**Peer Review History and Communication**

**MS Title:** The Influence of Similarity and Mimicry on Decisions to Trust

**Author Names:** Alexa S. Clerke and Erin A. Heerey

**Submitted:** May 5, 2020

**Editor First Decision—Revise & Resubmit**

July 31, 2020

Dear Dr. Clerke,

Thank you for submitting your work to *Collabra: Psychology*. I have now received reviews of your manuscript from two experts, and I have carefully read the manuscript myself. I am very grateful to the reviewers for their detailed and constructive comments. Both reviewers saw merit in your work, and they made several suggestions for improving the manuscript. Below, I will summarize what I consider their key points and I will add a couple of my own suggestions.

First, both reviewers ask that you clarify your definition of social reciprocity. In particular, the reviewers recommend that you clarify (a) how your conceptualization of social reciprocity relates to the concepts of ‘behavioral mimicry’, ‘synchrony’, and ‘social cue reciprocity’; (b) where your conceptualization of social reciprocity falls on the automatic- versus deliberate-response continuum; and (c) whether social reciprocity as manipulated in your study can be construed as just one type of similarity. The reviewers make additional suggestions to improve the literature review and ask that you provide an overview of hypotheses, along with the rationale underlying them, in the introduction.

Second, both reviewers express concerns about whether your manipulations have worked as intended. Specifically, Reviewer 2 points to a potential confound in your design, and both reviewers provide advice to address their concerns via (a) using manipulation check questions to assess how successfully you manipulated similarity and social reciprocity; and (b) assessing whether the same participants who were influenced by the similarity manipulation were similarly influenced by the social reciprocity manipulation. The reviewers additionally ask that you clarify the presentation of results in your figures, and that you provide more descriptive information. I would like to echo this last request and ask that you provide means and SDs for your DVs (in each experimental condition) wherever possible. Further, I ask that you provide statements of how you determined sample size for each of your studies in the Methods sections of the manuscript.

If you think that you can attend to the issues raised above as well as in the reviews, I invite you to submit a revised version of your manuscript. With your revision, please provide point-by-point responses to all reviewers’ suggestions.

Best regards,

Dr. Catherine Molho Associate Editor *Collabra: Psychology*

**Reviewer 1**

**Open response questions**

Your name will not be revealed unless you wish to sign your review. If you chose to show your name, please type it in the text field below. If you prefer your comments to be anonymous, leave the field blank.

This paper tested the effects of similarity and social reciprocity on trust decisions across four experiments. It consistently found that similarity and social reciprocity independently promoted trust decisions. Despite the importance of this topic, there are some issues in the paper that remains to be addressed. Below are some of my comments that the authors may find useful in further improving the paper.

1. My major concern is about the novel contribution of this research and why this was done. The introduction, particularly the definition of social reciprocity, is not very clear. For example, the authors seem to treat “social cue reciprocity”, “behavioral mimicry”, and “social synchrony” as the same construct when describing previous research. While behavioral mimicry reflects automatic and unintentional imitation of others, social reciprocity seems to be intentional reaction to others’ behavior (e.g., returning a smile to one’s social partner after receiving one from the partner). The authors argued that current mimicry literature suffers from several limitations (e.g., does not reveal actual genuine mimicry), but this research had nothing to do with mimicry, nor did it address the issues in the mimicry research. Taken together, the Introduction needs to do a better job in clarifying the key research question, why this question is important, and the hypotheses to be tested.
2. Related to my first point, it is until Experiment 3 that the authors mentioned the three hypotheses, among which the third one is that social reciprocity is a stronger predictor of trust decisions than similarity. Please outline these hypotheses in the Introduction with a clear rationale. For example, it remains unclear why partner reciprocity is apparent before information about that partner’s attitudes and beliefs is known. Even if social reciprocity may be the first interpersonal cue that people send (after appearance similarity), this does not necessarily suggest that social reciprocity should be a more important predictor of choices than similarity.
3. To manipulate social reciprocity, participants first gave like/dislike feedback after viewing each avatar’s response to the similarity items, and then receive the partner avatar’s response that was either reciprocal (matching) or non-reciprocal. Yet, the reciprocal responses (e.g., like & like, dislike & dislike) may be different across different level of similarity. This makes me wonder how participants actually gave the like/dislike feedback to the similarity items. Did they show more likes when there was a high similarity? In addition, did you check whether similarity and social reciprocity were manipulated successfully? The OSF datasets that come with this paper seem to include items on perceived similarity. If these were the manipulation checks, please also present these results in the paper.
4. Figures 2 and 3 are not presented in a clear way. Please add the figure legends for social reciprocity (high vs. low) in Figure 2, and use similar formats in Figure 3.
5. Experiment 2: How many options did participants have when choosing the proportion to invest in their partner? What is the overall mean and SD of investment in Exp. 2a and 2b? Did participants complete 40 trials of trust game in total (10 rounds per condition)? This needs to be clarified. Here, “participants’ beliefs about avatar trustworthiness” (p. 11) is better presented as “participants’ trust decisions”.
6. Regarding the data analysis in Experiment 3, I wonder why the authors chose to model each participant’s choices in separate logistic models. Given that the data has a nested structure (the trust items are nested within each of six conditions, which are nested within participants), why not conduct a multi-level logistic model to test the effects of similarity and social reciprocity? This analysis would give the statistics about whether similarity and social reciprocity significantly affect trust decisions. Also, as similarity is coded into a categorical variable with three levels, this variable should be re-coded into two dummy variables prior to conducting the logistic regression. Also, please use b0, b1, b2… to represent unstandardized regression coefficients.
7. How was the proportion of choices calculated in Figure 4? Were they calculated based on the logistic regression results? Please clarify this in the figure description.

