Replication of a Research Claim from Piwek et al. (2014)

from *Cognition*

Replication Team: Sean Hughes

Action Editor: Kai Jonas

Independent Reviewers

(add name below when you initiate review, comment “DONE” on your name when you finish):

Reviewer #1: Bill Chopik

Reviewer #2: Anna Szabelska

Review Period:

View-only links to:

[Original Paper](https://osf.io/khcw8/?view_only=dd5875fc6af641f1b99729cb51981425)

[Original Materials](https://osf.io/khcw8/?view_only=dd5875fc6af641f1b99729cb51981425)

[Replication Materials and Analysis](https://osf.io/7tr8a/?view_only=f13f35b5e9124b8f9914c0284e95cee4)

General information about preregistration is available at <https://cos.io/prereg> and you can reach out to the SCORE coordinators by emailing [scorecoordinator@cos.io](mailto:scorecoordinator@cos.io) or tagging that account in a comment in the document for additional assistance.

### Study Information

**\_\_\_\_\_\_**

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

**\_\_\_\_\_\_**

**AUTHORS:**

**Sean Hughes 1**

1 Ghent University

1. Description (optional)
   1. *Please give a brief description of your study, including some background, the purpose of the study, or broad research questions.*
   2. ***Example****: Though there is strong evidence to suggest that sugar affects taste preferences, the effect has never been demonstrated in brownies. Therefore, we will measure taste preference for four different levels of sugar concentration in a standard brownie recipe to determine if the effect exists in this pastry.*
   3. ***More info****: The description should be no longer than the length of an abstract. It can give some context for the proposed study, but great detail is not needed here for your preregistration.*

**DESCRIPTION:**

**Background**: The uncanny valley hypothesis refers to the idea that almost but not fully human-like artificial characters will trigger a profound sense of unease in people. This hypothesis is widely claimed in the popular media and scientific research. Yet, despite its popularity, empirical evidence for this claim remains inconsistent. In this pre-registered replication effort, we set out to replicate the findings reported by Piwek et al. (2014).

Piwek et al. (2014) argued that improving the motion quality of characters systematically improved the perceived acceptability of those characters. Specifically, the character classified in the deepest location of the uncanny valley (i.e., when the acceptability of an artificial character does not match its likeness to human form; when this discordance is greatest) became more acceptable when it was animated. They claimed that although an uncanny valley effect emerged for static characters, the deepening of the valley with motion, originally predicted by Mori (1970/2012), did not produce a stronger effect.

**Method**: The authors tested their claim by creating a set of seven computer characters (e.g., robot, zombie, human) that were assumed to vary in their human-likeness and acceptability (operationalised by the original authors as “*how acceptable one finds it to interact with a character on a regular basis*”; see p.271 of their manuscript). For each character a non-moving (static) and moving (animated) version was created. The latter (animated) version was also manipulated so that the movement varied in its assumed naturalness (from more to less natural).

**Procedure**: Participants were assigned to one of two experimental groups: human likeness ratings and acceptability ratings. In the human likeness ratings group, participants were shown a static image of a character in frontal orientation, facing the observer, for two seconds. After each image, participants gave ratings of human likeness on a 9-point Likert scale (1 – *very non-humanlike*, 9 – *very human like*). In the acceptability ratings group, participants were shown two-second-long videos of the complete set of animated characters performing a movement (knocking on the door with a right hand) at every level of movement distortion. This group was also shown a static image of each of the characters, either before or after the video. For both the static and moving character, after each presentation the participants were required to rate the character’s acceptability on a 9-point Likert scale (1 – *totally unacceptable*, 9 – *totally acceptable*).

**Results**: For the human likeness ratings the authors carried out a one-way ANOVA (with Greenhouse–Geisser correction) and reported a main effect of character (F(3.723, 70.737) = 33.277, *p* < 0.001, partial *η*2 = 0.637), such that human likeness ratings differed as a function of character.

For the acceptability ratings the authors first checked for a methodological confound (i.e., to see if acceptability ratings differed when ratings of the static characters had to be emitted before or after ratings of the moving characters). They reported no such confound; p = .59. They then examined if acceptability ratings differed as a function of character and motion using an ANOVA. They reported main effects for character (F(2.874,54.601) = 6.033, p = 0.001, partial *η*2 = 0.241) and motion (F(1.613,30.640) = 15.311, p < 0.001, partial *η*2 = 0.446), and a two-way interaction between these factors (F(4.683,88.985) = 3.928, p = 0.004, partial *η*2 = 0.171).

