Paladyn, J. Behav. Robot. 2019; 10:436–437

DE GRUYTER

Erratum

Gabriele Trovato*, Josue G. Ramos, Helio Azevedo, Artemis Moroni, Silvia Magossi,
Reid Simmons, Hiroyuki Ishii, and Atsuo Takanishi

Erratum to "A receptionist robot for Brazilian people: study on interaction involving illiterates"

Abstract: The original article was published in Paladyn, Journal of Behavioral Robotics, 2017, 8(1), 1-17, https://doi.org/10.1515/pjbr-2017-0001. The aim of this erratum is to report the results of a new analysis of the data.

Following a request by a colleague, we would like to report the revised results below. These corrections also contain more detailed data, in their proper APA notation, and do not substantively change the results and the discussion.

In **Section 3.3 ("Results")** of the original article one could read:

1. "being human does not seem to be significantly preferred to machine (p = 0.35 according to Mann-Whitney U-test, and p = 0.45 according to the Chi-squared goodness of fit test, confirming the null hypothesis of equal distribution)."

which should be replaced by:

"being human does not seem to be significantly preferred to machine ($X2$(1, N = 48) = 0.45, p = 0.502 according to the Chi-square goodness of fit test, confirming the null hypothesis of equal distribution)."

2. "No significant preference is either found in the category Voice (p = 0.11) through Kruskal-Wallis test, just that high pitch voice is less preferred (p = 0.03) through Chi-squared."

*Corresponding Author: Gabriele Trovato: Faculty of Science and Engineering, Waseda University, #41-304, 17 Kikui-cho, Shinjuku-ku, Tokyo 162-0044, Japan;
E-mail: gabriele@takanishi.mech.waseda.ac.jp
Josue G. Ramos, Helio Azevedo, Artemis Moroni, Silvia Magossi, Reid Simmons, Hiroyuki Ishii, Atsuo Takanishi: Faculty of Science and Engineering, Waseda University, #41-304, 17 Kikui-cho, Shinjuku-ku, Tokyo 162-0044, Japan

which should be replaced by:

"No significant preference is either found in the category Voice (3x2 contingency table Chi-square analysis: $X2$(2, N = 48) = 1.94, p = 0.38), just that high pitch voice is less preferred (4x2 contingency table Chi-square analysis: $X2$(3, N = 48) = 11.84, p = 0.008)."

3. "In the category Behaviour, Polite is significantly higher than all the other characteristics (p < 0.001). On the other hand, no preference is found for Direct answer against Detailed neither through Mann-Whitney (p = 0.29) or Chi-squared (p = 0.27))."

which should be replaced by:

"In the category Behaviour, Polite is significantly higher than all the other characteristics ($X2$(3, N = 48) = 59.69, p < 0.0001). On the other hand, no preference is found for Direct answer against Detailed ($X2$(1, N = 48) = 1.21, p = 0.271)."

In **Section 4.3.1 ("Semantic scales")** of the original article one can read:

1. "Gathered data were analysed using the Kruskal-Wallis test [43] in case of multiple groups of data and Mann-Whitney U-test [44] in case of two groups of data."

which should be replaced by:

"Gathered data were analysed using Wilcoxon test [43] or Mann-Whitney U-test [44]."

2. "Differential semantic scales were validated calculating Cronbach's alpha, a coefficient of internal consistency used to estimate the reliability of a psychometric test. Results were: 0.848 for Anthropomorpism; 0.904 for Animacy; 0.835 for Likeability; 0.883 for Perceived intelligence and 0.854 for Familiarity. We conclude that all scales are consistent."

which should be replaced by:

"Differential semantic scales were validated calculating Cronbach's alpha, a coefficient of internal consistency used to estimate the reliability of a psychometric test. Results were, respectively for Ana and Kobiana: 0.82 and 0.85 for Anthropomorphism; 0.89 and 0.92 for Animacy; 0.86 and 0.81 for Likeability; 0.87 and 0.91

for Perceived intelligence and 0.76 and 0.86 for Familiarity. We conclude that all scales are consistent."

