Individual Differences Are More Important Than The Emotional Category For The Perception Of Emotional Expressions

Elena Moltchanova, Christoph Bartneck

University of Canterbury

Private Bag 4800, 8140 Christchurch

New Zealand

elena.moltchanova@canterbury.ac.nz, christoph.bartneck@canterbury.ac.nz
Abstract

Emotional facial expression are an important communication channel between artificial characters and their users. Humans are trained to perceive emotions. Robots and virtual agents can use them to make their inner states transparent. Literature reported that some emotional types, such as anger, are perceived as being more intense than others. Other studies indicated that gender influences the perception. Our study shows that once the individual differences amongst participants are included in the statistical analysis, then the emotion type has no further explanatory power. Artificial characters therefore should adapt to their specific users.

Keywords: emotion, expression, gender, individual differences
Individual Differences Are More Important Than The Emotional Category For The Perception Of Emotional Expressions

**Introduction**

In recent years we have observed a dramatic increase in interactive technology that utilizes emotional facial expressions to communicate with the users. This includes highly human like androids, such as the Geminoid series, but also more comic-like characters, such as iCat (Bartneck, Reichenbach, & Breemen, 2004). Facial expressions are already part of our daily messaging culture in the form of Emojis. The assumption is that humans are already trained to recognize human facial expressions and hence artificial characters could utilize these to communicate their inner state or to form a social bond with the users.

Although facial expression recognition has been a subject of thorough study for many years, some issues still remain unresolved. For example, the universality of basic expressions is still under discussion (Russel, 1994). The factors influencing the perceived intensity of a particular emotion are also subject of many recent studies. For example, (Sonnenmans & Frijda, 1994) have found that the relation between the five specific intensity factors, namely (1) duration of the emotion and delay of its onset and peak; (2) the magnitude of perceived bodily changes; (3) frequency of recollection and re-experience of the emotion; (4) strength and severity of action tendency, and drasticness of actual behaviour; (5) magnitude of belief changes and influence on long-term behaviour, and the overall felt intensity differed among emotions.

Sonnenmans & Frijda, 1995 further reported that, for example, the negative emotions are felt more intensely, while (Hess, Blairy, & Kleck, 1997; Calvo & Nummenmaa, 2016; Wilhelm, Hildebrandt, Manske, Schacht, & Sommer, 2014) found that the overall recognition accuracy was highest for happy expressions and lowest for sad expressions, implying that some emotions need to be expressed more intensely to be recognized as such. Biele & Grabowska, 2006 have also found that the angry faces were generally judged to be more intense than the happy ones. Recognition of fear appears to be particularly problematic, as reported by (Rapcsak et al., 2000), for both, the
healthy subjects and for patients with focal brain damage. Bartneck et al. (2004) confirmed the difficulty of expressing fear and also described the relationship between the geometrical expression of emotions and their perception (Bartneck & Reichenbach, 2005). Kamachi et al., 2013 reported that suprised faces were better recognized than happy, sad or angry faces.

In addition to the differences in average intensity associated with each of the six basic facial emotional expressions (Anger, Disgust, Fear, Happiness, Sadness, Surprise), many studies have reported differences in facial expression processing and recognition for men and women as well as between people of different ages (McClure, 2000; Calder et al., 2003). Biele & Grabowska, 2006 report that the differences between girls and boys were apparent very early on. Although women in general tend to demonstrate greater accuracy when interpreting non-verbal clues, women were found to be better at recognizing facial expressions of fear and sadness, while men were reported to be superior at identifying expressions of anger. Similarly, while age in general seemed to correlated with less accurate perception of emotional expressions, recognition of sadness and disgust appeared to be better in older people. Calder et al., 2003 state that the above differences are small, but consistent.

There is also evidence for cultural differences in emotion recognition (Matsumoto & Ekman, 1989). Moreover the identification of emotions displayed by robots or avatars may depend on the cultural perception of robots’ helpfulness and friendliness in general (Becker-Asano & Ishiguro, 2011; Koda, 2007). Dynamic perception, i.e., facial emotion recognition from animation, appeared to be better than static perception (Biele & Grabowska, 2006). In addition to the above trends, there is also continuously accumulating evidence for individual differences as reported by, among others, (Sonnemans & Frijda, 1995) and (Suzuki, Hoshino, & Shigemasu, 2006).