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  | ✔ |  |  |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

**Reviewer 2**

**Open response questions**

Your name will not be revealed unless you wish to sign your review. If you chose to show your name, please type it in the text field below. If you prefer your comments to be anonymous, leave the field blank.

Thank you sending me this manuscript. I read it with great interest. The purpose of the study was to compare the influence of similarity and social reciprocity on various trust tasks. I like the premise of the study. I am a proponent of comparing the influence of different social modulators. As during every day social interactions, multiple social modulators are at play. The results seem robust – three experiments with decent sample sizes are reported (including some preregistrations) and the results seems consistent across these experiments. However, I do have some concerns about the paper, most importantly the social reciprocity manipulation which I am not convinced is a measure of social reciprocity. Similarly, I have concerns that the similarity manipulation may have been confounded by the reciprocity manipulation. I feel the authors need to address this before publication. My comments are as follows:

**Introduction**

p. 4 Although there is some evidence to suggest that behavioural mimicry leads to prosocial consequences (e.g. increased trust), these effects, particularly when studied within virtual environments, are not as robust as previously thought. This is worth highlighting in the manuscript. For example, in a review paper Hale and Hamilton (2016a) conclude “links from mimicry to liking, trust and other positive outcomes appear to be fragile” <https://doi.org/10.1016/j.neubiorev.2016.02.006> See also Hale and Hamilton (2016b) who did not find effects of mimicry on trust <https://www.nature.com/articles/srep35295>

p.5-6 When writing about temporal synchrony and the limitations of studying naturalistic interactions, the paper of Hale et al. (2019) is worth mentioning. The aim of the paper was to look specifically at (potentially fast-paced and hard to detect) social behaviours in naturalistic interactions. They discovered a new “fast nodding” behaviour, which could be an important marker of social reciprocity.

**Introduction/Methods**

I am not convinced by the social reciprocity manipulation. After participants discovered whether the partner was similar or not (e.g. your partner also likes rock music), they had to rate how much they like this (e.g. do you like that you partner also likes rock music?). The measure of social reciprocity was then the extent to which the partner (i.e. avatar) matched the liking of this information (using an emoji-style rating scale). The reciprocal avatars matched on 80% of trials (same emoji rating) and the non-reciprocal on just 20%. Non-reciprocal feedback was always + or - one on the emoji scale.

Was participants’ feedback generally more positive (i.e. higher on the emoji-scale) for the similar avatars? If so, then regardless of whether the avatar was reciprocal or not, the feedback the participants received from the similar avatar was always going to be more positive. For example, a similar but non-reciprocal avatar may respond to a full smiling emoji with a half smiling emoji (-1). Although different from the participants’ feedback, this is still positive feedback. Whereas, a non-similar but reciprocal avatar may be more likely to match a sad emoji with an equally sad emoji. Thus, could the similarity effects not just be driven by the different feedback participants received from these avatars, i.e. this feedback was more positive (although no necessarily reciprocal)? How often did participants respond with negative emojis when the avatar was similar and how often did they response with positive emojis when the avatar was dissimilar? It would be good to know the distribution of these ratings to rule out this potential confound.

For the social reciprocity manipulation participants were essentially asked “how much do you like that you partner also likes (or doesn’t like) X (e.g. rock music)?” and then they saw the extent to which the avatar also liked that they liked (or didn’t like) X. Is this social reciprocity or just a further similarity manipulation? The first similarity question (do you like rock music?) measures similarity in attitudes/interests whereas the second question (do you like that they also like it?) measures similarity in feelings. Thus, I feel both manipulations could be manipulations of similarity rather than similarity and social reciprocity. To strengthen the claim that the manipulations are manipulating different things (rather than just similarity and then more similarity), it would be good to report whether there are correlations between the impact of similarity vs. the impact of the social reciprocity manipulation. Are participants who are more influenced by similarity also more influenced by social reciprocity manipulation? If they are not correlated, then this could add weight to the view that different constructs were manipulated successfully. Alternatively, the authors should make a stronger claim for the idea that similarity in feelings is a form of social reciprocity. I would like to see a clear definition of exactly what the authors mean by social reciprocity and how this is distinct from similarity. Although examples are given in the introduction (e.g. smiling, mimicry), social reciprocity is never defined.

