Reply to PCIRR S1 decision letter reviews: Newman et al. (2011) replication

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response to each item. Editor's and reviewers' comments are in bold with our reply underneath in normal script.

A track-changes comparison of the previous submission and the revised submission can be found on: https://draftable.com/compare/UUjjFltIwKNM (https://osf.io/74hjf)

A track-changes manuscript is provided with the file: PCIRR-S1-RNR-Newman-etal-2011-replication-main-manuscript-trackchanges.docx (https://osf.io/d4ryn)

Reply to Editor: Prof. Chris Chambers

> **I have now received four timely and constructive evaluations of your Stage 1 submission. The reviews approach your manuscript from a number of different angles and also include a non-specialist review (MV) which I find can often be helpful for identifying areas where the structure and clarity of the presentation can be improved.**
>
> **In reading the reviews you will notice some headline issues to address, including clarification of hypotheses, specific areas of terminology, rationale for the proposed extensions, conditions for excluding data, and analysis plans.**
>
> **Reviewers SS and MV also raise questions about the justification and value of the replication, which I want to address directly. SS in particular notes: "While replication is important for ensuring the accuracy and generalizability of findings, but I feel that not all studies are equally worthy of this effort. The relative lack of continued research in this specific area of celebrity contagion suggests that its theoretical and practical significance may not justify the resources required for replication."**
>
> **This is perfectly valid opinion that reviewers are welcome to comment on, but I want to note that that judgments about the importance (or novelty) of the research question, over and above the scientific validity for the question, fall outside the scope of the Stage 1 evaluation criteria at PCI RR. MV raises the same point in a slighty different way, noting: "I think the justification of the RR can be improved. Pages 10-11 merely say that the contagion effect is important in the literature. But this doesn't feel like a good reason to repeat the study."**
>
> **Again, PCI RR does not consider judgements of replication value when evaluating RRs so there is no requirement to address this issue. That said, I think the fact that two reviewers bring up this issue provides an opportunity for you to maximise the eventual impact of your work by highlighting the rationale for the replication and why it matters. MV makes a number of related points where the clarity of communication - while sufficient for a specialist audience - could be enhanced to improve general readability.**
>
> **I hope you find these reviews helpful and look forward to receiving your revised manscript in due course.**

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit.

Reply to Reviewer #1: Dr./Prof. Saleh Shuqair

Thank you for inviting me to review this report, titled "Revisiting celebrity contagion and the value of objects: Replication and extensions Registered Report of Newman et al. (2011)." While the authors have done a commendable job in detailing the research design, the intended experiment, the measures, and the target sample, I have concerns regarding the importance and necessity of this replication effort.

Firstly, the concept of celebrity social contagion is not an extensively established concept within the literature, which raises questions about the significance of replicating this particular study. Although Newman et al. (2011) has been cited 388 times and is considered impactful, there are relatively few subsequent papers that have continued in this specific stream of literature.

For instance, to best of my knowledge, while there are some papers on social contagion in the context of celebrities, such as:

Newman, G. E., & Dhar, R. (2014). Authenticity is contagious: Brand essence and the original source of production. Journal of Marketing Research, 51(3), 371–386.
Huang, J. Y., Ackerman, J. M., & Newman, G. E. (2017). Catching (up with) magical contagion: A review of contagion effects in consumer contexts. Journal of the Association for Consumer Research, 2(4), 430-443.

Other papers have addressed related but distinct concepts, such as contamination, which differs from celebrity contagion
White, K., Lin, L., Dahl, D. W., & Ritchie, R. J. (2016). When do consumers avoid imperfections? Superficial packaging damage as a contamination cue. Journal of Marketing Research, 53(1), 110-123.
Hazée, S., Van Vaerenbergh, Y., Delcourt, C., & Warlop, L. (2019). Sharing goods? Yuck, no! An investigation of consumers' contamination concerns about access-based services. Journal of Service Research, 22(3), 256-271.
Smith, R. K., Newman, G. E., & Dhar, R. (2016). Closer to the creator: Temporal contagion explains the preference for earlier serial numbers. Journal of Consumer Research, 42(5), 653-668.
Argo, J. J., Dahl, D. W., & Morales, A. C. (2008). Positive consumer contagion: Responses to attractive others in a retail context. Journal of

> **marketing research, 45(6), 690-701.**
>
> **Having said that, I believe this is not an established theory that warrants replication. While replication is important for ensuring the accuracy and generalizability of findings, but I feel that not all studies are equally worthy of this effort.**
>
> **The relative lack of continued research in this specific area of celebrity contagion suggests that its theoretical and practical significance may not justify the resources required for replication.**
>
> **Therefore, I recommend not proceeding with this replication study.**
>
> **Best of luck**

Thank you for sharing your views on this.

The editor already noted the following:

> [...] judgments about the importance (or novelty) of the research question, over and above the scientific validity for the question, fall outside the scope of the Stage 1 evaluation criteria at PCI RR.

Please also see our reply to Reviewer 4's comment #9.

Addressing the broad view of the value of replications, some of our views on this topic are shared in:

> Feldman, G. (2025). The value of replications goes beyond replicability and is associated with the value of the research it replicates: Commentary on Isager et al. (2024). *Meta Psychology*. https://doi.org/10.17605/OSF.IO/BTNUJ

Summarized with the following quote:

> Replications are still often misunderstood and undervalued. Despite ongoing discussions regarding the importance of replications, there has been little to no progress in making replications mainstream. One of the strongest indicators is the replication-nonreplication publication ratio, with recent estimates putting the ratio of replications of publications at around 0.2% in psychology (Clarke et al., 2023) and 0.54% in education (Cook et al., 2024), unfortunately closely resembling the rates reported a decade ago with 0.1% in psychology (Makel et al., 2012) and 0.13% in education (Makel et al., 2014).