A better understanding of emotion perception, facial expression processing and recognition is important not only for the study of human psychology. It is also increasingly vital in the field of the human-computer and human-robot interaction, where the facial expression recognition has to occur in real time. Whereas traditionally
robots were created for physically demanding and dangerous tasks, and were meant to
operate far away from humans, now they have a range of applications in, for example,
health care and entertainment, which brings them into increasing contact with people
(Breazeal, 2003). It is thus important to understand how to make the robots to mimic
and to elicit various emotions, and therefore to determine the factors which affect their
perception (see, for example, (Broadbent, Stafford, & MacDonald, 2009). Hwang, Park,
& Hwang, 2013 have found that even the shape of a humanoid robot has an effect on
the human perception of the robot’s personality. Breazeal, 2003 has looked at the
emotions elicited either by static images or by video recording of the Kismet robot. The
dynamic assessment was found to be somewhat more accurate than the static one:
57%-86% correct vs. 47%-83% emotion-specific identification respectively. In a similar
experimental framework, McColl & Nejat, 2014 looked at the accuracy of perception of
emotion expressed by a human like robot and by a human actor. They also report
interpersonal variability in emotional perception as well as the difference between the
two agents.

Perhaps the best way to summarize all of the above is to use the words of Suzuki
et al., 2006: “one of the most widespread characteristics of emotional experience is the
striking nature of the variability among individuals”. In this study, we turn to the
emotional assessment of the faces of LEGO Minifigures. In their survey of socially
interactive robots, Fong, Nourbakhsh, & Dautenhahn, 2003 classify the robots into four
broad categories: anthropomorphic, zoomorphic, caricatured, and functional. The
cartoon-type faces of the LEGO Minifigures are the good example of the caricature.
There are several hundred different LEGO faces offering an extremely wide spectrum of
emotional expressions. Understanding the emotional perception of these LEGO
Minifigure faces can thus contribute to the understanding of the emotional perception of
robots and avatars from the caricatured category. In this study, we have used a sample
of LEGO Minifigure faces to addresses the following research questions:

1. Are some emotional categories perceived as more intense than others?

2. Do men perceive emotions differently than women?
3. To what extent do individual variations influence the facial expression processing?

In the process of answering these questions we do not only contribute to the existing body of knowledge on variability of emotions perception, but also highlight the importance of correct application of statistical methodology, namely accounting for person- and figurine-specific random effects.

**Method**

We conducted an experiment in which participants had to rate the perceived emotional expression of LEGO Minifigures. We then analyzed the responses received to answer the research questions. The advantage of using LEGO Minifigures is that they offer an extreme wide spectrum of emotional expression. There are several hundred different LEGO faces. By sampling from this population we ensure not to introduce any specific bias into the stimuli.

**Participants**

Sixty participants, comprising of 22 men and 38 women of an average age of 38.3 years (SD=12.5 yrs) were recruited for the study. The participants were recruited on the campus of the University of Canterbury.

**Process**

The participants were welcomed to the study and were seated in front of a computer. After reading the instructions and signing the consent form the participants could ask question to the experimenter. If the participants had no more questions the experiment started. After providing demographic information, the participants were tasked to rate the exact same set of 94 LEGO Minifigures using the computer in front of them (see figure 1). After completing the task the participant were debriefed and had again the opportunity to ask any questions they might have. The experiment took approximately 30 minutes to complete.
Stimuli

While many facial expression recognition studies use photographs of human facial expressions, in this study we have used a set of 94 LEGO Minifigures. These facial expressions are designed to cover a large variety of expressions and intensity levels. Figure 2 shows some examples of Minifigures used in the study. The Minifigures were randomly selected from a larger set of 722 Minifigures used in a previous study (Bartneck, Obaid, & Zawieska, 2013). The previous study focused on the historical development of the LEGO Minifigures and found that their expressions have become far more diverse and that in particular angry faces have become more frequent. This study turns the table and we are now investigating how the characteristics of the emotions and the participants influenced the perception of the LEGO Minifigures.

Measurements

Each Minifigure had to be rated by the participants to represent one of six emotions (Anger, Disgust, Fear, Sadness, Happiness, Surprise) with one out of five intensities (weak (1) to intense (5)). The questionnaire only allowed for one combination of emotional category and intensity. If, for example, the participant first considers a face to express fear at intensity level 4 and then changes his/her mind to surprise at intensity level 3, then only the later would be recorded. Thus, a total of $60 \times 94 = 5640$ responses with only 2 missing values were recorded.

