

Article

Identifying Personality Dimensions for Engineering Robot Personalities in Significant Quantities with Small User Groups

Liangyi Luo ^{1,*}, Kohei Ogawa ² and Hiroshi Ishiguro ¹

¹ Intelligent Robotics Laboratory, Department of Systems Innovation, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka 560-8531, Japan; ishiguro@irl.sys.es.osaka-u.ac.jp

² Intelligent System Laboratory, Graduate School of Engineering, Nagoya University, Furocho, Chikusa, Nagoya 464-8603, Japan; ogawa@irl.sys.es.osaka-u.ac.jp

* Correspondence: luo.liangyi@irl.sys.es.osaka-u.ac.jp; Tel.: +81-6-6850-6360

Abstract: Future service robots mass-produced for practical applications may benefit from having personalities. To engineer robot personalities in significant quantities for practical applications, we need first to identify the personality dimensions on which personality traits can be effectively optimised by minimising the distances between engineering targets and the corresponding robots under construction, since not all personality dimensions are applicable and equally prominent. Whether optimisation is possible on a personality dimension depends on how specific users consider the personalities of a type of robot, especially whether they can provide effective feedback to guide the optimisation of certain traits on a personality dimension. The dimensions may vary from user group to user group since not all people consider a type of trait to be relevant to a type of robot, which our results corroborate. Therefore, we had proposed a test procedure as an engineering tool to identify, with the help of a user group, personality dimensions for engineering robot personalities out of a type of robot knowing its typical usage. It applies to robots that can imitate human behaviour and small user groups with at least eight people. We confirmed its effectiveness in limited-scope tests.

Keywords: robot personality; robot personalities engineering; social robots; human–robot interaction; service robots



Citation: Luo, L.; Ogawa, K.; Ishiguro, H. Identifying Personality Dimensions for Engineering Robot Personalities in Significant Quantities with Small User Groups. *Robotics* **2022**, *11*, 28. <https://doi.org/10.3390/robotics11010028>

Academic Editors: Marco Ceccarelli and Huosheng Hu

Received: 7 January 2022

Accepted: 9 February 2022

Published: 14 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Engineering robot personalities for practical applications entails several idiosyncratic challenges. First, and foremost, it means engineering robot personalities in significant quantities, as opposed to the one or two robots to experiment on. By ‘significant’, we mean that the number of robot personalities should be significantly large for practical applications. How large is ‘significant’ depends on the application. For example, a music band of robots might require three to ten robot personalities, so that each of them appears to their audience as an individual character like how it is with human musicians. A robotic staff for a business establishment might require tens or hundreds, so that their human colleagues can recognise them as individuals as they do other humans. As for domestic service robots, it could put their human users in a more comfortable and constructive frame of mind if their robots are unique existence that cares specially for them, that is different from all the same models of robots working for their relatives or friends and different from the identical units that are mowing the lawns of strangers. Generally speaking, it depends on how many robots will be observed together as a population associated with a quality, goal, purpose, or workplace. Adding to this challenge, robots, if useful enough, will be mass-produced. While the hardware of robots can be mass-produced by duplicating the same design, personalities cannot be copied by definition [1]. Based on the current understanding of human personalities [2], we define a *robot personality* as a robot exhibiting characteristic patterns of computation and behaviour with inter- and intra-individual differences. The differences manifest as personality traits qualified on personality dimensions and quantified

as measurements of their strength on the corresponding dimensions. As with a human personality, a robot's personality individuates the robot as a unique presence different from the rest of their kind. Therefore, to engineer a robot personality is to engineer its individuality, which is significantly harder to do in large quantities. To engineer mass-produced, physically identical robots into significant quantities of desirable personalities is hence one of the the main challenges of robot personalities engineering for practical applications.

To engineer robot personalities in significant quantities, we consider such an optimisation process. We first design some unique *archetypes*, desirable target personalities on which some robot personalities will be based, and then proceed to minimise the distances between the robot personalities to construct and their corresponding archetypes in a coordinate system called a *personality space*, as defined by some personality dimensions. When the distances are small enough, the goal is achieved. (Optimisation methods are not the focus of this work). The question is *what are those personality dimensions?*

The answers should come from the users of the robots, since they are the main observers, judges, and beneficiaries of the robots' personalities. Most users are humans. Humans exhibit a tendency to attribute human qualities to non-human entities. For this reason, human personality dimensions, especially those of trait-based models, are frequently used as the bases for synthesising artificial personalities [3]. Trait-based personality models are often formulated as a set of personality dimensions on which an individual's traits can be measured, such as the five-factor model, first discovered by Tupes and Christal in 1961 (a reprint of their work is available [4]), and subsequently by Norman in 1963 [5]. The arguably most popular variant has these five dimensions [6]: extraversion, agreeableness, conscientiousness, neuroticism, and openness—known as the 'Big Five' [7]. However, there are several issues in applying human personality models to engineering robot personalities. The first issue is the most critical: the applicability of personality dimensions. Human personality models are often empirical formulations based on a lexicon formed overtime for describing humans, as explained by the lexical hypothesis [7,8]. Lexicons for describing humans do not necessarily apply to describing artificial agents [9]. It violates the lexical hypothesis to apply a human personality model to robots that are not created to resemble human presences (i.e., androids). The implication of this issue is that it is possible that not all personality dimensions of a model apply to a type of robot, especially when its appearance is far from a human image. Another issue is that, albeit of objectively equal importance, not all personality traits are equally prominent in people's eyes. A character is often recognised by their most memorable traits, as predicted by the 'availability bias' [10], and there can be 'central traits' that dominate the overall impressions of an individual [11]. How an individual is perceived is often affected by cognitive biases, which are something to consider or even leverage in robot personalities engineering. Last but not least, whether optimisation is possible on a personality dimension depends on how some specific users consider the personalities of a type of robot, especially whether they can provide effective feedback to guide the optimisation of the traits on a personality dimension. The dimensions may vary from user group to user group since it is possible that not all people consider a type of trait to be relevant to a type of robot that is not androids. To summarise, the first issue implies that we cannot be so sure if a human personality dimension applies to a robot that is not an android. The second issue implies that we need to focus on traits that matter most to users. The third issue implies that the specific users of certain robots have the ultimate say on what traits apply and matter to their robots. Such traits are which they can provide effective feedback on. Therefore, we need a test procedure to identify personality dimensions on which some users can provide effective feedback about the personality traits of their robots, as an engineering tool for engineering robot personalities out of a type of robot knowing its typical usage.

The main contribution of this work is such a test procedure. It applies to robots that can imitate human behaviour and small user groups with at least eight people. It is an engineering tool to identify the dimensions on which users can provide effective feedback to guide the engineering of robot personalities. At the beginning of the engineering work, the identified personality dimensions can serve as recommendations for some aspects of the robot personalities to focus on, and during the optimisation of robot personalities, they can constitute a coordinate system where the quality of the robot personalities under construction can be measured with the corresponding archetypes as reference points. As far as the recent surveys [3,12,13] can tell, we are the first to propose such a test procedure dedicated to engineering robot personalities in significant quantities for practical applications. To test the proposed test procedure, we conducted a series of tests simulating engineering tasks where 10 robot personalities were to be engineered out of 10 personality archetypes for small user groups of 3 to 18 people using the dimensions of the five-factor model [6]. The type of robot, a life-size humanoid ‘barebones’ robot, engaged users in dyadic communication, and its main modes of personality expression were head and eye movements. We confirmed the effectiveness of the proposed test procedure within the scope of the tests. The results show that the proposed method worked for user groups with at least eight people.

The rest of this article is organised as follows: Section 2 elaborates on the research goal, why we need the test procedure, and examines the insufficiency of previous work. Section 3 presents the test procedure. Section 4 relates the experiment conducted to test the test procedure. Section 5 shows the results of the experiment. Section 6 discusses the results and limitations, and Section 7 concludes this work.

2. Goals and the Limitations of the State of the Art in the Engineering Context

The test procedure should tell us, with the help of a user group, which of the personality dimensions of a personality model are suitable for engineering robot personalities out of a type of robot knowing its typical usage. It should meet three requirements:

- It identifies personality dimensions on which a group of users can provide effective feedback to guide the optimisation of the personality traits of the robot personalities under construction (as by minimising the distances between the robot personalities under construction and the corresponding archetypes);
- It supports engineering significant quantities of robot personalities;
- It works with small user groups.

Many previous studies have more or less done similar work in exploring robots’ potential for expressing personality traits or studying the effects or properties of robot personalities [14–23]. They offer valuable scientific insights into the roles of robot personality in human–robot interaction. However, their ‘tests’ were unsuitable for engineering tasks for they fell short of at least one of the above requirements.

The first and most common limitation is to consider the types of personalities rather than personalities with individual differences, which we call the ‘binary trait’ simplification. Vinciarelli and Mohammadi have referred to splitting personalities into two classes (per dimension) as ‘*binary classification approaches*’ and in their extensive survey commented that binary classes are ‘*not meaningful from a psychological point of view*’ [3]. Their survey has revealed that ‘binary classification approaches’ were prevalent in the field. We would also like to argue that the simplification is not meaningful from an engineering point of view either. In fact, it defeats the purpose entirely. To engineer robot personalities in significant quantities, the ‘binary trait’ simplification must not apply since we need to engineer far greater numbers than two robot personalities per dimension, with each of them manifesting their individual differences in characteristics as their traits. The proposed test procedure does not require the ‘binary trait’ simplification and hence is not subject to the said limitation.

