H Exploring Digital Realities empirically

Research

H.1 Who gets the fame, who is to blame? Empirical exploration of responsibility attribution in HCI

Lara Christoforakos, Ewgenia Borodkow, Sarah Diefenbach
Ludwig-Maximilians-Universität München, Department Psychology

1 Introduction

Innovative technologies, such as self-driving cars, social robots for assisted living or AI digital coaches, become increasingly autonomous and can be seen as active cooperation partners. Users cooperate with such technologies to achieve a certain outcome. This development places new emphasis on the question of responsibility: to what extent do users perceive themselves or the technology responsible for a certain interaction outcome? From a psychological perspective the subjective responsibility distribution in human-computer interaction (HCI) could have far reaching consequences. For example, if a fatal business decision can be attributed to the "smart" recommender software, employees may take higher risks than adequate. On the other hand, a lacking feeling of responsibility could reduce the experience of self-efficacy (e.g., Bandura, 1977) and lastly performance. Therefore, it is important to look into the attribution of responsibility as a phenomenon in HCI and investigate influencing factors. One of such factors could be the perceived autonomy of the technology and the user (e.g., van der Woerdt & Haselager, 2019), but also technology design factors such as human-likeness (e.g., Hinds, Roberts & Jones, 2004). Within our study we looked into a possible trade-off in responsibility attribution between user and technology regarding a certain interaction outcome as well as influencing factors for such an attribution, i.e. perceived autonomy of the user and the technology respectively and anthropomorphic technology design. In the following sections we summarize the theoretical background and hypotheses of our study, then present the methods and results, and finally discuss our findings, limitations and implications of our research.

2 Theoretical Background

According to the psychological self-determination theory, one factor influencing the responsibility people attribute to their counterpart for a behavioral outcome is perceived autonomy, i.e. the "inner endorsement of one's actions, the sense that they emanate from oneself and are one's own" (Deci & Ryan, 1987, p. 1025). This basic principle of human interaction also seems to apply for contexts of HCI. For example, studies have shown that participants attributed more responsibility to a computer that behaved autonomously (by providing real-time advice through an interface agent), compared to a computer that behaved non-autonomously (by providing a help menu) (Serenko, 2007).

In another study, a robot's perceived agency (measured by questions about the robot's control over the situation and its ability to make its own decisions) had a significant effect on the responsibility attributed to the robot for his actions (van der Woerdt & Haselager, 2019). Based on such findings and theory, we assume the following. **H1***:* The higher the perceived autonomy of the technology, the more responsibility users attribute to the technology for the interaction outcome.

When interacting with technologies, users' perceived own autonomy and responsibility for an outcome may vary relative to their perception of the technology's autonomy and responsibility (e.g., Berberian, Sarrazin, Le Blaye & Haggard, 2012; Kim, Chen & Zhang, 2016). As known from the phenomenon of diffusion of responsibility, the feeling of responsibility for an outcome can decrease with an increasing number of people involved in a social situation (Werth & Mayer, 2008). Assuming that social phenomena from human interaction are relevant for HCI (Nass & Moon, 2000; Reeves & Nass, 1996), this may also include diffusion of responsibility, i.e., if a technology is perceived a social counterpart, users will attribute some degree of responsibility for what is happening to the technology and feel less responsible themselves. Additionally, the perceived responsibility of others should correlate with perceived autonomy – if someone's behavior is externally controlled they may not be held responsible for the outcome. Based on the assumed interrelation of autonomy and responsibility, the following hypotheses are derived. **H2**: The more responsibility users attribute to the technology, the less responsibility they attribute to themselves. **H3**: The higher the perceived autonomy of the technology, the lower the users' own perceived autonomy.

HCI research furthermore implies that human-likeness in a technology can foster attribution of responsibility for an outcome. Anthropomorphic products have been found to be attributed more responsibility when compared to non-anthropomorphic products (e.g., Hinds et al., 2004). Based on these results, the following hypotheses are derived. **H4**: Users who interact with a technology with anthropomorphic design cues make higher responsibility attributions to the technology than users who interact with a technology without anthropomorphic design cues. **H5**: The interrelation of anthropomorphic design cues and responsibility attribution to the technology is mediated by the perceived autonomy of the technology.