> Accumulating meta-scientific evidence indicates replications are still an anecdote, at best, with most research not subject to independent direct replications. [...]
>
> At the moment, given how scarce replications are, almost any replication of published articles that were not yet replicated is of value. We need to conduct, submit, and publish more replications. [...] It would take years if not decades and a mindset shift to come anywhere close to a novel-replication publication ratio that makes sense for credible science.

We do not agree with the point about this area being neglected, and it would seem that atleast some of the other reviewers agree with the value in this replication, and we elaborate more on that below. But, even if one would argue that this research area is neglected, we offer the following anecdote about the value of replications from that same paper:

> **Replications go beyond replicability**
>
> [...] Finally, replications may help reignite interest in important research that was forgotten or pushed aside because of gatekeeping. Traditional imbalanced power structures in academia at times led to emphasizing a specific type or view of research and target topic, impacting academic discourse through biased publications and citations. Citation counts reflect these old imbalanced power structures. Replications can help bring attention to highly valuable neglected research and address these structural weaknesses. [...]

More on the value of replications and why they are needed, please also see the following talk:

- Feldman, G. (2024). "The value and importance of replications: Catching mistakes/fraud, clarifying theory, and testing measurement" talk given to the South Asian Journal Club. https://doi.org/10.17605/OSF.IO/2RXMY

As for the argument that there is a "relative lack of continued research", we will just briefly note that while the target article looked at celebrity contagion, it addresses that as an example embedded in the wider literature of "contagion". Contagion is still a very active line of research in marketing and psychology. Fairly recent well cited examples:

> Meng, L. M., Duan, S., Zhao, Y., Lü, K., & Chen, S. (2021). The impact of online celebrity in livestreaming E-commerce on purchase intention from the perspective of emotional contagion. *Journal of Retailing and Consumer Services*, *63*, 102733. [257 citations]
>
> Herrando, C., & Constantinides, E. (2021). Emotional contagion: a brief overview and future directions. *Frontiers in psychology*, *12*, 712606. [214 citations]

Reply to Reviewer #2: Dr./Prof. Lachlan Deer

> **This is a stage 1 replication and extension of Studies 1 and 2 from Newman et al (2011). The original paper's goal was to understand the value of celebrity possessions and the mechanisms driving their value. The original study posits three mechanisms: (1) association (sentimental value), (2) market forces (others having a higher willingness to pay at a future date) and (3) contagion (the transferal of immaterial quality/essence through physical contact) finding support for the latter. The authors plan to fully replicate the results from Study 1 and 2 that provide evidence for the contagion effect, with minor deviations, and extend across two dimensions: Physical proximity without physical contact; and Temporal proximity.**
>
> **<u>General Comments</u>**
>
> **The authors are undertaking a very interesting replication project. Overall, I am impressed by the level of planning and detail. All deviations from the original paper are carefully outlined and the manipulation checks, justification for sampling and power analysis are clear for the reader. My comments, intended to improve the manuscript, aim to strengthen the part of the study that provides extensions beyond the original Newman et al (2011) design. Each of these extensions' designs appear valid but grounding both hypotheses more firmly in literature and providing a clearer explanation and motivation of extension 1 would strengthen the manuscript.**

Thank you for the positive and supportive opening note and the constructive feedback.

> **Major Comments**
>
> **.1. As currently written, the extensions proposed in the study are geared around the existence of the contagion effect as the driver of celebrity object's value. This means that if one doesn't find support for this effect these extensions have lower value. However, each extension can have meaning/importance independent of which of the three mechanisms drives the result. Thinking through this in the manuscript would show the greater value of the extensions proposed, strengthen the validity of the extensions to be mechanism independent, and, in turn, strengthen the manuscript**

Thank you for this feedback.

Based on the cumulative reviewer feedback, we decided to remove one of the extensions that dealt with time, and to focus on the extension of the desire to have non-physical contact. We also realized that the way we refer to desire to have physical and non-physical contact as "contagion" is confusing. We originally did this to try and stay as close as possible to the target article's jargon, but in order to clarify the extension, we felt that it would be best to highlight the contrast between the Experiment 1 replication items of "Desire to have physical contact" (we added the clarification of ["contagion" in target article"]) versus the extension item of "Desire to have non-physical contact".

We also revised to make it clearer that in Experiment 1 the main dependent variable is the item valuation, and that the desire for both physical (replication) and non-physical (extension) contact is a dependent variable that serves more as a possible mechanism. We agree and see value in this extension item regardless of the findings regarding the replication item.

We appreciate the note about discussing the implications of our extension regardless of whether the replication item is supported or not. We therefore added a Stage 2 planned discussion of this point in our Discussion section:

> [By recommendation from reviewer Dr. Lachlan Deer: Discuss the findings and implications of the findings of our "desire for non-physical contact" extension independent of the results for the "desire for physical contact" (referred to as "contagion" in the target article) replication item.]

> **.2. Extension 1 as explained on pages 15 and 16 is unclear to me as a reader. The ideas become clearer in the Manipulations later in the paper section but are still too vague for me to clearly evaluate. I'd urge the authors to rephrase this section and be clearer on what the extension is, and further ground the hypothesis in more literature.**

Thank you for the feedback.

