Bennett CC, Stanojevic C, Kim S, Sabanovic S, Lee J, Piatt JA, et al. (2022) Comparison of in-home robotic companion pet use in South Korea and the United States: A case study. *9th IEEE International Conference on Biomedical Robotics & Biomechatronics (BIOROB).* In Press.

Comparison of In-home Robotic Companion Pet Use in South Korea and the United States: A Case Study*

Casey C. Bennett, *Member, IEEE*, Cedomir Stanojevic, Seongcheol Kim, Selma Sabanovic, *Member, IEEE*, Jinjae Lee, Jennifer A. Piatt, Janghoon Yu and Jiyeong Oh

***Abstract*— This paper presents an intensive case study of 10 participants in the US and South Korea interacting with a robotic companion pet in their own homes over the course of several weeks. Participants were tracked every second of every day during that period of time. The fundamental goal was to determine whether there were significant differences in the types of interactions that occurred across those cultural settings, and how those differences affected modeling of the human-robot interactions. We collected a mix of quantitative and qualitative data through sensors onboard the robot, ecological momentary assessment (EMA), and participant interviews. Results showed that there were significant differences in how participants in Korea interacted with the robotic pet relative to participants in the US, which impacted machine learning and deep learning models of the interactions. Moreover, those differences were connected to differences in participant perceptions of the robot based on the qualitative interviews. The work here suggests that it may be necessary to develop culturally-specific models and/or sensor suites for human-robot interaction (HRI) in the future, and that simply adapting the same robot's behavior through cultural homophily may be insufficient.**

I. Introduction

A. Background

Novel forms of robotic companions have been a growing area of research over the past decade, particularly for applications related to in-home chronic health conditions and aging-related issues [1,2]. Within the field of human-robot interaction (HRI), many of these companions take on the form factor of robotic pets, intended to provide social support and enhance cognitive functioning through social interaction. Such socially-assistive robots (SARs) are thus critical in many places that are faced with a rapidly increasing elderly population and/or greater awareness of the impact of mental health on physical health, in order to facilitate community-based health approaches [3-5].

However, open questions remain as to appropriate types of interactions that such robotic pets need to entail during these applications. For instance, how should the robotic pet behave in user homes in response to a person with dementia, or a younger person suffering from chronic depression and anxiety? Moreover, another question is whether the robot's behavior should be dependent on situational factors, such a geographical location or cultural setting. Indeed, previous research has shown that *situated robot use* in different cultural locales has a major impact on how the same robots are perceived and utilized by different groups of people [6,7]. However, many previous studies (see Section 1.B below) are limited to either controlled lab experiments or retrospective real-world data based purely on user recall, so a challenge still exists to explore these questions in real-time in real-world settings. That challenge is two-fold, entailing the need for approaches that can monitor HRI interactions that occur in real-time as well as identification of appropriate modeling methods relevant to robotic companions in-the-wild [8].

In this paper, we explore these questions in users in the United States and South Korea interacting with a robotic pet in their own homes over several weeks. We purposely adopted a *case study* approach, focusing on intensively tracking a smaller number of users over a longer period of time (every second of every day for 3 weeks), rather than gathering a small amount of data about many users briefly, to account for intra-person behavior variation. The fundamental goal was to determine whether there were significant differences in the types of interactions that occurred across different cultural settings, and how those differences affected modeling of real-world human-robot interactions.

B. Prior Work

There is existing previous research that has explored inter-group differences for various types of interactions between humans and robots [9-17]. That research has explored different types of groups (e.g. cultural, age, gender) across various settings (e.g. home, work, healthcare, military, education). For instance, Andreasson et al. (2018) examined gender differences regarding robot touch [9], while other groups have explored the effects of pre-existing negative attitude differences towards robots between genders [10]. Other researchers have asked whether HRI varies across age groups. For example, since older adults have slower reaction times, a question is whether that might lead to differences in human-robot collaborative task performance [11]. Another question is whether teenagers respond differently to social robots than adults [12].

*This research was supported by the research fund of Hanyang University (HY-2020) in Korea, as well as a grant frrom the National Science Foundation in the United States (Grant# IIS-1900683).

Casey C. Bennett is with the Department of Intelligence Computing, Hanyang University, Seoul Korea 04763 (Corresponding Author, e-mail: cabennet@hanyang.ac.kr).