3 Methods

Within our online study participants interacted with a design software to create a personalized birthday card. The software existed in two variations: an anthropomorphic (i.e. an avatar accompanying the usage of the design software) and a non-anthropomorphic (i.e. textual instructions during the usage of the design software), representing two experimental conditions.

Within the anthropomorphic condition, instructions were given by an avatar called Tom, presented in a speech bubble, and worded in first-person perspective (e.g. "Hi, I'm Tom! I am an easily operated design software. You can design birthday cards with my aid. Let's go"!). In the non-anthropomorphism condition, instructions were given as plain text and worded in third-person perspective (e.g. "This is an easily operated design software. You can design birthday cards with its aid. Let's go!"). A screenshot of the software in both conditions is presented in Figure 1.

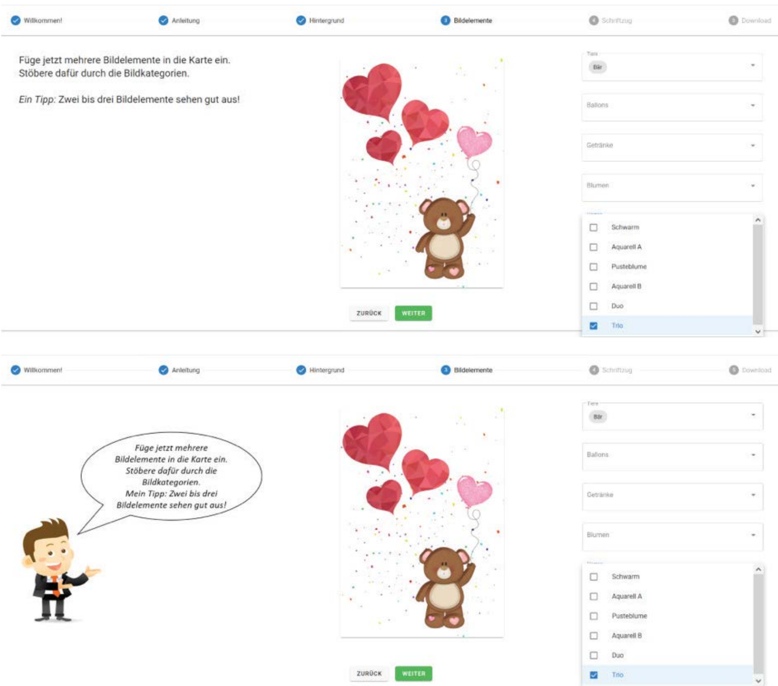

Figure 1: Screenshots of the design software in the non-anthropomorphic (upper image; Translated instructions: Now, add multiple visual elements to the card. To do so, browse through the categories. Tip: Two to three objects look good.) vs. anthropomorphic (lower image; Translated speech bubble: Now, add multiple visual elements to the card. To do so, browse through the categories. My tip: Two to three objects look good) condition.

3.1 Participants

266 participants (30,8 % male) between 18 and 70 years (*M* = 26.17; *SD* = 8.67) were recruited through an e-mail distribution lists and flyers at the Ludwig-Maximilians-Universität München, social media and an online platform. They were offered course credit or could participate in a raffle for five 20€ Amazon coupons.

3.2 Procedure

After giving informed consent, participants were directed to a website to design a personalized birthday card they could download at the end. They were randomly assigned to one of two experimental conditions (anthropomorphic, non-anthropomorphic). Depending on the condition, they were guided through the process either by an avatar (anthropomorphic condition) or neutral instructions (non-anthropomorphic condition). Participants were then redirected to the survey and rated the responsibility they attributed to the technology as well as themselves for the interaction outcome (i.e. the birthday card), and the perceived autonomy of the technology as well as their own. Furthermore, participants rated the perceived anthropomorphism of the design software. Finally, demographic data was collected.

3.3 Measures

Responsibility attribution. Based on the items used by Hur, Koo and Hofmann (2015), single items were used to assess responsibility attribution to the technology ("To what extent is the design software responsible for the result [finished birthday card]?") and the user ("To what extent are you responsible for the result [finished birthday card]?"), respectively. The items were rated on a 5-point Likert Scale (1 = not at all; 5 = fully).

Perceived autonomy. Based on the items used by Jung (2011), three items were used to assess each the perceived autonomy of the technology (e.g. "I feel that the design software had a lot of control over the design process.") as well as the user (e.g. "I feel that I had a lot of control over the design process."), respectively. The items were rated on a 5-point Likert Scale (1 = not at all; 5 = fully). Cronbach's α was .71 for the technology items and .74 for the user items.

