AN ONLINE EXPERIMENT ON THE STEREOTYPE CONTENT MODEL (SCM) AND CHATBOTS – DOES SWAPPING THE PICTURE MAKE A DIFFERENCE?

Roger Seiler and Sonja Wüest
Zurich University of Applied Sciences
St.-Georgen-Platz 2, CH-8400 Winterthur, Switzerland

ABSTRACT
Chatbots have become quite popular in recent years, and research on the topic has gained momentum. In this study, we address whether the stereotype content model applies in the context of chatbots. Furthermore, we consider whether swapping the image of a person in the chatbot interface makes a difference in terms of the SCM. An online experiment was conducted to A/B-test different chatbot versions, and the results suggest that the SCM applies. Furthermore, a chatbot can be perceived as warm and competent, with friendliness, in particular, fostering online trust. This, in turn, is a business-success-relevant construct in an online setting. Therefore, if a trustworthy chatbot is to be implemented, the SCM is considered relevant and necessary for implementing such a bot.

KEYWORDS
Chatbots, Stereotype Content Model, Trust, Online Experiment, A/B-Testing

1. INTRODUCTION
Since ELIZA (Weizenbaum, 1966) and A.L.I.C.E. (Wallace, 2009), chatbots have further evolved and become integrated as voice assistants in operating systems. Market size and growth estimates are high, and the field of application ranges from healthcare and banking to retail (Intelligence, n.d.). Furthermore, an extensive analysis of patents highlights their importance and how this technology is moving towards supporting and assisting customers (Pantano & Pizzi, 2020). This results in a shift from personal support to an automated and more technical touch (Pantano & Pizzi, 2020). A more virtual and digital customer service approach is emerging, so the question of how social aspects should be addressed arises. Personalisation and social presence are constructs related to this (Verhagen et al., 2014). The more human the bot appears, the more people can relate to it (Wilson et al., 2017). Because relational aspects (e.g. trust) can play a role in business success and e-commerce (Beatty et al., 2011; Kim et al., 2008; Kim & Peterson, 2017; Küktsch & Penez, 2015) and prior research has an emphasis on technical aspects (Pantano & Pizzi, 2020), this research contributes to the chatbot literature by testing whether the SCM applies in the given context. Furthermore, the question can be answered by swapping the chatbot picture to create a chatbot personality that users and potential customers perceive as positive.

This research builds on the work of Fiske et al. (2007, 2002) and Cuddy et al. (2008), which stated that universal dimensions of warmth and competence exist and when these meet robots, it is a double-edged sword (Tay et al., 2014). Furthermore, the SCM was introduced by Casciaro and Lobo (2005) based on the research previously mentioned.

This paper contributes to the current literature by applying the SCM in a chatbot context and conducting an online experiment (A/B-test) to ascertain whether the model also holds. This is especially relevant to e-commerce as purchase behaviours have changed due to the pandemic (Gao et al., 2020). Valaskova et al. (2021) commented that shopping patterns have changed, and 65% of consumers intend to maintain these new purchasing habits in the future. Furthermore, online service and support are applications where chatbots have replaced humans by drawing on deep learning, natural language processing (NLP), natural language
understanding (NLU), and natural language generation (NLG) to interact with customers (Mohamad Suhaili et al., 2021). Building online trust is essential, as is clarifying the relevance of the model, since it can be used to guide companies seeking to work with chatbots. The following section provides an overview of the relevant concepts through a literature review.

2. LITERATURE

Empirical research suggests that trust is a key success factor in online commerce (Beatty et al., 2011; Kim et al., 2008; Kim & Peterson, 2017; Köksal & Penez, 2015). This is especially relevant since the question of trusting the faceless and intangible takes place online (Beldad et al., 2010). Linked to this is the aspect of social presence that can arise when artefacts with social cues are perceived by users or customers (Verhagen et al., 2014). Social presence and social cues are both relevant when considered in the context of purchase intentions (Botha & Reynneke, 2015; Lu et al., 2016).