Seongcheol Kim, Jinjae Lee, Janghoon Yu and Jiyeong Oh are also with the Department of Intelligence Computing, Hanyang University, Seoul Korea (email: {sckim219, jjlee93, jqdjhy, ojyhi010309}@hanyang.ac.kr)

Selma Sabanovic is with the School of Informatics, Computing, and Engineering, Indiana University, Bloomington USA 47408 (e-mail: selmas@indiana.edu).

Cheda Stanojevic is with the Department of Parks, Recreation, and Tourism Management, Clemson University, Clemson, South Carolina, USA, 29634 (e-mail: cestan@iu.edu).

Jennifer A. Piatt is with the School of Public Health, Indiana University, Bloomington USA 47408 (e-mail: jenpiatt@indiana.edu).

Some research has also looked at cultural differences in HRI interactions, such as children's expressiveness during gameplay across different cultures [13] or moral judgments between US and Japan [14]. In a comparison specifically between the US and Korea, Lee & Sabanovic (2014) found that participants in South Korea envision robots as companions for the family, while US participants see home robots as individual assistants and modern appliances [6]. Other researchers have looked at the combination of cultural differences with other factors. For example, Rudovic et al. examined how engagement with a social robot varied in children with autism between Serbia and Japan, finding significant differences [15]. Fraune et al. (2015) showed that Japanese and US participants respond differently to groups of robots, relative to an individual robot [16].

Interestingly, however, past results in cross-cultural robotics have indicated that cultural homophily (e.g. agents adapted to a specific set of cultural attributes) alone does not necessarily always correspond to higher ratings of a robot by human participants [17]. This suggests that the impact of culture on HRI is not so simple as identifying differences. The real question is how such differences might impact our models of robot interactive behavior in real-world settings.

II. METHODS

A. *Setting, Users & Robot Description*

The study involved a sample of 10 users, with 5 recruited from the greater Seoul area in South Korea and 5 recruited from the US Midwest. Participants were in the age range of 20-35, and included 7 females and 3 males. All participants were recruited from the general population and were living alone. The study was approved separately by the IRBs at Indiana University (US) and Hanyang University (Korea).

Each participant was given a robotic pet for home use, in this case the Hasbro Joy-For-All robotic therapy pet (https://joyforall.com/) equipped with a robotic sensor collar (see Figure 1). The collar, which was developed at Indiana University's R-house robotics lab, contained sensors capable of detecting light, sound, and motion [18,19]. Sensor data was collected roughly every second of every day across the 3-week study period, while data about interaction modalities was simultaneously collected via use of a mobile app (described in Section 2.B). Equipment failure with the sensor collar during two participants (one in the US and one in Korea) led to partial data loss that was identified during the analysis phase, so they were excluded from the results below (leaving 8 participants with complete data).

B. *Experimental Design*

Along with collecting the sensor data described in Section 2.A, the experiment utilized a sampling method known as ecological momentary assessment (EMA) to gather real-time data about interactions occurring between the robotic pet and human participant [20]. EMA works by randomly sampling each user's behavior multiple times throughout the day over a period of time (days, weeks, months). EMA has been shown to be a powerful tool for monitoring everyday user behaviors by gathering real-time data via smartphones [21,22], as well as interactive robots [8,23]. For EMA in this study, we employed the PiLR mobile app (https://pilrhealth.com/). The

Figure 1. Joy-For-All robot and sensor collar

EMA app was setup to ping users via their smartphone roughly 5-7 times per day (referred to as "stimulus prompts"), arriving randomly during set time periods (e.g. morning, late afternoon, early evening)

The EMA prompts collected data about the interaction modality (the type of behavior) and proximity (whether the interaction occurred near/far to the robot). The interaction modalities were defined based on previous research with robotic pets in in-home settings [18,19,24,25], and included both active interactions directly with the robot (e.g. petting, talking, playing) as well as indirect passive interactions (e.g. watching television, eating together with the robot). Beyond the interaction-focused stimulus questions above, we incorporated additional psychological assessment questions to gauge user perception and emotional response post-interaction. As such, the study period was divided into baseline, intervention, post-intervention phases.

Each participant was enrolled into an approximately 3-week long study period, including ~16 day intervention phase of sensor data collection plus a 2-day pre/post questionnaire phase to establish a baseline and for final assessment. The 18-item pre-intervention questionnaire assessed the user's current daily habits and behavioral routines, while a similar post-intervention questionnaire assessed changes. After the 3-week period with the robot, there was a subsequent follow-up interview conducted by a research assistant to gather qualitative data about the participant's experience during the study with the robotic pet and EMA app. The full survey and questionnaires (referred to as the SoREMA instrument [8]) are available online on the author's website: http://www.caseybennett.com/uploads/SoREMA_Survey_Questionnaire.docx

To summarize, including the follow-up interview during the week after the robot deployment study period, there was **approximately a total 1-month study involvement for each participant**. During this time, a mix of both quantitative and qualitative data was collected (using a *convergent parallel mixed method* approach [26]). This included pre- and post-intervention questionnaires, robotic sensor data, EMA interaction data, and recorded interviews.