Perceived anthropomorphism. Two operationalizations were used to measure perceived anthropomorphism regarding the technology. A self-constructed single item ("How humanlike did the design software seem to you?") was rated on a 5-point Likert Scale (1 = not at all; 5 = fully). In addition, the Godspeed Anthropomorphism Subscale (Bartneck, Kulić, Croft, & Zoghbi, 2009) with a 5-point semantic differential (e.g., "machinelike/ humanlike") was used, except for one item ("moving rigidly/ moving elegantly") as it did not fit our operationalization. Cronbach's α was .83 for the remaining four items.

4 Results

IBM SPSS Statistics 26 was used for all data analyses. Additionally, PROCESS macro v3.4 (Hayes, 2017) was used to test the mediation hypotheses.

4.1 Descriptive Results

Relevant descriptive results as well as correlations are illustrated in Table 1 below.

Table 1: Descriptive statistics and correlations for the relevant variables

	Variable	M	SD	1	2	3	4	5	6
1.	Responsibility attribution (user)	3.18	0.90	1					
2.	Responsibility attribution (technology)	3.70	0.75	-.30**	1				
3.	Perceived autonomy (user)	3.02	0.87	.54**	-.27**	1			
4.	Perceived autonomy (technology)	3.32	0.85	-.35**	.39**	-.60**	1		
5.	Perceived anthropomorphism (self-constructed)	2.16	1.00	.19**	-.01	.22**	.00	1	
6.	Perceived anthropomorphism (Godspeed Scale)	2.13	0.84	.24**	-.05	.31**	-.07	.71**	1

Note. N = 266. **p < .01

4.2 Hypotheses testing

As expected, perceived anthropomorphism was significantly higher within the anthropomorphic condition (vs. non-anthropomorphic condition) for the self-constructed item ($t(261.06) = -3.16$, $p = .002$, *dCohen* = -0.38.) as well as for the Godspeed Anthropomorphism Subscale ($t(264) = -2.66$, $p = .009$, *dCohen* = -0.33), indicating that our manipulation was successful.

H1. A linear regression analysis was conducted. Results showed that perceived autonomy of the technology significantly predicted the responsibility attribution to the technology for the interaction outcome ($\beta = .39$, $t(264) = 6.93$, $p < .001$). It explained 15% of the variance ($R = .15$, $F(1,264) = 48.03$, $p < .001$.). Our hypothesis that the higher the perceived autonomy of the technology, the more responsibility users attribute to it for the interaction outcome was thus supported.

H2. The conducted Pearson-Correlation showed a significant negative correlation between the responsibility users' attributed to the technology and the responsibility they attributed to themselves regarding the interaction outcome (r = -.30, p < .001). The results supported our hypothesis that the more responsibility users attribute to the technology, the less responsibility they attribute to themselves.

H3. The conducted Pearson-Correlation showed a significant negative correlation between the users' own perceived autonomy and the perceived autonomy of the technology *(r* = -.60, p < .001). The results supported our hypothesis that the higher the perceived autonomy of the technology, the lower the users' own perceived autonomy.

H4. An independent-samples t-test was conducted. Results showed no difference in the responsibility attribution towards the technology between the anthropomorphic condition and the non-anthropomorphic condition ($t(264) = 0.80$, p = .425, *dCohen* = 0.11). Contrary to our hypothesis, responsibility attribution to the technology did not vary between the two conditions (anthropomorphic, non-anthropomorphic).

H5. A stepwise regression analysis was conducted. Variables were transformed to z-values. No significant main effect of the manipulation on responsibility attribution to the technology (β = -0.05, *SE* = 0.06, t = -0.80, p = 0.43) was found. Furthermore, the manipulation did not predict perceived autonomy of the technology (β = 0.02, *SE* = 0.06, t = 0.31, p = 0.76). Perceived autonomy of the technology did predict responsibility attribution to the technology (β = 0.39, *SE* = 0.06, t = 6.95, p < .001). Controlling for perceived autonomy of the technology, the direct effect of the manipulation on responsibility attribution to the technology remained not significant (β = -0.06, *SE* = 0.06, t = -1.00, p = 0.32). Thus, contrary to our hypothesis, perceived autonomy of the technology did not mediate the interrelation between the manipulation and responsibility attribution to the technology.