The SCM (Casciaro & Lobo, 2005) extends the work of Fiske and Cuddy by creating a model based on the two dimensions of sympathy and competence to categorize stereotypes (see Figure 1). Casciaro and Lobo proposed (according to the two dimensions in their model) a stereotypical “incompetent jerk” (low competence and low sympathy), a “competent jerk” (high competence but low sympathy), a “lovable fool” (high sympathy but low competence) and a “lovable star” (high sympathy and high competence) (Casciaro & Lobo, 2005). When picking who to work with, people tend to opt for the lovable star (Casciaro & Lobo, 2005). Furthermore, Casciaro and Lobo (2005) point out that the sympathy dimension affects the building up of trust. Indeed, this is in line with van der Holst who stated that this holds for e-commerce, but in health care, it is competence that fosters trust (Holst, 2021). This model has found its application in education (Niemiec & Ryan, 2009), brand management (Aaker et al., 2012), health care (Drevs, 2013), information technology (Franklin et al., 2013), nonprofit organisations (Aaker et al., 2010), and sales (Zawisza & Pittard, 2015). Furthermore, these dimensions also seem to hold in the context of human-robot interaction (HRI) (Christoforakos et al., 2021; Tay et al., 2014).

![Figure 1. SCM by Casciaro & Lobo (2005)](image)

![Figure 2. Conceptual Model](image)

Trust can develop when visiting websites (Ba & Pavlou, 2002; Lee et al., 2015; McAllister, 1995; D. H. McKnight et al., 2002; H. D. McKnight et al., 2002) and hence, can play a role when interacting with artefacts (Beatty et al., 2011; Kim et al., 2008; Kim & Peterson, 2017) or a specific technology (Mcknight et al., 2011). Furthermore, past research has shown that trust not only builds in human relationships but also when interacting with robots (Christoforakos et al., 2021). Even in mediated communication in general, trust can be formed (Pan & Steed, 2016). Trust builds upon first perceptions (Hampton-Sosa & Koufaris, 2005), and past empirical research suggests that images can help build trust (Steinbrück et al., n.d.). This is in line with Wood et al. (2008), who found evidence that verbal and non-verbal cues can foster trust in car sales based on a perception of the sympathy and competence of the salesperson.

Social presence plays a role in electronic mediated communication because an avatar displayed in a chat box has positive effects on perceived interactivity, social support, and trust (Chattaraman et al., 2014). Mediated communication can lead to the perception of social cues as well as the feeling of being close to the other party (Biocca et al., 2003).
Based on the work of Wood et al. (2008), Fiske et al. (2007, 2002), Cuddy et al. (2011, 2008, 2007) and Casciaro and Lobo (2005), we constructed the conceptual model (see Figure 2) for this research. Additionally, the hypotheses were derived from earlier research (see Table 1 for an overview).

Table 1. Overview of the hypotheses

<table>
<thead>
<tr>
<th>#</th>
<th>Text</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Text communication influences perceived sympathy.</td>
<td>Accept</td>
</tr>
<tr>
<td>2</td>
<td>Text communication influences perceived competence.</td>
<td>Accept</td>
</tr>
<tr>
<td>3</td>
<td>Social presence influences perceived sympathy.</td>
<td>Accept</td>
</tr>
<tr>
<td>4</td>
<td>Social presence influences perceived competence.</td>
<td>Reject</td>
</tr>
<tr>
<td>5</td>
<td>Avatar (picture) influences perceived sympathy.</td>
<td>Reject</td>
</tr>
<tr>
<td>6</td>
<td>Avatar (picture) influences perceived competence.</td>
<td>Reject</td>
</tr>
<tr>
<td>7</td>
<td>Sympathy influences trust.</td>
<td>Accept</td>
</tr>
<tr>
<td>8</td>
<td>Competence influences trust.</td>
<td>Accept</td>
</tr>
</tbody>
</table>

3. METHODOLOGY

The chatbot was implemented as a mock-up using BotSociety software. The pictures were tested in a previous study where students were asked to assign images to the corresponding stereotypes of the SCM. The “lovable fool” and the “incompetent jerk” pictures drawn from the earlier study were used here as stimulus material. Furthermore, text from a previous study by Rozumowski et al. (2017) was used. User responses were identical. The layout of the chatbot is designed to resemble a messaging app communication (e.g., Facebook Messenger). Both chatbots have the same name to avoid potential bias, and both chatbots introduce themselves as financial advisors.