Results

The frequency distribution of the responses is shown in Table 1. Most of the figurines were classified as either happy (49%) or angry (21%), and rated to be 2 or 3 on the intensity scale (27% and 24% respectively). The observed average intensity was highest for Fear (3.35) and lowest for Disgust (2.79). It should be noted, that assignment of emotions to Minifigures was not necessarily unanimous: 44 out of 92 Minifigures were assigned each of the six emotions at least once, and further 31 were assigned all but one emotion at least once. In fact, there was a lot of disagreement...
between raters with respect to both the emotion and the intensity of the emotion perceived. Most of the Minifigures (81 out of 94) had all possible emotional intensity ratings assigned at least once. However, the 'dominant' emotional intensity, i.e., the rating assigned most often, had on average 25 participants agreeing with it. The Fleiss’ Kappa statistic for emotions was evaluated at 0.369 and for intensities at 0.124, indicating weak agreement. This indicates, that it is important to account for inter-rater as well as inter-minifigure variability when performing statistical analysis.

**Statistical Methods**

In order to investigate association between intensity and type of emotion, a mixed-effects cumulative proportional odds ordinal logistic model was fitted. (Agresti, 2010). Let $Y_{ij}$ be the intensity assigned by rater $i$ to minifigure $j$. Then the cumulative probability distribution of the intensity conditional on emotion can then be modeled as:

$$h(Pr(Y_{ij} \leq y|X_{ij} = k)) = \alpha_y + \omega_k + \beta_s \text{sex} + \beta_a \text{age} + \xi_i + \phi_j,$$

(1)

where $k$ is the emotion assigned by rater $i$ to minifigure $j$, $\beta_s$ and $\beta_a$ account for the effects of sex and age respectively, $\xi_i$ and $\phi_j$ are rater- and minifigure-specific random effects, accounting for repeated measures, $\omega_k$ is the effect, capturing association between emotion and intensity, and $\alpha_y$ is the intensity category-specific intercept. The link function $h$ was chosen to be logit as is customary. For the purposes of model identifiability, $\omega_1 = 0$ and *Anger* is considered the reference emotion.

Since preliminary data analysis indicated the presence of substantial rater- and minifigure-specific variation, quantification of the sources of variance was of interest. (Martina Mittlböck & Shemper, 1996) and (Menard, 2000) discuss some ways to measure the proportion of variation attributable to various factors in binary and multinomial logistic regressions without specifically discussing the case of ordinal response. We have chosen to use the likelihood ratio based $R^2_L$ defined as follows:

$$R^2_L = \frac{-2(ln(L_0) - ln(L_M))}{-2 \ln(L_0)} = \frac{ln(L_0) - ln(L_M)}{ln(L_0)},$$

(2)

due to the fact that it naturally varies between 0 and 1 and has a proportional reduction error interpretation (Menard, 2000).
In order to analyse statistical significance of a particular factor, a model with and without the factor can be compared using the likelihood ratio test, in which case the $\chi^2$ statistic and the associated p-value can be reported.

All the analyses were implemented using R-software (R Core Team, 2014). The *ordinal* package was used for fitting the model 1 (Christensen, 2015).

The results of the estimation of the full model, adjusted for sex and age-group and including rater- and minifigure-specific random effects are shown in Table 2 and Figure 3. Women were found to perceive emotions as more intense than men ($p=0.0314$) and raters aged 30-49 were found to perceive emotions as more intense than raters aged 15-29. Although there were slight differences in the perceived intensity of different emotions, these differences were not found to be statistically significant ($\chi^2 = 1.2623, p = 0.9388$). It should be noted that in a fixed effects model, which does not take into account repeated assessment set-up, the differences in the perceived emotional intensity come out as highly statistically significant ($\chi^2 = 84.783, p < .0001$).

Besides the obvious lesson of the importance of correctly adjusting for random effects, this results brings out the importance of the inter-rater and inter-minifigure variation, which was then investigated using the $R^2_L$ coefficient, the results for which are shown in Table 3. The ratio $R^2_L$ indicates proportional improvement, i.e., increase in likelihood due to consecutive addition of various factors to the null model $h(Pr(Y_{ij} \leq y|X_{ij} = k)) = \alpha_y$. The largest proportional increase is due to the accounting for inter-rater and inter-minifigure variability. While accounting for the type of emotion after adjusting for age and sex leads to approximately 1% improvement in the likelihood, adding random effects leads to additional $(e^{0.1996} - 1) \times 100\% = 22\%$ improvement in likelihood.