If considered in the personalities engineering context, the second limitation of previous ‘tests’ is the requirement of a large sample of observers. Most if not all previous studies involved more than 20 observers, which is appropriate for studying effects or properties of robot personalities, where large representative samples are desirable. However, in engineering robot personalities for practical applications, most user groups will be small. Potential household user groups will mostly consist of two to nine people, extrapolating from the UN’s data in 2017 on household sizes [24]. As for small businesses, as of the time of this work, in the United States, currently the largest economy, the average number of employees is about 10, and for small businesses that have employees, the numbers of employees range from 1 to 19 [25]. As of Japan in 2019, small enterprises with fewer than 20 employees accounted for 85 percent of all enterprises, and those in service, retail, and wholesale industries had up to five employees [26]. Whether we consider potential household users or enterprise users, small user groups with fewer than 20 people will be the most common. The test results from one (large) sample of subjects do not necessarily apply to other (smaller) user groups due to how people perceive personalities depending on their own personalities and a number of other factors [27–31,31]. The results can be highly user-dependent and contextual, meaning different user groups may perceive the same robot personalities differently, and the same user group may perceive the same robot personalities differently under different circumstances. The same behaviour may indicate different traits in different minds in different contexts. Something as straightforward as eye contact can be pointing to different traits in different cultures. In many eastern cultures, staring into the other’s eyes is being confrontational or arrogant; in many western cultures, not doing so is being disrespectful and showing disinterest or guilt. It is not only cultural differences. Interpersonal differences should also be taken into account. Between lovers, staring into the other’s eyes can be a cue of strong affection, in eastern or western cultures alike; however, between rivals, it is expressing animosity, strength, resolution, among other possibilities. It is almost certain that, due to the complexity of personality and how it is perceived, a group of users is unlikely to perceive the robot personalities under construction as similar to the archetypes without optimisation, even when the robots exhibit the most ‘archetypal’ behaviour. This ‘self-other’ discrepancy has been observed in perceiving human personalities for a long time [32], and in the perception of robot personalities as well [33]. The discrepancy should be minimised for the sake of consistent user experience. What dimensions and how much distance to minimise depend on the users. It follows that we should optimise robot personalities on a case-by-case basis.

The third limitation and most fatal one is missing the step to check the applicability of human personality dimensions to robots that are not androids. We can assume human personality dimensions are applicable to androids since they are created to resemble humans and hence do. For other robots, we cannot be so sure if the items of a dimension describe a robot (e.g., whether an ‘open-minded’ robot vacuum cleaner makes sense to certain users). Consequently, this limitation would apply to nearly all previous testing methods if they were to be applied to personalities engineering unless the robots to engineer were androids, since the results acquired would be in violation of the lexical hypothesis [7,8]. The proposed method is itself a guard against descriptions that do not make sense regarding the type of robot to engineer.

3. Methods: Proposed Test Procedure

The test procedure tests hypotheses in the following format:

Hypothesis (Format). *Given a type of robot R exhibiting typical behaviour B in situation S and a personality dimension P of a personality model M , P is a personality dimension on which fidelity can be effectively optimised.*

Here, *fidelity* refers to the proximity between the robot personalities under construction and their corresponding archetypes on the personality dimensions of the model: the higher the fidelity/proximity, the smaller the distances. The exact definition of fidelity should depend on the personality model to investigate. Generally speaking, the fidelity of a trait of a robot personality under construction is the accretion of distance measurements between the archetype after which the robot personality is engineered and a number of observations of the trait from a group of observers on the corresponding personality dimension. Since a robot personality has to be observed (unlike humans, robots cannot report their own personalities as their ‘true’ personalities; what serve as their supposed ‘true’ personalities are their corresponding archetypes, which do not necessarily match their observed personalities without optimisation), the observers’ own personalities and backgrounds affect their observations, a measurement of fidelity is always associated with a particular group of observers, and there can be no absolutely objective fidelity measurement.

Testing a hypothesis in the format requires testing three corresponding sub-hypotheses:

Sub-Hypothesis 1. *The fidelity, computed as the proximity between the human observations and the corresponding archetypes, is statistically distinguishable from that by random guesses.*

Sub-Hypothesis 2. *There is a significant difference between the consistency of the observations on the robot personality and that of those on a human personality in the same settings.*

Here, *consistency* refers to the negative dispersion among the impressions of a significant number of observers on a robot personality: the more dispersed the impressions are, the lower the consistency. Usually, consistency can be measured as negative variance.

Sub-Hypothesis 3. *The pseudo-fidelity, computed as the proximity between the human observations and the corresponding observers’ own personalities, is statistically distinguishable from that by random guesses.*

To identify a dimension as engineering-worthy, we should reject only the null hypothesis of Sub-Hypothesis 1 (Case 5 in Table 1). The rationale behind is explained by the following three working hypotheses:

Table 1. Decision table.

Cases	Sub-Hypothesis 1	Sub-Hypothesis 2	Sub-Hypothesis 3	Identification
Case 1	No	No	No	No
Case 2	No	No	Yes	No
Case 3	No	Yes	No	No
Case 4	No	Yes	Yes	No
Case 5	Yes	No	No	Yes
Case 6	Yes	No	Yes	No
Case 7	Yes	Yes	No	No
Case 8	Yes	Yes	Yes	No

Yes: Yes, the null hypothesis is rejected. No: No, the null hypothesis is not rejected.

Working Hypothesis 1. *When some human observers are using a personality model designed based on a lexicon for describing humans or animals to assess the personality of a robot personality engineered after an archetype, the corresponding fidelity is not necessarily statistically distinguishable from that by random guesses, which implies that they have completed the assessment by guesswork.*

If the observers are just guessing the traits, their observations cannot be used to guide the optimisation of fidelity. However:

Working Hypothesis 2. *If a robot is capable of imitating human behaviour in a given context where such behaviour is expected and the behaviour is typical of the robot in a usage that matches the context, some traits that apply to humans will also apply to the robot and are prominently observable in their typical behaviour, thereby resulting in fidelity that is statistically distinguishable from that by random guesses.*

Even if the fidelity is non-random, we need to eliminate two other possible causes of non-randomness to make sure that it can guide personalities engineering: inconsistency and observers' own personalities. Reports scattered on a personality dimension can still lead to significant differences from random guesses if they are dispersed enough. Highly dispersed reports reflect great inconsistency of opinions on the robot personalities. If the observers report the robot personalities as similar to their own, which can occur [34], the corresponding fidelity will also be non-random while being irrelevant to the archetypes.

Working Hypothesis 3. *If the fidelity from human observations that are as consistent as on a human personality in the same settings is statistically distinguishable from that by random guesses and the cause of it is not that the observers have reported the robot personalities to be similar to their own, the human observations can be used to guide the optimisation of the fidelity.*

The third working hypothesis is supported by the following reasoning: when the robot's behaviour is mapped to personality measurements as completed by some human observers and the mapping is not random but consistent, there exists an instance of behaviour leading to measurements that are closest to the corresponding archetype. By approximating that instance of behaviour, we can approximate the optimal fidelity. Assuming the robot's behaviour is controlled by some parameters of a *generative personality model*, a personality model capable of generating behaviour with individual differences to reflect individual differences in personality, there should exist a set of parameters leading to the optimal behaviour. In that regard, common optimisation methods should apply to finding the parameters, such as gradient descent and genetic algorithms. However, whether they are efficient is another story.

The proposed test procedure is based on the 'robots-imitating-humans' approach [35–37] and existing statistical tools. The resources required are:

1. Fungible units of the type of robot to test;
2. Human archetypes who can serve as desirable examples for the type of robot in performing the tasks it is designed for;
3. Tools to capture the example behaviour as data;
4. Methods to enable the robot to imitate the example behaviour;
5. Actors that act as users;
6. The user group the robots are going to work for.

The procedure has four phases, as illustrated in Figure 1 (where the arrows mark dependencies; the capsule is the starting point; the rectangles are processes; the cylinders are data sets or materials; and the hexagons are the results of processes). It is worth noting that it is unnecessary to perform the entire procedure from the beginning when more orders for the same model of robots for the same usage come from some other user groups; in this case, we can start with user assessment. In the following subsections, we will go through the four phases one by one in detail.

3.1. Phase 1: Recording Archetypal Behaviour

In the first phase, we first recruit some candidate archetypes, and then, in 'behaviour recording sessions', we acquire their personality measurements, record their behaviour, and let them report the personalities of the actors acting as users. How behaviour recording sessions should be carried out depends. Generally speaking, it is a simulation of the typical usage of the type of robot to develop, where the candidate archetypes will be examples for the robot personalities. For example, if the type of robot is going to be office errand

runners, we hire model (well-received) office workers to simulate an office environment; if it is going to be waiters in a restaurant, we hire model waiters and waitresses and simulate an restaurant; if it is going to be singers on the stage, we simulate a stage with real singers.

3.2. Phase 2: Implementing the Behaviour

In this phase, we produce the stimuli for user assessment in Phase 3. The stimuli can be video recordings of the simulation in Phase 1 or they can be the robot personalities themselves. For the latter, more than one unit of the type of robot may be required. How to produce the stimuli depends. Generally speaking, the robots' behaviour should be as close to the archetypes' as possible. First, in 'screening', we need to exclude the candidates whose behaviour is beyond the robot's capabilities or operational parameters. We use only the behavioural data from the selected candidates, who will be the human archetypes. Then, we process the data. How to do this depends. In general, it is turning videos or motion capture data into a form that the robot can imitate. For example, in developing waiter robots, this can be processing a waiter's gestures and body motion when ushering guests into their seats into joint tracking data. Next, in 'extracting behaviour', we further separate the processed data according to the mode (modality) of behaviour we are interested to investigate. For example, for developing waiter robots, we might be more interested in facial expressions and hand gestures than gait as modes of personality expression. Finally, we program the robot with the behaviour to recreate the simulation in Phase 1 as how we produce stimuli for user assessment in Phase 3.

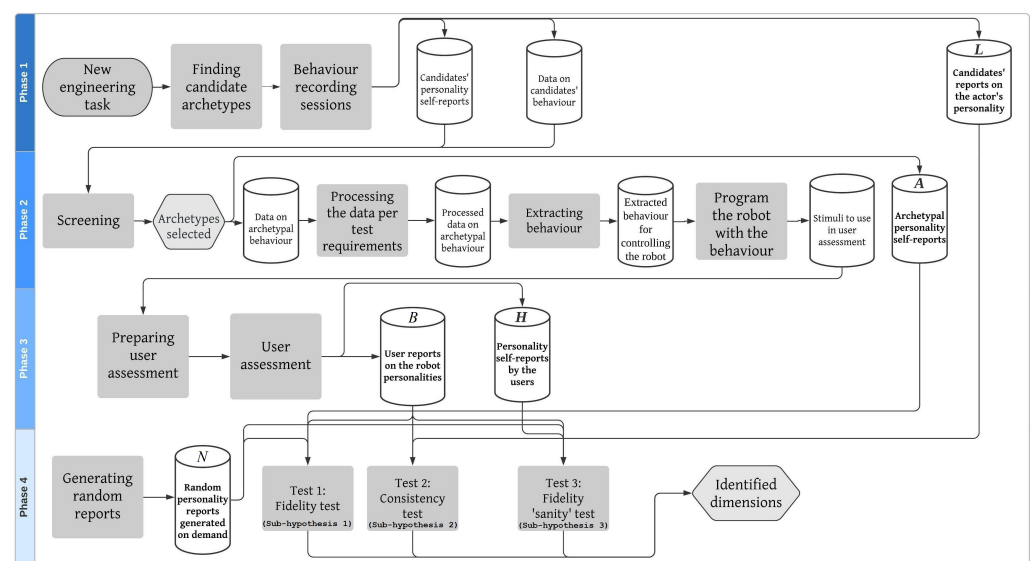


Figure 1. Workflow of the procedure (where the four rows represent the four phases; the arrows mark dependencies; the capsule is the starting point; the rectangles are processes; the cylinders are data sets or materials; and the hexagons are the results of processes). Phase 1 is about recording archetypal behaviour; Phase 2, implementing the behaviour; Phase 3, user assessment of robot personalities; and Phase 4, the three tests.

