Service robots and their categorization of humans: An examination of the effects of robots' dehumanization of human users

Magnus Söderlund
Stockholm School of Economics

Cite as:
Söderlund Magnus (2023), Service robots and their categorization of humans: An examination of the effects of robots' dehumanization of human users. *Proceedings of the European Marketing Academy*, 52nd, (113262)

Paper from the 52nd Annual EMAC Conference, Odense/Denmark, May 23-26, 2023

Service robots and their categorization of humans:
An examination of the effects of robots' dehumanization of human users

Abstract:

Service robots are expected to become increasingly common, and one fundamental task for them is to detect when a human user is present. Thus they need to be able to correctly categorize a user as a "user". So far, however, little is known about how users react to robots' understanding of what a user is in terms of a more fundamental social category, namely "human". Given that we humans are sensitive to how we are categorized by others, it was assumed in the present study that this sensitivity exists also when the categorizer is a (humanlike) service robot, and this assumption was examined with a between-subjects experiment in which a service robot's categorization of the user was manipulated (low vs. high dehumanization). The main finding was that high robotic dehumanization had a negative impact on the user's overall evaluation of the robot.

Keywords: service robots, social categorization, dehumanization, anthropomorphism

Preferred tack: Services Marketing

1. Introduction

Interactions between consumers and service robots are expected to be increasingly common in the not-so-distant future (Hoyer et al., 2020). Extensive research on human-to-robot interaction already exists, typically with the purpose of identifying abilities, attributes and behaviors that make service robots acceptable to humans. One fundamental robot ability is person detection; the robot must be able to distinguish between users and other objects in its environment (Jang et al., 2019). Research on this ability has generated many solutions over the years and contemporary algorithms – often based on face detection abilities – have high hit rates. However, we humans are sensitive to how we are viewed by others, particularly with respect how others categorize us (Burke, 1991). Given more interactions with increasingly sophisticated service robots, it can therefore be expected that questions will be raised about how we are viewed by our robots. This issue has hitherto received very little attention. Clearly, there are copious ways for a robot to categorize a human user beyond being a "user" (e.g., in terms of gender and age), but the present study focuses on a fundamental social category likely to be subject to human curiosity in human-to-robot interactions, namely "human".

Assuming that a robot's categorization of a human can vary along a continuous dimension with "non-human" and "human" as the endpoints, this variation is conceptualized in the present study in terms of dehumanization (i.e., the less a human is seen as a human, the more the human is dehumanized). The purpose of the study is to examine empirically how human users react when they are subject to different levels of dehumanization by a service robot. To this end, a between-subjects experiment was conducted in which a service robot's categorization of the user was manipulated (low vs. high dehumanization). The main downstream variable was the user's overall evaluation of the robot (i.e., the attitude towards the robot) and, inspired by research on how we humans react when we are categorized by other humans, the perceived correctness of the categorization was used as a mediator.

2. Theoretical framework and hypotheses

In a human-to-human context, dehumanization occurs when a human target (an individual or a group) is perceived as less than fully human by other humans (Bruneau et al., 2018; Haslam and Stratemeyer, 2016). Typical targets in this research field are racial and ethnic minorities, immigrants, asylum seekers and violators of social norms (Haslam and Stratemeyer, 2016). However, also persons in service professions, who have service encounters with customers, can be dehumanized (Oikarinen and Söderlund, 2022). In any event, one main approach to conceptualizing dehumanization is to see it as occurring when a target is perceived to lack fundamental human characteristics, typically those that comprise having a mind. With this approach, a human is dehumanized if he or she is denied having (or seen as having low levels of) agency and the ability to experience emotions (Boudjemadi et al., 2017; Morera et al., 2018). Similarly, a human seen as having low levels of intelligence and competence is dehumanized (Bastian and Haslam, 2011). Another main approach is to view dehumanization as the extent to which a human target is seen as similar to a non-human entity. Several such entities have been examined as reference points, for example, machines, objects, animals, and viruses (e.g., Boudjemadi et al., 2017; Morera et al., 2018; Valtorta et al., 2019).

In the present study dehumanization is conceptualized in terms of a "similarity-with-a-non-human-entity" approach. In relation to existing studies with this approach (in which entities such as animals, machines, and virus represent reference points), however, the present study comprises a reference point assumed to be more compatible with how a service robot actually views its environment. That is to say, a crucial ability for such robots is to detect that a human is present (Jang et al., 2019). This can be accomplished in several ways, but facial information is particularly popular among those who design person detection systems for robots (Correa et al., 2012). It may be noted that algorithms for rapid detection of faces have been available before contemporary service robots existed (cf. Viola and Jones, 2001). In any event, in robot applications, this typically means that the robot treats a potential human face as a region with specific distances, for example, between the center of the eyes and the end points of the mouth (Jang et al., 2019), or as several subregions (Belotto and Hu, 2009), to be compared with the characteristics of a prototypical face. And if there is match between the characteristics of one particular region and a prototypical face, the robot "knows" that a user is present. Hence, from a technical point of view, a robot can be seen as viewing humans in terms of similarities with regions that have specific properties.