**Results**

Figures. It would be nice to see the distribution of the scores rather than just the means e.g. with violin plots or something similar – bar charts are not very informative, see Weissgerber et al., (2015) <https://doi.org/10.1371/journal.pbio.1002128> I also found the labelling of the figures a bit confusing. For example, Fig2 and Fig3 are both showing the four conditions but with different colour schemes. And I think some labels may be missing from Fig2 (I assumed the darker bars are the reciprocal conditions?).

**Discussion**

p. 17 “knowledge of social reciprocity has temporal precedence over knowledge of attitude similarity in the real-world” - this might be true but in the current study participants first had knowledge of attitude/interest similarity and then reciprocity on each trial. So this is the opposite of how it operates in the real world. This should be acknowledged as a limitation.

Given the virtual nature of the interaction, the social reciprocity measure is very different to the social reciprocity measures used in previous “face to face” studies highlighted in the introduction, e.g. actual smiles (Heerey et al., 2013) or mimicry (Chartrand and Bargh, 1999). Reciprocal smiling and mimicry are largely seen as automatic (or predictive; Heerey et al., 2013), whereas the reciprocity measure used in the current study was an explicit rating of mood which is much less automatic. Although alluded to on p. 18, it would be nice to have a greater discussion comparing these automatic vs. deliberate forms of social reciprocity. I would recommend having a look at the cyber psychology literature e.g. Barak and Gluck-Ofri (2007) could be a good a starting point <https://doi.org/10.1089/cpb.2006.9938>.

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  | ✔ |  |  |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  | ✔ |  |  |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

**Author Response**

Jan 4, 2021

Dear Dr. Molho,

Thank you for the thorough and helpful feedback regarding our manuscript. An overarching request from both reviewers was to clarify our conceptualization of “social reciprocity” and its relation to “mimicry” and “synchrony”. As we discussed in the paper, social reciprocity is a form of mimicry that does not necessarily demand exact imitation but may also include complementary responses. The existing literature refers to social reciprocity as a form of mimicry and it is often labelled as such. In light of the confusion between the terms, and after careful consideration of our methodology, we’ve decided to change our terminology to “mimicry” instead of “social reciprocity”. We believe this change will be more cohesive with the current literature and make this paper more accessible to researchers who are interested in social processes. Furthermore, we believe our mimicry manipulation (previously known as our social reciprocity manipulation) falls neatly within the construct of mimicry. This change does not reflect a change in methods, results, or conclusion, and only acts to clarify our terminology within the existing literature.

You also requested that we include sample size considerations for our studies. We’ve revised the manuscript to include this information (page 9). Our responses to the rest of the reviewers’ suggestions are detailed below, with reviewer comments in black font and our responses below in blue font. Finally, we apologize for the length of time it has taken us to compile this response. The workload and childcare demands of the COVID pandemic has led to significantly reduced working hours in the context of substantially increased work-related demands for both authors.

Thank you again for your thoughtful comments on our manuscript. We hope that you find our response adequately address your concerns and those of the reviewers. We appreciate your time and consideration.

Sincerely,

Alexa Clerke

**Reviewer 1**

1. My major concern is about the novel contribution of this research and why this was done. The introduction, particularly the definition of social reciprocity, is not very clear. For example, the authors seem to treat “social cue reciprocity”, “behavioral mimicry”, and “social synchrony” as the same construct when describing previous research. While behavioral mimicry reflects automatic and unintentional imitation of others, social reciprocity seems to be intentional reaction to others’ behavior (e.g., returning a smile to one’s social partner after receiving one from the partner). The authors argued that current mimicry literature suffers from several limitations (e.g., does not reveal actual genuine mimicry), but this research had nothing to do with mimicry, nor did it address the issues in the mimicry research. Taken together, the Introduction needs to do a better job in clarifying the key research question, why this question is important, and the hypotheses to be tested.

After careful consideration, we have altered our terminology from “social reciprocity” to “mimicry” throughout the paper. Social reciprocity is a form of mimicry in which the returned cue may be a direct imitation of the eliciting cue but may also be a complementary cue. Though individuals can deliberately engage in reciprocal responding (i.e., deliberately mimicking or responding to a sender’s behaviour), social reciprocity can also be an automatic and unintentional process that naturally occurs at varying levels during interactions (Heerey & Crossley, 2013; Hale & Hamilton, 2019). Thus, social reciprocity is likely a specific type of behavioural mimicry, and when it is occurs automatically, one that the literature does not differentiate from unconscious mimicry. We therefore tend to agree with Reviewer 1 that these constructs are poorly differentiated in the literature.

Given the nature of the research design, mimicry is likely a more applicable term for what participants experienced. Specifically, in this research design, participants’ and avatars’ emoji feedback to one another is simultaneously viewed. That is, participants do not know which emoji an avatar will produce until after they have made their own selection. They believe that only after both players (themselves and the avatar) make their selections will the feedback be simultaneously revealed. Consistent with this design, our literature review now focuses on mimicry. We believe this change in terminology will minimize confusion amongst our readers and lead to a more cohesive paper. This change is reflected throughout the paper, but we specifically address the definitions of mimicry (versus social reciprocity) on page 5 of the manuscript.

