,* 304–323.https://doi.org/10.1016/j.cognition.2015.10.007

Cohn, N. (2018). Visual Language Theory and the scientific study of comics. In Wildfeuer, Janina, Alexander Dunst, Jochen Laubrock (Ed.). *Empirical Comics Research: Digital, Multimodal, and Cognitive Methods*. (pp. 305-328) London: Routledge.

De Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1535), 3475–3484.https://doi.org/10.1098/rstb.2009.0190

De Silva, P. R., & Bianchi-Berthouze, N. (2004). Modeling human affective postures: an

information theoretic characterization of posture features. *Computer Animation and Virtual Worlds*, *15*(34), 269–276. https://doi.org/10.1002/cav.29

Forceville, C. (2011). Pictorial runes in Tintin and the Picaros. *Journal of Pragmatics*, *43*(3), 875–890. https://doi.org/10.1016/j.pragma.2010.07.014

Gray, K. L. H., Adams, W. J., Hedger, N., Newton, K. E., & Garner, M. (2013).Faces and awareness: Low-level, not emotional factors determine perceptual dominance. *Emotion*, *13*(3), 537–544. https://doi.org/10.1037/a0031403

Griffiths, S., Rhodes, G., Jeffery, L., Palermo, R., & Neumann, M. F. (2018). The average facial expression of a crowd influences impressions of individual expressions. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(2), 311–319. https://doi.org/10.1037/xhp0000446

Halit, H., de Haan, M., Schyns, P., & Johnson, M. (2006). Is high-spatial frequency information used in the early stages of face detection? *Brain Research*, *1117*(1), 154–161. https://doi.org/10.1016/j.brainres.2006.07.059

Hyde, J., Carter, E. J., Kiesler, S., & Hodgins, J. K. (2016). Evaluating Animated Characters: Facial Motion Magnitude Influences Personality Perceptions. *ACM Transactions on Applied Perception*, *13*(2), 1–17.https://doi.org/10.1145/2851499

Ito, H., Seno, T., & Yamanaka, M. (2010). Motion Impressions Enhanced by Converging Motion Lines. *Perception*, *39*(11), 1555–1561. https://doi.org/10.1068/p6729

Itou, J., Matsumura, K., Munemori, J., & Babaguchi, N. (2019). A Comic-Style Chat System with Japanese Expression Techniques for More Expressive Communication. *Collaboration Technologies and Social Computing*, 172–187. https://doi.org/10.1007/978-3-030-28011-6_12

Jaiswal, D. P., Kumar, S., & Badr, Y. (2020). Towards an Artificial Intelligence Aided Design Approach: Application to Anime Faces with Generative Adversarial Networks. *Procedia Computer Science*, *168*, 57–64. https://doi.org/10.1016/j.procs.2020.02.257

Kawabe, T., & Miura, K. (2006). Representation of dynamic events triggered by motion lines and static human postures. *Experimental Brain Research*, *175*(2), 372–375. https://doi.org/10.1007/s00221-006-0673-6

Kendall, L. N., Raffaelli, Q., Kingstone, A., & Todd, R. M. (2016). Iconic faces are not real

faces: enhanced emotion detection and altered neural processing as faces become more iconic. *Cognitive Research: Principles and Implications*, *1*(1). https://doi.org/10.1186/s41235-016-0021-8

Kendall, L. N., Raffaelli, Q., Todd, R., Kingstone, A., & Cohn, N. (2020). Show me how you feel: Iconicity and systematicity in visual morphology. In P. Perniss, O. Fischer, & C. Ljungberg (Eds.), *Operationalizing Iconicity* (pp. 214-229). (Iconicity in Language and Literature series). John Benjamins Publishing Company. https://doi.org/10.1075/ill.17.13ken

Kennedy, J. M. (1982). Metaphor in Pictures. *Perception*, *11*(5), 589–605. https://doi.org/10.1068/p110589

Kukkonen, K. (2013). *Studying Comics and Graphic Novels*. Chichester: WileyBlackwell.