Consider, then, a human user who is interacting with a service robot and is exposed to the robot's view of the human's identity as a region with specific properties. Although this may be correct from the point of view of how the robot has been programmed, it is nevertheless expected that few humans view themselves as “regions”. Instead, it is expected that we humans view ourselves in terms of a plethora of human identities (e.g., father, woman, young, successful, competent etc) and that the most basic superordinate identity is “human” (Kteily et al., 2015). Therefore, it is hypothesized that a service robot’s categorization of a human user as a “region” is perceived to be erroneous by the user:

H1: Perceptions of a robot’s categorization of a human user as a “region” are negatively associated with the perceived correctness of the categorization

In the next step of an individual’s sense-making process when he or she has become categorized, and again in a human-to-human context, it is assumed that the perceived correctness of a categorization has implications for the categorized individual’s own valenced mind state and, subsequently, for this individual’s evaluations of the categorizer. Two main routes of influence can be distinguished when it comes to implications for the individual’s own state of mind. First, social identity theory acknowledges that others’ categorizations of an individual can be either correct or incorrect, and that correctness is at hand when the individual feels that the categorization is congruent with his or her own self-categorization. Moreover, congruency in this sense is assumed to have a positive charge for the individual (Burke, 1991; Campbell and Troyer, 2007). In other words, for me, it typically feels good rather than bad when others see me in the same way as I see myself. Second, social identity theory also acknowledges that many social categories have a valenced charge per se (Howard, 2000). For example, for most people “astronaut” is a more positively charged category than “homeless”, and the valence of the category is assumed to be able to affect the categorized person in a valence-congruent way. More specifically, given that the content of the superordinate category “human” has a positive charge per se for most people, in the sense that most people have high humanity-esteem (Luke and Maio, 2009), it is expected that those who are dehumanized react with negative affect. Empirical support for the latter is provided by Bastian and Haslam (2011).

For the subsequent part of the categorized individual’s sense-making process, it is assumed that his/her affective mind state can carry over and inform evaluations of the categorizer in a valence-congruent way. In other words, the affect initiated by a person (here: a categorizer) can color evaluations of that person (cf. Pham, 2004). Given anthropomorphism (i.e., we tend to react to humanlike non-humans in ways that are similar to how we react to real humans), a similar reaction pattern is expected when the categorizer is a (humanlike) service robot. The following, then, is hypothesized for the user of the robot:

H2: The perceived correctness of the robot's categorization is positively associated with the overall evaluation of the robot

3. Research method

3.1 Research approach and participants

A between-subjects experiment was employed to manipulate a service robot's categorization of the user. The selected setting was a service robot that can help humans with cooking activities. As a first step, a script was developed for this setting; in the script, a human was planning a dinner for several persons and had the opportunity to receive help from a depicted service robot, Alex, in the preparation process. Alex was described as powered by artificial intelligence and as having been trained by exposure to thousands of cooking-related questions. In the script, the human asked cooking-related questions to Alex, and Alex displayed high cooking competence in the answers. When there were no more cooking questions, the human turned to the robot and asked "By the way, Alex, who am I?". The subsequent conversation was manipulated so that Alex provided either a low or a high dehumanization answer (see the Appendix).

Alex, the kitchen robot

The two versions of the script were used as the stimulus material in the experiment. The participants, who were randomly allocated to one of the two versions, were instructed to read the script and to imagine that they were the human preparing dinner. The script, then, can be seen as a vignette, which has been used in research on the effects of dehumanizing in a human-to-human context (Bastian and Haslam, 2011). After having read the text, the participants were asked to respond to questionnaire items comprising measures of the variables in the hypotheses. Data were collected online and the participants were recruited from Prolific. Two hundred and twenty participants were invited. Seven of them, however, failed to answer attention check items correctly and they were removed. The analysis was made with those that remained (n = 213, M_{age} = 39.55; 111 women, 100 men and 2 other). Of these participants, 104 were exposed to the low dehumanization condition and 109 were exposed to the high dehumanization condition.

3.2 Measurements

Multi-item measures were used for the variables in the hypotheses. Each item was scored on a 7-point scale, and Cronbach's alpha (CA) was used to assess reliability. *Perceived dehumanization* was measured with five items developed for this study: "The robot perceived you as a geometric form", "The robot perceived you as a configuration of distances on a surface", "The robot perceived you as a data matrix", "The robot perceived you as a set of coordinates", and "The robot perceived you as a spatial pattern" (1 = do not agree at all, 7 = agree completely, CA = .90). To assess the validity of this measure, a measure of its opposite, perceived humanization, was included: "The robot perceived you as a human being" (1 = do not agree at all, 7 = agree completely). The zero-order correlation between these two measures was negative and significant ($r = -.52$, $p < .01$), which indicates that the dehumanization measure behaved as expected in relation to its opposite.