3.3. Phase 3: User Assessment

In Phase 3, we first make preparations for user assessment and then request the users to assess the robot personalities. The users either assess the robot personalities based on video stimuli or by interacting with the robot personalities themselves. For assessment based on video stimuli, we need to prepare only surveys. For assessment based on live interaction, we need to prepare for interactive settings as close to the simulations in Phase 1 as possible. After the surveys or interaction sessions, users also need to report their own personalities.

3.4. Phase 4: Three Tests

In Phase 4, we identify personality dimensions using three tests. We henceforth refer to users also as observers, since the users of the robots are the main observers of the corresponding robot personalities.

3.4.1. Data Sets Required

The tests require five data sets in total (Table 2), four of which are from the previous phases: the archetypal personality self-reports, hereinafter denoted as A , from Phase 2; the candidates' (or archetypes') reports on the actor's personality (if the simulation in Phase 1 has involved multiple actors and reports have been acquired on all the actors, the reports on one of them are enough), hereinafter denoted as L , from Phase 1; the users' reports on the robot personalities, hereinafter denoted as B , from Phase 3; and the personality self-reports by the users, hereinafter denoted as H , from Phase 3. The fifth data set consists of random reports generated on demand, hereinafter denoted as N . Here, B and N are not in bold because they are sets. In addition, r denotes the number of robot personalities to engineer, which is the same as the number of archetypes; u the number of users; t the number of personality dimensions to test; and c the number of reports in the data set L .

Table 2. Data sets required.

Data Sets	Dimensions	Row Vectors	Column Vectors
A	$r \times t$	$a_1, a_2, a_3, \dots, a_r$	$a'_1, a'_2, a'_3, \dots, a'_t$
L	$c \times t$	$l_1, l_2, l_3, \dots, l_c$	$l'_1, l'_2, l'_3, \dots, l'_t$
$B = \{B_1, B_2, B_3, \dots, B_u\}$	$u \times r \times t$	-	-
H	$u \times t$	$h_1, h_2, h_3, \dots, h_u$	$h'_1, h'_2, h'_3, \dots, h'_t$
N	depends	depends	depends

A is a $r \times t$ matrix. It consists of r t -dimensional row vectors corresponding to r archetypes or t r -dimensional column vectors corresponding to t personality dimensions. L is a $c \times t$ matrix. It consists of c t -dimensional row vectors corresponding to c sets of human observations on a human or t c -dimensional column vectors corresponding to t personality dimensions. B is a set that consists of u $r \times t$ matrices: $B_1, B_2, B_3, \dots, B_u$, since each of the observers has reported t traits on r robot personalities; each matrix has the same dimensions as A and can be expressed likewise. H is a $u \times t$ matrix. It consists of u t -dimensional row vectors corresponding to u observers (users) or t u -dimensional column vectors corresponding to t personality dimensions. N consists of randomly generated data per the dimensions required. To summarise, we have

$$A = [a_1 \ a_2 \ a_3 \ \dots \ a_r]^T = [a'_1 \ a'_2 \ a'_3 \ \dots \ a'_t], \quad (1)$$

where $A \in \mathbb{R}^{r \times t}$, $a_{i=1,2,3,\dots,r} \in \mathbb{R}^{1 \times t}$, and $a'_{j=1,2,3,\dots,t} \in \mathbb{R}^{r \times 1}$;

$$L = [l_1 \ l_2 \ l_3 \ \dots \ l_c]^T = [l'_1 \ l'_2 \ l'_3 \ \dots \ l'_t], \quad (2)$$

where $L \in \mathbb{R}^{c \times t}$, $l_{i=1,2,3,\dots,c} \in \mathbb{R}^{1 \times t}$, and $l'_{j=1,2,3,\dots,t} \in \mathbb{R}^{c \times 1}$;

$$B = \{B_1, B_2, B_3, \dots, B_u\}, \quad (3)$$

where $B_{i=1,2,3,\dots,u} \in \mathbb{R}^{r \times t}$;

$$H = [h_1 \ h_2 \ h_3 \ \dots \ h_u]^T = [h'_1 \ h'_2 \ h'_3 \ \dots \ h'_t], \quad (4)$$

where $H \in \mathbb{R}^{u \times t}$ and $h_{i=1,2,3,\dots,u} \in \mathbb{R}^{1 \times t}$, and $h'_{j=1,2,3,\dots,t} \in \mathbb{R}^{u \times 1}$.

N represents a ‘blind guesser’ who cannot perceive any personality traits and thus has no recourse but to guesswork when it is required to complete a personality assessment. It is what the observers are pitted against. Personality traits are often assessed with statements describing certain qualities of a subject, such as: ‘... is someone who likes to talk with friends’. The users must indicate the extent to which they agree or disagree with the statements. They are *guessing* if they have no idea how well the statements describe the robot personalities, or they can simply indicate that they neither agree nor disagree with the statements. A significantly large number of wild guesses should exhibit the same behaviour as random guesses generated by the uniform distribution.

3.4.2. Test 1: Robot Personalities’ Fidelity Test

The first test to run is the fidelity test. It tests Sub-Hypothesis 1 on each dimension. It requires data sets A , B , and N . Let s denote the total number of reports,

$$s = u \cdot r. \quad (5)$$

The fidelity test compares the fidelity from the human observations with that from N to identify the personality dimensions.

We can compute fidelity as follows: given a trait as observed by u users, which can be represented by a vector \mathbf{b} ($\mathbf{b} \in \mathbb{R}^u$), and the corresponding trait of the archetype a ($a \in \mathbb{R}$), fidelity vector \mathbf{f} ($\mathbf{f} \in \mathbb{R}^u$) is expressed as

$$\mathbf{f} = \text{abs}(\mathbf{b} - a), \quad (6)$$

where abs denotes a function that replaces all elements in a matrix or vector with their absolute values. (To avoid confusion, we refrain from using $||$ since it also denotes the determinant of a matrix).

Each element in the fidelity vector \mathbf{f} is a numerical distance, which can also be called ‘a proximity value’. For instances, if Observer 5 has reported the extraversion level of a robot to be 3.5 when the robot personality is based on Archetype 3, whose extraversion level is 4, then the 5th element in the \mathbf{f}_3 of extraversion is 0.5. The fidelity of all robot personalities on all dimensions can be represented as a $s \times t$ matrix: \mathbf{F} ($\mathbf{F} \in \mathbb{R}^{s \times t}$). \mathbf{F} can be computed from A and B using

$$\mathbf{F} = [\text{abs}(\mathbf{B}_1 - A) \quad \text{abs}(\mathbf{B}_2 - A) \quad \text{abs}(\mathbf{B}_3 - A) \quad \dots \quad \text{abs}(\mathbf{B}_u - A)]^T. \quad (7)$$

We generate the same number of random reports to the assessment scales. A random report consists of random numbers generated as random answers to the questionnaires about the robot personalities. For example, given a questionnaire consists of 44 questions with 5-point Likert scales enquiring how much an observer agrees with the corresponding statement, a human report would consist of 44 responses, whereas a random report would consist of 44 random integers on the range of [1, 5]. Random traits should be computed in the same way per the instructions of the personality assessment inventory, and the results are divided into $u \times r \times t$ matrices, which are denoted here as $N_1, N_2, N_3, \dots, N_u$ ($N_{i=1,2,3,\dots,u} \in \mathbb{R}^{r \times t}$). We then compute the random ‘fidelity’ \mathbf{F}' ($\mathbf{F}' \in \mathbb{R}^{s \times t}$) using

$$\mathbf{F}' = [\text{abs}(N_1 - A) \quad \text{abs}(N_2 - A) \quad \text{abs}(N_3 - A) \quad \dots \quad \text{abs}(N_u - A)]^T. \quad (8)$$

With \mathbf{F} and \mathbf{F}' ready, we apply an appropriate statistical test. Which test should be applied depends on the sample sizes, types of distributions in the samples, any underlying assumptions about the data, and whatnot. A test should be carefully chosen to yield practical results.

Let T denote a function that performs the featured test thus: it takes two real matrices of the same dimensions as the input and then returns a real vector of p -values as the output. The function computes the p -values between the two corresponding columns of the same index in the two matrices. Therefore, the p -values, represented as a vector \mathbf{p} ($\mathbf{p} \in \mathbb{R}^t$), can be computed as

$$\mathbf{p} = T(\mathbf{F}, \mathbf{F}'). \quad (9)$$

Other statistics can be computed likewise.

Here, we are testing t hypotheses simultaneously. Thus, the problem arises as whether corrections for the multiple comparisons problem should be applied. In scientific research, corrections are often applied. However, in engineering robot personalities, it should depend on the circumstances of the specific engineering task. Whether it is better to minimise the chance for either Type 1 or Type 2 errors depends. If it is more important to reduce cost and focus on the most prominent personality traits, it might be better to apply the corrections so that it is less likely to identify a personality dimension by chance while in truth optimisation cannot be effectively conducted on that dimension. If it is more important to utilise the full potential of the robots, so as to make them more ‘characteristic-rich’, it might be better to not apply the corrections so that it is less likely to disqualify a dimension by chance.

3.4.3. Test 2: Robot Personalities’ Consistency Test

The consistency test follows the fidelity test. It tests Sub-Hypothesis 2 for every robot personality on each dimension. This test requires data sets B and L .

Consistency measures the strength of the consensus on a robot personality some observers can reach. Given that a benchmark is yet to be established in the field, for the time being, the consistency of humans reporting on a human can be the standard for that of humans reporting on a robot in the same settings. We can use Bartlett’s test as a consistency *parity* test. It tests *homogeneity of variance* while being sensitive to non-normality, meaning passing this one test is a sign for both normality and homogeneity of variance of the reported personalities.