Reviewer 1 also questions the novelty of this contribution, which is not a criterion for manuscript acceptance at Collabra. Nonetheless, we did clarify our research question, the importance of our work, and our hypotheses on pages 8-9 of the manuscript.

1. Related to my first point, it is until Experiment 3 that the authors mentioned the three hypotheses, among which the third one is that social reciprocity is a stronger predictor of trust decisions than similarity. Please outline these hypotheses in the Introduction with a clear rationale. For example, it remains unclear why partner reciprocity is apparent before information about that partner’s attitudes and beliefs is known. Even if social reciprocity may be the first interpersonal cue that people send (after appearance similarity), this does not necessarily suggest that social reciprocity should be a more important predictor of choices than similarity.

Nonverbal cue exchange begins at the very start of an initial social interaction – in which participants typically introduce themselves, spontaneously exchange smiles and take the first steps toward getting to know one another. Because these initial interactions begin with polite “chit chat”, people’s deeply held personal attitudes and beliefs are unlikely to be disclosed this early in the process of becoming acquainted. However, nonverbal mimicry/reciprocity are apparent and may enhance the feelings of interaction fluency and interpersonal closeness that lead people to make such exchanges later on in the relationship. This rationale has been clarified in the introduction. In addition, we have emphasized our initial hypotheses (that both similarity and mimicry/reciprocity would play an important and measurable role in the context of trust decisions) earlier in the manuscript (page 9).

1. To manipulate social reciprocity, participants first gave like/dislike feedback after viewing each avatar’s response to the similarity items, and then receive the partner avatar’s response that was either reciprocal (matching) or non-reciprocal. Yet, the reciprocal responses (e.g., like & like, dislike & dislike) may be different across different level of similarity. This makes me wonder how participants actually gave the like/dislike feedback to the similarity items. Did they show more likes when there was a high similarity? In addition, did you check whether similarity and social reciprocity were manipulated successfully? The OSF datasets that come with this paper seem to include items on perceived similarity. If these were the manipulation checks, please also present these results in the paper.

Yes. Participants do give more ‘like’ feedback to similar than to dissimilar answers. Although this translates to higher levels of ‘likes’ for similar avatars, who by definition give more feedback that is similar to a participant’s, the feedback is not different across avatars after accounting for similarity. That is, low similarity avatars’ similar answers are equally liked as high similarity avatars’ similar answers; dissimilar answers are liked slightly less than similar answers, but this did not differ across the avatar types. Importantly, the computer was programmed to ensure that the non-matching feedback participants received was as likely to be lower on the rating scale than a participant’s as it was to be higher. Thus, there is no discrepancy across the avatars when they provide non-matching feedback (i.e., the mean difference between participants’ and avatars’ responses is ~0). Although this means that there is more positive feedback given to avatars that are more similar, simply because their replies are more similar, participants are no less positive toward similar feedback across the set of avatars. To illustrate this, we now include supplementary plots of participants’ feedback when responding to similar (Figure S1a) and dissimilar statements (Figure S1b) in the supplementary materials.

We also included the manipulation check data within the manuscript (see Figures 2a, 3, and 4). In Experiment 1, this involved a single item (“similar to me”) and we expanded this in Experiments 2 and 3 to include a second item (“in sync with me”). Because Spearman-Brown coefficients (see Eisinga, Grotenhuis & Pelzer, 2013; now cited in the manuscript) suggested that these items measure the same construct, *similarity*, we averaged them in Experiments 2 and 3. Data/results are reported in the manuscript (see Results; Experiments 2 and 3).

1. Figures 2 and 3 are not presented in a clear way. Please add the figure legends for social reciprocity (high vs. low) in Figure 2, and use similar formats in Figure 3.

We have reproduced many of our figures as violin plots (per Reviewer 2’s suggestions) and enhanced the legends/captions. We believe these changes have increased both the descriptiveness and the readability of our figures.

1. Experiment 2: How many options did participants have when choosing the proportion to invest in their partner? What is the overall mean and SD of investment in Exp. 2a and 2b? Did participants complete 40 trials of trust game in total (10 rounds per condition)? This needs to be clarified. Here, “participants’ beliefs about avatar trustworthiness” (p. 11) is better presented as “participants’ trust decisions”.

Participants were able to choose any amount of their endowment to invest on each trial (0 to 10 points). They simply typed the number of points they wished to select and pressed the ENTER key. They played 40 trials of the trust game (10 with each avatar). This has been clarified in the manuscript (page 15). The means and standard deviations of participants’ investments with each avatar in Experiments 2a and 2b are now included within the results (Table 2). Finally, we have also reworded “participants’ beliefs about avatar trustworthiness” to “participants’ trust decisions” as suggested by Reviewer 1 (page 16).