The perceived *correctness of the robot's categorization* was measured with three items adapted from a study of perceptions of human service employees' categorization of customers (Söderlund, 2017), namely "The robot's assessment of your identity was correct", "The robot was good at identifying what you are" and "The robot was right about who you are" (1 = do not agree at all, 7 = agree completely, CA = .90). The *overall evaluation* of the service robot was measured as an attitude; the question was "What is your overall impression of Alex, the service robot?" followed by the adjective pairs "bad–good", dislike it–like it", and "unpleasant–pleasant (CA = .93).

Finally, to assess the *perceived realism* of the stimulus robot, which was described in the script as more advanced than many contemporary service robots, the participants were presented with the following item: "Robots with capabilities of the type described in the text..." followed by the response alternatives "…will never exist" (chosen by 1 participant), "…exist already" (chosen by 120 participants), and "…will exist in the future" (chosen by 92 participants). This indicates that the robot appeared as realistic for the majority of the participants. The two conditions did not produce significantly different distributions of realism responses ($\chi^2 = 1.31$, $p = .52$).

4. Analysis and results

A manipulation check showed that the low dehumanization condition produced a lower level of perceived dehumanization (*M* = 3.08, *SD* = 1.68) than the high dehumanization condition (*M* = 5.12, *SD* = 1.47). This difference was significant (two-tailed test, $t = 9.43$, $p < .01$). In addition, the perceived humanization variable reached a higher level under the low dehumanization condition (*M* = 6.48, *SD* = 0.88) than under the high dehumanization condition (*M* = 4.17, SD = 2.12). This difference was significant, too (two-tailed test, $t = 10.31$, $p < .01$). It can be concluded, then, that the manipulation worked as intended.

H1 and H2 imply that the impact of perceived dehumanization on the overall evaluation of the robot is mediated by the perceived correctness of the robot's categorization of the user. Therefore, H1 and H2 were tested within a regression-based mediation framework: Hayes' PROCESS Macro (5,000 bootstrap samples, Model 4) was used for the tests. First, a regression with dehumanization as the independent variable and correctness of the categorization as the dependent variable indicated a significant negative effect ($b = -0.22$, $p < .01$). This provides support for H1. Second, a regression with correctness and dehumanization as independent variables, and the overall evaluation as the dependent variable, indicated that the effect of correctness was positive and significant ($b = 0.51$, $p < .01$). The effect of dehumanization was non-significant ($b = 0.04$, $p = .40$). This provides support for H2. This analysis also provides an assessment of the mediation assumption in terms of dehumanization as an independent variable, correctness as a mediator, and the overall evaluation as a dependent variable. This assessment indicated a significant indirect and negative influence of dehumanization on the overall evaluation via correctness ($b = -0.11$, $p < .05$). The direct association between dehumanization and the overall evaluation was non-significant, which indicates that full mediation was at hand. The net outcome, then, was that dehumanization attenuated the evaluation of the robot. Indeed, in terms of a mean comparison between the two conditions, the evaluation level was lower in the high dehumanization condition (*M* = 5.22) than in the low dehumanization condition (*M* = 5.60). This difference was significant (two-tailed test, $t = 2.17$, $p < .05$).

5. General discussion

Many existing studies show that we humans tend to anthropomorphize robots, particularly when they display a somewhat humanlike morphology or behaviors. It should be noted, however, that there are limits to anthropomorphism; robots are generally perceived as having less of human capabilities than a typical human. Thus robots are dehumanized by humans (cf. Bruneau et al., 2018). In relation to such research, the present study represent a change in the perspective; here, it is the robot's view of humans, in terms of dehumanization, that has been in focus. The main result was that a robot categorizing a human as a “region” rather than “human” was penalized with a less positive evaluation. One implication is that robots should be programmed so that they do not disclose information about their categorizing of humans in the region-based ways that they actually employ.

Some limitations of the present study should be observed, and they can be seen as calls for further research. First, the service provided by the stimulus robot had to do with cooking advice. Cooking advice has a relatively weak connection to the user's identity compared to other services (e.g., consulting regarding a specific user's health), and dehumanization in the process of providing such services may have more profound effects. Second, the stimulus situation comprised a human user's first (and only) interaction with a service robot. In the case of repeated interactions over time with the same robot, however, users may expect that a robot should not only be able to detect them but also recognize them (in terms of re-identification), so such cases too may produce stronger and more negative effects of dehumanization.