We revised and simplified the background and rationales behind the extension:

> The theoretical model that Newman et al. (2011) uses to define the contagion effect specifies that physical contact is a necessary prerequisite for the effect to occur — the "essence" of a person is imbued into an object through physically touching the item. This model is reflected in the measures for contagion in the original study; they all involve physically touching a person or object (e.g. "How much would you want to give this person a hug or shake their hand?")
>
> However, more recent studies have posited that physical contact is not necessary for the contagion effect to occur (Huang et al., 2017; Morales et al., 2018). They posit that,

> among other vectors of "contamination", just being close to an object is enough for a person to "contaminate" it. For example, Kim and Kim (2011) found that an object can become "infected" just by being in the general vicinity of a source of "contamination", without the source of "contamination" ever having to come into actual physical contact with the object itself. Furthermore, Stavrova et al. (2016) found that contagion can affect objects that do not even physically exist: even a piece of music can be "contaminated" by the intentions of the person who made them.
>
> Therefore, in order to study whether adding a dimension of contagion that does not involve physical contact would still cause an effect in the context of the manipulations of this study, we added a measure of contagion that features contagion but does not feature physical contact.

> **.3. The hypothesis behind Extension 1 in Table 3 is unclear as currently written. Please adopt a more precise statement**

We amended our description of the extension hypothesis and added at the end of the "Extension: Desire to have non-physical contact" section:

> We meant this as an exploratory extension, yet our baseline was to compare physical to non-physical contact, and so our expectations mirrored that of the findings for physical contact.

We also revamped Table 1 to more explicitly write all hypotheses, including the hypothesis for the desire for physical contact (H2), and for the extension we wrote we expected it to mirror the same findings as for H2:

> Exploratory extension: Our expectations mirror H2. Interaction: Positive: celebrity higher Mixed/negative: celebrity lower

> **.4. The hypothesis behind extension 2 currently argues that for positive morality celebrities' contagion is positive when there's less time since contagion occurs and the opposite for negative contagion. These arguments are plausible, under the assumption that a "contagion effect" decays over time. However, the literature cited to support this is indirectly related. Reconsider whether the directionality of this hypothesis is needed.**

> **.4.1. Remark: This hypothesis as written is also contingent on a contagion effect being the main finding in the replication. As suggested above, one might reconsider this and thus the directionality on the hypothesis too**

Thank you. We appreciate the feedback.

We decided to simplify things and removed the second extension.

Minor Comments

The comments below are designed to help the reader better understand the ideas being presented:

.5. What is a contagion belief? (page 8/9)

This was supposed to mean “contagion, the type of belief”, but as it does seem redundant we revised it to read just “contagion” instead.

.6. The explanation of contagion effects at the beginning of "contagion theory" on p8 seem inconsistent at first glance with the examples discussed in the paragraph "the importance of the contagion effect..." these examples don't speak directly to contagion as you defined it. I’d encourage the authors to rework this paragraph or potentially remove it as it isn’t essential to the manuscript.

Thank you for the feedback. We removed the paragraph beginning with “The importance of the contagion effect…” to hopefully enhance the flow of this section.

.7. Who are these living people who aren't celebrities that participants list? It might be useful to provide the reader with an overview of what’s reported when the data comes in.

Thank you. We added a planned description of what non-celebrities were elicited. Along the same line of thought, we think some sort of overview of what celebrities were elicited would also be of help. We therefore added a section in the results section.

Description of the elicited persons

[By recommendation from reviewer Dr. Lachlan Deer: We aim to broadly describe the elicited celebrity and non-celebrity figures.]

.8. These comments are further thoughts that came to mind when reading the manuscript. The authors need not act on them in this study, but may find them useful at a later date:

.8.1. This study focuses on valence around moral values. There are interesting dimensions beyond morality that could be explored.

.8.2 A study that attempts to better separate association and contagion (as the authors mention, these are somewhat intertwined) can offer a cleaner insight on the role of contagion vs association. What the authors currently do sticks to the approach of Newman et al (2011), which is fine for what they are trying to do in this paper but as I read through both, I pondered the value of an experiment design that better separates.

.8.3. When reading this study and Newman et al (2011), I wondered how much a participant self-reporting the celebrities matters as opposed to a pre-specified list that could be crowd sourced or taken from a different source. What might be interesting there, is (although likely non-randomized) how different levels of exposure to a celebrity or different views about their morality interplay with the kind of effects found in this type of study.

Thank you for these suggestions. We feel that this is already a very ambitious project, and so we will keep those in mind for a discussion as future directions in Stage 2. We added the following to our Discussion section:

> [By recommendation from reviewer Lachlan Deer: Potentially discuss as limitations and future directions - 1) valence beyond morality, 2) association beyond physical connection [beyond the extension], and 3) advantages and disadvantages of the elicitation procedure as compared to a fixed list of celebrities.]

.9. Based on the evaluation criteria provided:

.9.1. The scientific validity of the research question(s).

Sufficient. More work needed on extension 1 as detailed above

.9.2. The logic, rationale, and plausibility of the proposed hypotheses, as applicable.

Both hypotheses need better grounding in literature. Extension 1 needs to be clearer for me to provide an accurate evaluation.

.9.3. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable).

Well executed

.9.4. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses.

Well executed

.9.5. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s).

Extensions are currently explained conditional on finding a contagion effect. Please generalize to be mechanism independent.

Thank you very much for the detailed review.

