Cultural Difference in Back-Imitation's Effect on the Perception of Robot's Imitative Performance

Yasser Mohammad[1(✉)] and Toyoaki Nishida[2]

[1] Assiut University, Asyut, Egypt
yasserm@aun.edu.eg
[2] Kyoto University, Kyoto, Japan
nishida@i.kyoto-u.ac.jp

Abstract. Cultural differences have been documented in different aspects of perception of robots as well as understanding of their behavior. A different line of research in developmental psychology has established a major role for imitation in skill transfer and emergence of culture. This study is a preliminary cross–cultural exploration of the effect of imitating the robot (back imitation) on human's perception of robot's imitative skill. In previous research, we have shown that engagement in back imitation with a NAO humanoid robot, results in increased perception of robot's imitative skill, human–likeness of motion, and willingness of future interaction with the robot. This previous work mostly used Japanese university students. In this paper, we report the results of conducting the same study with subjects of two cultures: Japanese and Egyptian university students. The first finding of the study is that the two cultures have widely different expectations of the robot and interaction with it and that some of these differences are significantly reduced after the interaction. The second finding is that Japanese students tended to attribute higher imitation skill and human likeness to the robot they imitated while Egyptian students did not show such tendency. The paper discusses these findings in light of known differences between the two cultures and analyzes the role of expectation in the differences found.

1 Introduction

Attitude toward robots is one of the major factors determining the success or failure of future social robots that are expected to occupy our homes, offices, hospitals and schools. One important factor that affects these attitudes is culture.

Culture is a multifaceted and complex concept that may have different meanings for different researchers [19]. In this work, we follow Samani et al. [19] and Taras et al. [21] and define culture as a group's shared set of specific basic beliefs, values, practices and artefacts that are formed and retained over a long period of time. This includes communicative aspects (e.g. nonverbal behaviors including gestures and proximities).

J.T.K.V. Koh et al. (Eds.): Cultural Robotics 2015, LNAI 9549, pp. 17–32, 2016.
DOI: 10.1007/978-3-319-42945-8_2

Previous studies have shown that culture plays an important role in shaping people's attitudes toward robots in several contexts. For example, Bartneck [1] studied the perception of robot anthropomorphism and likability for United States and Japanese subjects and found that Japanese subjects tended to like conventional robots more than US subjects while the reverse was observed for androids (e.g. robots with highly human–like appearance covered with artificial skin) [1]. Finding differences between eastern and western cultures in cross–cultural HRI research is common. Lee and Sabanović [9] studied the acceptability of different robot designs (appearance) by subjects from Turkey, South Korea, and United States. They found that religious belief and media exposure are not enough to explain the discovered differences between people from these countries in their preferences which suggests a specific role of culture. Both of these studies involved measuring people's response to robot representations (e.g. images) rather than actual interactions with them.

It is commonly held that westerners perceive robots differently than easterners because of the difference of their portray in media. A common example is comparing "The Terminator" with "Astro Boy". While the first is a killing machine the later is a helping child–like robot with human–like curiosity and emotions. This conception though is challenged by some research findings. For example Bartneck et al. compared Dutch, Chinese, German, Mexican, American (USA) and Japanese participants based on the Negative Attitude towards Robots Scale (NARS) and found no particularly positive attitudes for Japanese participants [2]. Wang et al. found that Chinese participants expressed more negative attitudes toward robots than American participants [23]. Shibata et al. reported no difference between UK and Japanese participants when subjectively reporting about a Paro robot and found in both cases that physical interaction improves subjective evaluations of the robot [20]. These results taken together does not support the simplistic commonly held belief that eastern people are more accepting of robots than their western counterparts but shows a complicated interaction between several factors including appearance, culture, interaction quality, etc.

Cultural transfer may be mediated by imitation. Nielsen [14] argues that emergence of imitation and play in children was a precursor for the emergence of culture as a complex construct in human life. Imitation is not always a conscious process in humans. For example, Chartrand and Bargh experimentally showed that behavioral mimicry has a significant effect on the interaction and increases empathy towards the interaction partner [3] which is usually referred to as the "chameleon effect". Several HRI studies looked for similar effects when people interact with robots. Riek et al. showed that real–time head gesture mimicry improves rapport between a human and a robot [16].

HRI studies of imitation have focused on the effect of robot's imitative ability on human's perception of the traits of this robot and convincingly argued for a positive effect [16]. In a series of previous studies [11–13], we investigated the opposite case in which a human imitates the robot. The main hypothesis was that this form of back-imitation will have positive effects on the perception of robot's imitative skill and may also lead to more acceptance [11]. We found that back-imitation leads indeed to increased perception of robot's imitative skill and human–likeness of motion and may lead to increased intention of future interaction with it [13]. For the purposes of this study we define back imitation following Mohammad and Nishida [13] as *the imitation of the learner by the teacher during, before or after the demonstration of a new task*.