Table 2. Overview of the chatbots and the online experiment procedure

| 1. E-mail with an invitation to the experiment. |
| 2. Starting page with welcome and scenario. |
| 3. Ice-breaker questions to test knowledge about chatbots. |
| 4. Randomization (Chatbot 1 [Lovable Star] / Chatbot 2 [Incompetent Jerk]). |
| 5. Sympathy |
| 6. Competence |
| 7. Trust |
| 8. Social presence |
| 9. Affinity to technology |
| 10. Socio-demographic questions |

Chatbot “lovable star” Chatbot “incompetent jerk”

The scales in the conceptual model are drawn from existing literature for reasons of rigour as well as comparability to prior research. Sympathy, competence, and trust are based on the scales used in McCroskey and Teven (1999) and the scale for social presence as in Bailenson et al. (2001).
In addition, control questions were used (e.g. hair colour and if the chatbot image was wearing glasses or not). Additionally, affinity to technology was measured as a control variable using the scales from Grabner et al. (2016). A manipulation check was implemented using the SCM, and participants had to categorise their Avatar according to this. If they matched, the manipulation check was rated as successful.

A pre-test was conducted before the final invitation e-mail was sent to participants. Before the experiment started, they were given a short text about their scenario. Here, they were asked to watch the interaction with a chatbot as they would like to know more about a particular pension plan. Randomisation was used to determine which chatbot the participant saw (an overview of the procedure is listed in Table 2).

Data analysis was conducted with SPSS 25 and before data analysis, cleansing and recoding of reverse-coded items took place. Missing values were labelled as such to exclude them from further analysis, and the answer given was compared to the scales and values below or above the scale (seven-point scales were used). One case was identified and deleted from the data set.

4. RESULTS AND DISCUSSION

The survey was online for 18 days and 936 student participants were contacted, of which 156 opened the invitation link contained in the e-mail. On the welcome page, 25 participants aborted, and a further 19 did not continue the survey after being asked the knowledge questions about chatbots. A further 32 people chose not to continue after being randomly assigned to a chatbot. This left 76 participants that completed the questionnaire (N=76).

Of the 76 participants, 47 are females and 29 are males. 53.9% of the participants are students (of which 65.8% were employed part-time). 75% of the participants have a Bachelor’s degree or higher, and 57.9% live in an urban region. 47.4% already have a pension plan. 68.4% of the participants consider themselves technology-competent, while 80.1% understand what “chatbot” means (both answers were self-reported); however, only 21.5% use chatbots (this low figure is not necessarily out of anxiety since 73.7% reported no security concerns regarding the technology).

The “lovable star” group had an average age of 27.4 years (SD = 4.2), and 57.9% were female. On the other hand, the average age of the “incompetent jerk” group was 28.9 years (SD = 5.3), and 65.8% were female. A Chi-square test did not show any significant group differences regarding age, gender, type of employment, education, place of residence (rural/urban), or having a pension plan. Therefore, the randomised assignment to the groups (Chatbots 1 and 2) was successful, and further analysis could be conducted.