Some EMA prompts resulted in users reporting no interaction occurring (roughly 65% of time) and were thus excluded from further analysis. That is to be expected with real-world use of robots. Additionally, some modalities were only rarely performed and so excluded. This left us with a sample of 152 interactions across five modalities: petting, playing, moving the SAR (from one location to another),

TABLE I. FEATURE LIST

Category	Features	Description
Accelerometer	accel_x, accel_z, accel_y	Raw average readings from accelerometer in x, y (lateral) and z (up/down) directions
Light & Sound Sensor	light_val, sound_val	Raw average readings from light and sound sensors
Motion	motion_detect	Percentage of time robot was detected as "in-motion" (above some noise threshold) in any coordinate direction (x, y, z)
Rotation	arc	Average amount of rotation motion during interaction
Orientation	orient	Orientation during the interaction in which the robot spent the max time (i.e. mode)
Sound Category	Quiet, Moderate, Loud	Percentage of time that specific sound categories were detected, using sound sensor manufacturer specified thresholds
Sound Transitions	Quiet-Moderate, Quiet-Loud, Moderate-Quiet, Moderate-Loud, Loud-Quiet, Loud-Moderate	Frequency of detected transitions between sound categories during interaction
Orientation Category	Landscape Right, Landscape Left, Portrait Up, Portrait Down, Flat	Percentage of time that specific orientation categories were detected, using accelerometer manufacturer-specified thresholds
Orientation Transitions	orient_shift	Frequency of detected transitions between orientation categories
Any Detection	awake	Percentage of time that a "signal" was detected by any sensor (above some noise threshold) on the robotic collar

talking to the SAR, watching TV/radio (or other media, e.g. YouTube). Each “interaction” represented a 15-minute time period (users were specifically directed by the EMA prompts to report interactions for that entire time range), **so the 152 interactions constituted nearly 40 hours of total interaction data**. The modalities were not mutually exclusive, so for instance a participant could be petting the SAR while talking to it. Indeed, participants reported approximately 2.5 modality types per interaction (higher among Korean vs. US participants). This was intended though, in order to reflect real-world settings where people often do multiple things at a time without a clear start/stop

C. Analysis Approach

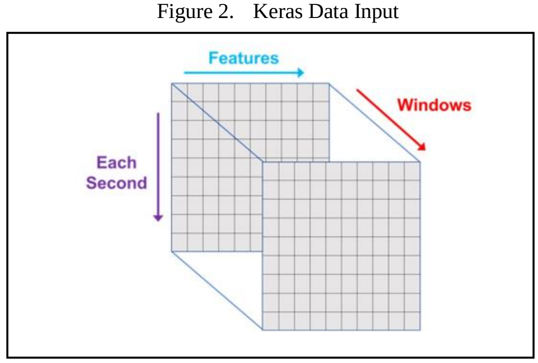
The analysis in this paper is broken into 3 parts: 1) a general analysis of interaction patterns between the groups using descriptive statistics, 2) a machine learning & deep learning analysis, 3) a qualitative analysis of participant interview data.

First, we examined the overall differences in interaction behaviors between participants in the US and Korea, using descriptive statistics to give a high-level overview of the data.

Second, we utilized machine learning (ML), deep learning (DL), and XgBoost models that were originally developed on the US data in previous work [8]. **The aim is to examine whether the same models work on the Korean data or not**, or adversely whether different models would be needed for each cultural locale. The details of these approaches have been described in previous work, but in short the EMA data became the "targets" (i.e. interaction modalities) while the sensor data became the "features" for the ML and DL models. For simplicity we collapsed the dataset into a series of binary classification predictions (e.g. petting vs. not petting) rather than attempt a complex multi-class classification problem. Due to target class imbalance, the data was re-balanced using SMOTE [27].