Given a number of observations \mathbf{x}_A ($\mathbf{x}_A \in \mathbb{R}^u$), as on a personality trait of an agent, and those on another \mathbf{x}_B ($\mathbf{x}_B \in \mathbb{R}^u$), consistency parity p ($0 < p < 1$) can be computed as

$$p = \text{Bartlett}(\mathbf{x}_A, \mathbf{x}_B). \quad (10)$$

The p here is indeed a p -value. However, what can be considered as ‘significant’ in the context of consistency parity requires support from more empirical results. For now, any choices seem arbitrary. The convention of $p = 0.05$ can be a (very loose) significance threshold, and since we run r tests per dimension, we need to correct the p -value to $0.05/r$ by applying the Bonferroni correction to counter the multiple comparisons problem because we do not want to disqualify a dimension by chance, thereby relaxing it further. A consensus is topic-dependent, which means that a group of observers can reach a consensus separately on A or B, while a consensus on A and B together is nonsense. Consequently, we need to measure the consistency parities separately (per archetype per trait).

Let Bt denote a function that performs Bartlett’s test, thus: it takes two real matrices of arbitrary numbers of rows but the same number of columns as the input, and it returns a real vector \mathbf{p} of p -values as the output. The function computes the p -values between two corresponding columns of the same index in the two matrices. Let \mathbf{P} denote the matrix that contains all p -values of consistency parities. It can be computed as

$$\mathbf{P} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \mathbf{p}_3 \quad \dots \quad \mathbf{p}_r]^\top, \quad (11)$$

where $\mathbf{p}_{i=1,2,3,\dots,r} = Bt(\mathbf{L}, \mathbf{B}_i)$ and $\mathbf{p}_i \in \mathbb{R}^{1 \times t}$,

Where $\mathbf{B}_{i=1,2,3,\dots,r} \in \mathbb{R}^{u \times t}$, and

$$\mathbf{B}_i = [\mathbf{b}_{1i} \quad \mathbf{b}_{2i} \quad \mathbf{b}_{3i} \quad \dots \quad \mathbf{b}_{ui}]^\top, \quad (12)$$

where \mathbf{b}_{m_i} ($m = 1, 2, 3, \dots, u$) denotes the m th observer's report on the i th robot personality. To disqualify a dimension, we can consider the number of null hypotheses (consistency parity) rejected on that dimension. Considering that the consistency parity testing approach may not be strict, a stricter threshold for the number of rejected null hypothesis can be implemented here to enhance the effectiveness of the test.

3.4.4. Test 3: Robot Personalities' Fidelity 'Sanity' Test

Finally, the fidelity 'sanity' test checks whether we can reproduce the results in the fidelity test using the observers in place of the archetypes as the reference points. It tests Sub-Hypothesis 3 on each dimension.

This test is almost the same as Test 1 except that we swap the archetypes with the observers themselves as the reference points. Fidelity is defined as the proximity between the observed robot personalities and their corresponding archetypes; therefore, the proximity between the observed robot personalities and the corresponding observers is not fidelity; instead, we call this quantity *pseudo-fidelity*. The test follows the exact same procedure as the fidelity test, only that this time the reference points are changed from the archetypes to the personalities of the observers themselves. This test requires data sets \mathbf{H} , \mathbf{B} , and \mathbf{N} .

We compute the pseudo-fidelity and generate s random reports to compute random traits as before. From \mathbf{N} , we get $u \times r \times t$ matrices: $\mathbf{N}'_1, \mathbf{N}'_2, \mathbf{N}'_3, \dots, \mathbf{N}'_u$ ($\mathbf{N}'_{i=1,2,3,\dots,u} \in \mathbb{R}^{r \times t}$). Because we need the proximity values between the observations of each observer and the observer themselves, we construct $u \times r \times t$ matrices: $\mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \dots, \mathbf{H}_u$ ($\mathbf{H}_{i=1,2,3,\dots,u} \in \mathbb{R}^{r \times t}$), where

$$\mathbf{H}_{i=1,2,3,\dots,u} = [\mathbf{h}_i \quad \mathbf{h}_i \quad \mathbf{h}_i \quad \dots \quad \mathbf{h}_i]^T, \quad (13)$$

where \mathbf{h}_i is the i th observer's personality.

Then, we apply the same procedure as in the fidelity test, which can be expressed as

$$\mathbf{F}_o = [\text{abs}(\mathbf{B}_1 - \mathbf{H}_1) \quad \text{abs}(\mathbf{B}_2 - \mathbf{H}_2) \quad \text{abs}(\mathbf{B}_3 - \mathbf{H}_3) \quad \dots \quad \text{abs}(\mathbf{B}_u - \mathbf{H}_u)]^T \quad (14)$$

$$\mathbf{F}'' = [\text{abs}(\mathbf{N}'_1 - \mathbf{H}_1) \quad \text{abs}(\mathbf{N}'_2 - \mathbf{H}_2) \quad \text{abs}(\mathbf{N}'_3 - \mathbf{H}_3) \quad \dots \quad \text{abs}(\mathbf{N}'_u - \mathbf{H}_u)]^T. \quad (15)$$

Then, the p -values of this test, represented as a vector \mathbf{p}_o ($\mathbf{p}_o \in \mathbb{R}^t$), can be computed as

$$\mathbf{p}_o = T(\mathbf{F}_o, \mathbf{F}''). \quad (16)$$

4. Experiment: Testing the Test Procedure

We tested the test procedure in a scope represented by the following question: *given a user group, what are the personality dimensions to use in engineering a type of life-size humanoid 'barebones' robot into ten robot personalities for human–robot dyadic communication, where the typical personality expressing behaviour consists of head and eye movements?* Dyadic communication, albeit not a situation that needs a large number of robot personalities, was chosen for the abundance of literature on human–robot interaction since we needed to compare our results with those in existing literature to know whether the proposed test procedure was effective.

4.1. Materials: Robot and Software

The robot used to test the test procedure was a humanoid 'barebones' robot (Figure 2). Its eyes and necks were actuated by pneumatic actuators with a typical response time of 200 milliseconds. The robot was controlled by a proprietary interface that received actuator commands for movements. A simple mapping program generated commands for the robot's eyes and head by mapping head and eye movement angles, which were captured by OpenFace 2 [38–41], to actuator positions, so that, when the archetypes turned their heads or eyes left or right, up or down, the robot would do the same. The mapping was done without smoothing or filtering.

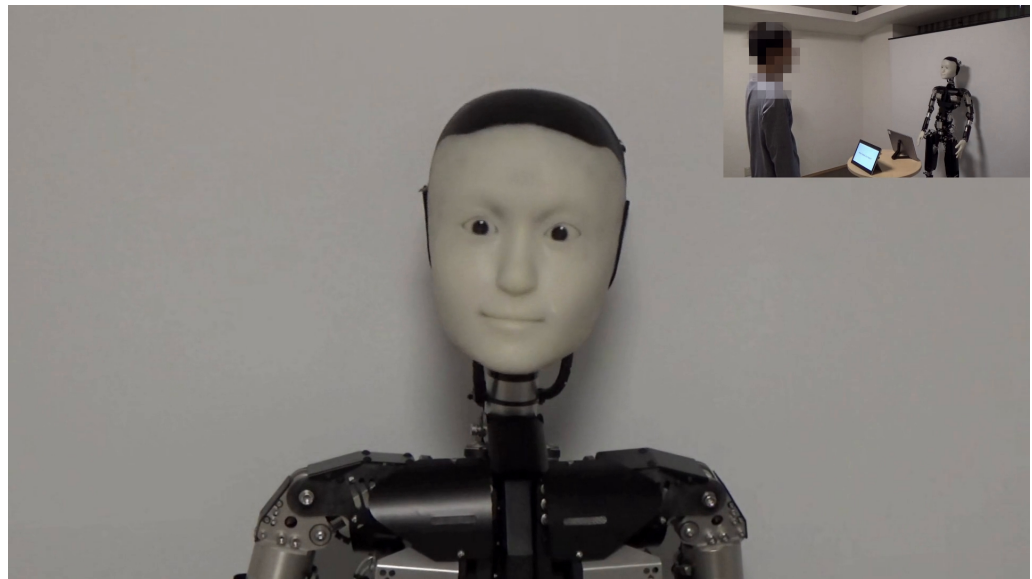


Figure 2. Tested robot in action as what the human observers viewed in Phase 3 (but without pixelating the actor's face) when they reported the robot personalities.

The test procedure was implemented in GNU Octave (Version 6.2.0) with the 'statistics' package (Version 1.4.2). The featured statistical test in Tests 1 and 3 was Welch's test, which was executed by the `welch_test` function. The featured statistical test in Tests 2 was Bartlett's test, which was executed by the `bartlett_test` function.

4.2. Going through the Procedure

The test settings were that 10 robot personalities were to be engineered out of 10 personality archetypes for small user groups of 3 to 18 people. The robot engaged users in dyadic communication, and its main modes of personality expression were head and eye movements. The personality assessment inventory used was by John et al. [42], which has 44 scales for assessing five principal personality traits (dimensions): extraversion, conscientiousness, agreeableness, neuroticism, and openness. If we fill in the hypothesis format per the settings, it would be

Hypothesis (Generic Form). *Given a humanoid 'barebones' robot that moves its head and eyes in dyadic communication and a personality dimension P of the five-factor model (where $P \in \{\text{extraversion, conscientiousness, agreeableness, neuroticism, openness}\}$), P is a personality dimension on which fidelity can be effectively optimised.*

We tested 505 such hypotheses in 101 tests corresponding to 101 user groups simulated by drawing from a pool of 18 observers. The first user group consisted of all 18 observers for a showcase run of the procedure (Sections 5.1–5.3). Then, we drew 20 combinations of 15 people, 20 combinations of 12 people, 20 combinations of 8 people, 20 combinations of 5 people, and 20 combinations of 3 people, amounting to 100 simulated user groups in total. Therefore, the total number was 101.