1. Regarding the data analysis in Experiment 3, I wonder why the authors chose to model each participant’s choices in separate logistic models. Given that the data has a nested structure (the trust items are nested within each of six conditions, which are nested within participants), why not conduct a multi-level logistic model to test the effects of similarity and social reciprocity? This analysis would give the statistics about whether similarity and social reciprocity significantly affect trust decisions. Also, as similarity is coded into a categorical variable with three levels, this variable should be re-coded into two dummy variables prior to conducting the logistic regression. Also, please use b0, b1, b2… to represent unstandardized regression coefficients.

Reviewer 1 appears to have slightly misunderstood several things about the conditions and the analysis here. Because this is an unusual analysis in the field of psychology, we think that others may have similar misunderstandings and have therefore clarified several points.

We begin our explanation by reiterating the fact that in the basic design for this experiment similarity and mimicry each have two levels (low and high). This means that each of our four avatars (not six as Reviewer 1 seems to suggest) has both a value for similarity and a value for mimicry. As in a typical 2x2 ANOVA design, similarity and mimicry are fully crossed over the set of avatars. However, the tricky bit of this design is that participants experience each avatar meaning that a standard repeated-measures analysis of the effects of similarity and mimicry cannot disentangle the individual contributions of each of these factors to trust decisions (even though the sums of squares for the factors are orthogonal, per ANOVA definitions). Experiments 1 and 2 show that both factors are important. Our aim in Experiment 3 is to understand the individual contributions of similarity and mimicry to trust decisions.

To answer this question using the exact same manipulation parameters as in Exps 1 and 2, we adopt a design from experimental economics. In this design participants make choices between pairs of items that differ in value on several integrated factors, allowing a researcher to determine how those factors contribute to participants’ preferences. For example, participants choose between item A & item B, item B & item C, and finally item A & item C. Based on how they attribute their choices across all the possible pairings, we can learn how participants’ preferences relate to this decision space, and therefore can disentangle the differential contributions of each factor to a given participant’s decisions. Because we use four avatars, there are six possible avatar pairings. A multi-level analysis of the six pairings, as Reviewer 1 suggests, would tell us only about differences in the pairings – not about differences in the decision-space as a whole. That is, knowing how participants as a group respond to pairing 1 versus pairing 2 does not help to disentangle Similarity from Mimicry because the different pairings include avatars that may or may not differ on both factors.

Here, we ask participants to respond to each possible pairing 14 times (once for each trust item). This allows us to determine the proportion of these 14 trials, within a given pairing, in which they chose one avatar over another (represented in Fig 4b). An individual participant’s choices across all trials of all pairings (the full decision-space) become the dependent variable in our logistic analysis (i.e., these are the data the model predicts, using the independent variables of Similarity and Mimicry). For example, when choosing between the high-similarity/high-mimicry avatar and the high-similarity/low-mimicry avatar (which participants do 14 times), a participant whose decisions are strongly influenced by the presence of reciprocity will show a strong preference for one avatar over the other. If they like the mimicry element of the high-mimicry avatar’s behaviour, they will almost always select that avatar (if they dislike it, they will almost always choose the low-mimicry avatar). If that preference carries throughout the full set of choices, then the model for that participant will show a coefficient that is farther from zero (the sign on the coefficient denotes the direction of preference). A participant who is indifferent to this avatar feature would be as likely to select the high-mimicry avatar as the low-mimicry avatar (i.e., they would select each of these avatars ~7 times). If their indifference is consistent across the full choice set, they would receive a coefficient near 0, meaning that mimicry did not influence their choices to any significant degree.

Taken together, this analysis tells us about the subjective utility of similarity and mimicry for each individual participant by returning a coefficient for similarity and one for mimicry for each person (note that these individual-level data are never examined in the context of any hypothesis test). It is this full set of data (the modeled coefficients) that we subsequently analyze to determine the degree to which similarity and mimicry shape choice behaviour at the group level. This is not a dissimilar process to, for example, computing a composite score for a factor for each participant and then performing a subsequent NHST using the resulting composite scores as the dependent variable. We have clarified these points in the manuscript on page 19.

Finally, Reviewer 1 indicates that we mistakenly used “β” to represent the unstandardized regression coefficients in the Experiment 3 logistic regression equation. Note that the term “β” here refers to a model term, and not to model output. When referring to model output, (i.e., reporting the values of unstandardized coefficients from a model) it is correct that one should use the term “*b*”. Here however, these terms refer to model definitions, rather than output. As is typical in that case, we retain the definitional term “β” in the manuscript.

1. How was the proportion of choices calculated in Figure 4? Were they calculated based on the logistic regression results? Please clarify this in the figure description.