Finally, they carried out a post-hoc Tukey analysis of the moving vs. static zombie character. This latter analysis showed that the naturally moving zombie was rated as significantly more acceptable than the static zombie (*q* = -6.38, *p* < 0.01).

1. Hypotheses (required)
   1. *List specific, concise, and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here. If a specific interaction or moderation is important to your research, you can list that as a separate hypothesis.*
   2. ***Example****: If taste affects preference, then mean preference indices will be higher with higher concentrations of sugar.*
   3. *The hypothesis for SCORE will be indicated as* ***H\*****. If you plan to do confirmatory research beyond the focal claim and test selected for the SCORE program, please make sure they are clearly delineated from the SCORE claim by numbering further hypotheses beginning with* ***H1****.*

**HYPOTHESIS:**

**Focal SCORE hypothesis test H\*: for the character with the lowest acceptability rating in the static condition (zombie), the naturally moving version of that character will be rated as more acceptable than the static version.**

**Confirmatory Hypotheses:**

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 tests will examine if Piwek et al.’s (2014; Fig. 2) findings replicate. Specifically, if the (battle and toy) robots will be rated lower in human likeness than a mannequin (while not differing from one another); that the mannequin will be rated lower in human likeness than a skeleton or zombie (again these two latter characters will not differ from one another); and the skeleton and zombie will be rated as lower in human likeness than the two human figures (low- and high-resolution man).

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 accepted least, characters that are most (e.g., humans) and least like humans (e.g., robots) will be accepted 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 lower acceptability ratings for all characters relative to those obtained for characters 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 as a function of character. Naturally moving zombies are predicted to be significantly more acceptable than the static zombie.

### Design Plan

*In this section, you will be asked to describe the overall design of your study. Remember that this research plan is designed to register a single study, so if you have multiple experimental designs, please complete a separate preregistration.*

**\_\_\_\_\_\_**

1. *Study type (required; multiple choice, select one)*
   1. ***Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.***
   2. *Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, natural experiments, and regression discontinuity designs.*
   3. *Meta-Analysis - A systematic review of published studies.*
   4. *Other*

**STUDY TYPE:** Experiment: This is an experimental study and will be conducted online. Participants will be recruited via the Prolific website (<https://prolific.co/>). The original authors recruited a rather young cohort of participants (*Mean age* = 25, *SD* = 4.7). We will select a sample of participants with a similar age profile.

**\_\_\_\_\_\_**

1. Blinding (required; multiple choice, highlight applicable option(s))
   1. *Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.*
      1. *No blinding is involved in this study.*
      2. *For studies that involve human subjects, they will not know the treatment group to which they have been assigned.*
      3. *Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as “double blind”)*
      4. *Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.*
2. Is there any additional blinding in this study?

**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).

**\_\_\_\_\_\_**

1. Study design (required)
   1. *Describe your study design. Examples include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Describe any counterbalancing required. Typical study designs for observation studies include cohort, cross sectional, and case-control studies.*
   2. *Example: We have a between subjects design with 1 factor (sugar by mass) with 4 levels.*
   3. *More info: This question has a variety of possible answers. The key is for a researcher to be as detailed as is necessary given the specifics of their design. Be careful to determine if every parameter has been specified in the description of the study design. There may be some overlap between this question and the following questions. That is OK, as long as sufficient detail is given in one of the areas to provide all of the requested information. For example, if the study design describes a complete factorial, 2 X 3 design and the treatments and levels are specified previously, you do not have to repeat that information.*

**STUDY DESIGN:**

The Piwek et al. (2014) study contained two outcome measures (human likeness ratings & acceptability ratings).

With respect to the human likeness ratings, participants will rate seven different characters (i.e., a battle robot, toy robot, mannequin, skeleton, zombie, low-resolution man, and high-resolution man) in terms of their likeness to humans (i.e., when it comes to human likeness ratings there will be a single factor [*Character Type*] with seven levels).

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) x 5 (*Levels of Movement*) within-participants design will be employed, such that participants will rate all character types across all levels and types of motion in terms of how acceptable those characters are.

**\_\_\_\_\_\_**

1. Randomization
   1. *If you are doing a randomized study, how will you randomize, and at what level?*
   2. *Example: We will use block randomization, where each participant will be randomly assigned to one of the four equally sized, predetermined blocks. The random number list used to create these four blocks will be created using the web applications available at http://random.org.*
   3. *More info: Typical randomization techniques include: simple, block, stratified, and adaptive covariate randomization. If randomization is required for the study, the method should be specified here, not simply the source of random numbers.*

**RANDOMIZATION:** Participants will be randomly assigned to either the human likeness or the acceptability condition. Presentation of the stimuli will be randomly counterbalanced in each condition (with a confound of presentation of still images first being tested; see above).