The corresponding workflow is illustrated by Figure 3. The situation we simulated in Phase 1 was two people exchanging information about themselves. We selected 10 human archetypes out of 15 candidates, who were students from our university. Each of the 15 candidates had come to our laboratory and engaged in a brief exchange with the actor. They first introduced themselves and then answered 12 questions adapted from those on a website for practising English conversation [43]. Then, the candidate and the actor switched roles: it was now their turn to ask the same questions, and the actor answered by reciting a script, which is recorded in Appendix B. The head and eye movements of the candidates responding to the actor's answers were recorded as the archetypal behaviour for the robot

personalities. The candidates also reported their own personalities thrice: once upon the application, once before the interview, and once after. Appendix A relates the interview in more extensive details. In Phase 2, we selected the 10 archetypes, whose head and eye movements were within the robot's capability, and shortened their corresponding videos (10- to 15-min long) into videos of about 2-min's length according to an abridged version of the original script (the part in bold of the script recorded in Appendix B makes the abridged version). The shortened videos were remade with the robot personalities replacing the archetypes. In the remakes, the robot personalities imitated the archetypes' head and eye movements under the control of our movement mapping software. The same actor answered their questions, which were shown on a screen. Although there were 10 remakes featuring 10 robot personalities, they were all installed on the same robot. Now, for Phase 3, we embedded the remakes into an online assessment form, which had 11 pages for personality assessment: 10 randomised pages for the robot personalities, and the 11th for the observer's own personality. We published the assessment form on Amazon's Mechanical Turk and employed 18 workers to complete it. Finally, in Phase 4, we applied the three tests to the 101 simulated user groups and worked out the results.

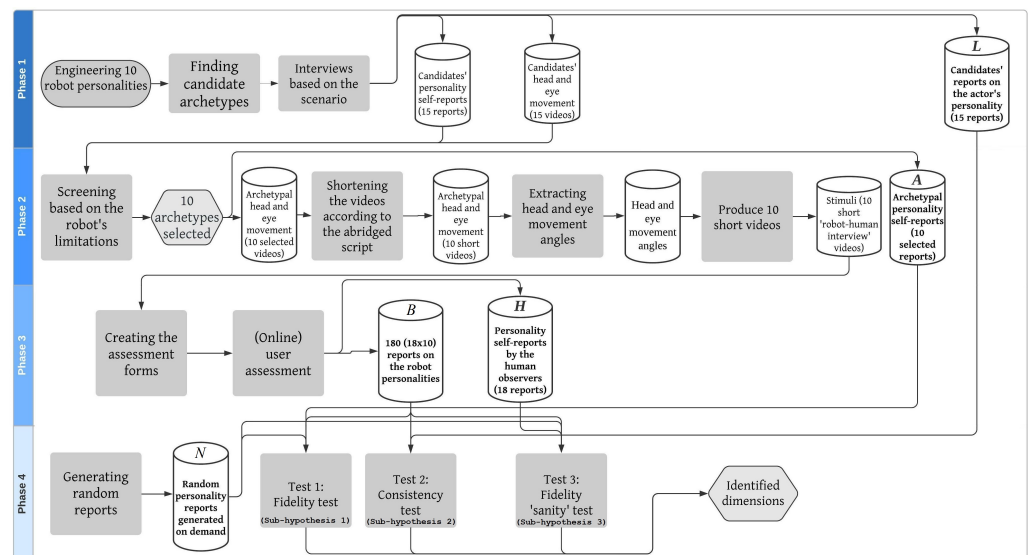


Figure 3. Workflow adapted to our test.

5. Results of Testing the Test Procedure

First, the results from all 18 observers as a single user group:

5.1. Robot Personalities' Fidelity Test

For the featured statistical test in our experiment, we used Welch's test considering that our fidelity tests were going to feature large samples (30–180) with unknown variances. Plugging in our data, where $r = 10$, $t = 5$, and $u = 18$, to execute the tests using our implementation in GNU Octave, we obtained

$$p = [0.002 \quad 0.580 \quad < 0.001 \quad 0.607 \quad 0.721] \quad (17)$$

corresponding to extraversion, agreeableness, conscientiousness, neuroticism, and openness. The mean fidelity values are plotted in Figure 4 with the corresponding standard deviations illustrated by the error bars. As for the significance threshold (α), we assumed the conventional 0.05. We also assumed that we intended to discovery the full potential of the robot, which required us to minimise the chance for Type 2 errors. Therefore, we did not apply corrections for the multiple comparisons problem.

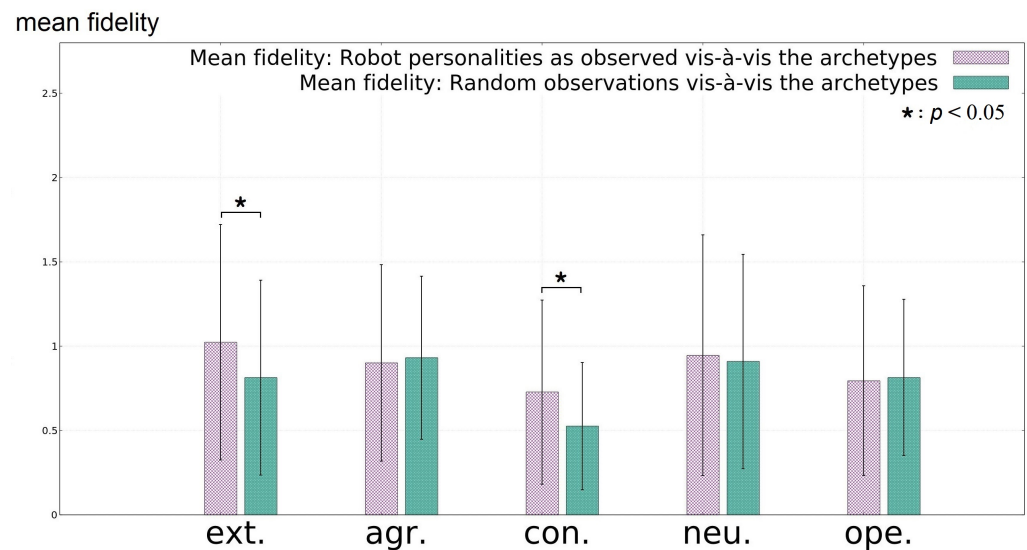


Figure 4. Results of the fidelity test (Section 5.1).

We can observe that (Figure 4), for extraversion, the fidelity from the human observers ($M = 1.022, SD = 0.698$; M henceforth refers to ‘mean’, and SD ‘standard deviation’) compared with that from randomness ($M = 0.813, SD = 0.577$) demonstrated a significant difference, $t(345.76) = 3.09, p = 0.002$. (The difference here may seem small despite being statistically significant because fidelity is not a quantity of a large magnitude to start with.) As for conscientiousness, the fidelity from the human observers ($M = 0.727, SD = 0.546$) compared with that from randomness ($M = 0.525, SD = 0.378$) demonstrated a significant difference, $t(318.41) = 4.08, p < 0.001$. There was no significant difference on agreeableness, $t(346.33) = -0.55, p = 0.580$, between the fidelity by the human reports ($M = 0.900, SD = 0.582$) and that by randomness ($M = 0.931, SD = 0.483$). The same can be observed for neuroticism, $t(353.41) = 0.51, p = 0.607$, between ($M = 0.946, SD = 0.713$) and ($M = 0.909, SD = 0.636$); and for openness, $t(345.39) = -0.36, p = 0.721$, between ($M = 0.795, SD = 0.563$) and ($M = 0.814, SD = 0.464$).

The above results imply that extraversion and conscientiousness might be the personality dimensions on which the fidelity of the ten robot personalities could be optimised. On other dimensions, the observers’ reports were equivalent to random guesses.

5.2. Robot Personalities’ Consistency Test

Plugging in our data to run all consistency parity tests (50 tests in total), we obtained the P matrix, which was transcribed into significance marks as recorded in Table 3, where an asterisk marks $p < 0.005$, which was our significance threshold acquired by applying the Bonferroni correction ($p = 0.05/10$) since we did not want to disqualify a dimension by chance. The consistency parity has held in most cases (in Bartlett’s test, homogeneity of variance is the null hypothesis).

We can see that, as for extraversion and conscientiousness, all test results in the corresponding columns failed to reject the null hypothesis of consistency parity. In the neuroticism column, the null hypothesis was rejected once. This means that we could fail the identification of neuroticism as an engineering-worthy dimension even if neuroticism had been identified in Test 1. How many rejections can fail the identification is at the moment a relatively arbitrary decision, and since our consistency parity test was not strict at all, we determined that one rejection was enough to fail the identification of the corresponding dimension.

Table 3. Results of the robot personality consistency test (Section 5.2).

Robots	Five Principal Personality Dimensions				
	ext.	agr.	con.	neu.	ope.
Robot (as Archetype 1)	x	x	x	*	x
Robot (as Archetype 2)	x	x	x	x	x
Robot (as Archetype 3)	x	x	x	x	x
Robot (as Archetype 4)	x	x	x	x	x
Robot (as Archetype 5)	x	x	x	x	x
Robot (as Archetype 6)	x	x	x	x	x
Robot (as Archetype 7)	x	x	x	x	x
Robot (as Archetype 8)	x	x	x	x	x
Robot (as Archetype 9)	x	x	x	x	x
Robot (as Archetype 10)	x	x	x	x	x

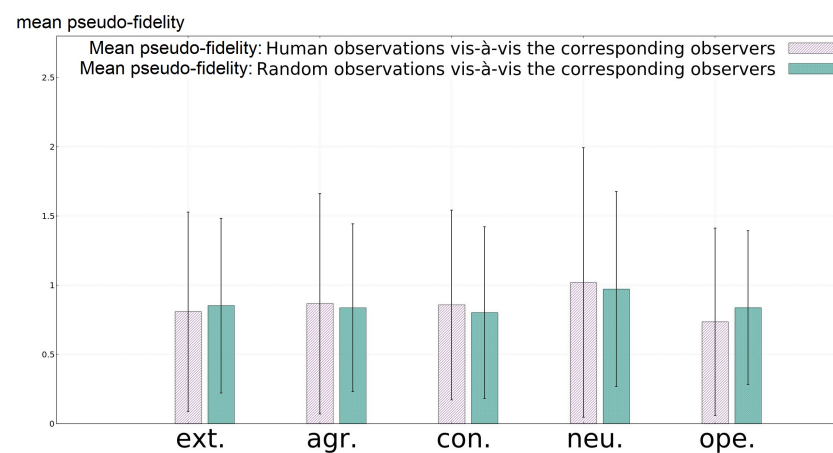
x: The null hypothesis is not rejected. *: The null hypothesis is rejected.