**Discussion**

Sixty men and women of different age were asked to assign a particular emotion and the associated intensity to each of the 94 LEGO Minifigures. The results confirm earlier findings that women tend to perceive emotional expressions generally more
intensely, and that older people (aged over 30 y.o.) tend to perceive emotions as less intense than the younger people (under 30 y.o.). However, once the model accounted for the Minifigure- and rater-specific effects, no difference was found in the emotion-specific intensity distribution. The rater-specific variation was found to constitute a substantial part of the variance observed in the response.

This striking variability has already been reported earlier by, for example, (Sonnemans & Frijda, 1995) and (Suzuki et al., 2006). Our study confirms the necessity of always taking it into account when analysing or otherwise modeling facial expression perception. In our case, failure to adjust for the random effects would have resulted in incorrect conclusion of statistically significant effect of emotion on intensity perception.

Although we have mentioned the previously reported importance of culture on perception of emotions (Matsumoto & Ekman, 1989; Becker-Asano & Ishiguro, 2011; Koda, 2007), we were unable to account for the cultural background in this study. Extending the number of participants and controlling for ethnical and cultural background may provide further insights into the extent of interpersonal variation in emotional perception.

In some studies, such as (Breazeal, 2003) and (McColl & Nejat, 2014), it is possible to speak of accuracy of emotional perception in the sense of the user correctly perceiving the emotion that the agent, whether the human actor or the robot, was meant to express. However, we need to point out that there is no ground truth in the expression and perception of emotions. In fact, as our results show, the emotion appears to be really in the eye of the beholder. We used the established approach of inviting a large sample of participants to rate stimuli using a Likert scale. It should also be noted that the stimuli used in our experiment were static photographs of a certain style of faces. We were somewhat surprised to find the weak agreement between the respondents as to the emotion and the associated intensity represented by each LEGO Minifigure despite the fact that the facial expressions of Minifigures are fairly simple and highly stylized. The results for animated facial expression or for static, but more realistic human faces (e.g. photographs) could be different. It would be enlightening to see
whether the variability in user-specific perception is likely to be a general phenomenon
independent of the agent (human actor, humanoid robot, stylized avatar etc.) and the
manner of presentation (static recording, dynamic recording, real time interaction).
While there are some indications that it is so (see, for example (Breazeal, 2003) and
(McColl & Nejat, 2014)), more studies are needed to provide a definitive conclusion.

Robots and computers are becoming more and more part of our lives. Their roles
differ widely but are expected to include teaching and caring (Broadbent et al., 2009).
Even though robots might not be specifically designed to express emotions, the users
are likely to perceive them anyway. Facial emotional expression can play an important
role in communicating the robots emotional states. Hence it does pay to consider the
emotional messages robots send, and that is what many researchers are doing (see, for
example, (Bonarini, 2016)). However, given the users’ high individual variability of the
perception of emotions, which this study serves to confirm, it does seem necessary for
the artificial character to adapt to each specific user instead of attempting generic
expressions.
References


Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S.


<table>
<thead>
<tr>
<th>Emotion</th>
<th>Intensity</th>
<th>Total (%)</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>181 271 264 251 245</td>
<td>1212 21%</td>
<td>3.08 (1.35)</td>
</tr>
<tr>
<td>Disgust</td>
<td>88 191 173 124 56</td>
<td>632 11%</td>
<td>2.79 (1.17)</td>
</tr>
<tr>
<td>Fear</td>
<td>33 66 67 67 89</td>
<td>322 6%</td>
<td>3.35 (1.35)</td>
</tr>
<tr>
<td>Happiness</td>
<td>463 771 664 504 340</td>
<td>2742 49%</td>
<td>2.81 (1.27)</td>
</tr>
<tr>
<td>Sadness</td>
<td>56 77 67 63 30</td>
<td>293 5%</td>
<td>2.77 (1.27)</td>
</tr>
<tr>
<td>Surprise</td>
<td>56 123 63 82 62</td>
<td>437 8%</td>
<td>2.93 (1.24)</td>
</tr>
<tr>
<td>Total</td>
<td>877 1499 1349 1091 822</td>
<td>5838 100%</td>
<td></td>
</tr>
<tr>
<td>(%)</td>
<td>15% 27% 24% 19% 15%</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