These studies were conducted using mostly Japanese university student participants and no cultural evaluation was conducted. In this paper, we repeat one of these experiments with participants from Japan and Egypt and show that the positive effects of back-imitation were lacking in Egyptian subjects. We discuss this results in terms of the effect of prior expectation and cultural aspects.

A few studies reported the response of Egyptian subjects to robots. For example, Trovato et al. [22] compared the response of Egyptian and Japanese subjects to a humanoid robot speaking in Arabic (native language of Egypt) and Japanese and found that people from each nationality preferred robots that spoke in their native language and used the culture-specific greeting gestures. The experiment was conducted using only videos of the robot. One problem of this study is that the effect of language understanding may overshadow other cultural differences. Salem et al. [18] conducted a cross–cultural study in which a humanoid robot (Ibn Sina) was displayed in a major exhibition (Dubai's GITEX) and compared the response of people from different nationalities including African Arabs and South eastern Asians. The study focused on the order of robot applications and found significant interplay between religion, age and cultural origin and acceptance of robots in different applications.

This work differs from the aforementioned studies in that it focuses on actual interaction with the robot (a NAO humanoid robot in our case) and measures the effect of a behavioral aspect of the robot instead of its appearance or design. We believe that behavior and motion are as important as appearance in attribution of skill and human–likeness and in general acceptance of the robot for different roles.

Imitative skill in this paper is defined as the objective accuracy in copying limb motions demonstrated by the human. As such, it is related to motion human–likeness which describes the degree by which motion trajectories of robot limbs resemble human motion in general not necessarily the demonstrated behavior. For example, a robot that closes its hand during demonstrating a waving gesture will have low imitative skill but the motion can still be human–like in the sense that it is similar in form to normal human motion in terms of smoothness and respecting human joint range limits. A concept related to human–likeness that we discuss later in this paper is humanness which is defined as the degree by which humanity is ascribed to an agent [5]. Our previous studies found that two factors contribute to this overall assessment of humanness clustering positive traits (e.g. curiousity, sociability, friendliness) and negative traits (e.g. jeouleousy, impatience, distractibility) [13]. These two clusters of features consitute the positive and negative humanness scores in this study. Interaction quality is defined here as the participant's overall subjective evaluation of her interaction with the robot.

The rest of the paper is organized as follows: Sect. 2 details the experimental design used in this study and comments on different design choices. Section 3 reports the results of the study and Sect. 4 discusses their implications. The paper is then concluded.

2 Experimental Design

The design of this experiment is similar to the main study reported in [13]. The main difference is that participants came from two different nationalities (Egyptian and Japanese). This entailed employing appropriately different statistical analysis of the questionnaires.

The experiment was conducted in Japan which allowed us to recruit 36 Japanese subjects but only 10 Egyptian subjects. We used the data of only 10 Japanese subjects who participated in the experiment reported in [13] selected to match the gender, age and education level of the 10 available Egyptian participants. This is achieved by removing all female Japanese subjects (as all Egyptian subjects were males), we then removed younger Japanese subjects until we had 15 subject of which we picked 10 subjects randomly. This led to 20 participants in total for this study. All participant were male with average age of 26 years for Japanese participants and 30 years for Egyptian participants. Sixteen of the participants were studying STEM subjects and the other four were majoring in humanities (one from Egypt and three from Japan). It should be noted that we found no difference based on educational background (STEM/humanities) in any of the aspects studied in [13] or this paper. None of the participants had previous interaction with robots and none of them had previous exposure to the robot used in the experiment (NAO).

The robot used in this paper was NAO V3.3 [4] which is a small humanoid robot (Height = 57.3 cm, Width = 27.5 cm) produced by Alderbaran Robotics. Only four of the seven DoFs of each arm were controlled in this study (2DoFs in the shoulder and 2DoFs in the elbow). The lower body of the robot was fixed in a stable pose. Participant motion was collected using a Kinect sensor and the data was fed to the robot software in real time.

The experimental procedure was identical to the main study in Mohammad and Nishida [13]. We provide a brief description of the procedure here for completeness. The three conditions for the interaction were NI (No Imitation), BI (Back Imitation) and MI (Mutual Imitation) that will be explained in detail shortly. Participants had two conditions either Egyptian (EGY) or Japanese (JPN).

The experiment involved interactions between the NAO robot, the participant and a physically realistic NAO simulator (called WAN throughout the study) that was projected on a standard computer screen using Choregraphe [15]. The NAO robot and the simulator were controlled using the same software developed based on the C++ NAOqi SDK which allowed us to elicit the same motions with the same speeds from the robot and the simulator.

The experiment was designed as two rounds of a game called follow–the–leader where either the NAO robot, its simulated agent, or the participant was assigned the leader's role and the other two players tried to just copy his/its arm motion as fast and as accurately as possible.