Lankes, M., & Bernhaupt, R. (2011). Using embodied conversational agents in video games to investigate emotional facial expressions. *Entertainment Computing*, *2*(1), 29–37. https://doi.org/10.1016/j.entcom.2011.03.007

Lipp, O. V., Price, S. M., & Tellegen, C. L. (2009). No effect of inversion on attentional and affective processing of facial expressions. *Emotion*, *9*(2), 248– 259. https://doi.org/10.1037/a0014715

Liu, K., Chen, J. H., & Chang, K. M. (2019). A Study of Facial Features of American and Japanese Cartoon Characters. *Symmetry*, *11*(5), 664. https://doi.org/10.3390/sym11050664

McCloud, S. (2006) *Making Comics: Storytelling Secrets of Comics, Manga andGraphic Novels*. New York: HarperCollins.

Meeren, H. K. M., van Heijnsbergen, C. C. R. J., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences*, *102*(45), 16518–16523. https://doi.org/10.1073/pnas.0507650102

Ojha, A., Forceville, C., & Indurkhya, B. (2021). An experimental study on the effectof emotion lines in comics. *Semiotica*, *2021*(243), 305–324. https://doi.org/10.1515/sem-2019-0079

Piepers, D. W., & Robbins, R. A. (2012). A Review and Clarification of the Terms "holistic," "configural," and "relational" in the Face Perception Literature. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00559

Prazak, E. R., & Burgund, E. D. (2014). Keeping it real: Recognizing expressions in real compared to schematic faces. *Visual Cognition*, *22*(5), 737–750. https://doi.org/10.1080/13506285.2014.914991

Prince, S., & Hensley, W. E. (1992). The Kuleshov Effect: Recreating the Classic Experiment. *Cinema Journal*, *31*(2), 59. https://doi.org/10.2307/1225144

Reschke, P. J., & Walle, E. A. (2021). The Unique and Interactive Effects of Faces, Postures, and Scenes on Emotion Categorization. *Affective Science*, *2*(4), 468– 483. https://doi.org/10.1007/s42761-021-00061-x

Sullivan, D., & Critten, J. (2014). Adventures in Research: Creating a video game textbook for an information literacy course. *College & Research Libraries News*, *75*(10), 570–573. https://doi.org/10.5860/crln.75.10.9215

Schouwstra, S. J., & Hoogstraten, J. (1995). Head Position and Spinal Position as Determinants of Perceived Emotional State. *Perceptual and Motor Skills*, *81*(2), 673–674. https://doi.org/10.1177/003151259508100262

Seno, T., Ueda, S., Takeichi, M., Palmisano, S., & Ohtsuka, S. (2013). Adding Vertical Lines to a Face Increases Perceived Sadness. *VISION: The Journal of the Vision Society of Japan*, *25*(1), 1–7.

Stamenković, D., Tasić, M., & Forceville, C. (2018). Facial expressions in comics: an empirical consideration of McCloud's proposal. *Visual Communication*, *17*(4), 407–432.

Volkova, E. P., Mohler, B. J., Dodds, T. J., Tesch, J., & Bülthoff, H. H. (2014). Emotion categorization of body expressions in narrative scenarios. Frontiers in Psychology, 5. https://doi.org/10.3389/fpsyg.2014.00623

Wessler, J., & Hansen, J. (2021). Facial mimicry is independent of stimulus format: Evidence for facial mimicry of stick figures and photographs. Acta Psychologica, 213, 103249. https://doi.org/10.1016/j.actpsy.2020.103249

Zhang, S., Liu, X., Yang, X., Shu, Y., Liu, N., Zhang, D., & Liu, Y. J. (2021). The Influence of Key Facial Features on Recognition of Emotion in Cartoon Faces. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.687974

Contact details

Olga Rubtsova

National Research University Higher School of Economics (Moscow, Russia), Laboratory for Cognitive Psychology of Digital Interfaces Users. Junior Research Fellow

olga.rubtsova98@gmail.com

Elena S. Gorbunova

National Research University Higher School of Economics (Moscow, Russia), School of Psychology, Laboratory for Cognitive Psychology of Digital Interfaces Users. Associate professor, Laboratory head

esgorbunova@hse.ru

Financing

This article is an output of a research project implemented as a part of Basic Research Program at the National Research University Higher School of Economics (HSE University).

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.