Table 1

*Observed distribution of emotion and intensity ratings of 92 LEGO Minifigures by 60 raters*
<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thresholds:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α₁</td>
<td>-2.2106</td>
<td>0.5375</td>
<td>-4.113</td>
<td></td>
</tr>
<tr>
<td>α₂</td>
<td>-0.1215</td>
<td>0.5367</td>
<td>-0.226</td>
<td></td>
</tr>
<tr>
<td>α₃</td>
<td>1.4903</td>
<td>0.5370</td>
<td>2.775</td>
<td></td>
</tr>
<tr>
<td>α₄</td>
<td>3.2632</td>
<td>0.5386</td>
<td>6.059</td>
<td></td>
</tr>
<tr>
<td><strong>Other coefficients:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digust vs. Anger</td>
<td>ω₂</td>
<td>0.0682</td>
<td>0.1029</td>
<td>0.662</td>
</tr>
<tr>
<td>Fear vs. Anger</td>
<td>ω₃</td>
<td>0.0333</td>
<td>0.1460</td>
<td>0.228</td>
</tr>
<tr>
<td>Happiness vs. Anger</td>
<td>ω₄</td>
<td>-0.0240</td>
<td>0.1001</td>
<td>-0.240</td>
</tr>
<tr>
<td>Sadness vs. Anger</td>
<td>ω₅</td>
<td>0.0970</td>
<td>0.1398</td>
<td>0.694</td>
</tr>
<tr>
<td>Surprise vs. Anger</td>
<td>ω₆</td>
<td>0.0376</td>
<td>0.1368</td>
<td>0.275</td>
</tr>
<tr>
<td>Women vs. Men</td>
<td>βₛ</td>
<td>0.6442</td>
<td>0.2993</td>
<td>2.152</td>
</tr>
<tr>
<td>aged 30-49 vs. 15-29</td>
<td>β₁</td>
<td>-0.7447</td>
<td>0.3358</td>
<td>-2.218</td>
</tr>
<tr>
<td>aged 50+ vs. 15-29</td>
<td>β₂</td>
<td>-0.6796</td>
<td>0.4460</td>
<td>-1.524</td>
</tr>
</tbody>
</table>

Table 2

*The results of estimating model 1.*
Table 3

Variance explained by various models as compared to the null model

\[ h(Pr(Y_{ij} \leq y|X_{ij} = k)) = \alpha_y + \beta_{sex} + \beta_{age} \]

\[ h(Pr(Y_{ij} \leq y|X_{ij} = k)) = \alpha_y + \omega_k + \beta_{sex} + \beta_{age} \]

\[ h(Pr(Y_{ij} \leq y|X_{ij} = k)) = \alpha_y + \omega_k + \beta_{sex} + \beta_{age} + \xi_i \]

\[ h(Pr(Y_{ij} \leq y|X_{ij} = k)) = \alpha_y + \omega_k + \beta_{sex} + \beta_{age} + \xi_i + \phi_j \]

<table>
<thead>
<tr>
<th>Model</th>
<th>( R^2_L )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h(Pr(Y_{ij} \leq y</td>
<td>X_{ij} = k)) = \alpha_y + \beta_{sex} + \beta_{age} )</td>
</tr>
<tr>
<td>( h(Pr(Y_{ij} \leq y</td>
<td>X_{ij} = k)) = \alpha_y + \omega_k + \beta_{sex} + \beta_{age} )</td>
</tr>
<tr>
<td>( h(Pr(Y_{ij} \leq y</td>
<td>X_{ij} = k)) = \alpha_y + \omega_k + \beta_{sex} + \beta_{age} + \xi_i )</td>
</tr>
<tr>
<td>( h(Pr(Y_{ij} \leq y</td>
<td>X_{ij} = k)) = \alpha_y + \omega_k + \beta_{sex} + \beta_{age} + \xi_i + \phi_j )</td>
</tr>
</tbody>
</table>

The ratio indicates proportional improvement, i.e., increase in likelihood due to addition of various factors. The largest proportional increase is due to the accounting for inter-rater and inter-minifigure variability.

---

Figure 1. Screenshot of the computer based questionnaire.
(a) A standard Minifigure  
(b) A movie actor

*Figure 2.* Example Minifigures
Figure 3. Observed proportions of emotional intensity, on a scale from 1 to 5, by the attributed emotion.