The experimental procedure consisted – after the orientation – of three sessions of this game in the three conditions to be explained soon. A pre–experimental questionnaire (PREQ) and a post–experimental questionnaire (POSTQ) were employed as well as one questionnaire after each session (Q1, Q2, Q3). See Fig. 1 for examples from these questionnaires.

Each session consisted of two rounds. In the first round, either the robot or the simulated agent was assigned the leader's role and in the second round, the participant was always the leader and was imitated only by the robot.

The first round was the manipulated part of the experiment. Three conditions were used: BI (Back Imitation) condition in which the leader was the robot. MI (Mutual Imitation) condition in which the robot was the leader as long as the participant is accurately imitating its motion but when the participant fails in this imitation, the robot imitates the participant once then reverts to become the leader again. NI (No Imitation) condition in which the simulated agent is the leader.

The participant imitated something in the first round in all conditions (even the NI condition). What is meant by *no imitation* in the NI condition is that the participant did not imitate *the robot*. The MI condition is an extreme simplification of mother–infant mutual imitation in early years of life [7]. The only difference between the NI and BI conditions was the order by which the robot and the simulated agent moved. In the NI condition, the simulated agent moved first which made it the leader and in the BI condition the robot moved first which made it the leader. The MI condition differed from the BI condition only in that the robot occasionally (when participant's imitation was far from perfect) imitated the participant. For more information on the experimental design and justification for design decisions, please refer to Mohammad and Nishida [13].

The second round was identical in all the sessions and only the robot (not the on–screen agent) copied the pose of the subject in real time using the system proposed by Mohammad and Nishida in [10] with minor modifications. For more details on the imitation engine used please refer to [10,13]. This second round was conducted for 5 min for every participant. The same algorithm was employed with the same parameters in the three sessions which means that *objectively* the imitative skill of the robot was the same in the three conditions.

Session and pre–experimental questionnaires measured 22 independent variables (shown in Table 1) on a Semantic Differential Scale (see Fig. 1 for an example) while the POSTQ questionnaire measured the preferences of the subjects on the nine independent variables corresponding to measurement of robot skill and interaction quality. As Fig. 1 shows, the participant had the choice to select no best/worst condition.

The first five items – in all questionnaires – measured robot skill (i.e. accuracy, speed, naturalness of movement, human–likeness of motion and overall performance). One item measured participant's self evaluation of his imitative skill during the first part of the session. The remaining 16 items were the same as

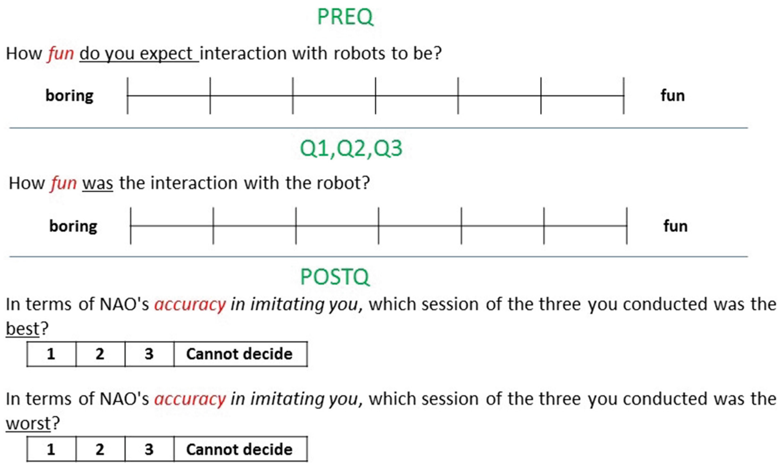

Fig. 1. Sample questions from the five questionnaires used in this study. The same questionnaires were used in [13].

Table 1. Dimensions of Evaluation employed in this study (see [13] for internal consistency evaluation).

Dimension(cronbach's α)	Dependent variable (% of variance)	Indep. variables(Loadings)
Robot skill(0.94)	Imitative skill(86)	Accuracy(0.6), overall(0.45), naturalness(0.66)
	Human–likeness of motion(6)	Human–likeness(0.995)
	Speed(5)	Speed(0.97)
Robot humanness(0.94)	Positive(70)	Curious(0.5), friendly(0.6), fun loving(0.4), sociable(0.4), *trusting(0.2)*[a]
	Negative(10)	Distractible(0.8), impatient(0.4), jealous(0.4), *nervous(0.0)*[a]
Interaction quality(0.97)	(98)[b]	Pleasant(0.7), fun(0.7)
Intention of future(0.89) interaction	(90)[b]	Closeness(0.7), living with the robot(0.7)
Likability(0.89)	Likability (84)	Polite(0.6), *sympathetic(0.02)*[a], humble(0.8)

[a]Items removed because they had small (<0.3) loadings.
[b]Dependent variable name is the same as the dimension name.

the ones used by Salem et al. to measure humanness (based on the scale designed by Haslam et al. [5]), shared-reality and likability [17].