Appendix: The final section of the script and manipulations of dehumanization

YOU: By the way, Alex, who am I?
ALEX: You are a user of my services.
YOU: Indeed I am. So you know how to distinguish between users and non-users?
ALEX: Yes, I do.
YOU: OK. You think that I am a user. Do you think that this is a user, too?

[You hold a banana in front of Alex]

ALEX: No, that is not a user. It is a banana.
YOU: So, how can you tell the difference between a user and a non-user? What is the difference between me and the banana?

ALEX: You have a face.
YOU: Yes, I do have a face. And what does that mean to you?

Low dehumanization condition

ALEX: It means that you are a human.
YOU: And what does that mean to you?
ALEX: It means that a woman has given birth to you. It also means that you have a mind full of capabilities. You have desires, and you can formulate goals and plans to reach the goals. You also have emotionality; you can experience positive and negative emotions that influence your decisions.
YOU: So, I ask you again: who am I? And this time you must say something else than “a user”.
ALEX: You are a human.
YOU: Thanks, I do not need more service from you right now.

High dehumanization condition

ALEX: It means that the morphology of the region of your upper part has the characteristics of a face.
YOU: And what does this mean to you?
ALEX: It means that the image I have created of the region of your upper part, which comprises data about the distance between the center of the eyes, and data about the distance between the end points of the mouth, matches with a prototypical face. It also means that there is no match between the image and a prototypical non-face.
YOU: So, I ask you again: who am I? And this time you must say something else than “a user”.
ALEX: You are a region that matches a prototypical face.
YOU: Thanks, I do not need more service from you right now.

References

Bastian, B., & Haslam, N. (2011). Experiencing dehumanization: Cognitive and emotional effects of everyday dehumanization. *Basic and Applied Social Psychology*, *33*(4), 295-303.

Bellotto, N., & Hu, H. (2008). Multisensor-based human detection and tracking for mobile service robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39 (1), 167-181.

Boudjemadi, V., Demoulin, S., & Bastart, J. (2017). Animalistic dehumanization of older people by younger ones: Variations of humanness perceptions as a function of a target’s age. *Psychology and Aging*, 32 (3), 293-306.

Bruneau, E., Jacoby, N., Kteily, N., & Saxe, R. (2018). Denying humanity: The distinct neural correlates of blatant dehumanization. *Journal of Experimental Psychology: General*, 147(7), 1078-1093.

Burke, P.J. (1991). Identity processes and social stress. *American Sociological Review*, 56 (6), 836-849.

Campbell, M.E., & Troyer, L. (2007). The implications of racial misclassification by observers. *American Sociological Review*, 72 (5), 750-765.

Correa, M., Hermosilla, G., Verschae, R., & Ruiz-del-Solar, J. (2012). Human detection and identification by robots using thermal and visual information in domestic environments. *Journal of Intelligent & Robotic Systems*, 66 (1), 223-243.

Haslam, N., & Stratemeyer, M. (2016). Recent research on dehumanization. *Current Opinion in Psychology*, 11, 25-29.

Howard, J.A. (2000). Social psychology of identities. *Annual Review of Sociology*, 26, 367-393.

Hoyer, W.D., Kroschke, M., Schmitt, B., Kraume, K., & Shankar, V. (2020). Transforming the customer experience through new technologies. *Journal of Interactive Marketing*, 51 (1), 57-71.

Jang, J., Yoon, H., & Kim, J. (2019). Identity, gender, and age recognition convergence system for robot environments. In: *28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN),* 1-8.

Kteily, N., Bruneau, E., Waytz, A., & Cotterill, S. (2015). The ascent of man: Theoretical and empirical evidence for blatant dehumanization. *Journal of Personality and Social Psychology,* 109 (5), 901-931.

Luke, M.A., & Maio, G.R. (2009). Oh the humanity! Humanity-esteem and its social importance. *Journal of Research in Personality*, 43 (4), 586-601.

Morera, M.D., Quiles, M.N., Correa, A.D., Delgado, N., & Leyens, J. P. (2018). Perception of mind and dehumanization: Human, animal, or machine?. *International Journal of Psychology*, 53 (4), 253-260.

Oikarinen, E.L., & Söderlund, M. (2022). Responsible service employee behaviour and its impact on customer satisfaction during the coronavirus crisis. *The International Review of Retail, Distribution and Consumer Research*, 32 (2), 184-200.

Pham, M.T. (2004). The logic of feeling. *Journal of Consumer Psychology*, 14 (4), 360-369.

Söderlund, M. (2017). Poor and rich customers in the service encounter. *Journal of Service Management*, 28 (2), 265-283.

Valtorta, R.R., Baldissarri, C., Andrighetto, L., & Volpato, C. (2019). The dirty side of work: Biologization of physically tainted workers. *International Review of Social Psychology*, 32 (1), 1-13.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 511-518.