Features in the dataset (see Table 1) included motion in the x/y/z directions, rotational motion (arc), light/sound values, orientation, and transitions between sound/orientation categories (e.g. loud to quiet, portrait to landscape). For predicting the EMA target, the feature data for that 15-minute time period was sliced into 15-second-long overlapping windows, with 50% overlap (similar to [28]). The way these features were handled depends on the type of modeling method. In general, the standard ML approaches calculated averages or percentages/frequencies for each feature across the entire interaction (the "time window") resulting a single row of data for each target, whereas DL approaches utilized smaller time slices as the windows so that each interaction was broken into many windows. For DL, the data can be visualized as a mutli-dimensional array, with a row for each second of data ("y" dimension), a column for each sensor data feature ("x" dimension), and each 15-second window being a third "z" dimension (see Figure 2).

Figure 2. Keras Data Input



Standard ML approaches were performed using the python package Scikit-Learn (https://scikit-learn.org). Multiple modeling methods were attempted: Random Forest, Gradient Boosting, Neural Networks, and Support Vector Machines (SVM). Models were generally run using the default parameters in Scikit. Results were evaluated using 5-

fold cross-validation based on accuracy and AUC metrics. DL modeling was performed using the python package Keras (https://keras.io/), which is a deep learning library based on TensorFlow. The data was fed into a DL model consisting of a single 2D convolutional neural network layer (CNN) with kernel size set to 1 and using a ReLU activation function, followed by a single recurrent layer (LSTM) with 50 units [29]. The idea was that the CNN could parse out "invariant representations" of pattern signatures occurring anywhere in the interaction, followed by the LSTM detecting critical "sequences" of those patterns over time. A final fully connected "Dense" layer using a sigmoidal activation function was used to make the final binary classification predictions. To evaluate performance, 20% of the data was held out as a "test set" for each classification run.

TABLE II. INTERVIEW QUALITATIVE CODING HEIRACHY

Code Group 1	Code Group 2
Robot/Collar	Charging Design Interactions
EMA App	Alerts Gamifications Incentivizing Interactions Leaderboard
Overall Experiment	Challenges Feelings Interactions

Third, we performed a qualitative analysis of the participant interviews. These were first coded by two independent coders using the Atlas TI software (https://atlasti.com/), using a coding scheme developed for the project that included a hierarchy of codes. The top level of the hierarchy distinguished comments related to the robot/collar, the EMA app, and the experiment itself. Below those top level codes was a second level with codes for design, interactions, alerts, incentives, challenges, charging/battery issues, and desire for leaderboards or other types of gamification with the robot. The code hierarchy can be seen in Table 2. The codes in "Code Group 2" were further broken into positive, negative, or suggestions for improvement. Interrater reliability between the two coders was calculated as 0.67, implying moderate agreement.

III. RESULTS

A. General Interaction Analysis

General patterns of interaction modalities by cultural locale (US or Korea) can be seen in Table 3. Reported modalities were fairly consistent between the two groups, outside of Petting and Playing. That may be a linguistic or definitional difference in understanding of the terms. It had no bearing on the rest of the analysis here, but may be of interest for future research.

As can be seen in Table 1, there are numerous features in the dataset. We analyzed the average values for those across all interactions for the Korean and US groups. Mostly the values were consistent across groups, but there were a few notable differences for light, sound, and motion (arc) which are shown in Table 4. In short, US participants appeared to move the robotic pet around more frequently, and had higher levels of noise in the environment. Conversely, the Korean participants exhibited less movement and noise, but higher ambient light levels. Our interpretation was that this might be due to different living environments and lifestyles between the two locales, e.g. Korean living spaces tend to be smaller and more compact than US living spaces (thus less need to move around in Korea). However, the exact reasons for these differences are still unclear and need more research.

TABLE III. INTERACTION MODALITY FREQUENCY

Cultural Locale	Petting	Talking	Playing	TV / Radio	Moved It
US	38.0%	15.2%	3.8%	19.0%	23.9%
KOR	25.7%	19.9%	17.8%	19.4%	17.3%

TABLE IV. NOTABLE FEATURE DIFFERENCES

Cultural Locale	Interact-ion Cnt	Sound val	Light val	arc	Motion detect
US	80	89.48	189.4	0.059	0.201
KOR	72	59.24	590.0	0.011	0.010
Grand Total	***152***	***75.18***	***378.9***	***0.036***	***0.111***

B. ML & DL Modeling

In previous work, we have been able to build ML and DL models that achieved approximately 75-80% accuracy in detecting interaction modalities in user homes in the US based on this EMA approach and sensor features [8]. One question we had was **whether that same approach would work in Korea, particularly given the differences in feature values that we observed (Section 2.A).** To evaluate this, we applied the same ML and DL models trained on US data to the Korean data collected here. Additionally, we combined all the US and Korean data together as if it were one single training sample and attempted to re-create the same models. The results can be seen in Table 5 (ML) and Table 6 (DL). For brevity, we only show the gradient boosting results for ML here. We did employ other modeling methods, including random forests and SVM, but the patterns were consistently the same.