The order of exposure to the three conditions (NI, BI, MI) was randomized between subjects. This is one difference from the study in [13] for which the higher number of participants allowed for a balancing of all ordering possibilities (12 in total). Nevertheless, the same orderings were used for Egyptian and Japanese subjects.

3 Results and Discussion

The goal of this study is to assess cultural differences between Egyptian and Japanese subjects related to back imitation. We analyzed the questionnaire data from different angles as will be shown in this section.

3.1 Effect of Imitation Condition

The first analysis step was multivariable ANOVA with nationality, imitation condition and session order as independent variables. The results are shown in Table 2. There were statistically significant effects of nationality (culture) on imitative skill ($F = 5.9094(1, 46)$, $p = 0.019$) and human–likeness of motion ($F = 14.5562(1, 46)$, $p = 0.0004$). There was no statistically significant effect for imitation condition which is in line with the results reported in [13] in which only preferences showed a statistically significant difference between conditions probably due to cognitive mediation.

Table 2. Multivariable ANOVA Analysis. Only dependent variables that showed statistically significant results are reported.

Dependent	Independent	F	p
Imitative skill	Condition	0.0328(2, 46)	0.9677
	Nationality	**5.9094(1, 46)**	**0.0190**
	Order	0.4694(2, 46)	0.6283
	Condition*Nationality	0.3492(2, 46)	0.7071
	Condition*order	0.9805(4, 46)	0.4275
	Nationality*order	0.4300(2, 46)	0.6531
Human–likeness of motion	Condition	1.1278(2, 46)	0.3325
	Nationality	**14.5562(1, 46)**	**0.0004**
	Order	0.9032(2, 46)	0.4123
	Condition*Nationality	**3.1434(2, 46)**	**0.0525**
	Condition*order	0.5581(4, 46)	0.6942
	Nationality*order	0.1025(2, 46)	0.9028

More interestingly for our current study there is an interaction between the experimental condition (NI, BI, MI) and nationality ($F = 3.7400(2, 46)$, $p = 0.0313$) in human–likeness of motion. This interaction suggests that nationality affects the way participants' perception of human–likeness of motion was affected by the experimental condition. We found no ordering effect or interactions and this was confirmed by factorial Wilcoxon rank sum test.

We then compared the three conditions using Wilcoxon rank sum test and found no statistically significant difference between the three conditions when looking at all participants. For this test and for all factorial tests of the three experimental conditions, we use Sidak's multi comparison correction formula. Instead of the standard modification of the significance level α, we increase the individual *p-values* according to Sidak's formula assuming three tests:

$$p' = 1 - (1 - p)^3$$

This is reported as *adj.p* in all tables in this paper and is used as the basis for accepting or rejecting hypotheses. We also report the Hedge's g effect size [6] and 95 % confidence intervals for all tests.

Table 3. Results of Wilcoxon rank sum test comparing Egyptian and Japanese participants (EGY vs. JPN) based on answers to session questionnaires. Only variables that showed statistically significant results are reported.

Dimension	p	Ranksum	z	Hedges' g	95 % CI
Imitative skill	**0.007**	1096.50	2.68	**0.743**	[**0.205, 1.261**]
Human likeness	<**0.001**	1174.00	3.91	**1.088**	[**0.526, 1.622**]

To confirm the effect of nationality on subjective evaluations in session questionnaires found in the aforementioned ANOVA analysis, we conducted factorial Wilcoxon rank sum test comparing Egyptian and Japanese subjects (independent of session condition) in their evaluation of the robot. Dependent variable results are shown in Table 3 which shows statistically significant effect for imitation skill and human–likeness of motion. Both were higher for Egyptian subjects (M/SD = 4.23/1.28 and 3.53/1.26 in order) than for Japanese subjects (M/SD = 3.26/1.25 and 2.19/1.13 in order). The relation between this finding and expectations of these subjects will be discussed later in this section.

As a final check of the effect of nationality, we repeated the factorial Wilcoxon rank sum test for participants of each nationality. Egyptian and Japanese participants showed no statistically significant difference between the three experimental conditions (NI, BI, MI) in any of the measured dimensions.

Given the failure of direct participant evaluation of the robot in detecting any difference between conditions, we analyzed the preference data from participants. To analyze the preferences collected in POSTQ, we calculated a score for every session as follows (see Fig. 1 for an example question from this questionnaire): If the subject selected one session as best in some dimension, it received a +1 score. If a session was selected as worst it received a −1 score. If only a best session was selected, the remaining two sessions received a −0.5 score. If only a worst session was selected the remaining two sessions received a +0.5 score.

Finally, if no sessions were selected as best or worst (30 % of the subjects), the three sessions received a zero score.

Using Wilcoxon rank sum test to analyze this preference data, we found no statistically significant differences for the 20 participants. Analyzing the data for participants of each nationality separately showed a different story. While Egyptian participants did not show any statistically significant difference between the three conditions on any of the evaluation dimensions (Table 5), Japanese participants showed statistically significant preference for the BI and MI conditions over the NI condition for naturalness and imitative skill (Table 4).