The reliability analysis was conducted by calculating Cronbach’s Alpha values. All the Cronbach’s Alpha values of the constructs used in this study were above the recommended cut-off value of .7 stated in Nunnally (1978). Hence, a group comparison was conducted using the Mann-Whitney U-Test. Significant group difference can be reported for sympathy: U = 516, r = .25, and p = .032; competence: U = 242, r = .57 and p < .001; and trust: U = 425, r = .36, and p = .002. Therefore, the “lovable star” was perceived as more sympathetic, competent, and trustworthy than the “incompetent jerk”. However, the social presence construct shows a non-significant group difference: U = 671, r = .06, and p = .592.

<table>
<thead>
<tr>
<th>Construct (# items)</th>
<th>Cronbach’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sympathy (6)</td>
<td>.872</td>
</tr>
<tr>
<td>Competence (6)</td>
<td>.863</td>
</tr>
<tr>
<td>Trust (5)</td>
<td>.823</td>
</tr>
<tr>
<td>Social presence (5)</td>
<td>.735</td>
</tr>
<tr>
<td>Affinity to technology (5)</td>
<td>.725</td>
</tr>
</tbody>
</table>

An overview of the hypotheses is given in Table 1. Regarding sympathy a cross-table analysis (>= 5 = high sympathy, <5 = low sympathy) and Chi-square test were conducted \( \chi^2 (1, N = 76) = 13.625, p = .000 \). Cramer’s V was calculated, and the value of .423 p = .000 is in the middle range, and in terms of the Odds Ratio the “lovable star” was 6.2 times more sympathetic than the “incompetent jerk”. Hence, H1 is accepted.
Competence was analysed in the same way. $\chi^2 (1, N = 76) = 21.111$, $p = .000$. Cramer’s V was calculated, the value of .527 $p = .000$ was in a strong range, and in terms of the Odds Ratio the “lovable star” was 10.5 times more competent than the “incompetent jerk”. Hence, H2 is also accepted.

Spearman’s correlation coefficients were calculated to test for the effects of social presence on perceived sympathy. The Spearman’s correlation coefficient (two-tailed) was $r(74) = .240$, $p = .036$, which is a weak but significant correlation. Consequently, H3 is accepted.

Spearman’s correlation coefficients were also calculated to test for the effects of social presence on perceived competence. The Spearman’s correlation coefficient (two-tailed) was $r(74) = -.044$, $p = .708$, which is very weak and there was no significant correlation. Hence, H4 is rejected.

The question regarding the Avatar and his appearance was used to test H5. Participants who did not know or were unable to state the Avatar’s hair colour or if he was wearing glasses were put into the group that did not perceive the Avatar. All participants answering the questions correctly were put into the group that consciously perceived the Avatar. The Chi-square test was not significant for either the “lovable star”: $\chi^2 (1, N = 76) = .016$, $p = .584$ or the “incompetent jerk”: $\chi^2 (1, N = 76) = 3.031$, $p = .298$. Hence, no effect of the appearance of the Avatar on sympathy was reported, and H5 is accordingly rejected.

The same test was then conducted for H6 to determine whether the Avatar’s appearance influenced perceived competence. The Chi-square test was calculated for the “lovable star”: $\chi^2 (1, N = 76) = 0.215$, $p = .643$ and for the “incompetent jerk”: $\chi^2 (1, N = 76) = 1.406$, $p = .236$. Because no significant effect was reported, H6 must also be rejected.

Regression analysis was used in a further analysis of the data. Sympathy as an independent variable and trust as a dependent variable were included in the model ($R^2 = .53$, $\beta = .675$, $p < .001$). Hence, H7 is accepted. Competence as an independent variable and trust as a dependent variable were included in the model ($R^2 = .41$, $\beta = .493$, $p < .001$). Hence, H8 is accepted.

The null hypothesis that communication does not influence the perception of chatbots can be rejected. However, this study suggests an effect on sympathy and competence and, in turn, trust. The text used seems to influence perceived sympathy as well as competence, while social presence has a significant effect on sympathy but not on competence.

The rejection of H5 and H6 is not in line with empirical evidence from earlier research and Avatar appearance was deemed not to influence perceived sympathy or competence. Since sympathy and competence positively affect trust, the SCM applies in the context of chatbots.