5 Discussion

In sum, our study explored autonomy as an influencing factor of responsibility attribution between user and technology in HCI as well as a possible trade-off in responsibility attribution and perceived autonomy between user and technology regarding an interaction outcome. We also looked into effects of anthropomorphic technology design on autonomy perception of and responsibility attribution to a technology. Results showed that the more autonomous a technology was perceived, the more responsibility users attributed to it regarding the interaction outcome. Furthermore, we found that the more responsibility users attributed to the technology, the less responsibility they attribute to themselves.

The same was found for autonomy. Yet, we could not find an effect of anthropomorphic technology design on the perceived autonomy of the technology nor the responsibility attributed to the technology for an interaction outcome. Furthermore, we did not find a mediating effect of autonomy regarding the latter interrelation. In the following sections, we discuss these findings with regards to previous research.

Regarding the interrelation of perceived autonomy and attributed responsibility, our study results imply that the more autonomous users perceived the technology to be, the more responsibility they attributed to the technology for the interaction outcome. Our results are compatible with previous research that manipulated autonomy of (autonomous) vehicles (McManus & Rutchick, 2019) and measured responsibility attributed to the driver for a certain outcome. Having measured a subjective perception of autonomy, our results imply that already slight variances in perception regarding the autonomy of a technology can come along with significant variances in the attribution of responsibility to such for a certain interaction outcome. Still, our results are of correlational nature and should be further investigated in a systematical manner.

In addition to previous research, our study also highlighted two trade-off effects, namely, the division of attributed responsibility for the created birthday card between technology and user as well as the division of perceived autonomy between technology and user. The more responsibility users attributed to themselves, the less responsibility they attributed to the technology, and vice versa. Also, the more autonomous they perceived themselves to be in the design process, the less autonomy was attributed to the technology, and vice versa. In line with the theoretical concept of autonomy and its two opposite poles, namely autonomy versus controlled behavior (Deci & Ryan, 1985), an increase of autonomy for one of the social actors goes along with a decrease for the other.

Finally, our study looked into the role of anthropomorphic technology design in attributed responsibility and perceived autonomy of the technology. Our successful experimental manipulation highlights human features in technologies as a possible effective way to manipulate the perceived degree of anthropomorphism. Yet, contrary to recent findings in HCI research (e.g., Hinds et al., 2004) we found no effect of anthropomorphic design on attributed responsibility of the technology. A possible explanation could be that previous studies have focused on positive respectively negative outcomes of HCI and have thus raised the question of attributing blame or credit to the technology vs. the users themselves (e.g. Serenko, 2007; Waytz et al., 2014).

In such situations anthropomorphic design may have triggered psychological attribution phenomena known from human interaction, such as the self-serving bias, i.e. "an ego-biased attribution," where "we try to explain our behavior in terms that flatter us and put as in a good light" (Miller & Ross, 1975, p. 213), and hence had a significant influence on responsibility attribution to the technology. Our study focuses on a more neutral operationalization (design of a birthday card). Thus, such an attribution phenomenon might not have been activated. Further research should look into this interrelation applying an accordingly "neutral" interaction outcome. In accordance, no mediation effect of perceived autonomy of the technology regarding this interrelation could be found. Furthermore, contrary to previous studies (Hinds et al., 2004; Puzakova et al., 2013; Waytz et al., 2014), there was no interrelation found between anthropomorphic design cues and perceived autonomy of the technology. This observation could root in the averagely high rating of perceived autonomy of the technology in both conditions, which might have caused a restriction in variance of such.

6 Limitations

One main limitation of our study is the correlational nature of our results regarding autonomy. As we did not manipulate the autonomy of the technology we can only report results of Pearson correlations focusing on the subjectively perceived autonomy of the technology by the users. Furthermore, as descriptive results indicate an averagely high perceived autonomy of the technology in both conditions, variances of perceived autonomy might have been restricted and affected the results. Further studies should look into the reported interrelations more systematically, e.g. by manipulating the degree of the technology's autonomy or choosing a technology that allows more variance in the perception of the technology's autonomy.