Table 4. Results of Wilcoxon rank sum test comparing the preferences of the three conditions for Japanese participants

Dimension	Conditions	adj. p	Ranksum	Hedges' g	95 % CI
Accuracy	BI vs. NI	0.565	31.0	0.818	[−0.490, 2.001]
	MI vs. NI	0.165	29.0	0.935	[−0.400, 2.126]
	MI vs. BI	0.842	43.0	−0.196	[−1.372, 1.010]
Naturalness	BI vs. NI	0.088	25.50	1.685	[0.138, 2.974]
	MI vs. NI	**0.045**	23.50	**1.922**	**[0.294, 3.254]**
	MI vs. BI	0.987	41.5	−0.108	[−1.289, 1.089]
Human–likeness of motion	BI vs. NI	0.763	33.0	0.565	[−0.694, 1.737]
	MI vs. NI	0.709	32.0	0.642	[−0.631, 1.816]
	MI vs. BI	1.000	39.5	−0.000	[−1.187, 1.187]
Overall	BI vs. NI	0.444	30.0	0.980	[−0.366, 2.174]
	MI vs. NI	**0.022**	24.00	**2.023**	**[0.359, 3.376]**
	MI vs. BI	0.975	36.0	0.387	[−0.844, 1.558]
Speed	BI vs. NI	0.999	41.0	−0.164	[−1.341, 1.039]
	MI vs. NI	0.981	43.0	−0.351	[−1.523, 0.874]
	MI vs. BI	1.000	41.0	−0.186	[−1.362, 1.019]
Imitative skill	BI vs. NI	**0.045**	24.00	**1.582**	**[0.068, 2.853]**
	MI vs. NI	**0.026**	23.00	**2.117**	**[0.419, 3.490]**
	MI vs. BI	0.999	40.0	−0.049	[−1.232, 1.143]

For Japanese participants, Table 4 shows a statistically significant difference in naturalness (a component of the human–likeness of motion independent variable according to Table 1) between the MI and NI conditions (adj. p = 0.045) and insignificant difference with adj. p = 0.088 for BI and NI conditions. Nevertheless, only statistically significant differences in imitative skill were found in this study in preferences data. Mohammad and Nishida [13] reported – on the other hand – statistically significant differences between the same conditions for both imitative skill and human–likeness of motion. The inability to reproduce

Table 5. Results of Wilcoxon rank sum test comparing the preferences of the three conditions for Egyptian participants

Dimension	Conditions	adj. p	Ranksum	Hedges' g	95 % CI
Accuracy	BI vs. NI	0.724	22.5	0.679	$[-0.751, 1.977]$
	MI vs. NI	0.957	24.5	0.393	$[-0.978, 1.688]$
	MI vs. BI	0.930	30.5	−0.305	$[-1.603, 1.051]$
Naturalness	BI vs. NI	0.611	31.0	−1.095	$[-2.511, 0.565]$
	MI vs. NI	0.939	20.0	0.310	$[-1.296, 1.835]$
	MI vs. BI	0.636	12.5	1.011	$[-0.941, 2.643]$
Human–likeness of motion	BI vs. NI	0.977	25.5	0.333	$[-1.028, 1.629]$
	MI vs. NI	0.951	30.0	−0.305	$[-1.603, 1.051]$
	MI vs. BI	0.636	33.5	−0.808	$[-2.113, 0.653]$
Overall	BI vs. NI	0.092	38.00	−1.979	$[-3.460, -0.115]$
	MI vs. NI	0.648	33.5	−0.857	$[-2.165, 0.617]$
	MI vs. BI	0.611	21.5	0.748	$[-0.698, 2.049]$
Speed	BI vs. NI	0.370	19.5	1.150	$[-0.409, 2.485]$
	MI vs. NI	0.993	27.0	−0.409	$[-1.792, 1.065]$
	MI vs. BI	0.136	34.00	−2.097	$[-3.701, -0.027]$
Imitative skill	BI vs. NI	0.763	29.0	−0.496	$[-1.878, 0.997]$
	MI vs. NI	1.000	22.5	−0.056	$[-1.603, 1.505]$
	MI vs. BI	0.716	13.0	0.463	$[-1.293, 2.073]$

the difference in human–likeness of motion may be due to the small sample size in this study.

The conclusion we can draw from these results is that Egyptian and Japanese participants in this study differed in their response to the BI and MI conditions in comparison with the NI condition. While back and mutual imitation was associated with an increase in the perception of robot's imitative skill for Japanese subjects (as reported previously in [13]), no such differences were found for Egyptian subjects.

3.2 Analysis of Expectations and Post-Experimental Questionnaires

Table 3 shows statistically significant differences in subjective evaluations of human–likeness of motion and imitative skill between Egyptian and Japanese participants (independent of the experimental condition). This hints at a general difference in the perception of these variables by participants of the two nationalities.