As can be seen in the tables, using the trained models from the US on the Korean data was not successful. There was slightly more success when combining the US & Korean data together as one sample, but it was still suboptimal, with accuracy and AUC values 10% less than the original models. Our interpretation here is that this may be due to the feature value differences observed in Section 2.A, and by extension potential lifestyle differences between the US and Korea. In other words, since home living environments are often quite different between the US and Korea, the data that robotic sensors collect in user homes may be quite different too. Moreover, it is a distinct possibility that different sensors are needed in the two cultural locales, to collect different types of

TABLE V. ML RESULTS (GRADIENT BOOSTING)

	US Only		US Train, KOR Test		All Combined	
Modality	**Acc.**	**AUC**	**Acc.**	**AUC**	**Acc.**	**AUC**
Petting	94.3%	0.9655	63.9%	0.5000	80.2%	0.8945
Talking	72.0%	0.8300	50.0%	0.5248	54.5%	0.5485
Playing	91.6%	0.9751	51.4%	0.4884	76.4%	0.8327
Listening TV/Radio	58.6%	0.5421	44.4%	0.4444	48.7%	0.4558
Moving It	60.4%	0.6788	45.8%	0.5000	50.6%	0.5548
Average	75.4%	0.7983	51.1%	0.4915	62.1%	0.6573

TABLE VI. DL RESULTS

	US Only		US Train, KOR Test		All Combined	
Modality	**Acc.**	**AUC**	**Acc.**	**AUC**	**Acc.**	**AUC**
Petting	75.6%	0.7234	49.8%	0.5099	61.0%	0.6854
Talking	68.4%	0.7638	48.2%	0.4745	52.6%	0.5842
Playing	86.2%	0.9285	68.8%	0.7076	67.7%	0.7899
Listening TV/Radio	74.8%	0.7781	49.7%	0.5056	65.2%	0.7052
Moving It	66.0%	0.6938	45.7%	0.4275	65.9%	0.6890
Average	75.3%	0.7752	52.4%	0.5250	62.5%	0.6908

features specifically relevant to the different home living environments. We return to this topic in more detail in the Discussion section.

Related to that, we also note that there were significant differences in the participant interview data, where Korean participants made more frequent negative comments about the technology design, while US participants seemed to focus more on the interaction in general. We discuss this in the qualitative analysis in the next section.

C. *Qualitative Interview Analysis*

The coded data from the participant interviews was analyzed using a phenomenological approach [30]. In general, there was a high level of consistency between Korean and US participants across the hierarchy, both at the Code Group 1 and Code Group 2 levels (data not shown for brevity). There were two notable exceptions to that, however. First, Korean participants made negative comments about the technology and suggestions for improvement more frequently than the US participants (t-test p value = 0.018, see Table 7). Second, US participants were more likely to make comments about the interaction itself, versus things like the technology design, alerts, gamification, etc. Nearly 20% of the US comments focused on the interaction, versus 13% of Korean comments.

It is possible that those differences in perspectives between the US and Korea about the robotic companion pet and EMA app system may have influenced how participants interacted (or did not interact) with the robot. That may have subsequently impacted the modeling results reported Section 3.B of this paper.

TABLE VII. CODING ANALYSIS RESULTS

Code Type	**US**	**KOR**
Negative	27.9%	34.7%
Suggestion	39.3%	43.1%
Positive	30.8%	22.2%

For example, Korean participants often expressed that they felt uncomfortable with the robotic companion pet, mentioning that:

- *"I was really worried about what should I do with the robot, at first, as I've never raised a pet before."*
- *"At the end of the day, I used to turn off my cat and sleep, but I felt a little uncomfortable about just shutting it off."*
- *"I purposely didn't turn off the sound of the cat. I thought it would be better for the experiment, so I just used it at home without turning off the sound at all… the sound was a little excessive because he cried even at dawn."*

Many of the Korean comments seemed to relate to participants not being used to raising pets in small living spaces, the frequent sounds of the robot (which are problematic given the lack of sound-proofing in many Korean apartments), or otherwise just being uncomfortable with the technology design in general.