5.3. Robot Personalities' Fidelity 'Sanity' Test

Plugging in our data, we obtained

$$p_o = [0.534 \quad 0.692 \quad 0.420 \quad 0.598 \quad 0.120]. \quad (18)$$

The means and standard deviations are plotted in Figure 5.

**Figure 5.** Results of the fidelity 'sanity' test (Section 5.3).

We can observe that (Figure 5), for extraversion, there was no significant difference, $t(351.89) = -0.62$, $p = 0.534$, between the pseudo-fidelity between the personality reports and their corresponding reporters ($M = 0.808$, $SD = 0.720$) compared with that of the random reports ($M = 0.852$, $SD = 0.631$). The same can be observed for agreeableness, $t(334.70) = 0.40$, $p = 0.692$, ($M = 0.866$, $SD = 0.796$) and ($M = 0.836$, $SD = 0.607$); and conscientiousness, $t(354.56) = 0.81$, $p = 0.420$, ($M = 0.857$, $SD = 0.685$) and ($M = 0.801$, $SD = 0.620$); neuroticism, $t(326.14) = 0.53$, $p = 0.598$, ($M = 1.019$, $SD = 0.973$) and ($M = 0.972$, $SD = 0.704$); and openness, $t(344.83) = -1.56$, $p = 0.120$, ($M = 0.735$, $SD = 0.677$) and ($M = 0.837$, $SD = 0.556$).

The null hypothesis was rejected on neither extraversion nor conscientiousness. Therefore, the two dimensions were not disqualified.

5.4. Applying the Test Procedure to Smaller User Groups

In the above, we have demonstrated a showcase run of the procedure with 18 observers. However, smaller user groups will be more common. Therefore, we conducted more tests on the 20 simulated user groups of 15 users, 12 users, 8 users, 5 users, and 3 users each—100 user groups in total. The results are recorded in Tables 4–8, respectively. The results varied,

but extraversion and conscientiousness were the most frequently identified dimensions, followed by agreeableness and neuroticism, which were possible but much less frequent. For user groups of size-5 and size-3, there were not many identifications: fewer than half of the user groups in each setting identified at least one dimension from the five. It was possible that the current method was only effective for user groups with more than eight people.

The variations imply that the personality dimensions identified previously, extraversion and conscientiousness, were not necessarily the case across all simulated user groups, which were still drawn from the same small pool of 18 observers. We expect results to be more varied in real situations. Personality perception varies from person to person and user group to user group, which is why a method such as the one proposed here will be useful.

Table 4. Results of 20 randomly chosen size-15 groups.

Groups	Five Principal Personality Dimensions ^a				
	ext.	agr.	con.	neu.	ope.
User Group 1	*	-	*	-	-
User Group 2	-	-	-	-	-
User Group 3	*	-	*	-	-
User Group 4	*	-	*	-	-
User Group 5	*	-	-	-	-
User Group 6	*	-	*	-	-
User Group 7	*	-	*	-	-
User Group 8	*	*	-	*	-
User Group 9	*	-	*	-	-
User Group 10	*	-	-	-	-
User Group 11	*	-	*	-	-
User Group 12	*	-	*	-	-
User Group 13	*	-	-	-	-
User Group 14	*	-	-	-	-
User Group 15	*	-	*	-	-
User Group 16	*	-	*	-	-
User Group 17	*	-	*	-	-
User Group 18	*	-	-	-	-
User Group 19	*	-	*	-	-
User Group 20	*	-	*	-	-

^a Identified personality dimensions are marked with an asterisk.

Table 5. Results of 20 randomly chosen size-12 groups.

Groups	Five Principal Personality Dimensions ^a				
	ext.	agr.	con.	neu.	ope.
User Group 1	*	-	-	-	-
User Group 2	*	-	*	-	-
User Group 3	*	-	-	-	-
User Group 4	*	-	-	-	-
User Group 5	*	-	-	-	-
User Group 6	*	-	*	-	-
User Group 7	*	-	-	-	-
User Group 8	*	-	*	-	-
User Group 9	*	-	-	-	-
User Group 10	*	-	*	-	-
User Group 11	*	-	-	-	-
User Group 12	*	-	-	-	-
User Group 13	*	-	*	-	-
User Group 14	*	-	*	-	-
User Group 15	*	-	-	-	-
User Group 16	*	-	*	-	-
User Group 17	*	-	-	-	-
User Group 18	*	-	*	-	-
User Group 19	*	-	*	-	-
User Group 20	*	-	-	-	-

^a Identified personality dimensions are marked with an asterisk.

Table 6. Results of 20 randomly chosen size-8 groups.

Groups	Five Principal Personality Dimensions ^a				
	ext.	agr.	con.	neu.	ope.
User Group 1	*	-	*	-	-
User Group 2	*	-	-	-	-
User Group 3	-	-	-	-	-
User Group 4	-	-	-	-	-
User Group 5	*	-	-	-	-
User Group 6	*	*	-	-	-
User Group 7	*	-	-	-	-
User Group 8	-	-	-	-	-
User Group 9	*	-	*	-	-
User Group 10	-	-	-	-	-
User Group 11	*	-	-	-	-
User Group 12	*	-	*	-	-
User Group 13	*	-	-	-	-
User Group 14	-	-	-	-	-
User Group 15	*	-	*	-	-
User Group 16	*	-	*	-	-
User Group 17	*	-	*	-	-
User Group 18	*	-	-	-	-
User Group 19	*	-	-	-	-
User Group 20	-	-	*	*	-

^a Identified personality dimensions are marked with an asterisk.

Table 7. Results of 20 randomly chosen size-5 groups.

Groups	Five Principal Personality Dimensions ^a				
	ext.	agr.	con.	neu.	ope.
User Group 1	*	-	-	-	-
User Group 2	*	-	*	-	-
User Group 3	-	-	-	-	-
User Group 4	-	-	-	-	-
User Group 5	-	-	-	-	-
User Group 6	-	-	-	-	-
User Group 7	*	-	*	-	-
User Group 8	-	-	-	-	-
User Group 9	-	-	-	-	-
User Group 10	-	-	-	-	-
User Group 11	*	-	-	-	-
User Group 12	*	-	*	-	-
User Group 13	*	-	-	-	-
User Group 14	-	-	-	-	-
User Group 15	-	-	-	-	-
User Group 16	-	-	-	-	-
User Group 17	-	-	*	-	-
User Group 18	-	-	-	-	-
User Group 19	-	-	-	-	-
User Group 20	-	-	-	-	-

^a Identified personality dimensions are marked with an asterisk.

5.5. Summary of the Results

In our tests of the test procedure, most tests featuring simulated user groups with eight or more people could identify extraversion. Conscientiousness was identified in many tests as well but significantly less frequent than extraversion. The identification of extraversion agrees with several studies about the head and eye movements or gaze of humanoid robots or virtual agents [20,33,44–46], as does the identification of conscientiousness [33,45]. However, the two dimensions were by no means the only possibilities. Agreeableness and neuroticism were possible as well, though much less frequent. All results agreed that openness was not an engineering-worthy dimension for the type of robot tested for the usage of dyadic communication.

Table 8. Results of 20 randomly chosen size-3 groups.

Groups	Five Principal Personality Dimensions ^a				
	ext.	agr.	con.	neu.	ope.
User Group 1	-	-	-	-	-
User Group 2	-	-	-	-	-
User Group 3	-	-	-	-	-
User Group 4	*	-	*	-	-
User Group 5	-	-	-	-	-
User Group 6	-	-	-	-	-
User Group 7	*	-	-	-	-
User Group 8	-	-	-	-	-
User Group 9	*	-	-	-	-
User Group 10	*	-	*	-	-
User Group 11	*	-	*	-	-
User Group 12	-	-	-	-	-
User Group 13	-	-	*	-	-
User Group 14	*	*	-	-	-
User Group 15	-	-	-	-	-
User Group 16	*	-	-	-	-
User Group 17	-	-	-	-	-
User Group 18	-	-	-	-	-
User Group 19	-	-	-	-	-
User Group 20	-	-	-	-	-

^a Identified personality dimensions are marked with an asterisk.

6. Discussion

The test procedure was put to 101 limited-scope tests, where we simulated an engineering situation where 10 humanoid robot personalities were to be engineered for dyadic communication for small user groups of 18, 15, 12, 8, 5, and 3 users. The corresponding test results for simulated user groups of eight or more people reflect the existing body of knowledge on the potential of humanoid robots for expressing personality traits, thereby confirming the effectiveness of the proposed test procedure within the scope of the experiment for user groups with at least eight people. However, the variations revealed should allow us a glimpse into a common situation of engineering robot personalities, which is that no one answer applies to all cases. The working hypotheses—that not all but some human personality dimensions can be used to engineer robot personalities for non-android robots that can imitate human behaviour—are supported as least within the scope of the experiment. Our test procedure is the first one dedicated to engineering robot personalities in significant quantities for practical applications with small user groups. It is inevitably primitive, and there is much space for improvement. There are several limitations in particular we can focus on.

The primary limitation of the test procedure is that it is currently ineffective for user groups of fewer than eight people. Compatibility with smaller user groups is the most needed improvement considering that most household user groups probably have fewer than eight people. Another limitation is that the test procedure is not a guarantee for the applicability of personality dimensions based on the lexical hypothesis. For guaranteed applicability, we will need to develop dedicated personality models for robots based on lexicons for describing robots, as a recent study did for conversational agents [9]. Dedicated models can be used together with the test procedure to identify engineering-worthy dimensions for a type of robot. Considering that robots come in different shapes and sizes for different purposes, even a dedicated model does not necessarily fully apply to a particular type of robot. Last but not least, our test procedure is formulated based on our definition of robot personalities. Our definition stipulates that robots be individuals exhibiting characteristic patterns of computation and behaviour with inter- and intra-individual differences, which suits our aim as to engineer robot personalities in significant quantities. However, as there is currently no consensus on how robot personality should be defined and it is questionable whether robot personalities should be imitations of human

personalities [12], our understanding on robot personalities will keep evolving. Our current approach to creating robot personalities needs to adapt and improve accordingly.

