Study information

**Titel**:

The Uncanny Valley Hypothesis: Examining the Moderating Role of Motion and Appearance. A Registered Replication of Piwek, McKay, & Pollick (2014)

**Authors:**

Sean Hughes

Ghent University  
Department of Experimental Clinical and Health Psychology

Stephanie Timperman

(Ghent University)

**Description:**

The uncanny valley hypothesis refers to the idea that almost but not fully humanlike artificial characters will trigger a profound sense of unease in people. This hypothesis is widely acknowledged both the popular media and scientific research. Despite its popularity, empirical evidence surrounding this hypothesis remains inconsistent. In this pre-registered effort, we sought to replicate the findings reported by Piwek et al. (2014).

They argued that improving the motion quality of characters systematically improved the acceptability of the characters. In particular, the character classified in the deepest location of the uncanny valley became more acceptable when it was animated. They claimed that although an uncanny valley effect was found for static characters, the deepening of the valley with motion, originally predicted by Mori (1970/2012), was not obtained.

**Confirmatory Analyses:**

**H1**: Ratings of human likeness (i.e., how similar to a human a character is) will differ as a function of *Character Type*.

**H1a.** Follow-up comparison testswill examine if Piwek et al.’s (2014) findings replicate, such that robots will be rated lower than a mannequin (but not each other), the mannequin lower than skeleton or zombie, and these two figures will not differ from one another but both will be rated as lower in human likeness than the two human figures.

**H2:**  Acceptability ratings will vary as a function of *Character Type*.

**H2a**. Follow-up comparison tests will examine if Piwek et al.’s findings replicate, such that characters close in likeness to humans but not themselves human (e.g., zombie, skeleton), will be liked least, characters that are most (e.g., humans) and least like humans (e.g., robots) will be liked relatively more.

**H3**: Acceptability ratings will vary as a function of *Motion Type*.

**H3a**: Follow-up comparison tests will examine if Piwek et al.’s findings replicate, such that the uncanny valley effect (see H2) will be larger for static characters than moving characters. Increasing movement distortion is predicted to lead to a lowering of acceptability ratings for all characters relative to those ratings obtained in the natural motion condition.

**H4**: There will be an interaction effect between *Character Type* and *Motion* *Type* for acceptability ratings.

**H4a**: Follow-up comparison tests will examine if Piwek et al.’s findings replicate. Specifically we will examine if degrading motion has a generally negative effect on all characters, with the change in acceptability ratings from static to naturally moving differing depending upon the character. Naturally moving zombies are predicted to be significantly more acceptable than the static zombie.

**Exploratory Analyses**

Design plan

**Study Type:**

This is an experimental study that will be carried out in an online setting. Participants will be recruited via the Prolific website (https://prolific.co/).

**Blinding**:

Participants will be blind to the purpose of the study when taking part. However, they will be informed about the purpose of the study after completing it (i.e., during the debriefing phase).

**Study design:**

The Piwek et al. (2014) study contained one between participants factor (*Outcome Type*: human likeness ratings vs. acceptability ratings).

With respect to the human likeness ratings, all participants encountered a single factor (*Character Type*) with seven levels (i.e., battle robot vs. toy robot vs. mannequin vs. skeleton vs. zombie vs. low res man vs high res man).

With respect to acceptability ratings, a 7(*Character Type*) x 6 (*Motion Type*: static vs. natural motion vs. distorted A, vs. distorted B vs. distorted C vs. distorted D) within participants design was employed, such that all participants rated all character types across all motion types in terms of their acceptability.

**Randomization:**

Participants were randomly assigned to either the human likeness ratings condition or the acceptability rating condition. Presentation of the stimuli will also be randomly counterbalanced within each condition.

Sampling Plan

**Existing Data:**

The original authors provided data from their original study (we are making this available via XXX). In our study we will collect entirely new data. These data have not yet been collected, created, or realized.

**Explanation of existing data:**

Piwek et al. (2014) recruited forty students (M*age* = 25, *SD* = 4.7) from a Scottish university and paid them for their participation. They divided those participants into two different experimental conditions: human likeness ratings and acceptability ratings.

**Data collection procedure:**

Participants will be recruited through the online data collection platform Prolific and provided with a monetary reward for their efforts.

**Sample size:**

Our sample size is targeted at 100 participants, while the original study of Piwek et al. (2013) had a sample size of 40. We will attempt to recruit a 160 participants in case of drop-outs. The criteria for replication is an effect in the same direction as the original study and a p-value < 0.05. BEREKEN steekproefgrootte adhv POWER!

