Reply to PCIRR S1 RNR2 decision letter reviews: Newman et al. (2011) RRR

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response to each item. The 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/XuhGNuBfgjod (https://osf.io/5mz7w)

A track-changes manuscript is provided with the file: "PCIRR-S1-RNR2-Newman-etal-2011-replication-main-manuscript-track-changes.docx" (https://osf.io/fs2a5)

Reply to Editor: Dr./Prof. Chris Chambers

> **I have now received two re-reviews of your submission. As you will see, the evaluations are positive and we are now approaching a final Stage 1 recommendation. In this round, the reviewers offer some valuable suggestions for further clarifying and refining a number of conceptual and methodological aspects of the study. Provided you are able to respond comprehensively to these points in a revision and response, I anticipate being able to award Stage 1 in-principle acceptance without further review.**

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

Reply to Reviewer #1: Dr./Prof. Lachlan Deer

.1. Summary

This revised Stage 1 Registered Report presents a well-considered replication and extension of Newman et al. (2011), aiming to clarify the psychological mechanism of contagion in valuing celebrity possessions. I commend the authors for their thoughtful and transparent revisions in response to prior feedback. The manuscript now presents a tighter, more theoretically grounded design, with clearer hypotheses and an articulated analysis plan. I believe it is close to being ready for in-principle acceptance.

My remaining comments are offered to further sharpen the conceptual and methodological clarity of the manuscript, especially in terms of theoretical scope, statistical transparency, and interpretive precision.

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

2. Comments

I provide detailed comments below aimed at improving the paper, with relative importance indicated by the 0-3 stars (***). I use a three-star rating for points that are essential to resolve before inprinciple acceptance, two stars for issues that I believe should be clarified or revised to enhance the manuscript's clarity and rigor, and one star for suggestions that are more interpretive or stylistic in nature and offered at the editor and/or authors' discretion.

.2.1. Conceptual Overreach in Practical Implications *

• The paragraph linking contagion to examples such as luxury branding, place-of-origin effects, and dark tourism overextends the theoretical construct. While these are interesting domains, they likely also reflect distinct mechanisms—such as authenticity, symbolic association, or identity signaling—rather than contagion as defined in this study (i.e., essence transfer via physical or imagined contact). I recommend clarifying how these examples are connected to the specific contagion mechanism being tested, or alternatively finding other examples that are more clearly linked to the construct under study.

Thank you for the feedback.

The examples we provided are not essential to our replication or the specific effects, but were meant to show wider implications of "contagion" going beyond the narrower scope in the target article of "celebrity contagion", that this effect is not merely about celebrities.

As this section is not strictly necessary, and we already have a section on the wider literature on contagion in the "Contagion effect" section immediately above this one, to address your request we decided to remove that text.

> **.2.2. Circularity of Market Demand Explanation ****
>
> **• The manuscript currently describes the market demand explanation as circular—suggesting that people are willing to pay more simply because others are expected to pay more. While I see the point being raised, I'd suggest rethinking the use of the term "circular" here. From a behavioral economics perspective, second-order beliefs ("I believe others will value this highly") are often central to valuation—especially in contexts like collectibles, fashion, or luxury goods. These beliefs don't need an exogenous "first cause" to be meaningful, nor are they inherently circular. They're part of how markets function, particularly in socially constructed value settings.**
>
> **• If the concern is about what triggers the *initial* valuation spiral, that's a valid theoretical point—but it's more about the origins of demand or belief cascades, rather than circular logic. You might consider reframing this as a limitation of explanation depth (i.e., it stops at prediction rather than causal understanding), rather than logical inconsistency. • Clarifying this would strengthen your theoretical framing and help avoid unintentionally undermining a plausible and widely accepted valuation mechanism.**

We appreciate the suggestion, and modified this section to read the following:

> However, *this fails to explain where the initial demand and valuation of the item came from, or the reason why others would be willing to pay more*.

> **3. Ambiguity Around Correction for Multiple Comparisons ****
>
> **• It's great to see the use of Holm correction for post-hoc comparisons and a stricter alpha level for certain exploratory or follow-up analyses. These steps show a clear effort to reduce the risk of false positives. That said, from a reader's perspective—particularly in the context of a Registered Report—it's still not entirely transparent how many statistical tests are**

planned overall and how the correction procedures map onto the full set of analyses.