The proportion of choices data in Figure 4b (previously Figure 4) represents participants’ trust decisions in relation to relative differences between avatar characteristics within a given pair. For example, when choosing between avatars that were identical in similarity (i.e., both low or both high similarity) but different in mimicry, results show that participants selected the high versus low mimicry avatar about 63% of the time (Figure 4b, top bar). Therefore, this graph represents the average proportion of trials in which participants chose one avatar over another, depending on the relative differences between the avatars. These choice data are the dependent variable in the logistic regression (aggregated over all participants). Thus, this graph is not based on the logistic regression results, but rather shows the aggregated data patterns upon which the regression estimates were calculated. We have clarified this in the manuscript.

**Reviewer 2**

1. p. 4 Although there is some evidence to suggest that behavioural mimicry leads to prosocial consequences (e.g. increased trust), these effects, particularly when studied within virtual environments, are not as robust as previously thought. This is worth highlighting in the manuscript. For example, in a review paper Hale and Hamilton (2016a) conclude “links from mimicry to liking, trust and other positive outcomes appear to be fragile” https://doi.org/10.1016/j.neubiorev.2016.02.006 See also Hale and Hamilton (2016b) who did not find effects of mimicry on trust <https://www.nature.com/articles/srep35295>

The introduction to the manuscript has been substantially revised. This has included a shift in terminology, and clarification of research questions. We have cited Hale & Hamilton 2016a & 2016b, along with recent work from Tamir and Thornton that clarifies the differences between these online paradigms and suggests both a boundary condition and a social cognitive rationale for why mimicry should achieve its effects (pages 3-8).

1. p.5-6 When writing about temporal synchrony and the limitations of studying naturalistic interactions, the paper of Hale et al. (2019) is worth mentioning. The aim of the paper was to look specifically at (potentially fast-paced and hard to detect) social behaviours in naturalistic interactions. They discovered a new “fast nodding” behaviour, which could be an important marker of social reciprocity.

We have now cited Hale et al., (2019) in the introduction as this paper makes several important points about both the limitations of using confederates to study mimicry and provides evidence of the timescale of mimicry at low-frequency behaviours (e.g., the slow postural head movements interlocutors make). Although we think that their finding of hypo-coherence in high frequency movements (rapid head nodding when listeners agree with/understand speakers) is highly intriguing as a form of social reciprocity, our refocused introduction does not leave much room for discussion of this idea – though we do mention it in our discussion.

1. I am not convinced by the social reciprocity manipulation. After participants discovered whether the partner was similar or not (e.g. your partner also likes rock music), they had to rate how much they like this (e.g. do you like that you partner also likes rock music?). The measure of social reciprocity was then the extent to which the partner (i.e. avatar) matched the liking of this information (using an emoji-style rating scale). The reciprocal avatars matched on 80% of trials (same emoji rating) and the non-reciprocal on just 20%. Non-reciprocal feedback was always + or - one on the emoji scale.

Was participants’ feedback generally more positive (i.e. higher on the emoji-scale) for the similar avatars? If so, then regardless of whether the avatar was reciprocal or not, the feedback the participants received from the similar avatar was always going to be more positive. For example, a similar but non-reciprocal avatar may respond to a full smiling emoji with a half smiling emoji (-1). Although different from the participants’ feedback, this is still positive feedback. Whereas, a non-similar but reciprocal avatar may be more likely to match a sad emoji with an equally sad emoji. Thus, could the similarity effects not just be driven by the different feedback participants received from these avatars, i.e. this feedback was more positive (although no necessarily reciprocal)? How often did participants respond with negative emojis when the avatar was similar and how often did they response with positive emojis when the avatar was dissimilar? It would be good to know the distribution of these ratings to rule out this potential confound.

Reviewer 1 has a similar concern about the feedback. We have now added a supplementary section to the manuscript that includes violin plots showing these data, split by whether the response is similar/dissimilar to the participant’s own. Generally, these data show that participants, who all genuinely believed that they were interacting with other participants, gave neutral to mild positive feedback to the avatars on non-similar responses and mild positive to highly positive feedback to avatars when they were similar. This is consistent with politeness norms. Importantly, participants were no more likely to give negative feedback to any avatar than another, after accounting for answer similarity. Thus, the vast majority (over 75%) of emoji feedback is positive. However, Reviewer 2 is correct in noting that this might cause slightly less positive feedback to be received from the avatars whose responses are low in similarity. We note that sad emojis make up only about 12% of the avatar responses across all participants in all experiments (with very sad emojis comprising less than 3%) and the median participant received sad emoji feedback fewer than 7 times. We also note that sad emoji feedback is not always interpreted as “negative” (see Kralj Novak, et al., 2015; https://doi.org/10.1371/journal.pone.0144296). For example, one might give sad emoji feedback to express sympathy or disappointment on behalf of another, both of which are viewed positively by recipients. On that basis, the small number of sad emojis participants saw is unlikely to have significantly affected their responses to the avatars. However, we have added a note in the discussion to address this possible limitation.