Reply to Reviewer #3: Dr./Prof. Susanne Adler

> **Thanks a lot for the opportunity to review this stage 1 submission. Overall, I think that the article is well-chosen and that the theoretical background as well as the justification for choosing this article is reasonable. I also consider the methodology of the replication to be thorough, wellrounded, and close to the original studies. I have however noticed some aspects that may help advance the stage 1 manuscript.**

Thank you for the positive and supportive opening note and the constructive feedback.

> **.1. First, I have a few comments concerning the background sections, specifically regarding the alignment of your manuscript vis-à-vis the original paper.**
>
> **.1.1. The beginning of the paper and the theoretical setup are a bit hard to follow. Specifically, it was not always clear which elements in the original study you focused on in your manuscript since the "Main hypotheses and key findings in the target article" section is rather short and mostly lists the constructs. Providing more information on the original study vis-à-vis the objectives of the replication could help clarify this section. For example, you mention some analyses (e.g., mediations, moderations) that you do not plan to test. For the contagion sensitivity moderation specifically (H2a and H2b in the original manuscript), you only later explain why you will not test it. The hierarchical regression that is mentioned in the original paper, for instance, is not mentioned at all. Maybe you can specifically expand this section to be clearer about which constructs, relationships, hypotheses, and tests you consider for our replication, which you consider for exploratory purposes, and which you discard from your manuscript.**

Thank you for raising this point.

The target article had many hypotheses and analyses, and we realized that we may have fallen short of clarifying the target article for the readers so that they could follow our identified key hypotheses and analyses.

We first revamped Table 1 to include all the needed information and bolded the identified key hypothesis and analysis. We then rewrote the "Main hypotheses and analyses in the target article for replication" section to explain and clarify which analyses are within the scope of our replication, and to state which studies and hypotheses we are replicating and what we identified as the key findings.

> **.1.2. As a related, but minor point: In Table 1 you specify the main hypotheses by Newman et al. (2011). Experiment 1 however does not directly specify hypotheses in the original article while Experiment 2 overall lists four hypotheses.**

Thank you for the feedback. We aimed to reconstruct the hypotheses from the target article.

We decided to overhaul our Table 1 to include all the effects tested and findings reported in the target article, so that it is clearer what the target article tested and found. We then redid our power analysis based on our identified key hypotheses (which had no impact on our target sample size, given that it was based on Simonsohn, 2015) and added a sensitivity analysis.

> **.1.3. There are also instances in which the terminology is not consistent. For example, on p. 13 for Experiment 2, you write about "item valuation" which is "purchase intention" in the original manuscript; or in Table 2, you write market value instead of demand. Can you clarify the construct names in your manuscript vis-à-vis the original paper?**

One of the challenges with the target article was that the terms used for the variables did not always match what they actually measured and with naming that was confusing.

We decided to address this comment in a more comprehensive way, and to change all the variable names to what we thought better captured what those measures were meant to measure. We then applied the same naming for those measures throughout the manuscript.

We now clarify this explicitly in the section "Main hypotheses and analyses in the target article for replication" as a footnote of "We summarized all hypotheses in Table 1, with our identified core hypotheses and analyses targeted for replication in bold":

> We note that we thought the terminology used by the target article did not represent what was manipulated or measured well and lacked consistency, and so we decided to change the terminology to more accurately capture what was done. In Experiment 1, the target article referred to measures of "contagion" and "market value", which we changed to "desire for physical contact" and "market demand". In Experiment 2, they had a manipulation of the domains referred to as "contagion" versus "market value", which we changed to "physical contact" and "market demand", and dependent measures of "purchase intentions" and "pleasure from wearing", which we changed to "willingness to purchase" and "pleasantness of wearing". We noted those changes in the design tables Tables 3 and 4.

As noted, we also included those clarifications in design tables Tables 3 and 4 with labels stating "in the target article".

> **.2. Second, concerning the extensions: I think that the hypotheses overall make sense since contagion should depend on psychological distance but I noticed a few things about their operationalization.**
>
> **.2.1. Concerning the physical contact extension: The item "How much would you like to meet this person?" does not specify whether there will be physical contact or not. Since many people do shake hands or hug whenever they meet in person, the item wording could be ambiguous since it does not directly exclude physical contact and therefore might miss its objective. Instead, maybe specify that the meeting would be a video meeting or a phone call to ensure that the respondents would be unambiguously aware of the missing physical contact.**

Thank you for this valuable suggestion. We agree.

We changed the question for the extension into:

> "How much would you like to meet this person through a video call?"

And renamed it to "desire for non-physical contact", instead of this being about physical proximity, and updated our rationale for the extension accordingly in section "Extension: Desire to have non-physical contact".

> **.2.2. Concerning the temporal proximity extension: Much of the theoretical background for this section rather refers to physical proximity and not just to temporal proximity. There is also a small mismatch in your argument. On p. 16 you write that "implying a weaker impact of time on negative contagion" which suggests that temporal proximity may not be as effective for negatively (vs. positively) perceived individuals. This is however not reflected in the hypothesis in Table 3.**
>
> **.2.3. The manipulation of temporal proximity is also quite obvious. Since the assessment of both close and distance temporal contact will be on the same page, this position could trigger carryover effects from one measure to the other or at least prompt participants to compare their ratings. I know that you may not like to introduce another between-subject factor here but it might already help to separate the measure on two different pages.**

Thank you for the feedback. Much appreciated.

We decided to remove the temporal proximity extension to keep things a lot simpler and focus on the replication. This is already a very complex replication.

We now only include the "desire to have non-physical contact" to try and mirror the contagion hypothesis "desire to have physical contact" mechanism.