To check this possibility we compared the post–experimental questionnaires and expectations of the participants of each country. Table 6 shows the results of

Table 6. Results of Wilcoxon rank sum test comparing the expectations of Egyptian and Japanese participants

Dimension	p	Ranksum	z	Hedges' g	95 % CI
Imitative skill	0.520	96.0	−0.6	0.323	[−0.597, 1.216]
Human–likeness of motion	**0.016**	74.00	−2.40	**1.090**	**[0.073, 2.016]**
Speed	0.847	102.0	−0.2	0.062	[−0.840, 0.960]
Interaction quality	**0.012**	72.00	−2.50	**1.309**	**[0.252, 2.255]**
Intention of future interaction	**0.002**	145.50	3.08	**−2.101**	**[−3.155, −0.869]**
Humanness (Positive)	**0.002**	146.00	3.06	**−1.923**	**[−2.948, −0.734]**
Humanness (Negative)	**0.007**	141.00	2.69	**−1.426**	**[−2.385, −0.347]**
Likability	0.306	119.0	1.0	**−0.610**	**[−1.507, 0.339]**

this analysis for expectations measured during the pre–experimental questionnaire using Wilcoxon's rank sum test. Egyptian subjects showed higher expectations of human–likeness of motion ($p = 0.016$), and interaction quality ($p = 0.012$) and lower expectation of both positive humanness and negative humanness ($p = 0.002$ and $p = 0.007$ respectively) and intention of future interaction ($p = 0.002$) compared with Japanese subjects. There was no difference in the expectation of robot speed, imitative skill or likability.

Analysis of the post–experimental questionnaire shows a slightly different pattern. Egyptian subjects now differed from Japanese subjects in imitative skill ($p = 0.025$), positive and negative humanness ($p < 0.001$ and $p = 0.031$ respectively). These results are summarized in Table 7. The main difference between these results and subject expectations (in the pre–experimental questionnaire) is the disappearance of the difference in human–likeness of motion, interaction quality, and imitation skill (see Table 6 compared to Table 7)

Comparing the post–experimental and pre–experimental questionnaires directly using paired t-test for both nationalities revealed no differences for Japanese subjects but a statistically significant reduction of the subjective

Table 7. Results of Wilcoxon rank sum test comparing the posterior evaluation of Egyptian and Japanese participants

Dimension	p	Ranksum	z	Hedges' g	95 % CI
Imitative Skill	**0.025**	135.00	2.23	**−1.138**	**[−2.068, −0.113]**
Human–likeness of motion	0.416	94.0	−0.8	0.298	[−0.620, 1.191]
Speed	0.190	122.5	1.3	−0.561	[−1.457, 0.382]
Interaction quality	0.125	125.5	1.5	**−0.642**	**[−1.541, 0.310]**
Intention of future interaction	0.878	107.5	0.2	−0.039	[−0.937, 0.863]
Humanness (Positive)	**<0.001**	153.00	3.59	**−2.768**	**[−3.943, −1.358]**
Humanness (Negative)	**0.031**	134.00	2.16	**−1.233**	**[−2.171, −0.191]**
Likability	**0.001**	149.50	3.33	**−2.091**	**[−3.143, −0.862]**

perception of robot's imitative skill ($p = 0.008$) and interaction quality ($p = 0.006$) for Egyptian subjects.

These results taken together suggest that Egyptian participants had higher expectations for robot's human–likeness of motion and interaction quality compared with Japanese participants, yet this difference disappeared after the experiment. On the other hand, there was a consistent difference in attribution of humanness (both positive and negative). Egyptian subjects in general attributed less humanness to the robot than Japanese subjects and this was not affected by the experiment (i.e. it appeared in both the pre–experimental and post–experimental questionnaires).

4 Discussion and Limitations

The statistical analysis reported in Sect. 3 revealed several differences between Egyptian and Japanese participants in this experiment that we will try to discuss and understand in light of known cultural differences between the two countries.

Considering expectations, Egyptian subjects showed higher initial expectation of robot's human–likeness of motion and interaction quality. This was not caused by a difference in previous experience either with the NAO itself or with other robots as all our participants reported never to have interacted with any robots before. Expectations of speed (another measure of skillfulness) did not show any difference between the two groups. Future investigation may be necessary to find an explanation for this difference or role it out on a larger sample.

A more interesting finding is that Egyptian subjects gave the robot consistently lower scores in humanness compared with Japanese subjects before and after the experiment even though they changed their scores in other factors (e.g. imitative skill and interaction quality).

Two cultural factors may be related to this difference. Firstly, the common belief that Japanese people have a more positive attitude toward robots [1] is sometimes attributed to being more secure regarding the challenge to human specialty posed by robots [8]. The common explanation is that the Shinto belief in *kami* with their mobility and existence in nature attributes to the inanimate spiritual features reserved usually for humans and the all powerful God in religions of the west (Judaism, Christianity, and Islam). This same factor may explain our finding that Japanese subjects were willing to give high humanness scores to the robot before and after interacting with it. Bartneck [1] also found that Japanese subjects prefer conventional robots over androids and the NAO robot used in this experiment was more of the first kind.