**\_\_\_\_\_\_**

### 

### Sampling Plan

*In this section we’ll ask you to describe how you plan to collect samples, as well as the number of samples you plan to collect and your rationale for this decision. Please keep in mind that the data described in this section should be the actual data used for analysis, so if you are using a subset of a larger dataset, please describe the subset that will actually be used in your study.*

**\_\_\_\_\_\_**

1. Existing data (required; multiple choice, highlight one option)
   1. *Preregistration is designed to make clear the distinction between confirmatory tests, specified prior to seeing the data, and exploratory analyses conducted after observing the data. Therefore, creating a research plan in which existing data will be used presents unique challenges.* ***Please select the description that best describes your situation.*** *Please do not hesitate to contact us if you have questions about how to answer this question (*[*scorecoordinator@cos.io*](mailto:scorecoordinator@cos.io)*).*
      1. **Registration prior to creation of data: As of the date of submission of this research plan for pre registration, the data have not yet been collected, created, or realized.**

**\_\_\_\_\_\_**

1. Explanation of existing data (optional)
   1. *If you indicate that you will be using some data that already exist in this study, please describe the steps you have taken to assure that you are unaware of any patterns or summary statistics in the data. This may include an explanation of how access to the data has been limited, who has observed the data, or how you have avoided observing any analysis of the specific data you will use in your study.*
   2. ***Example****: An appropriate instance of using existing data would be collecting a sample size much larger than is required for the study, using a small portion of it to conduct exploratory analysis, and then registering one particular analysis that showed promising results. After registration, conduct the specified analysis on that part of the dataset that had not been investigated by the researcher up to that point.*
   3. ***More info****: An appropriate instance of using existing data would be collecting a sample size much larger than is required for the study, using a small portion of it to conduct exploratory analysis, and then registering one particular analysis that showed promising results. After registration, conduct the specified analysis on that part of the dataset that had not been investigated by the researcher up to that point.*

**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 experimental groups: human-likeness condition and acceptability condition.

**\_\_\_\_\_\_**

1. Data collection procedures (required)
   1. *Please describe the process by which you will collect your data. If you are using human subjects, this should include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool (e.g. inclusion and exclusion rules), and your study timeline. For studies that donÍt include human subjects, include information about how you will collect samples, duration of data gathering efforts, source or location of samples, or batch numbers you will use.*
   2. ***Example****: Participants will be recruited through advertisements at local pastry shops. Participants will be paid $10 for agreeing to participate (raised to $30 if our sample size is not reached within 15 days of beginning recruitment). Participants must be at least 18 years old and be able to eat the ingredients of the pastries.*
   3. ***More information****: The answer to this question requires a specific set of instructions so that another person could repeat the data collection procedures and recreate the study population. Alternatively, if the study population would be unable to be reproduced because it relies on a specific set of circumstances unlikely to be recreated (e.g., a community of people from a specific time and location), the criteria and methods for creating the group and the rationale for this unique set of subjects should be clear.*

**DATA COLLECTION PROCEDURES:**

This direct replication attempt will use the same set of stimuli and record the same set of responses as the original authors. Whereas the original study was carried out with Scottish university students in the lab, this replication attempt will use a more heterogeneous set of participants who will be recruited and complete the study online (via the Prolific Academic recruiting platform:<https://prolific.ac/>) in exchange for a monetary reward (at a rate of 6 British pounds per hour: the exact amount will depend on the time taken in each case).

Prior to the study participants will be informed that they will complete an experiment. Privacy and consent information will be provided which they will read and agree to prior to initiating the experiment (informed consent and study information provided to the participants can be found on the OSF page associated with this project). This information will indicate that their data will be stored publically, used by the lead author and others for research purposes, and stored separately from their demographic information. Participants will be informed that they can send the researchers an email if they have any questions about their rights and privacy and that they can terminate their participation in the study at any point.