> **.3. Third, concerning the analyses and reporting:**
>
> **.3.1. You chose to apply two approaches to determine the required sample size but discarded the analysis that resulted in n = 736 participants. Is there a specific reason to leave the discarded analysis in the manuscript? If you decide to keep the power analysis in the manuscript, can you elaborate a bit more on the methods? Currently, you only mention the R packages (which is good to ensure that the developers get the appropriate credit), but the rationale for the analysis as well as why you decided on a target sample of n = 1200 remains implicit.**

Thank you. This has helped us realize we can do much better in how we did and explained our power analysis, and so have worked to overhaul the entire approach and our description.

We would like to note that in the previous draft we did not discard analyses but instead adopted the analysis that required the larger sample size of the two methods we applied - power analyses and the small telescopes Simonsohn (2015) we used to calculate a minimum sample size.

We expanded on our description of what we did, added sensitivity analyses, and modified our "Power and sensitivity analyses" to the following:

> TThe target article's studies had many hypotheses and many dependent variables, and conducted many analyses. It was not always clear which of the analyses were considered to be the key analyses of interest, and so we flagged what we considered to be the main hypotheses and analyses. In Study 1, for the key dependent variable of item valuation they found support for the fame and valence main effects (our H1a and H1b), and argued that the main effect for valence is related to a main effect for valence in the desire physical contact (our H2b) and that the main effect of valence is related to the main effect of fame (H3a). In Study 2, the key dependent variable was the willingness to purchase a celebrity item, with the main analysis for the purpose of the study was examining the impact of the manipulation of physical contact showing a valence by level of physical contact interaction (our H9c). Our power analysis for these five key hypotheses was that the smallest effect of those required 344 participants in order to detect.
>
> However, to account for the likelihood that the target article's effects are an overestimation, we used the small telescopes approach as described in Simonsohn (2015) to aim for enough power to detect effects much weaker than those reported by the original study, by using a general rule of thumb of multiplying the target article's original samples by 2.5 to obtain the required replication sample size. The largest sample size in the original Experiments 1 and 2 was 455 for Experiment 2 (which is an overestimation by around two times, given that the key hypothesis for Experiment 2 only tested on half of the sample looking at physical contact). We therefore multiplied 455 by 2.5 to result in 1137.5, which we rounded up to 1200. We felt that targeting 1200 rather than 344 would give the target article much better chances for a successful replication, if the effect indeed exists.
>
> We ran a sensitivity analysis using GPower and found that a sample of 1200 would allow us to detect (95% power; alpha of 5%) a one-way main effect of $f = 0.11$ with three conditions (for H1a and H2b), a one-way main effect of $f = 0.10$ with two conditions (for H1b and H3a), an interaction effect of $f = 0.10$ with 4 conditions in a 2 by 2 design (H5c), equivalent to $\eta 2$ lower than 0.01, considered tiny effects, far smaller than the effects detected by the target article.

> **.3.2. In the manuscript, you specify that you do not plan to exclude any participants which strikes me as odd since excluding inattentive participants may increase data quality. Further, you are collecting multiple variables that could serve as a quality check. Do you, for example, plan to check variables such as the time each participant took to complete the questionnaire? Do you plan to do consistency checks (e.g., if participants indicate that a person is "Not at all famous" in the celebrity condition)? How do you proceed if participants indicate that they were not filling out this questionnaire seriously?**

We include a detailed description of our quality measures, based on our extensive experience in running similar replications. In all of our replications so far with Prolific, their online participants have shown to be very attentive and serious.

Regarding the recruitment of the participants on Prolific, we note:

> We targeted the general US American population sample using Prolific's filters: we restricted the location to the US using "standard sample", and set the participant filters to "Nationality: United States", "Country of birth: United States", "Place of most time spent before turning 18: United States", "Minimum Approval Rate: 95, Maximum Approval Rate: 100", "Minimum Submissions: 100, Maximum Submissions: 10000".

We also included our description of how we plan to address attentiveness even before the study begins:

> Participants first indicated their consent with four questions confirming their eligibility, understanding, and agreement with the study terms, to which they needed to answer "yes" to proceed to the rest of the study. Three of the four questions also served as attention checks, with the options order being rotated (yes, no, not sure) indicating that the participant confirmed that they would: (1) pay close attention to details and answer subsequent questions carefully, (2) agree to having to answer attention and comprehension checks, and (3) that they are a native English speaker born, raised, and currently located in the US. Failing any of the three attention questions meant that the participants did not indicate consent and therefore could not continue to the rest of the study. These were followed by a question that requested participants to copy and paste a statement indicating that they understood and agreed to the terms of the study, and they were allowed to try that as many times as needed to get it right. The two experiments were then presented in a random order.

Beyond that - no, we do not include additional checks and measures because these include subjective measures, flexibility, and forking paths of analysis subsequently lead to debates

regarding which of the analyses are the most reflective and issues of power. Given the very large sample, we consider those to be random noise.

Consider what the definition of "inattentive" means and whether there is an agreed-upon objective criteria for attentiveness. We are unaware of such a criteria. You suggested one such subjective arbitrary measure - time. If we take that as an example, like a decision that the minimum time should be set to 2 minutes. One might ask - why is the participant who completed the study in 1:59:00 minutes any better or more attentive than the participant who completed it within 2:00:01 minutes? Multiple criteria for exclusion makes that even more arbitrary.