Another cultural aspect related to this finding is the traditional Islamic negative view of human–like (or even animal-like) pictures and sculptures. Most scholars of Islam ban images (*sowrah* in Arabic) depicting animate beings (e.g. humans and animals) from ritual places and many Muslim Egyptian families do not hang such pictures in their homes until today. The main justification of this negative image is that the artist when depicting animate objects is taking the role of the creator and this is considered a grave sin in Islam. This clear distinction between God-created animate beings and Human-created artificial beings may have contributed to the low expectations of humanness assigned by Egyptian participants (all Muslims) to the robot even though they had higher expectations of human–likeness of *motion* because human–like motion is not at odds with the traditional ban on human–like appearance in the artificial.

Another important finding of this paper is the interaction between culture and evaluation of the three conditions in the experiment. Egyptian participants were not affected by back or mutual imitation while Japanese subjects showed higher perception of the imitative skill and human–likeness of motion for the robot they imitated confirming the results reported in [12,13].

Two possible psychological factors were given in [13] as contributers to the effect of back imitation on perception of imitative skill and human–likeness of motion: effort justification and increased perception of agency. Effort justification refers to people's tendency to attribute greater value or importance to whatever they invest effort in and is stemming from Festinger's theory of cognitive dissonance. Effort justification should be the same for both Japanese and Egyptian participants in this experiment. This implies that the increased perception of agency may have been the critical factor. Again, culture may have something to do with this difference. Japanese people are more accepting of assigning higher humanness scores probably because of accepting a fuzzier boundary between the natural/animate and artificial/inanimate as discussed earlier. Egyptian people are less generous here probably because of the stricter boundary between the natural/animate and artificial/inanimate. Based on that, Egyptian subjects may have been less inclined to assign higher agency to the robot they imitated. It is still an artificial creation which should not and cannot have agency in this strict boundary view.

Another interesting finding of the paper is that even though Egyptian subjects had higher expectations of interaction quality and human–likeness of motion before the experiment, the short interaction with the robot reduced these unrealistic expectations causing no difference between Egyptian and Japanese participants' evaluations of these two facets of the robot after the experiment. This shows that while some cultural differences may need to just be taken into account when designing HRI scenarios (e.g. humanness assignment discussed earlier), other differences may be reduced by giving people of different cultures enough time to interact with the robot.

The findings of this paper and the previous discussion should not be taken at face value because of the limitations of this study. Firstly, the number of participants was too small to draw conclusions on the two cultures at large. Moreover, the sample was not representative of the two cultures: all participants were university students and most of them were graduate students. Moreover, all participants were students in Kyoto University which is one of the top universities in Japan. This means that they are not even fully representative of university students in these countries. This is especially true for Egyptian participants who were all, except one, graduate students living outside their country.

Nevertheless, and despite these limitations, the reported experiment hints at cultural differences between Japanese and Egyptian people in interacting with robots. The importance of these findings is enhanced by the fact that they were based on actual interaction with an autonomous robot instead of being based on pictures, videos or interactions with a WOZ (Wizard of Oz) operated robot. It enforces the message from several other studies in the HRI community that cultural aspects must be taken into account when designing robots but extends that concept to the design of the interaction instead of only appearance and robot behavior. For example, while a short session of back-imitation may be a good idea for familiarizing Japanese subjects with a robot (as we suggested in [13]), this same manipulation is not expected (based on the results of this study) to have any effect on Egyptian subjects.

5 Conclusion

In this study, expectations and effects of interaction with a NAO robot between Egyptian and Japanese subjects in a follow–the–leader game were compared. A statistically significant interaction between the participant's culture and the effect of back and mutual imitation on them was found. While Japanese subjects tended to assign higher scores of imitative skill and human–likeness of motion to the robots they imitated themselves, no such effect was found for Egyptian subjects. Moreover, analysis of the expectations and post–experimental questionnaires of the two groups revealed patterns of difference that are interesting for a follow-up study. For example, Egyptian subjects consistently assigned lower humanness scores to the robot compared with Japanese subjects. This difference may have been affected by the specific religious views common in these two cultures. In the future we will consider an expanded questionnaire that measures religious leanings of participants to provide more insight into this possibility.