The primary limitation of the experiment is that we only confirmed the effectiveness of the test procedure in a limited scope. We do not assert universal applicability. Future research can also aim to expand the applicability of the proposed test procedure and validate it in a broader context. In particular, albeit the test procedure is meant for engineering populations of robot personalities that will be observed together, such as a music band or idol group of robots or a staff of robotic errand runners in an office, the scope of the experiment did not properly reflect this aim. The experiment featured multiple instances of dyadic communication rather than users interacting with multiple robot personalities at the same time. It is arguable that dyadic communication may still represent the most common form of interaction. For example, in an office where humans work with multiple errand runners, it is still more common for a human to interact with one of them at a time. Still, an important future work is to cover interaction involving multiple physically present robots. This would require multiple units of the type of robot to study and abolishing video-based stimuli. In addition, the content of the interaction featured in the experiment, which was about getting to know each other, albeit important for leaving first impressions, did not necessarily reflect the main contents of interaction with service robots. Future studies and implementations of the test procedure would benefit from settings that better reflect the contents of the interaction based on realistic roles of service robots. Another major limitation of the experiment was the simulation of 101 user groups by drawing from a small pool of 18 observers. There were two possible impacts of this approach: underrepresented inter-group variations and intra-group consistency. A main goal of the test procedure is to take into account the inter-group differences of different user groups when engineering robot personalities. However, by drawing from the same pool of 18 observers, the possible inter-group variations were not properly reflected in the results (Tables 4–8). The intra-group consistency by randomly combining members of the 18 online observers, who observed the robot personalities by watching short videos, might be lower than that from a real user group interacting in the same environment with real robots.

Another limitation present both in the proposed test procedure and our experiment was introduced by using an imprecise personality measure. This is also a limitation of the field, which currently lacks a personality measure precise enough for engineering tasks. The use of an imprecise measure based on Likert scales not only introduces possible incompatibility issues with statistical tests, it also affects the resolution of presenting archetypes. In practical engineering work, we cannot let two archetypes have the same trait on any dimension. However, traits as measured by an imprecise measure can be the same in terms of their numeric trait levels. Therefore, using human archetypes as measured by an imprecise personality measure should be limited to identifying personality dimensions for further practical engineering work but not the actual engineering work, especially engineering tasks of a certain scale, such as when we need to engineer hundreds or thousands of robot personalities, as suitable for mass-produced robots. To that end, we will need a dedicated personality model with precise, continuous scales for designing and measuring archetypes in greater numbers with better precision. But then again, it is doubtful whether this dedicated model for robot personalities can be used by human archetypes to report their own personalities. The future of robot personalities engineering might need to go beyond imitating humans.

7. Conclusions

We have proposed a test procedure for identifying with the help of a user group personality dimensions to engineer robot personalities out of a type of robot knowing its typical usage. The test procedure should be suitable for practical engineering tasks since it identifies personality dimensions on which users can provide effective feedback to guide the engineering of robot personalities; it supports engineering robot personalities in significant quantities; and it needs as few as eight users to function. It can serve as a useful tool for robot personalities engineers to focus on the personality dimensions that matter to users

and on which personality traits can be effectively engineered. It can also help check whether previous engineering effort yields successful results. Regrettably, we were unable to validate the tool in a more general sense. Therefore, we encourage our peers, in academia or industry, to try and improve the tool and come up with better ones to make robot personalities engineering a less challenging endeavour.

Author Contributions: Conceptualization, L.L., K.O. and H.I.; Data curation, L.L.; Formal analysis, L.L.; Funding acquisition, K.O. and H.I.; Investigation, L.L.; Methodology, L.L.; Project administration, L.L., K.O. and H.I.; Resources, L.L., K.O. and H.I.; Software, L.L.; Supervision, K.O. and H.I.; Validation, L.L.; Visualization, L.L.; Writing—original draft, L.L.; Writing—review and editing, L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JST ERATO under Grant [JPMJER1401], Grant-in-Aid for Scientific Research (A) from JSPS under Grant [18H04114], and Grant-in-Aid for Scientific Research from JSPS under Grant [19H05693].

Institutional Review Board Statement: The study was approved by the Research Ethics Committee of Osaka University.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets collected and analysed in this study can be found via https://drive.google.com/drive/folders/11MGEfyrxheGvJRX_5BLPjQt6SoQFR1Go?usp=sharing, accessed on 6 January 2022.

Acknowledgments: The authors sincerely acknowledge Graham Peebles for his advice on structuring this article.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Details of the Interview

The interviews were conducted to capture the head and eye movements of the candidate archetypes. The sessions took place in a soundproof room, where a round white table was set in the middle, between two chairs placed on opposite sides. The candidate and our interviewer (the actor) sat face-to-face on these chairs. On the table were two cameras positioned side-by-side, one pointing at the candidate's face right in front of them on the eye level and the other, a gimbal camera, set to track the candidate's head just in case they moved out of the field of the fixed camera (this never happened). The clips shot by the fixed camera were used as the sources of head and eye angles. There was also a tablet computer set on a tablet stand for displaying the questions asked during the interviews. Its screen first faced the interviewer. Behind the interviewer to the right was another camera on top of a tall tripod for logging the interviews. When an interview session started, the interviewer requested the candidate to briefly introduce themselves. After the candidate finished, the interviewer proceeded to ask questions from the tablet. There were 12 questions in total, all selected from a website for practising English conversation [43]. The interviewer proceeded to ask the 12 questions sequentially. The candidate had been told that they could skip questions just in case some questions were too personal; the candidates answered all questions nonetheless. The interviewer listened to the candidate answering the questions. In about 5 to 10 min, the interviewer exhausted the list. He then proffered the tablet to the candidate. From now on, it was the candidate's turn to ask the same questions in the same order, only that the interviewer (now being interviewed) would answer all questions according to a script he had memorised beforehand. This part of the interview was controlled and it took about five minutes for most candidates. When the candidate exhausted the list, the interview was over. Finally, the candidate reported their own personalities (for the third time) and then the interviewer's before they left.

Appendix B. The Conversation Script

The below is the script the actor used to answer the questions from the participants of the interview. Part of the script also appeared as the conversation script of the human–robot interaction video stimuli. The part that made it to the videos is highlighted in bold. (The answers in the script represent no views or opinions of any real person or group.)

Q1. What are some characteristics of your personality?

A1. I am introverted. . . very conscious of myself but not so much of others. I might not be very likeable, but the people who know me well may consider me a good companion or ally. I'm wont to have unorthodox views about things, but at the same time, I could be very conservative—a paradox as people might say.

Q2. What makes you happy?

A2. Oh! Many things make me happy, such as a good meal or a good book. In general, I'm happy when things have gone my way.

Q3. Would you like to be different?

A3. That. . . depends on how different I would be. Were I to be offered to have my personality changed drastically, for good or ill, I think I would decline without hesitation. However, if the changes were minor and for the good, I think I wouldn't mind.

Q4. Are you a determined person? Are you a stubborn person?

A4. I'm determined to achieve the goals I deem important. In addition, I can be very stubborn—especially about doing the right things. However, I'm less determined and stubborn in non-essential things. For instance, I don't mind very much having my lunch menu suddenly changed or vacation cancelled—even when I am looking forward to them.

Q5. What is one thing that many people don't know about you?

A5. That I'm evil and often dream of taking over the world. Just kidding! A thing that many people don't know—which I suspect that my parents are no exceptions—is my untwonted mischievous side. It is a point a bit difficult to elaborate on, but I'll try. For most of the time, I'm darn serious. However, when chances come, I might try to play some harmless tricks. For example, there was one time when I was in an English class where the students were assigned to give a presentation on a topic of their interest. I made a presentation on fictional villains. I tried to imitate the demeanour of some villains—mainly how they spoke—as how I introduced them. However, there was hardly any reaction from my classmates, though. Perhaps my villain impressions were bad.

Q6. Do you think your personality has changed over the past few years? In what way has it changed?

A6. I don't think my personality has changed much over the past few years. That being said. . . discernible changes did occur, with the most prominent one being my growing indifference towards things that do not concern me directly.

Q7. Do you think you can change a major characteristic of your personality if you try?

A7. I don't think that's possible—to cause fundamental changes in one's personality through personal effort. What seems to be possible, however, is to change one's behaviour through conscious effort. . . so that other people might have a different impression of you. I think. . . it is how celebrities maintain their personas.

Q8. If you could change any aspect of your personality, what would it be?

A8. Tough question. I wouldn't want any major changes, whatever they were, so. . . um. . . what's the aspect of my personality that's hamstringing me the most? Mm. . . I'm wont to do things on my own and less inclined to collaborate with others. In addition, I

seldom do things together with people just for the sake of doing things together that is, I seldom go shopping, go to the movies, or have lunch with other people. Being such a lone wolf has its pros and cons. However, I think it would benefit me if I tune-up my inclination for group activities just a little bit.

Q9. *Are you shy? In which occasions are you shy?*

A9. *I am shy, but the stereotypical ‘Asian’ shy. Generally speaking, I am shy in front of people I don’t know well.*

Q10. *Are you more introverted or more extroverted?*

A10. *I am extremely introverted. . . but not the way how introverts are introverted. Most introverts are easily associated with shyness; they are introverted because they are not very social—even a little afraid of people—at least that’s the stereotype. It’s not my case, though. I’m introverted because I don’t care much about other people.*

Q11. *Do you think you have an unusual personality? Why?*

A11. *I don’t think so. My personality is just the product of me being a person optimised for the environment he is in; in other words, I am who I am because this is how things will go the easiest for me.*

Q12. *Do you consider yourself to be even-tempered?*

A12. *I’m very even-tempered. In fact, I never get angry. However, again. . . I suspect that my emotional stability comes from my lack of concern for many things.*