**\_\_\_\_\_**

1. Sample size (required)
   1. *Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, interactions, or countries included. If the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, how many units are you collecting at each level of the analysis?*
   2. ***Example****: Our target sample size is 280 participants. We will attempt to recruit up to 320, assuming that not all will complete the total task.*
   3. ***More information****: For some studies, this will simply be the number of samples or the number of clusters. For others, this could be an expected range, minimum, or maximum number.*

**SAMPLE SIZE:**

**This study is comprised of two conditions - the acceptability and human likeness conditions**

**Acceptability condition**

- Stage 1 sample size: designed to detect 75% of the originally observed effect at 90% power. The original interaction effect reported in Piwek et al. (2014) (acceptability condition) was *Partial Eta Squared ƞ²p = 0.171*. We used a formula to transform Partial Eta Squared to Effect Size *f* = 0.45 (*f=√ ƞ²p/1- ƞ²p*). 75% of that effect is *f* = 0.3375. Given a 90% power, an error probability of 5%, and the original design, the sample size requires **95** participants.

- Stage 2 sample size: designed to detect 50% of the originally observed effect at 90% power. 50% of that effect is *f* = 0.225. Given a 90% power, an error probability of 5%, and the original design, the sample size requires **210** participants.

To achieve the target analytic sample, the target sample size for recruitment is 231 (required sample + 10% to allow for attrition or incomplete data). If necessary, data collection will continue in batches of 10 participants until the required sample size is collected.

**Likeness condition**

- Stage 1 sample size: designed to detect 75% of the originally observed effect at 90% power. The original main effect reported in Piwek et al. (2014) (likeness condition) was *Partial Eta Squared ƞ²p = 0.637*. We used a formula to transform Partial Eta Squared to Effect Size *f* = 1.324. (*f= √ ƞ²p/1- ƞ²p*). 75% of that effect is *f* = 0.993. Given a 90% power, an error probability of 5%, and the original design, the sample size requires **13** participants.

- Stage 2 sample size: designed to detect 50% of the originally observed effect at 90% power. 50% of that effect is *f* = 0.662. Given a 90% power, an error probability of 5%, and the original design, the sample size requires **26** participants.

To achieve the target analytic sample, the target sample size for recruitment is 30 (to allow for potential loss of data due to attrition and incomplete data). If necessary, data collection will continue in batches of 5 participants until the required sample size is collected.

**\_\_\_\_\_\_**

1. Sample size rationale
   1. *This could include a power analysis or an arbitrary constraint such as time, money, or personnel.*
   2. ***Example****: We used the software program G\*Power to conduct a power analysis. Our goal was to obtain .95 power to detect a medium effect size of .25 at the standard .05 alpha error probability.*
   3. ***More information****: This gives you an opportunity to specifically state how the sample size will be determined. A wide range of possible answers is acceptable; remember that transparency is more important than principled justifications. If you state any reason for a sample size upfront, it is better than stating no reason and leaving the reader to “fill in the blanks.” Acceptable rationales include: a power analysis, an arbitrary number of subjects, or a number based on time or monetary constraints.*

**SAMPLE SIZE RATIONALE:**

Power calculations were done in accordance with the guidelines of the [Social Sciences Replication Project (SSRP).](http://www.socialsciencesreplicationproject.com/) The first round of data collection achieves 90% power to detect 75% of the original effect size. The pooled sample, if necessary after testing the effect on the first round of data, achieves 90% power to detect 50% of the original effect size.

**\_\_\_\_\_\_**

1. Stopping rule
   1. *If your data collection procedures do not give you full control over your exact sample size, specify how you will decide when to terminate your data collection.*
   2. ***Example****: We will post participant sign-up slots by week on the preceding Friday night, with 20 spots posted per week. We will post 20 new slots each week if, on that Friday night, we are below 320 participants.*
   3. ***More information****: You may specify a stopping rule based on p-values only in the specific case of sequential analyses with pre-specified checkpoints, alphas levels, and stopping rules. Unacceptable rationales include stopping based on p-values if checkpoints and stopping rules are not specified. If you have control over your sample size, then including a stopping rule is not necessary, though it must be clear in this question or a previous question how an exact sample size is attained.*

**STOPPING RULE:**

We will terminate data collection whenever we have complete data that allows us to detect 50% of the originally observed effect at 90% power in both the acceptability (n = 210) and human-likeness conditions (n = 26).