3. "Comparing Ana with Kobiana, significant differences are found for Group EV in categories Anthropomorphism (p = 0.003); Likeability (p = 0.021); Familiarity (p = 0.002); Good as receptionist (p = 0.001); and Appropriate voice (p = 0.005), all in favour of Ana. Conversely, for Group CL only category Anthropomorphism is significant (p < 0.0001)."

which should be replaced by:

"Comparing Ana with Kobiana with Wilcoxon test, significant differences are found for Group EV in categories Anthropomorphism (Z = 2.59, p = 0.01); Likeability (p = 0.021); Animacy (Z = 2.01, p = 0.044); Likeability (Z = 2.98, p = 0.003); Familiarity (Z = 4.05, p < 0.001); Good as receptionist (Z = 2.76, p = 0.006); and Appropriate voice (Z = 3.09, p = 0.002), all in favour of Ana. Conversely, for Group CL only category Anthropomorphism is significant (p < 0.0001)."

4. "Apparently, Group CL perceived significantly higher than EV the categories Likeability, Perceived intelligence, Familiarity and Good as receptionist for both robots (p < 0.001 in all cases).

Instead, Anthropomorphism was a different case, as Kobiana was judged significantly (p < 0.001) less anthropomorphic by Group CL compared to group EV."

which should be replaced by:

"Apparently, Group CL perceived significantly higher than EV the categories Likeability, Perceived intelligence, Familiarity and Good as receptionist for both robots. Through Mann-Whitney test for Ana: Likeability: U = 82.5, p < 0.0001; Perceived intelligence: U = 176.5, p = 0.0004; Familiarity: U = 133.5, p < 0.0001. For Kobiana: Likeability: U = 81, p < 0.0001; Perceived intelligence: U = 90, p < 0.0001; Familiarity: U = 98, p < 0.0001.

Instead, Anthropomorphism was a different case, as Kobiana was judged significantly (U = 40, p < 0.0001) less anthropomorphic by Group CL compared to group EV."

In **Section 4.3.3 ("Voice")** of the original article one can read:

1. "Human voice was rated significantly higher (p < 0.001) than the robotic one. However, it was judged less appropriate for the appearance of Kobiana (p = 0.03), while appropriate for Ana (p = 0.001)."

which should be replaced by:

"Human voice was rated significantly higher (Ana: U = 98.5, p = 0.004; Kobiana: U = 124, p = 0.031) than the robotic one. However, it was judged less appropriate for the appearance of Kobiana (U = 117.5, p = 0.021), while appropriate for Ana (U = 80, p = 0.001)."

2. "The second analysis consists in crossing results regarding voice with results on explicit preference. Calculations of Spearman's Rho reveal a significant (r = 0.46; p = 0.002) correlation between the ratings of explicit preference with the differences of the ratings of the voices of the two receptionists.

Finally, another correlation was found between Familiarity and Good voice: Spearman's test highlighted a significant correlation (r = 0.29; p = 0.007). This means that any of the two receptionists, whenever associated with a bad voice, were rated more uncanny. This correlation is stronger in case of Kobiana (r = 0.52; p = 0.006)"

which should be replaced by:

"The second analysis consists in crossing results regarding voice with results on explicit preference. Calculations of Spearman's Rho reveal a significant ($r(38)$ = 0.46, p = 0.0028) correlation between the ratings of explicit preference with the differences of the ratings of the voices of the two receptionists.

Finally, another positive correlation was found between Good Voice and Familiarity: This means that a receptionist, whenever associated with a bad voice, was rated more uncanny. This correlation is strong in case of Kobiana ($r(38)$ = 0.52, p = 0.0006)."

Finally, in **Section "References"** of the original article there is the following Ref. [43]:

[43] W. H. Kruskal and W. A. Wallis, Use of Ranks in One-Criterion Variance Analysis, Journal of the American Statistical Association, 47, 260 (1952) 583–621, 1952.

which should be replaced by:

[43] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bulletin, 1945, 1(6), 80–83

References

[1] G. Trovato et al., A receptionist robot for Brazilian people: study on interaction involving illiterates, Paladyn, Journal of Behavioral Robotics, 2017, 8(1), 1–17