In addition, when we do exclusions we run into other issues such as possibly much smaller samples with lower power, and multiple possible interpretations of the result. Consider, for example, a scenario where the exclusion of 20% of the participants led to a smaller sample with lower power where the replication detected a signal with the full sample but did not detect a signal with the exclusion strict criteria. Did the replication fail? One might also raise the opposite scenario where the full sample with no exclusions does not detect a signal, but the sample after 20% exclusions of those under 2 minutes. Did the replication succeed? In the first case, it could be a simple issue of power. In the second case, it would depend on what that threshold does. What if authors or auditors then run additional analyses showing that when instead of excluding those under 2:00:00 you exclude those under 2:01:00 or those under 1:59:00 it's a failed replication, what then? One possible way to address all of that is a multi-verse analysis, but that also has its weaknesses and requires very large samples. Is the 2:00:00 quality? always? according to whom?

We would very much prefer to avoid such debates regarding subjective criteria. Therefore, once participants have met the Prolific quality criteria and passed the attention checks in the consent form we include all participants.

> **.3.3. I further really like that you provide the R markdown files with all the code and results. Going through the files, however, I think that there should be more comments or explanations about the analyses. For example, concerning the reliability analyses in the Extensions, it is unclear for which construct the reliability is calculated unless one checks the code directly.**

Thank you for this suggestion. We moved the comments into the plaintext section, and revised them for improved readability.

.4. I have some minor points which I noticed when reading the manuscript:

.4.1. In the hypotheses, you write about "willingness to contact celebrities and their possessions." "Willingness to contact" sounds a bit like initiating communication via mail, Messenger apps, etc., and not like contagion or actual physical contact. Can you rephrase the wording here?

Thank you. We revised our references to all variables to better capture what they stand for. In this case we revised to "Desire to have physical contact". We also amended our Table 1 and references to the hypotheses.

.4.2. On the bottom of p. 11, you forgot to mention "liking" as a DV of the original study. You also excluded liking from Table 2. Is there a specific reason for it?

This was because the raw statistics and mean comparison tests for "liking" were not reported in the original study; all that was reported was a single graph.

We added liking to the tables (Table 1), results reporting, and our description of the original article. Because the target article did not report those stats, it is reported at "N/A".

.4.3. Is there a reason not to include the interaction fame x valence from Newman's Experiment 1 in Table 2?

Previously Table 2, now Table 1, did include the fame x valence interaction.

In this revision, we revamped Table 1 to try and make things even clearer and easier to follow to include all the effects.

.4.4. Table 5 (Experiment 2) could be a bit misleading on the experimental design. Since IV2 (physical contact) and IV3 (demand) are placed below each other it seems as if each participant receives information for both IVs.

We added "[2x2x2 between-subject design]" to the table title, and a note:

> "Note. Design is 2 (valence: positive vs. negative) x 2 (manipulation: physical contact vs. market demand) x 2 (direction: highlighted vs. decrease) between-subjects design"

We now also clarify the design at the beginning of the sections "Experiment 1" and "Experiment 2" under Method->"Design and procedure".

.4.5. For Table 2, could you please add a brief note that explains the multiplier for the required sample size?

The supplementary materials had a few paragraphs which explained how the sample size calculations work ("As some analyses were conducted with a certain subset of the total participant pool…") in the section "Power analysis of the target article effects to assess required sample for replication".

We also added to the Table note:

> Multiplier is used when the analysis was conducted on a subsample (*N*/multiplier).

.4.6. On p. 31, you indicate different effect size estimates for the original effect compared to Table 1.

Thank you for spotting the oversight. We revamped Table 1 and removed the references to effect sizes from the "Method->Evaluation criteria for replication findings" section.

.4.7. Figures 1 (and similar Figures): Could you add to the Figure note what the error bars represent (a 95% CI, I assumed)? It also seems as if the colors for the error bars are not easily distinguishable from the points.

We added "Error bars represent 95% CI" to all figure notes throughout the manuscript.

Unfortunately, the jmv R package we used to create those plots does not allow us to change the error bar colors, and so we left it as is. We will return to examine the readability of the figures after data collection, and if there's still an issue we will revert to using a different plotting package.

.4.8. For Figures 6 and 7, can you please clarify that the y-axis represents the difference in purchase intentions (pleasure)?

The labels for those Figures have now been updated to read "Change in purchase intentions/pleasure from wearing".

.4.9. The manuscript includes two Tables 5 (p. 22 and 23)

The tables' number ordering has now been fixed.

Reply to Reviewer #4: Dr./Prof. Miguel Vadillo

> **I must begin my review acknowledging that I am not an expert in the topic of the paper. I accepted to review the registered report on the assumption that the manipulations, dependent variables and analyses were unlikely to be complicated (and they aren't!) but reading the manuscript I was soon overwhelmed by the number of manipulations and dependent variables and eventually lost track of what the experiment is really about, what are the crucial effects to be tested and why they are important. All this information is admittedly in the manuscript, but it is scattered in different places, making the reading of the proposal complicated for nonexperts like me. I do have suggestions for the next version of the manuscript, but this are mostly directed towards making it more accessible and focused.**

Thank you for the constructive feedback.

> **.1. The current pdf file is 73 pages long, and this doesn't include the General Discussion, to be written when we have the results! I think the text is too long and this makes it easier to get lost at some point. I would suggest the authors to remove everything that is not essential and, most importantly, avoid repetitions. There are few places where it is easy to delete text without affecting the integrity of the ms. But I would urge the authors to apply this logic to the rest of the manuscript.**

We appreciate the feedback. We moved what we can to tables and figures, and now keep the results section very brief and to the point. We will aim to improve on that further for Stage 2 as the real results come in and we add a General Discussion.