**\_\_\_\_\_\_**

### 

### Variables

*In this section you can describe all variables (both manipulated and measured variables) that will later be used in your confirmatory analysis plan. In your analysis plan, you will have the opportunity to describe how each variable will be used. If you have variables which you are measuring for exploratory analyses, you are not required to list them, though you are permitted to do so.*

**\_\_\_\_\_\_**

1. Manipulated variables
   1. *Describe all variables you plan to manipulate and the levels or treatment arms of each variable. This is not applicable to any observational study.*
   2. ***Example:*** *We manipulated the percentage of sugar by mass added to brownies. The four levels of this categorical variable are: 15%, 20%, 25%, or 40% cane sugar by mass.*
   3. ***More information****: For any experimental manipulation, you should give a precise definition of each manipulated variable. This must include a precise description of the levels at which each variable will be set, or a specific definition for each categorical treatment. For example, “loud or quiet,” should instead give either a precise decibel level or a means of recreating each level. 'Presence/absence' or 'positive/negative' is an acceptable description if the variable is precisely described.*

**MANIPULATED VARIABLES:**

*Confirmatory Analyses*

Outcome Type: Two dependent variables will be assessed

· 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*).

**Character Type**: 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 against a gray background.

**Motion Type**: Participants will encounter each character (either statically and/or 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. 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.

*Exploratory Analyses:*

**\_\_\_\_\_\_**

1. Measured variables
   1. *Describe each variable that you will measure. This will include outcome measures, as well as any predictors or covariates that you will measure. You do not need to include any variables that you plan on collecting if they are not going to be included in the confirmatory analyses of this study.*
   2. ***Example****: The single outcome variable will be the perceived tastiness of the single brownie each participant will eat. We will measure this by asking participants ‘How much did you enjoy eating the brownie’ (on a scale of 1-7, 1 being ‘not at all’, 7 being ‘a great deal’) and ‘How good did the brownie taste’ (on a scale of 1-7, 1 being ‘very bad’, 7 being ‘very good’).*
   3. ***More information****: Observational studies and meta-analyses will include only measured variables. As with the previous questions, the answers here must be precise. For example, 'intelligence,' 'accuracy,' 'aggression,' and 'color' are too vague. Acceptable alternatives could be 'IQ as measured by Wechsler Adult Intelligence Scale' 'percent correct,' 'number of threat displays,' and 'percent reflectance at 400 nm.'*

**MEASURED VARIABLES:**

*Confirmatory Analyses:*

Human likeness and acceptability ratings will be measured (see above).

**\_\_\_\_\_\_**

1. Indices
   1. *If any measurements are going to be combined into an index (or even a mean), what measures will you use and how will they be combined? Include either a formula or a precise description of your method. If your are using a more complicated statistical method to combine measures (e.g. a factor analysis), you can note that here but describe the exact method in the analysis plan section.*
   2. ***Example****: We will take the mean of the two questions above to create a single measure of ‘brownie enjoyment.’*
   3. ***More information****: If you are using multiple pieces of data to construct a single variable, how will this occur? Both the data that are included and the formula or weights for each measure must be specified. Standard summary statistics, such as “means” do not require a formula, though more complicated indices require either the exact formula or, if it is an established index in the field, the index must be unambiguously defined. For example, “biodiversity index” is too broad, whereas “Shannon’s biodiversity index” is appropriate.*

**INDICES:** raw human likeness and acceptability ratings will be used.

**\_\_\_\_\_\_**

### 

### Analysis Plan

*You may describe one or more confirmatory analysis in this preregistration. Please remember that all analyses specified below must be reported in the final article, and any additional analyses must be noted as exploratory or hypothesis generating.*

*A confirmatory analysis plan must state up front which variables are predictors (independent) and which are the outcomes (dependent), otherwise it is an exploratory analysis. You are allowed to describe any exploratory work here, but a clear confirmatory analysis is required.*

**\_\_\_\_\_\_**

1. Statistical models (required)
   1. *What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, multiple regression, SEM, etc) and the specification of the model (this includes each variable that will be included as predictors, outcomes, or covariates). Please specify any interactions, subgroup analyses, pairwise or complex contrasts, or follow-up tests from omnibus tests. If you plan on using any positive controls, negative controls, or manipulation checks you may mention that here. Remember that any test not included here must be noted as an exploratory test in your final article.*
   2. ***Example****: We will use a one-way between subjects ANOVA to analyze our results. The manipulated, categorical independent variable is 'sugar' whereas the dependent variable is our taste index.*
   3. ***More information****: This is perhaps the most important and most complicated question within the preregistration. As with all of the other questions, the key is to provide a specific recipe for analyzing the collected data. Ask yourself: is enough detail provided to run the same analysis again with the information provided by the user? Be aware for instances where the statistical models appear specific, but actually leave openings for the precise test. See the following examples:* 
      * 1. *If someone specifies a 2x3 ANOVA with both factors within subjects, there is still flexibility with the various types of ANOVAs that could be run. Either a repeated measures ANOVA (RMANOVA) or a multivariate ANOVA (MANOVA) could be used for that design, which are two different tests.*
        2. *If you are going to perform a sequential analysis and check after 50, 100, and 150 samples, you must also specify the p-values you’ll test against at those three points.*