1. For the social reciprocity manipulation participants were essentially asked “how much do you like that you partner also likes (or doesn’t like) X (e.g. rock music)?” and then they saw the extent to which the avatar also liked that they liked (or didn’t like) X. Is this social reciprocity or just a further similarity manipulation? The first similarity question (do you like rock music?) measures similarity in attitudes/interests whereas the second question (do you like that they also like it?) measures similarity in feelings. Thus, I feel both manipulations could be manipulations of similarity rather than similarity and social reciprocity. To strengthen the claim that the manipulations are manipulating different things (rather than just similarity and then more similarity), it would be good to report whether there are correlations between the impact of similarity vs. the impact of the social reciprocity manipulation. Are participants who are more influenced by similarity also more influenced by social reciprocity manipulation? If they are not correlated, then this could add weight to the view that different constructs were manipulated successfully. Alternatively, the authors should make a stronger claim for the idea that similarity in feelings is a form of social reciprocity. I would like to see a clear definition of exactly what the authors mean by social reciprocity and how this is distinct from similarity. Although examples are given in the introduction (e.g. smiling, mimicry), social reciprocity is never defined.

As Reviewer 2 suggests, mimicry may be perceived as a form of either behavioural similarity or similarity in feeling, which we now clarify in the introduction (page 5). We have now included the data from the manipulation check item(s) in the relevant results sections, including figures. Reviewer 2 has also suggested that we examine the relationship of similarity and mimicry in the design. Unfortunately, because similarity and mimicry are fully crossed variables in a 2x2 design, this is not really possible to do in Experiments 1 and 2. Experiment 3, which disentangles the relative contributions of both variables to trust decisions, does allow us to look more closely at these variables as it includes a true interaction term. We note that the interaction term is not significantly different from zero, suggesting that they achieve their effects in an additive, rather than an interactive manner. However, as Hale and Hamilton (2016) review, in designs in which participants have no opportunity to exchange information about themselves (e.g., the “describe a picture” task), the absence of information about the self may attenuate the effects of mimicry on perceptions of similarity. However, we suggest that designs in which objective similarity can be precisely manipulated, the effects of mimicry may be “added” to objective similarity data, thereby enhancing feelings of closeness and perceived similarity. This suggests an interesting direction for future research and is one that we now address in the discussion section.

1. Figures. It would be nice to see the distribution of the scores rather than just the means e.g. with violin plots or something similar – bar charts are not very informative, see Weissgerber et al., (2015) https://doi.org/10.1371/journal.pbio.1002128 I also found the labelling of the figures a bit confusing. For example, Fig2 and Fig3 are both showing the four conditions but with different colour schemes. And I think some labels may be missing from Fig2 (I assumed the darker bars are the reciprocal conditions?).

We agree with Reviewer 2 that other types of plots are more informative than simple bar plots. We have therefore converted most of the graphs to violin plots, with the individual data points included. We have also clarified the figure legends/captions to provide more information and make them easier to read.

1. p. 17 “knowledge of social reciprocity has temporal precedence over knowledge of attitude similarity in the real-world” - this might be true but in the current study participants first had knowledge of attitude/interest similarity and then reciprocity on each trial. So this is the opposite of how it operates in the real world. This should be acknowledged as a limitation.

We beg to differ slightly with Reviewer 2 on this point. In real world interactions, information (both verbal statements and nonverbal behaviour) are transmitted in an exchange format. That is, individuals reveal both types of information in a relatively simultaneous manner – cues and gestures co-occur with explicit verbal information. Here, although the infrastructure is not as smoothly coordinated as in life, the same process is happening. That is, participants gain explicit information that gives a sense of the degree of objective similarity with another person, together with nonverbal cues, prior to the next exchange of information/cues. Although this process is desynchronized relative to real-world interaction, there is nonetheless an exchange of both information types that unfolds over time. Given the popularity of social media in society today, and participants’ responses to debriefing, we suspect that participants had no difficulty in treating their experience in the study as similar to both conversation and life. Nonetheless, we have now explicitly addressed this idea in the manuscript (pages 22-23).

1. Given the virtual nature of the interaction, the social reciprocity measure is very different to the social reciprocity measures used in previous “face to face” studies highlighted in the introduction, e.g. actual smiles (Heerey et al., 2013) or mimicry (Chartrand and Bargh, 1999). Reciprocal smiling and mimicry are largely seen as automatic (or predictive; Heerey et al., 2013), whereas the reciprocity measure used in the current study was an explicit rating of mood which is much less automatic. Although alluded to on p. 18, it would be nice to have a greater discussion comparing these automatic vs. deliberate forms of social reciprocity. I would recommend having a look at the cyber psychology literature e.g. Barak and Gluck-Ofri (2007) could be a good a starting point <https://doi.org/10.1089/cpb.2006.9938>.

Emoji feedback is not necessarily viewed as an explicit rating of mood – although it can be. In this set of experiments, we purposefully did not suggest that participants use these emojis as mood indicators so that a much broader spectrum of interpretations might occur – an idea we now include in the discussion.

Although the experimental emoji-exchange process is substantially less automatic than live interaction, from the participant perspective, the emojis (of self and avatar) appear onscreen at the same time, making it feel like mimicry is occurring. We have now added discussion points addressing these issues (page 23).