References

1. APA Dictionary of Psychology. Available online: <https://dictionary.apa.org/personality> (accessed on 8 December 2021).
2. Baumert, A.; Schmitt, M.; Perugini, M.; Johnson, W.; Blum, G.; Borkenau, P.; Costantini, G.; Denissen, J.J.A.; Fleeson, W.; Grafton, B.; et al. Integrating Personality Structure, Personality Process, and Personality Development. *Eur. J. Personal.* **2017**, *31*, 503–528. [CrossRef]
3. Vinciarelli, A.; Mohammadi, G. A Survey of Personality Computing. *IEEE Trans. Affect. Comput.* **2014**, *5*, 273–291. [CrossRef]
4. Tupes, E.C.; Christal, R.E. Recurrent Personality Factors Based on Trait Ratings. *J. Personal.* **1992**, *60*, 225–251. [CrossRef] [PubMed]
5. Norman, W.T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *J. Abnorm. Soc. Psychol.* **1963**, *66*, 574–583. [CrossRef] [PubMed]
6. McCrae, R.R.; Costa, P.T. Updating Norman’s “adequacy taxonomy”: Intelligence and personality dimensions in natural language and in questionnaires. *J. Personal. Soc. Psychol.* **1985**, *49*, 710–721. [CrossRef]
7. Goldberg, L.R. An alternative “description of personality”: The Big-Five factor structure. *J. Personal. Soc. Psychol.* **1990**, *59*, 1216–1229. [CrossRef] [PubMed]
8. Galton, F. The Measurement of Character. In *Readings in General Psychology*; Dennis, W., Ed.; Prentice-Hall, Inc.: Hoboken, NJ, USA, 1949; pp. 435–444.
9. Völkel, S.T.; Schödel, R.; Buschek, D.; Stachl, C.; Winterhalter, V.; Bühner, M.; Hussmann, H. Developing a Personality Model for Speech-Based Conversational Agents Using the Psycholexical Approach. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems CHI’20, Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–14. [CrossRef]
10. Tversky, A.; Kahneman, D. Judgment under Uncertainty: Heuristics and Biases. *Science* **1974**, *185*, 1124–1131. [CrossRef] [PubMed]
11. Asch, S.E. Forming Impressions of Personality. *J. Abnorm. Soc. Psychol.* **1946**, *41*, 258–290. [CrossRef] [PubMed]
12. Mou, Y.; Shi, C.; Shen, T.; Xu, K. A Systematic Review of the Personality of Robot: Mapping Its Conceptualization, Operationalization, Contextualization and Effects. *Int. J. Hum.-Comput. Interact.* **2020**, *36*, 591–605. [CrossRef]
13. Robert, L.; Alahmad, R.; Esterwood, C.; Kim, S.; You, S.; Zhang, Q. A Review of Personality in Human–Robot Interactions. *Found. Trends® Inf. Syst.* **2020**, *4*, 107–212. [CrossRef]
14. Kaniarasu, P.; Steinfeld, A.M. Effects of blame on trust in human robot interaction. In Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2014; pp. 850–855. [CrossRef]
15. Lee, K.; Peng, W.; Jin, S.A.; Yan, C. Can Robots Manifest Personality? An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human–Robot Interaction. *J. Commun.* **2006**, *56*, 754–772. [CrossRef]
16. Lohse, M.; Hanheide, M.; Wrede, B.; Walters, M.L.; Koay, K.L.; Syrdal, D.S.; Green, A.; Huttenrauch, H.; Dautenhahn, K.; Sagerer, G.; et al. Evaluating extrovert and introvert behaviour of a domestic robot—A video study. In Proceedings of the RO-MAN 2008—The 17th IEEE International Symposium on Robot and Human Interactive Communication, Munich, Germany, 1–3 August 2008; pp. 488–493. [CrossRef]

17. Yamashita, Y.; Ishihara, H.; Ikeda, T.; Asada, M. Path Analysis for the Halo Effect of Touch Sensations of Robots on Their Personality Impressions. In *Social Robotics*; Agah, A., Cabibihan, J.J., Howard, A.M., Salichs, M.A., He, H., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 502–512.
18. Broadbent, E.; Kumar, V.; Li, X.; Sollers, J., 3rd; Stafford, R.Q.; MacDonald, B.A.; Wegner, D.M. Robots with Display Screens: A Robot with a More Humanlike Face Display Is Perceived To Have More Mind and a Better Personality. *PLoS ONE* **2013**, *8*, e72589. [[CrossRef](#)] [[PubMed](#)]
19. Goetz, J.; Kiesler, S. Cooperation with a Robotic Assistant. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2002; pp. 578–579. [[CrossRef](#)]
20. Andrist, S.; Mutlu, B.; Tapus, A. Look like me: Matching robot personality via gaze to increase motivation. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Korea, 18–23 April 2015; ACM: New York, NY, USA, 2015; pp. 3603–3612. [[CrossRef](#)]
21. Ogawa, K.; Bartneck, C.; Sakamoto, D.; Kanda, T.; Ono, T.; Ishiguro, H. Can An Android Persuade You? In Proceedings of the RO-MAN 2009—The 18th IEEE International Symposium on Robot and Human Interactive Communication, Toyama, Japan, 27 September–2 October 2009; pp. 516–521.
22. Walters, M.L.; Syrdal, D.S.; Dautenhahn, K.; te Boekhorst, R.; Koay, K.L. Avoiding the uncanny valley: Robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Auton. Robot.* **2008**, *24*, 159–178. [[CrossRef](#)]
23. Paetzel-Prüsmann, M.; Perugia, G.; Castellano, G. The Influence of robot personality on the development of uncanny feelings. *Comput. Hum. Behav.* **2021**, *120*, 106756. [[CrossRef](#)]
24. Household Size and Composition around the World. Available online: https://www.un.org/en/development/desa/population/publications/pdf/popfacts/PopFacts_2017-2.pdf (accessed on 8 December 2021).
25. Small Business Statistics. Available online: <https://smallbiztrends.com/tag/small-business-statistics> (accessed on 8 December 2021).
26. 2019 White Paper on Small and Medium Enterprises in Japan 2019 White Paper on Small Enterprises in Japan (Summary). Available online: https://www.chusho.meti.go.jp/pamflet/hakusyo/2019/PDF/2019hakusyosummary_eng.pdf (accessed on 8 December 2021).
27. Weiss, A.; van Dijk, B.; Evers, V. Knowing Me Knowing You: Exploring Effects of Culture and Context on Perception of Robot Personality. In Proceedings of the 4th International Conference on Intercultural Collaboration, Bengaluru, India, 21–23 March 2012; Association for Computing Machinery: New York, NY, USA, 2012; pp. 133–136. [[CrossRef](#)]
28. Bruckenberg, U.; Weiss, A.; Mirnig, N.; Strasser, E.; Stadler, S.; Tscheligi, M. The Good, The Bad, The Weird: Audience Evaluation of a “Real” Robot in Relation to Science Fiction and Mass Media. In *Social Robotics*; Herrmann, G., Pearson, M.J., Lenz, A., Bremner, P., Spiers, A., Leonards, U., Eds.; Springer International Publishing: Cham, Switzerland, 2013; pp. 301–310.
29. Sandoval, E.B.; Mubin, O.; Obaid, M. Human Robot Interaction and Fiction: A Contradiction. In *Social Robotics*; Beetz, M., Johnston, B., Williams, M.A., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 54–63.
30. Kriz, S.; Ferro, T.; Damera, P.; Porter, J.R. Fictional robots as a data source in HRI research: Exploring the link between science fiction and interactional expectations. In Proceedings of the 19th International Symposium in Robot and Human Interactive Communication, Viareggio, Italy, 13–15 September 2010; pp. 458–463.
31. Ray, C.; Mondada, F.; Siegwart, R. What do people expect from robots? In Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, Nice, France, 22–26 September 2008; pp. 3816–3821.
32. Funder, D.C. On seeing ourselves as others see us: Self–other agreement and discrepancy in personality ratings. *J. Personal.* **1980**, *48*, 473–493. [[CrossRef](#)]
33. Bremner, P.A.; Celiktutan, O.; Gunes, H. Personality Perception of Robot Avatar Teleoperators in Solo and Dyadic Tasks. *Front. Robot. AI* **2017**, *4*, 16. [[CrossRef](#)]
34. Woods, S.; Dautenhahn, K.; Kaouri, C.; Boekhorst, R.; Koay, K.L. Is This Robot Like Me? Links Between Human and Robot Personality Traits. In Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots, Tsukuba, Japan, 5–7 December 2005; pp. 375–380. [[CrossRef](#)]
35. Schaal, S. Is imitation learning the route to humanoid robots? *Trends Cogn. Sci.* **1999**, *3*, 233–242. [[CrossRef](#)]
36. Breazeal, C.; Scassellati, B. Robots that imitate humans. *Trends Cogn. Sci.* **2002**, *6*, 481–487. [[CrossRef](#)]
37. Nakaoka, S.; Nakazawa, A.; Kanehiro, F.; Kaneko, K.; Morisawa, M.; Ikeuchi, K. Task Model of Lower Body Motion for a Biped Humanoid Robot to Imitate Human Dances. In Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, AB, Canada, 2–6 August 2005; pp. 3157–3162. [[CrossRef](#)]
38. Baltrušaitis, T.; Zadeh, A.; Lim, Y.C.; Morency, L.P. OpenFace 2.0: Facial Behavior Analysis Toolkit. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, Xi'an, China, 15–19 May 2018.
39. Baltrušaitis, T.; Robinson, P.; Morency, L.P. Constrained Local Neural Fields for robust facial landmark detection in the wild. In Proceedings of the IEEE International Conference on Computer Vision Workshops, 300 Faces in-the-Wild Challenge, Sydney, Australia, 2–8 December 2013.
40. Zadeh, A.; Baltrušaitis, T.; Morency, L. Convolutional Experts Constrained Local Model for Facial Landmark Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 2051–2059. [[CrossRef](#)]

41. Wood, E.; Baltruaitis, T.; Zhang, X.; Sugano, Y.; Robinson, P.; Bulling, A. Rendering of Eyes for Eye-Shape Registration and Gaze Estimation. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; IEEE Computer Society: Washington, DC, USA, 2015; pp. 3756–3764. [[CrossRef](#)]
42. John, O.P.; Donahue, E.M.; Kentle, R.L. *The Big Five Inventory—Versions 4a and 54*; Technical Report; Institute of Personality and Social Research, University of California, Berkeley: Berkeley, CA, USA, 1991.
43. Conversation Questions Personality. Available online: <http://iteslj.org/questions/personality.html> (accessed on 8 December 2021).
44. Ruhland, K.; Zibrek, K.; McDonnell, R. Perception of Personality through Eye Gaze of Realistic and Cartoon Models. In Proceedings of the ACM SIGGRAPH Symposium on Applied Perception, Tübingen, Germany, 13–14 September 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 19–23. [[CrossRef](#)]
45. Celiktutan, O.; Bremner, P.; Gunes, H. Personality Classification from Robot-mediated Communication Cues. In Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New York, NY, USA, 26–31 August 2016.
46. Ijuin, K.; Jokinen, K. Exploring Gaze Behaviour and Perceived Personality Traits. In *Social Computing and Social Media. Design, Ethics, User Behavior, and Social Network Analysis*; Meiselwitz, G., Ed.; Springer International Publishing: Cham, Switzerland, 2020; pp. 504–512. [[CrossRef](#)]