Acknowledgments. This study has been partially supported by JSPS Grant-in-Aid for JSPS Postdoctoral Fellows P12046, JSPS KAKENHI Grant Number 24240023 and 15K12098, the Center of Innovation Program from JST, and AFOSR/AOARD Grant No. FA2386-14-1-0005

References

1. Bartneck, C.: Who like androids more: Japanese or us americans? In: The 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2008, pp. 553–557. IEEE (2008)
2. Bartneck, C., Nomura, T., Kanda, T., Suzuki, T., Kato, K.: Cultural differences in attitudes towards robots. In: Symposium on Robot Companions (SSAISB 2005 Convention), pp. 1–4 (2005)
3. Chartrand, T.L., Bargh, J.A.: The chameleon effect: the perception-behavior link and social interaction. J. Pers. Soc. Psychol. **76**(6), 893–902 (1999)

4. Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J., Lafourcade, P., Marnier, B., Serre, J., Maisonnier, B.: Mechatronic design of NAO humanoid. In: IEEE International Conference on Robotics and Automation, ICRA 2009, pp. 769–774, May 2009
5. Haslam, N., Loughnan, S., Kashima, Y., Bain, P.: Attributing and denying humanness to others. Eur. Rev. Soc. Psychol. **19**(1), 55–85 (2008)
6. Hedges, L.V.: Distribution theory for Glass's estimator of effect size and related estimators. J. Educ. Behav. Stat. **6**(2), 107–128 (1981)
7. Jones, S.S.: The development of imitation in infancy. Philos. Trans. R. Soc. B: Biol. Sci. **364**(1528), 2325–2335 (2009)
8. Kaplan, F.: Who is afraid of the humanoid? investigating cultural differences in the acceptance of robots. Int. J. Humanoid Rob. **1**(03), 465–480 (2004)
9. Lee, H.R., Sabanović, S.: Culturally variable preferences for robot design and use in south korea, turkey, and the united states. In: Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction, HRI 2014, pp. 17–24. ACM (2014)
10. Mohammad, Y., Nishida, T.: Tackling the correspondence problem. In: Yoshida, T., Kou, G., Skowron, A., Cao, J., Hacid, H., Zhong, N. (eds.) AMT 2013. LNCS, vol. 8210, pp. 84–95. Springer, Heidelberg (2013)
11. Mohammad, Y., Nishida, T.: Effect of back and mutual imitation on human's perception of a humanoid's imitative skill. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2014, pp. 788–795. IEEE (2014)
12. Mohammad, Y., Nishida, T.: Human-like motion of a humanoid in a shadowing task. In: International Conference on Collaboration Technologies and Systems, CTS 2014, pp. 123–130. IEEE (2014)
13. Mohammad, Y., Nishida, T.: Why should we imitate robots? effect of back imitation on judgment of imitative skill. Int. J. Social Robot. **7**(4), 497–512 (2015). http://dx.doi.org/10.1007/s12369-015-0282-2
14. Nielsen, M.: Imitation, pretend play, and childhood: essential elements in the evolution of human culture? J. Comp. Psychol. **126**(2), 170–178 (2012)
15. Pot, E., Monceaux, J., Gelin, R., Maisonnier, B.: Choregraphe: a graphical tool for humanoid robot programming. In: The 18th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2009, pp. 46–51. IEEE (2009)
16. Riek, L.D., Paul, P.C., Robinson, P.: When my robot smiles at me: enabling human-robot rapport via real-time head gesture mimicry. J. Multimodal User Interfaces **3**(1–2), 99–108 (2010)
17. Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., Joublin, F.: To err is human (-like): effects of robot gesture on perceived anthropomorphism and likability. Int. J. Soc. Rob **5**(3), 1–11 (2013)
18. Salem, M., Ziadee, M., Sakr, M.: Marhaba, how may i help you?: effects of politeness and culture on robot acceptance and anthropomorphization. In: Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction, HRI 2014, pp. 74–81. ACM (2014)
19. Samani, H., Saadatian, E., Pang, N., Polydorou, D.: Newton Fernando, O.N., Nakatsu, R., Valino Koh, J.T.K.: Cultural robotics: The culture of robotics and robotics in culture. Int. J. Adv. Robot. Syst. **10**, 400–409 (2013)
20. Shibata, T., Wada, K., Tanie, K.: Statistical analysis and comparison of questionnaire results of subjective evaluations of seal robot in Japan and UK. In: IEEE

International Conference on Robotics and Automation, ICRA 2003, vol. 3, pp. 3152–3157. IEEE (2003)
21. Taras, V., Rowney, J., Steel, P.: Half a century of measuring culture: review of approaches, challenges, and limitations based on the analysis of 121 instruments for quantifying culture. J. Int. Manag. **15**(4), 357–373 (2009)
22. Trovato, G., Zecca, M., Sessa, S., Jamone, L., Ham, J., Hashimoto, K., Takanishi, A.: Cross-cultural study on human-robot greeting interaction: acceptance and discomfort by egyptians and japanese. Paladyn, J. Behav. Rob. **4**(2), 83–93 (2013)
23. Wang, L., Rau, P.L.P., Evers, V., Robinson, B.K., Hinds, P.: When in rome: the role of culture & context in adherence to robot recommendations. In: ACM/IEEE International Conference on Human-Robot Interaction, HRI 2010, pp. 359–366. IEEE Press (2010)