• In particular, it would be helpful to distinguish between the total number of confirmatory tests (e.g., hypothesis-driven tests across multiple DVs and experiments) and exploratory ones, and to clarify:

– Which specific tests are included under Holm correction,

– Whether corrections are applied within DVs or across the full set of tests,
– And which tests (if any) are excluded from adjustment.

• I raise this not because I believe your approach is flawed—it's thoughtful and aligned with best practices—but because it would strengthen inferential transparency. Without this summary, readers may struggle to assess the evidential weight of any significant results, especially if multiple related tests are presented together. A brief table or paragraph outlining this structure would really help clarify how to interpret results at Stage 2.

Everything that we plan to do in Stage 2 is transparent and laid out very clearly. All our planned analyses are provided with our Rmarkdown code. We also provided an example of running that Rmarkdown code on random simulated data of our Qualtrics survey to show what the output would look like. To give readers a better sense of what that would look like in the manuscript without having to look at our code, in our results section we provided examples of our planned analysis on the simulated data from our Rmarkdown code. In case we missed something or the data had something that we did not foresee or could not prepare for in advance using the stimulated data and mock analysis, we will add it and then document that as exploratory and/or a deviation in Stage 2.

Regarding the Holm correction, we wrote (words in italics were added in this revision to make things clearer):

> "We applied the Holm p-values adjustment to all analyses that include *ANOVA* post-hoc comparisons."

This means the application of the Holm correction to ANOVA post-hoc contrasts of conditions in independent variables that includes more than two conditions (or their interactions). This is specified in the accompanying Rmarkdown code and output. We did not mean this to indicate a correlation to any other adjustments.

Given that the Stage 1 results section is on mock data that is random noise, we only report main effects and interactions, there are simply no interesting contrasts to report or discuss. Yet, in Stage 2, if we see the need to examine specific post-hoc contrasts, then this pre-registration is meant to say that we will be using the Holm adjustment for that result.

4. Mixed Use of 90% and 95% Confidence Intervals **

• I noticed both 90% and 95% confidence intervals reported throughout, but it's not entirely clear why. Is this tied to preregistered thresholds, directional hypotheses, or exploratory framing? A brief clarification would help the reader interpret these intervals appropriately.

Our original explanation for this disparity was placed in the Note section for Table 1:

> We report 90% CIs for η^2 effect sizes throughout this article as these are equivalent to an .05 alpha (Steiger, 2004; Lakens, 2014).

You have helped us realize that it might be beneficial to add another explicit mention in the manuscript. Therefore, we added a new section to "Data analysis strategy" to better clarify the differences in confidence intervals:

> **Confidence intervals**
>
> In this study, alphas for all studies are set to be 0.05. As F-test values (and thus the corresponding eta-squared effect sizes) can only be positive, an 0.05 alpha is equivalent to a 90% CI, not a 95% CIs (Steiger, 2004; Lakens, 2014). Therefore, throughout the results, we report 90% CIs instead of 95% CIs for all eta-squared confidence intervals where applicable.

5. Language: "Magical Thinking" **

• While this term has precedent in the literature, it can come across as dismissive or culturally loaded, even though it is a term used in the literature. Something like "intuitive belief" or "essentialist reasoning" might better capture the construct without unintended implications.

Thank you for the feedback. We removed the use of that phrase.

6. Language: "Contaminate" and "Infected" **

• Describing objects as "contaminated" or "infected" by cultural or symbolic sources might carry unintended negative connotations. I'd suggest using more neutral terms like "symbolic transfer" or "association,"

> **especially when referring to music, art, or identity-linked objects. Again, the terms likely originate within the literature, but one can take extra care with their modern day usage.**

Interesting point, we appreciate the suggestion.

As you pointed out, these are terms commonly used by this literature despite the negative connotation, likely purposefully trying to convey a certain message. We agree this is not ideal, yet that it might be better to keep the references' use of that language to convey their sentiments, and then discuss those as possible improvements in the literature in future research.

Therefore, to address your request, we now use quotes to mark the phrases "contaminated" and "infected" to indicate that these terms are what was used by the literature but that they are not phrasings we necessarily agree with. We also added a footnote:

> The literature on contagion commonly used phrases of negative connotation such as "infect*" "contaminat*". We include them here as is given the historical context, yet feel that it would be best for this literature to come up with more neutral alternatives. We discuss this in the Discussion section.