**Editor Second Response—Revise & Resubmit**

Apr 12, 2021

Dear Dr. Clerke,

Thank you for submitting the revised version of your manuscript to Collabra: Psychology. I have received advice from one of the first round reviewers (Reviewer 2), and I have carefully read the revised manuscript myself. I appreciate your careful attention to the concerns the reviewers and I raised. I am happy to provisionally accept your manuscript for publication. However, prior to the final acceptance of your manuscript, I would like to ask you to address a few small things.

Minor revisions:

1. Please amend the discussion according to the reviewer’s suggestions. The reviewer asks that you acknowledge the fact that a considerable number of participants did not believe they were really interacting with other participants. Please describe the characteristics of excluded participants as requested (in the main manuscript or in the Supplementary Information).
2. Thank you for adding information regarding the criteria you used to determine the sample size of your studies. I would like to ask you to further clarify which analysis you specified when conducting your a priori power analyses. Were you interested in obtaining 95% power to detect main effects of both manipulations on trust? Did you focus on the main effect of only the mimicry manipulation? Please provide the relevant details in the Methods sections of your experiments.

I very much look forward to receiving your final revision and accepting it for publication in Collabra: Psychology.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Catherine Molho

# Reviewer 1

##### Open response questions

### **Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.**

I very much enjoyed reading this revised manuscript. Thank you for sending it to me.

The introduction is greatly improved - it ‘flows’ much better and sets up the research questions very nicely. Moreover, the decision to frame the paper more generally in terms of mimicry rather than social reciprocity was a good one. Thank you also for addressing my concerns about the emoji mimicry manipulaiton in the discussion and for replacing the bar plots with violin plots.

I feel the manuscript it worthy of publication and makes a solid contribution to the field. However, one minor thing I would like to see added to the limitations/future directions is the deception failure. In total, 44 participants (if my maths is correct!) across all the experiments did not believe they were interacting with other participants. Althought the numbers are relatively small compared to the overall sample, it would be good to know a bit more about these participants. Do they share any characteristics? For example, are they more likely to be of a particular age or gender? As deception is often used in social interaction research, especially in the virtual domain, it would be interesting to know a bit about these participants as this could help in the design of future studies.

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  |  | ✔ |

Author Response

Apr 26, 2021

Dear Dr. Molho,

We wish to thank you for the helpful feedback and the conditional acceptance of our manuscript.

To address the reviewer’s suggestion of providing more information about the participants who did not believe our manipulation, we included a set of comparison analyses in the supplemental materials (see page 35 of the manuscript). Here we have compared those participants included in the analysis and those who were excluded for deception failure on age, gender, and personality traits. We found no significant differences in age or personality traits but did find a significant difference in gender. Specifically, male participants were more likely to be excluded from the main analysis due to deception failure than were female participants. We referred to this supplementary analysis in the main manuscript on page 23 with the following phrasing: “Nonetheless, even though these avatar interactions were not as realistic as true interactions, all the participants we included in our analyses genuinely believed they had interacted with other people (*for a comparison of those included in the analysis and those excluded for deception failure on age, gender, and personality traits see Supplementary Materials*)”. Furthermore, we’ve added this dataset to the Open Science Framework.

To address your request of clarifying our a priori power analysis, we added the following language to page 9 of the manuscript: “A G\*Power (v3.1) analysis revealed that, assuming a small to moderate effect size *for the main effects of similarity and mimicry* of ηp2 =.04, 55 participants would achieve 95% power (at α=.05).” Notably, in Studies 2 and 3, we based our sample size estimates on effect sizes obtained in Study 1.

We’ve also corrected two minor errors in the manuscript. In Experiment 2b, we originally wrote that there were 17 males when it should have read 19 males. We’ve corrected this error on page 14 of the manuscript. In Experiment 3, we wrote that we discarded 9 participants from the analysis for deception failure and 5 for inattentive task performance. However, we mistakenly listed the participants who were discarded for both deception failure and inattentive performance in both figures. We’ve corrected this error on page 18 of the manuscript, and it now reads: “Of these 97, we discarded a total of 12 participants: seven due to deception failure, three due to inattentive task performance, and two for both deception failure and inattentive task performance.”

Thank you again for your thoughtful comments on our manuscript. We hope that this response adequately addresses your concerns and those of the reviewers. We appreciate your time and consideration.

Sincerely,

Alexa Clerke

**Editor Final Decision—Accept**

Apr 26, 2021

Dear Dr. Clerke,

Thank you very much for addressing the reviewer’s comments and my own requests in your revision. I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology.

As there are no further revisions to make, you do not have to complete any tasks at this point. Our managing editor will contact you in case there are any pre-production file related questions. You will have an opportunity to check the page proofs before we publish your article.

Thank you for choosing Collabra: Psychology as an outlet for your work. I think this is a valuable contribution to the literature and I look forward to seeing it published!

Best regards,

Catherine Molho