We now report all descriptives and statistical analysis in figures and tables. For example, all statistical analyses and effects are now in Table 1, to more easily compare to the hypotheses and the findings of the target article. In the text we simply report the effect sizes and confidence intervals for the main effects and interactions.

If needed, in Stage 2 we will move some of the figures and analyses, for example regarding liking and historical significance, to the supplementary materials, but for the Stage 1 we feel it important we are all aligned on what is to be analyzed.

> **.2. The introduction includes a lengthy presentation of contagion theory. But I didn't think that this was important to understand the text. I think it can be deleted or reduced to no more than 4-5 lines, directing the reader towards recent reviews.**

We appreciate the feedback, but this is a subjective matter. Some reviewers want to see more in the introduction, whereas others want it shorter. We aimed to strike a balance, and hope that it would be helpful to readers as background information on contagion theory.

To address your point, we cut out some aspects like "The importance of the contagion effect…" and hope that it now flows a bit better and feels more concise.

> **.3. Table 2 includes all the main effects and interactions of the original experiments, but surely not all of these are crucial. Why not focusing just on the two crucial results? This is something that can be done in the text itself. In fact, right now this information is presented in the main text, in Table 1 and in Table 2. I think the text would become much clearer presenting this information just once, possibly in the main text, but more clearly (I miss more information about the original experiments and the interpretation of effects; see below).**

Thank you. Given your feedback on length, we felt that it would be better to report all those in the tables and keep the text concise. We prefer to err on the side of reporting too much, than reporting too little, given that we faced many challenges in understanding and deducing what the target article did and its reports and statistics. To aid the readers we now include sections in the introduction and bolding and notes in the tables to emphasize what are to be considered the main analyses to focus on.

> **.4. On p. 15 there is first a 2-paragraph summary of the extensions and then two full sections explaining the two extensions in more detail. I think this can be summarized to a single paragraph. And if this is well explained in the text, Table is not needed.**

Thank you. We removed one of the extensions, and revised our explanation of the remaining hypothesis to keep it more concise.

> **.5. The power and sensitivity analyses have been conducted for every effect in Table 2, but wouldn't it make more sense to focus just on the crucial effects? The Supplementary Material for further information, but if the analysis is focused on just the crucial effects, maybe this explanation in the SM can be summarized as well and inserted in the main text. (Also, it is possibly easiest for the reader to provide the direct link to the RMarkdown file, instead of giving the name of the file in the OSF folder.)**

Thank you for the feedback. We agree.

We now more clearly identify the key analyses, and conduct the power analysis on those. We also revamped the power analysis section to make it clearer and to supplement it with a sensitivity analysis.

The Rmarkdown files are likely to change several times before publication, and so to avoid having to update this every time and to avoid broken links and mistakes, we refer to the main OSF where the readers should be able to easily find and identify the Rmarkdown code and HTML outputs.

> **.6. Pre-registered protocols often include a section with the analysis plan. An alternative to this is to actually write the results section with random values in places where the numerical values will the reported. But I do not think it is necessary to do both. In other words, I think it would suffice to write either the "Data analysis strategy" or the "Results" section, but not both at this stage. And this extra space can be used to expand the explanations that will allow the reader to understand why each effect is important. When I got to the current Results section, I had already lost track of which results where important and why. Whatever section you decide to keep, I would urge you to remind the reader about which analyses are crucial and why and how they should be interpreted.**

Thank you, that is helpful advice and appreciated. We now integrated planned analyses from "Data analysis strategy" with the results section. That section now only clarifies outliers, order effects, and manipulation checks, which are less relevant for the results section.

> **.7. Figures S1 and S2 in the Supplementary Material can be replaced by very brief verbal descriptions or simply deleted with minor amendment in the text.**

Thank you. We intended this as a service to readers, but now removed those figures from the supplementary. Indeed, readers can simply refer to LeBel et al. (2019) for those.

> **.8. The previous changes would make a lot of space that the authors can then use to include information that right now can be missed, especially by non-expert readers. For instance, it takes a lot of time to understand what Newman did and why. On p. 7-8 we are told that Newman contrasted three explanations and found evidence supporting the contagion hypothesis. I would help a lot if, at this point, the manuscript presented an overview of what participants were asked to do in Newman's experiments, what they did, and how those results lead to the conclusion that contagion was the main factor driving the effect. This general explanation of what Newman's experiment was about is currently missing and it forces the reader to link the bits of information about Newman's study scattered throughout the ms. Note also that if Newman's experiments, results and discussion are presented with some detail here, this would allow the authors to simplify the explanation of their own experiments, because they will only need to remind the reader what's the same as in Newman and what is new in the current experiments.**

Thank you for the feedback. This is a tricky dilemma, how much to repeat from what the target article did but without having to simply repeat all the target article.

Our intention was to present a short overview of the research behind contagion, and present Newman et al. (2011) as one of the main articles underpinning our understanding of this theory. We had to juggle either (1) presenting a detailed analysis of the original experiments and then having a section where we refer to this section for our replication setup, or (2) presenting an overview of the target article and its hypotheses, then explaining our own replication methods (which by definition is a close replication of the one that the target article used). Our original idea was to follow the latter method.

We worked to do better in this revision, and we hope that it reads better and is clearer.