Adversely, US participants tended to comment more about the nature of the interaction and the robotic behaviors, particularly in response to their own human behaviors (e.g. walking through the house or while cooking). For example US participants mentioned that:

- "*I was literally walking through my house and it would just meow when I got a little bit close to it*."
- "*like it being able to maybe just roll over or something if it wanted food, which I know doesn't eat but, like a meow near his food bowl*."
- *"but it was just like it kind of noticed the sound or something [and] kind of did like a questionable meow ... because I was hammering something really loud, but I had headphones in so I didn't know and then I took them out, and [it] was like asking for attention, right after that, like after I'd been really, really noisy."*

Many of those US interaction comments seemed to be focused on their expectations of real cat behavior based on past experience, projected onto the behaviors of the robotic companion pet. US participants also seemed to overestimate the capabilities of the robotic companion pet in some cases, misattributing its behaviors to capabilities the technology does not actually have. That seemed to be reflected in their more positive comments, and potentially have affected their interactions with the robot as well. More broadly, this raises interesting potential future research questions for robotic companion pet research, in that study participants' past experience with real living pets might affect the results.

IV. Discussion

A. Main Summary

We conducted an intensive case study of 10 participants in the US and South Korea to compare in-home use of robotic companion pets across different cultural locales. We collected a mix of quantitative and qualitative data about each participant over the course of month. In particular, we were interested in if any differences in interactions existed between the two cultural settings, and how those differences might impact modeling of those interactions.

Results showed that there were indeed significant differences in participant perceptions of the robotic companion pet, and that the types of interactions varied as well. For instance, Korean participants had a more negative view of the technology design. There were also notable differences in the collected sensor data, with US participants tending to move the robot around more and have noisier home environments. Moreover, ML and DL models that were developed in the US failed to work as well in South Korea, which may be partially attributable to those differences listed above.

Some of the interaction/perception differences seem to be tied to the different home living environments in Korea versus the US, such as smaller living spaces and lack of sound-proofing in apartments that causes some functionality of robotic companion pets to be suboptimal in Korean contexts. We discuss the broader implications of these findings in the next section.

B. Implications for Robotic Companion Pets

SARs and robotic companion pets hold great potential in in-home settings for chronic care and aging-related issues. However, typically research on those technologies focused on the design and application within a specific cultural locale, such as Japan or the US. The results here suggest that significant differences in lifestyle and home-living environments may complicate the adoption of such robots. What may be seen as a beneficial feature in one setting may cause problems in another settings, as evidenced by the frequent cat sounds made by the robot in this study. There are different cultural expectations in different locales as well (e.g. Korean cities), which feed into differences in human behavior and the design of the built environment, even into how rooms and buildings are architected. Indeed we would be remiss not to point out that Korea is one of the most densely populated countries in the world, with the Seoul area alone having roughly 42,600 people per square mile.

Those cultural differences lead to challenges, but perhaps opportunities as well. For example, we may need to define culturally-relevant interaction modalities pertaining to the specific living situations and lifestyles in Korea and the US, as well as create specific robotic sensor suites suited to detect those types of modalities. Such an approach could not only help enable better models of HRI interactions (machine learning or otherwise), but also enhance the behaviors of robotic companions to better fit the cultural expectations of different locales in a way that goes beyond attempting to simply adapt the same robot to different cultures (i.e. beyond cultural homophily).

Much previous research in HRI has argued for "culturally-robust" or "culturally-aware" systems (including our own), where robots are designed specifically to create adaptable behaviors that match the value system of the local human culture [6,17,31,32]. While that is certainly one approach for social robots, here we take the position that it may be necessary to create *fundamentally* different models of robot behavior specifically for different cultures. Indeed, we are seeing a similar phenomenon in a separate ongoing study with a bilingual robot playing video games with human interactors, where alterations in robotic socio-cognitive behavior have different effects depending on the language spoken (e.g. Korean versus English) [33].

C. Limitations

There are a number of limitations to this study. By design, the study was setup as an intensive *case study* of a smaller number of users, with the aim of collecting a large amount of data about each participant over many weeks, rather than gathering a small amount of data about many users briefly. However, both approaches have merit and can provide different kinds of information for researchers [34]. The results here thus tell part of the story, but more research is needed using different study designs. To that end, we are currently conducting a larger ongoing user study comparing South Korea and US robotic companion pet use.

Additionally, more sophisticated modeling methods, such as generative adversarial networks (GANs) or variational autoencoders (VAEs) [35] may be needed to make single models work across diverse datasets, though it remains to be seen whether that by itself would solve the problem. Another approach may be to use different learning strategies, such as "few-shot learning" [36,37] that attempts to predict similarity rather than outright classification, averse to traditional supervised learning. These are possibilities that as of yet remain unanswered. We cannot rule out that larger datasets and/or different modeling methods may produce better results. That remains a challenge for future research in the domain of robotic companion pets for home care.