**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.

**\_\_\_\_\_\_**

1. Transformations (optional)
   1. *If you plan on transforming, centering, recoding the data, or will require a coding scheme for categorical variables, please describe that process.*
   2. ***Example****: The “Effect of sugar on brownie tastiness” does not require any additional transformations. However, if it were using a regression analysis and each level of sweet had been categorically described (e.g. not sweet, somewhat sweet, sweet, and very sweet), ‘sweet’ could be dummy coded with ‘not sweet’ as the reference category.*
   3. ***More information****: If any categorical predictors are included in a regression, indicate how those variables will be coded (e.g. dummy coding, summation coding, etc.) and what the reference category will be.*

**TRANSFORMATIONS:**

No transformations will be required.

**\_\_\_\_\_\_**

1. Inference criteria (optional)
   1. *What criteria will you use to make inferences? Please describe the information youÍll use (e.g. p-values, bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?*
   2. ***Example****: We will use the standard p<.05 criteria for determining if the ANOVA and the post hoc test suggest that the results are significantly different from those expected if the null hypothesis were correct. The post-hoc Tukey-Kramer test adjusts for multiple comparisons.*
   3. ***More information:*** *P-values, confidence intervals, and effect sizes are standard means for making an inference, and any level is acceptable, though some criteria must be specified in this or previous fields. Bayesian analyses should specify a Bayes factor or a credible interval. If you are selecting models, then how will you determine the relative quality of each? In regards to multiple comparisons, this is a question with few “wrong” answers. In other words, transparency is more important than any specific method of controlling the false discovery rate or false error rate. One may state an intention to report all tests conducted or one may conduct a specific correction procedure; either strategy is acceptable.*

**INFERENCE CRITERIA:**

For this replication we will use the original authors criteria (*p*-value < 0.05) for the confirmatory analyses.

**\_\_\_\_\_\_**

1. Data exclusion
   1. *How will you determine what data or samples, if any, to exclude from your analyses? How will outliers be handled? Will you use any awareness check?*
   2. ***Example****: No checks will be performed to determine eligibility for inclusion besides verification that each subject answered each of the three tastiness indices. Outliers will be included in the analysis.*
   3. ***More information****: Any rule for excluding a particular set of data is acceptable. One may describe rules for excluding a participant or for identifying outlier data.*

**DATA EXCLUSIONS:**

***Confirmatory Analyses***

To control for the possibility that static images shown before the dynamic characters could influence participants ratings, we will counterbalance the order of those images in the acceptability condition (i.e., half will encounter the images before and the other half will encounter them after the moving videos of the characters). Note that Piwek et al. (2014) did not find such an order effect in their original study (but did assess for it).

**\_\_\_\_\_\_**

1. Missing data
   1. *How will you deal with incomplete or missing data?*
   2. ***Example****: If a subject does not complete any of the three indices of tastiness, that subject will not be included in the analysis.*
   3. ***More information****: Any relevant explanation is acceptable. As a final reminder, remember that the final analysis must follow the specified plan, and deviations must be either strongly justified or included as a separate, exploratory analysis.*

**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. As noted in the sampling section, we will first check to ensure that we have recruited participants with complete data to detect 50% of the original effects with 90% power. If this is not the case we will continue data collection in-line with what we specified above.

**\_\_\_\_\_\_**

1. Exploratory analysis (optional)
   1. *If you plan to explore your data set to look for unexpected differences or relationships, you may describe those tests here. An exploratory test is any test where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time.*
   2. ***Example****: We expect that certain demographic traits may be related to taste preferences. Therefore, we will look for relationships between demographic variables (age, gender, income, and marital status) and the primary outcome measures of taste preferences.*

**EXPLORATORY ANALYSIS:**

**—**

### 

### Other

1. Deviations from the original study (required)
   1. Please describe any known deviations from the original study. Include discussion of whether and how these deviations might impact the results of this replication/reproduction attempt and its comparison to the original study.
2. Other (optional)
   1. If there is any additional information that you feel needs to be included in your preregistration, please enter it here. Literature cited, disclosures of any related work such as replications or work that uses the same data, or other context that will be helpful for future readers would be appropriate here.

**DEVIATIONS AND OTHER INFORMATION:**

**\_\_\_\_\_\_**