And a planned discussion in Stage 2:

> [Planned discussion for Stage 2 recommended by reviewer Dr. Lachlan Deer: Discuss the negative connotation phrases commonly used in the contagion literature ("infect*" and "contaminat*"), the possible downsides, and a possible alternative.]

> **7. Consider Outlining Planned Robustness or Sensitivity Checks ***
>
> **• It might be useful to note whether you plan to run any basic robustness checks—like excluding borderline manipulation check failures—especially in the case of ambiguous results. This isn't essential, but could strengthen the transparency of your Stage 2 plan. As is, I think the authors plan on using all data as collected. There's pros and cons here that I'd recommend thinking over once more.**

We currently have a "Manipulation checks and other measures" in the results section detailing our planned reporting of the manipulation checks.

Regarding exclusions, we have extensive experience thinking about and tackling this issue in the context of replications and seeing first hand its pros and cons and we have come down to deciding to only run exploratory supplementary analyses when there is a failure to replicate one of the core hypotheses. This makes clear what the main confirmatory analysis is (as close as

possible to the one reported in the target, which did not detail what was planned in case manipulation checks were not working as intended) and what is supplementary and exploratory.

8. Provide a Brief Overview of Elicited Celebrities and Non-Celebrities *

• Since the figures are self-elicited, it could be helpful to provide a brief summary of who participants tend to name—just thematically or by category. This would help readers assess generalizability without needing raw data. An appendix table would suffice.

Thank you for the suggestion. We already added your suggested section in the Results section for this purpose in the previous submission:

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

9. Acknowledge Interplay Between Market Beliefs and Personal Valuation *

• In the theoretical setup, personal valuation and perceived market demand are presented as distinct mechanisms. That distinction is useful for experimental clarity, but in practice—especially in consumer contexts—these two often inform each other. People's willingness to pay may be shaped by what they believe others will value, and vice versa. • This doesn't undermine your theoretical categorization, but it may be helpful to briefly acknowledge this interplay, either in the theory section (to show conceptual nuance) or in the discussion. Doing so would add realism to the interpretation and help clarify why the mechanisms might not produce wholly divergent outcomes.

Thank you for the feedback. We added a mention to this interaction effect in the theory section:

> Therefore, people may be more likely to pay higher prices for these items simply due to how scarce these items are, with the assumption that other people will pay more for these items later on, *which in turn causes the perceived market demand of the item to rise* [...]

10. Consider Exploratory Heterogeneity Analyses in Stage 2 *

• Given the self-elicited figures, you might find it informative to explore whether effects vary by familiarity, moral alignment, or perceived celebrity status. Clearly labeling this as exploratory in Stage 2 would make such analyses a nice bonus without affecting the core design.

Thank you for the suggestions. These are interesting, and we see many other interesting directions we could potentially take in Stage 2. Yet, this is already a very complex project with many extensions and analyses, so to keep things focused and concise, we would rather not add possible exploratory directions to our Stage 1. Once we conclude the Stage 2 analyses we committed to in Stage 1, we can re-examine whether there are directions that would help make the manuscript stronger with meaningful insights for the follow-up literature.

Reply to Reviewer #2: Dr./Prof. Susanne Adler

> **thanks a lot for the thorough revision of your Stage-1 manuscript. I feel that the front end is much clearer now and I see the value in removing the temporal distance extension.**

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

> **I have a few minor points left:**
>
> - **In Footnote 1 (““Social” priming replications have so far had a very poor replication record”), please add one or more references (e.g., Mac Giolla et al., 2024).**

Thank you for the suggestion. We added the citation.

> - **For Table 1, please add a note that clarifies the meaning of N/A.**

Thank you for the suggestion.

To the Note section for Table 1 we have added:

> N/A = Not available, not reported in the target article.

> - **On p. 19 of the revised manuscript (section “Power and sensitivity analyses), I find the second paragraph (starting with “The target article’s studies had many hypotheses…”) a bit confusing. Specifically, the part: “[they] 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).” The descriptions sound more like an interaction effect than a main effect. Rewording this section could improve clarity (please discard the comment, if it is just my reading of the section).**

Thank you for catching that, that was an oversight. We changed to (changes in italics):

> [...] and argued that the main effect for valence is related to a main effect for valence in the desire physical contact (our H2b) *and a main effect of fame on market demand* (H3a).