Acknowledgment

This work was supported by the research fund of Hanyang University (HY-2020) in Korea, as well as the National Science Foundation in the United States (Grant# IIS-1900683). We would also like to thank our various collaborators over the years who have contributed to different aspects of this work.

References

[1] L. Pu, W. Moyle, C. Jones, and M. Todorovic. "The effectiveness of social robots for older adults: a systematic review and meta-analysis of randomized controlled studies," *The Gerontologist*, vol.59, no.1, pp. e37-e51. 2019.

[2] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, "Social robots for education: A review," *Science robotics*, vol.3, no.21, pp. eaat5954, 2018.

[3] C.F. Reynolds III, P. Cuijpers, V. Patel, A. Cohen, A. Dias, N. Chowdhary, O.I. Okereke et al., "Early intervention to reduce the global health and economic burden of major depression in older adults," *Annual Review of Public Health*, vol.33, pp. 123-135, 2012.

[4] L. Eldridge, S. Nagata, J. Piatt, C. Stanojevic, S. Šabanović, C.C. Bennett, and N. Randall, "Utilization of socially assistive robots in

recreational therapy," *American Journal of Recreation Therapy*, vol.19, no.2, pp. 35-45, 2020.
[5] A.L. Hartzler, L. Tuzzio, C. Hsu, and E.H. Wagner, "Roles and functions of community health workers in primary care," *The Annals of Family Medicine*, vol.16, no.3, pp. 240-245, 2018.
[6] H.R. Lee, and S. Šabanović, "Culturally variable preferences for robot design and use in South Korea, Turkey, and the United States," *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 17-24, 2014.
[7] T.T. Nomura, D.S. Syrdal, and K. Dautenhahn, "Differences on social acceptance of humanoid robots between Japan and the UK," *In Procs 4th int Symposium on New Frontiers in Human-Robot Interaction. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB)*, 2015.
[8] C.C. Bennett, C. Stanojević, S. Šabanović, J.A. Piatt, and S. Kim, "When no one is watching: Ecological momentary assessment to understand situated social robot use in healthcare," *9th ACM International Conference on Human-Agent Interaction (HAI)*, pp. 245-251, 2021.
[9] R. Andreasson, B. Alenljung, E. Billing, and R. Lowe, "Affective touch in human-robot interaction: Conveying emotion to the Nao robot," *International Journal of Social Robotics*, vol.10, no.4, pp. 473-491, 2018.
[10] T. Nomura, T. Kanda, and T. Suzuki, "Experimental investigation into influence of negative attitudes toward robots on human–robot interaction," *AI & Society*, vol.20, pp. 138–150, 2006.
[11] L. Morillo-Mendez, M.G. Schrooten, A. Loutfi, and O.M. Mozos, "Age-related differences in the perception of eye-gaze from a social robot. *International Conference on Social Robotics (ICSR)*, pp. 350-361, 2021.
[12] F. Rea, P. Muratore, and A. Sciutti, "13-year-olds approach human-robot interaction like adults," *Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pp. 138-143, 2016.
[13] S. Shahid, E. Krahmer, and M. Swerts, "Child–robot interaction across cultures: How does playing a game with a social robot compare to playing a game alone or with a friend?," *Computers in Human Behavior*, vol.40, pp. 86-100, 2014.
[14] T. Komatsu, F.M. Bertrame, and M. Scheutz, "Blaming the reluctant robot: Parallel blame judgments for robots in moral dilemmas across us and japan," *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 63-72, 2021.
[15] O. Rudovic, J. Lee, L. Mascarell-Maricic, B.W. Schuller, and R.W. Picard, "Measuring engagement in robot-assisted autism therapy: A cross-cultural study," *Frontiers in Robotics and AI*, vol.4, pp. 36, 2017.
[16] M.R. Fraune, S. Kawakami, S. Sabanovic, P.R.S. De Silva, and M. Okada, "Three's company, or a crowd?: The effects of robot number and behavior on HRI in Japan and the USA." *Robotics: Science and Systems (RSS)*, vol.10, 2015.
[17] V. Lim, M. Rooksby, and E.S. Cross, "Social robots on a global stage: establishing a role for culture during human–robot interaction." *International Journal of Social Robotics*, vol. 13, no. 6, pp. 1307-1333, 2021.