**Sample size rationale:**

Power calculations were done based on “The Replication Recipe: What makes for a convincing replication?” by Brandt et al. (2013) and the guidelines of the [Social Sciences Replication Project (SSRP).](about:blank)

1. While having a high statistical power is crucial to the replication study – up to 0.95 -
2. Power calculations were done in accordance with The first round of data collection achieves 90% power to detect 75% of the original effect size.

**Stopping rule:**

Data collection will continue until we recruit AA participants. We will collect an additional BB participants to allow for attrition and incomplete data so that we still have CC participants as required by our power analysis.

Variables

**Manipulated variables:**

**Outcome Type**: Two dependent variables will be assessed in the study in a between participants fashion: half of the participants will provide *human likeness ratings* for each character while the other half will provide *acceptability ratings* for each character completing each movement type.

* Human likeness ratings will be assessed using a 9 point Likert scale ranging from 1 (*Very Non-Humanlike*) to 9 (*Very Humanlike*).
* Acceptability ratings will be assessed using a 9 point Likert scale ranging from 1 (*Totally Unacceptable*) to 9 (*Totally Acceptable*). Participants will be encouraged to go with their first impressions when making their choice.

**Character Type**: The type of characters will be manipulated within participants. Seven 3D computer characters will be used: a battle robot, toy robot, mannequin, skeleton, zombie, low resolution man, and a high resolution man. Each image will be located in the middle of the screen and presented in the frontal orientation, facing the viewer.

**Motion Type**: The manner in which a character moves will also be manipulated within participants (for those in the acceptability condition). Specifically, participants will encounter each character statically as well as engaging in five different movements: natural movement, distorted movement A, B, C, and D. The movement will always involve a knocking motion (e.g., knocking on a door) with the character’s right hand. Whereas the natural movement is generally smooth and accomplished by moving multiple joints simultaneously to create an end-effector (e.g., wrist) trajectory with a bell-shaped speed profile (Flash & Hogan, 1985; Rosenbaum, 2009).

The distorted movements were designed to make the motion less natural by disrupting simultaneous movement of the joints of the arm. This was achieved by locking the shoulder joint angle constant whilst allowing the elbow to move, and then switching to locking the elbow joint angle constant whilst letting the shoulder joint move, and continuing this alternation for the duration of the movement.

**Measured variables:**

Human likeness and acceptability ratings will be measured.

**Indices**:

NA

Analysis Plan

**Statistical models:**

*Confirmatory Analyses*

We will use the following models (and corrections as in Piwek et al. [2014]):

* Human likeness ratings will be subject to a one-way ANOVA with Character Type as a within participant factor. If a main effect of Character Type emerges follow-up testing will be carried out to interpret that effect.
* Acceptability ratings will be subject to a 7(*Character Type*) x 6 (*Motion Type*) within participants ANOVA. If the main effects of Character Type or Motion Type, or the interaction between the two emerge, then follow-up tests will be carried out to interpret the findings.

Note: Greenhouse–Geisser corrections will be used for all tests in cases where violations of the sphericity assumption occur.

*Exploratory Analyses*

**Transformations:**

No transformations will be required.

**Inference criteria:**

For this replication we will use the original authors criteria (*p*-value < 0.05) for the confirmatory analyses. We will also compute Bayesian factors in accordance with procedures outlined by Rouder, Speckman, Sun, Morey, and Iverson (2009) to estimate the amount of evidence for the hypothesis that likeness and acceptability ratings differ as a function of Character Type or Motion Type (alternative hypothesis) or that there is no such difference (null hypothesis).

**Data exclusions:**

As the static images shown before the dynamic characters could influence participants ratings, we will conduct an control experiment, as suggested and executed in original study. If the results show an interference for certain participants, the data cannot be used to make assumptions regarding the hypothesis.

Participants will be excluded from analyses if they fail to provide complete data for all measures, or if they fail the attention or suspicion checks.

**Missing data:**

Participants will be excluded from analyses in any case where they fail to provide complete data for all measures. Therefore missing data treatment will not be necessary.

**Exploratory analyses:**

In both conditions, the human likeness rating as well as the acceptability rating condition, participants will be asked to answer questions regarding eeriness on a 5- point Likert scale:

1. Not at all eery
2. A little eery
3. Moderate eery
4. Strongly eery
5. Extremely eery

The level of significance will be standard (p<0.05).