7 Implications

Our study results come with essential implications in theory and practice. With regard to theoretical implications and future HCI research, perceived autonomy of the technology as a construct seems to play an important role for responsibility attribution regarding an interaction outcome between human and technology. Specifically, our results showed a positive interrelation of such. Furthermore, perceived autonomy of the user vs. the technology as well as the responsibility the users attribute to themselves vs. the technology respectively for an interaction outcome seem to lay on a scale with opposite poles. Thus, attributed responsibility for an interaction outcome and perceived autonomy appear to be divided between user and technology. On a practical level, with innovative technologies becoming increasingly autonomous, their perception as such might come with an increased responsibility attributed to the technologies for interaction outcomes and consequently less responsibility attributed to the users by themselves.

While at first glance this might seem comfortable from a user perspective, it might foster carelessness of users as well as their dissatisfaction due to low perceived self-efficacy and thus impair HCI. Therefore, the found correlational results should be looked into more systematically and perceived valence of HCI outcome should be taken into account as studies have shown that it can play an important role regarding the question of responsibility attribution (e.g., Moon, 2003).

Acknowledgements:
This research has been funded by the German Federal Ministry of Education and Research (BMBF), project GINA (FKZ: 16SV8097)

Literature

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84(2), 191–215. doi:10.1037/0033-295X.84.2.191

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics, 1, 71–81. doi:10.1007/s12369-008-0001-3

Berberian, B., Sarrazin, J.-C., Le Blaye, P., & Haggard, P. (2012). Automation technology and sense of control: A window on human agency. PLoS One, 7(3), e34075. doi:10.1371/journal.pone.0034075

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. Journal of Research in Personality, 19, 109–134. doi:10.1016/0092-6566(85)90023-6

Deci, E. L., & Ryan, R. M. (1987). The support of autonomy and the control of behavior. Journal of Personality and Social Psychology, 53, 1024–1037. doi:10.1037/0022-3514.53.6.1024

Hayes, A. F. (2017). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (2nd ed.). New York, NY, USA: Guilford Publications.

Hinds, P. J., Roberts, T. L., & Jones, H. (2004). Whose job is it anyway? A study of human-robot interaction in a collaborative task. Human-Computer Interaction, 19, 151–181. doi:10.1207/s15327051hci1901&2_7

Hur, J. D., Koo, M., & Hofmann, W. (2015). When temptations come alive: How anthropomorphism undermines self-control. Journal of Consumer Research, 42, 340–358. doi:10.1093/jcr/ucv017

Jung, Y. (2011). Understanding the role of sense of presence and perceived autonomy in users' continued use of social virtual worlds. Journal of Computer-Mediated Communication, 16, 492–510. doi:10.1111/j.1083-6101.2011.01540.x

Kim, S., Chen, R. P., & Zhang, K. (2016). Anthropomorphized helpers undermine autonomy and enjoyment in computer games. Journal of Consumer Research, 43, 282–302. doi:10.1093/jcr/ucw016

McManus, R. M., & Rutchick, A. M. (2019). Autonomous vehicles and the attribution of moral responsibility. Social Psychological and Personality Science, 10, 345–352. doi:10.1177/1948550618755875

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? Psychological Bulletin, 82(2), 213–225. doi: 10.1037/h0076486

Moon, Y. (2003). Don't blame the computer: When self-disclosure moderates the self-serving bias. Journal of Consumer Psychology, 13, 125–137. doi:10.1207/S15327663JCP13-1&2_11

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of Social Issues, 56, 81–103. doi:10.1111/0022-4537.00153

Puzakova, M., Kwak, H., & Rocereto, J. F. (2013). When humanizing brands goes wrong: The detrimental effect of brand anthropomorphization amid product wrongdoings. Journal of Marketing, 77(3), 81–100. doi:10.1509/jm.11.0510

Reeves, B., & Nass, C. I. (1996). The media equation: How people treat computers, television, and new media like real people and places. Cambridge, UK: Cambridge University Press.

Serenko, A. (2007). Are interface agents scapegoats? Attributions of responsibility in human-agent interaction. Interacting with Computers, 19, 293–303. doi:10.1016/j.intcom.2006.07.005

van der Woerdt, S., & Haselager, P. (2019). When robots appear to have a mind: The human perception of machine agency and responsibility. New Ideas in Psychology, 54, 93–100. doi:10.1016/j.newideapsych.2017.11.001

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. Journal of Experimental Social Psychology, 52, 113–117. doi:10.1016/j.jesp.2014.01.005

Werth, L., & Mayer, J. (2008). Sozialpsychologie. Heidelberg: Spektrum.