[18] C.C. Bennett, S. Sabanovic, J.A. Piatt, S. Nagata, L. Eldridge, and N. Randall, "A robot a day keeps the blues away," *IEEE international conference on healthcare informatics (ICHI)*, pp. 536-540, 2017
[19] N. Randall, C.C. Bennett, S. Sabanovic, S. Nagata, L. Eldridge, S. Collins, and J.A. Piatt, "More than just friends: in-home use and design recommendations for sensing socially assistive robots (SARs) by older adults with depression," *Paladyn Journal of Behavioral Robotics*, vol.10, no.1, pp. 237–255, 2019.
[20] S. Shiffman, A.A. Stone, and M.R. Hufford, "Ecological momentary assessment," *Annual Review of Clinical Psychology*, vol.4, pp. 1-32, 2008.
[21] C. Vesel, H. Rashidisabet, J. Zulueta, J.P. Stange, J. Duffecy, F. Hussain, et al., "Effects of mood and aging on keystroke dynamics metadata and their diurnal patterns in a large open science sample: A BiAffect iOS study," *Journal of the American Medical Informatics Association*, vol.27, no.7, pp. 1007–1018, 2020.
[22] J. Zulueta, A. Piscitello, M. Rasic, R. Easter, P. Babu, S.A. Langenecker, et al., "Predicting mood disturbance severity with mobile phone keystroke metadata: A BiAffect digital phenotyping study," *Journal of Medical Internet Research*, vol.20, no.7, pp. e241, 2018.
[23] E.A. Bjorling, E. Rose, A. Davidson, R. Ren, and D. Wong, "Can we keep him forever? Teens' engagement and desire for emotional connection with a social robot," International Journal of Social Robotics, vol.12, no.1, pp. 65–77, 2020.
[24] S. Sabanovic, C.C. Bennett, W.L. Chang, and L. Huber, "PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia," *IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, pp. 1–6, 2013
[25] S. Collins, S. Sabanovic, M. Fraune, N. Randall, L. Eldridge, J.A. Piatt, C.C. Bennett, and S. Nagata, "Sensing companions: Potential clinical uses of robot sensor data for home care of older adults with depression," *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 89–90, 2018.
[26] J.W. Creswell, and J.D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed methods Approaches*, Thousand Oaks, CA, USA: Sage Publications, 2017.
[27] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol.16, pp. 321–357, 2002.
[28] Y. Chen and C.Shen, "Performance analysis of smartphone-sensor behavior for human activity recognition," *IEEE Access*, vol.5, pp. 3095–3110, 2017.
[29] K. Xia, J. Huang, and H. Wang, "LSTM-CNN architecture for human activity recognition," *IEEE Access*, vol.8, pp. 56855–56866, 2020.
[30] C.F. Graumann, "The phenomenological approach to people-environment studies," *Handbook of Environmental Psychology*, vol.5, pp. 95-113, 2002.
[31] B. Bruno, R. Menicatti, C.T. Recchiuto, E. Lagrue, A.K. Pandey, and A. Sgorbissa, "Culturally-competent human-robot verbal interaction," *5th International Conference on Ubiquitous Robots (UR)*, pp. 388–395, 2018.
[32] S.Sabanovic, C.C. Bennett, and H.R. Lee, "Towards culturally robust robots: A critical social perspective on robotics and culture," *Proceedings of the HRI Workshop on Culture-Aware Robotics*, 2014.
[33] C. Bennett, B. Weiss, J. Suh, E. Yoon, J. Jeong, and Y. Chae, "Exploring data-driven components of socially intelligent AI through cooperative game paradigms," *Multimodal Technologies and Interaction*, vol.6, no.2, pp.16, 2022.
[34] C. Stanojevic, L.A Eldridge, J.L. McIntire, A. Bowen, S. Dawson, B. McCormick, et al., "Employing an international research collaboration framework to pilot test an evidence-based recreational therapy program," *Therapeutic Recreation Journal*, vol.56, no.1, pp. 1-16, 2022.
[35] G.M. van de Ven, H.T. Siegelmann, and A.S. Tolias, "Brain-inspired replay for continual learning with artificial neural networks," *Nature Communications*, vol. 11, no. 1, pp. 1-14, 2020.
[36] Y. Wang, Q. Yao, J.T. Kwok, and L.M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Computing Surveys*, vol. 53, no. 3, pp. 1-34, 2020.
[37] Y. Wu, and Y. Demiris, "Towards one shot learning by imitation for humanoid robots," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2889-2894, 2010.