> **.9. I think the justification of the RR can be improved. Pages 10-11 merely say that the contagion effect is important in the literature. But this doesn't feel like a good reason to repeat the study. Is there anything that cast doubts on Newman's original study? Any reason to assume that the results might not be robust or generalizable? Any measure/manipulation that the original study did not include but is worth including? The authors in fact extended Newman's experiment. This would be an excellent place to explain why this was timely, relevant…**

Our view on replications is that replications are needed and should be mainstream. No singular finding with one sample at one point in time is sufficient for us to draw definitive conclusions for eternity for all participants. We do not feel like replications should be explained for each target, and we hope that they would come to be taken for granted, especially so for high impact well-cited findings. We reflect briefly on these points in the following:

- Feldman, G. (2024). "The value and importance of replications: Catching mistakes/fraud, clarifying theory, and testing measurement" talk given to the South Asian Journal Club. https://doi.org/10.17605/OSF.IO/2RXMY
- Feldman, G. (2025). The value of replications goes beyond replicability and is associated with the value of the research it replicates: Commentary on Isager et al. (2024). *Meta Psychology*. https://doi.org/10.17605/OSF.IO/BTNUJ retrieved from https://osf.io/35fh8

Replications are especially important when the original findings did not share all their materials, data, and code, were not pre-registered and well-powered, and were not comprehensive in their reporting.

We do not mean to cast doubt on any specific original findings. We personally set out to try and replicate findings that we hope would replicate well, to clarify what they did, update those to current times and standards, and allow the community a better understanding and a more accurate effect size of the phenomenon.

All that said, we worked to improve to try and incorporated more of what you suggested. We include references to "Whether and how contagion predicts valuation is still under debate." and "the potential in improving on its reproducibility, clarity, methods, and reporting".

.10. "Main hypotheses and key findings in the target article".

The reader gets to this point without knowing very well the procedure/design/results of Newman et al. Without this previous information, it is almost important to understand why the two interactions in Table 1 are important. Much more information about Newman et al. must be included before reaching this point. And perhaps it would be convenient to present these two crucial results themselves much earlier in the ms. For instance, before explaining why replicating Newman is important. I must confess I have not completely understood why these effects are crucial and why they mean. Possibly because when I got to this point, I still new nothing about Newman's procedure/design/logic. Note also that pages 13 and Table 2 contain the same information.

Thank you for the valuable suggestion.

The "Main hypotheses and analyses in the target article for replication" and Table 1 were overhauled to try and do a better job at explaining what the target article did and the main analyses were.

.11. The explanation of the procedure and design (pp. 20-30) can be summarized and simplified substantially. The text is complicated substantially by having different sections for procedure, design, manipulations, dependent variables… I think it would be much easier to simply present what participants were asked to do and in the same order in which they are asked to do it, presenting along the way the manipulations and dependent variables (not in a separate section). That is, when you explain that participants were asked to give the name of an individual, you can say that depending on the experimental condition they were asked to give the name of a positive, negative or mixed individual. And at that point you can say that this was manipulation X. Then you do not need to have a separate section with the manipulations. Otherwise, the reader is forced to go forward and backward in the text to unite the different pieces of information. The same applies to the dependent variables.

Thank you for the feedback.

We overhauled the methods section to address your feedback. We integrated design and procedure and combined other sections into one design section and removed redundancies. We hope it is now easier to follow.

> **.12. This is just a suggestion (like all the previous), but I wonder if it would be useful to report Experiment 1 and Experiment 2 independently. That is, having the Method section of Experiment 2 after the results of Experiment 1. This would reduce working-memory load a lot for the reader.**

Thank you for the suggestion. This is a matter of personal taste, and we tried different options, each with its own strengths and weaknesses. After making things more concise, we hope that it would be easier to follow, and have opted to keep the current structure as is.

> **.13. "Evaluation criteria for replication findings" Wouldn't it make more sense to present this after the analysis plan? That is, once the reader knows what will be analyzed, at the very point where you explain what's the crucial result, you can alert they that you will compare this to the results of the original study using method X.**

Thank you for this suggestion. Per feedback received we integrated the data analysis into the results section, but we still moved the evaluation criteria section to the end of the Methods section.

> **.14. In the results section it might be best to first report all the analyses that pertain to the suitability of data and then the crucial inferential analyses. In other words, we want to know if the basic requirements were met before interpreting the crucial results. This would imply mentioning first whether outliers will be removed or not and whether the manipulation checks were as intended. Here you also mentioned Cronbach's alpha for some dependent measures, but until this point it was not clear to me that you would merge response to different items for different dependent variable. I must confess I got to this point with very little understanding of the procedure, manipulations and dependent variables.**

We made several changes to address your feedback.

We now begin the results section of Experiment 1 with a description of our conducting reliability tests and how we compute scores. We now make all analyses regarding the three objects in Experiment 1 explicit before we begin reporting the results of all analyses.

We do not plan to use any exclusions or use the manipulation checks, in a similar way to what was reported in the target article which used the full sample. We only plan to conduct additional exploratory analyses in case we fail to find support for our predictions, which are not typically included in Stage 1.

> **.15. p. 49 repeats much of the same information as p. 31. Surely these can be summarized/merged.**

You made references to the section “Comparing replication to target article’s findings” which you felt repeated the section “Evaluation criteria for replication findings” in the methods section.

We revamped Table 1 to make the comparison between the target article’s findings and our findings easier to follow, and so the section “Comparing replication to target article’s findings” is now very short and simple:

> We summarized the replication statistics and effects in Table 1 to allow for easier comparison to the target article’s findings.

With an added note below that:

> [Stage 2 plan: We will expand further once the real data comes in]

> **.16. SM, “Materials and scales used in the replication + extension experiment” This section can be replaced by a simple link in the main text.**

As with item 5, the root directory may change according to the version of the manuscript the RMarkdown files are attached to, to avoid having to change the link for every version, we have to keep them as names in the OSF folder.