5. IMPLICATIONS, LIMITATIONS, AND FURTHER RESEARCH

The results of this research suggest that social presence and the chatbot text have a positive effect on the perceived sympathy of a chatbot. Therefore, companies seeking to implement a chatbot perceived as competent and likeable should consider the SCM when implementing and testing their chatbots. Furthermore, they are advised to pay attention to details such as text, which affects sympathy and competence. In times of artificial intelligence and platforms to implement conversational agents (e.g. Azure, Amazon Web Services, Google Cloud, IBM Bluemix, and others), it could make a difference to customer perception of chatbots if the text is adapted or trained in line with the SCM. This aspect affects both sympathy and competence – and ultimately trust. Therefore, according to empirical evidence (Beatty et al., 2011; Kim & Peterson, 2017), this is business-success-relevant because it influences online customer decisions (Kim et al., 2008).

The data analysed in this study suggest that swapping a picture is insufficient since the Avatar’s image influenced neither sympathy nor competence. However, further analysis is needed to explain why this was the case here as past research suggests that changing a picture may be all that is required. One possible reason is that social presence has a more significant effect than the image itself.

It must be noted that the sample in our study comprising students is not representative of the entire population. However, one advantage for us was that this younger age group was generally familiar with the concept of chatbots. Indeed, earlier research does not categorically reject the use of student sampling such as MTurk or convenience adult samples (Krupnikov & Levine, 2014). Students samples are considered legitimate provided they are adequate (e.g. in early-stage research) as conducted in this paper (Ferber, 1977). Students samples can also be used for initial marketing research (Ashraf & Merunka, 2017).
This study presented chatbot interaction to participants in an online experiment, and the results may have been different in a field study (in the context of an actual business transaction and interaction). Furthermore, the results may differ depending on the products used in the exchange. In this study, a financial setting and scenario were used for prior research, and these were not changed for reasons of comparability. Despite this, later findings may differ since the importance of sympathy and competence (and their effects on trust) depend on the context (see van der Holst (2021) for differences in e-commerce and health care).

As this study did not rely on real-life customer interaction, further research could conduct a field study where user-interaction is actively implemented into the study design (or by using a live chatbot). Since past literature suggests that pictures are quite powerful when evaluating sympathy and competence in the context of the SCM, a possible further experiment could extend the research by adding an additional group of chatbots without images. This could shed light on this subject and may lead to insights into why some studies show effects and others do not.

Furthermore, this research used statistical methods that do not account for direct and indirect effects. Consequently, a possible further research opportunity would be to replicate this study and use structural equation modelling (SEM) and other suitable statistical analysis methods to account for direct and indirect effects that might be present when researching in the context of the SCM.

6. CONCLUSION

The results of this study suggest that the SCM holds in the context of chatbots. Furthermore, perceived competence and sympathy are antecedents to building trust. This is important when humans interact with artefacts, such as visiting an online store or seeking assistance from a company’s chatbot. As we have already stressed, trust is generally relevant to business success, so it makes sense to have artefacts fostering trust in line with the SCM.

This study suggests that swapping a picture is insufficient. Indeed, constructs such as social presence have to be considered as well when implementing a customer experience (CX) in line with the SCM. This study also suggests social cues and text play a role; therefore, it is advisable to introduce these aspects into a chatbot in addition to an image. Because cloud platforms provide a fast and easy way to do this, it is essential to pay attention to the whole interaction and, for example, to customise the text in terms of the SCM.

As subjective and complex customer perception involves testing whether the chatbot and its cues are perceived, these can make the difference between an artefact being perceived as a “lovable star” or an “incompetent jerk”. The latter would have a negative impact on business, whereas the former would contribute to business success.

REFERENCES


Holst, E.M. van der, 2021. Designing for Trust: A neuroscientific and design perspective on trust within a professional context. https://doi.org/10.3990/1.9789036551762


