Print ISSN: 3104-932X
Online ISSN: 3104-9338

Collaboration Frequency and Character IP Brand Equity

Wanting Zhao*

Faculty of Economics and Business, Complutense University of Madrid, Madrid, Spain

Received: Nov 4, 2026

Revised: Nov 8, 2026

Accepted: Nov 19, 2026

Published online: November 20, 2026

To appear in: *International Journal of Advanced AI Applications,* Vol. 3, No. 2 (February 2027)

* Corresponding Author: Wanting Zhao (wtinggz@gmail.com)

Abstract. This paper separates co-branded product evaluation from parent-brand evaluation and develops a planning protocol for character intellectual property (IP) research. In 195 public observations, the co-branded product mean was 0.418 points below the pre-exposure parent-brand mean on a nine-point composite (paired-bootstrap 95% CI −0.565 to −0.277). This cross-object comparison is not a decline in the parent brand. The separate within-parent evaluation change was −0.067 (−0.160 to 0.029). We propose persona-conditioned free-text generation and semantic similarity rating (SSR) to explore character-IP equity distributions and inform a future factorial experiment. Generic power simulations are retained separately as sensitivity analyses; no LLM generation or SSR embedding analysis was conducted. The contribution is an explicit distinction between measured evaluation, proposed character-IP equity and simulation-based planning.

Keywords: *character intellectual property; co-branding; brand dilution; brand evaluation; semantic similarity rating; experimental planning.*

1. Introduction

A character can appear on clothing, drinks, stationery and digital merchandise without becoming more valuable to its audience. A successful collaboration may sell an attractive product while leaving the character's own associations unchanged. Conversely, repeated licensing may increase recognition while making the character seem interchangeable. The managerial problem is therefore whether additional collaborations strengthen the character brand, rather than whether an individual licensed product receives a favorable response.

Customer-based brand equity concerns the value associated with brand knowledge [1], whereas dilution concerns changes to beliefs about a parent brand [2]. Alliance outcomes depend on partner relationships and fit [3], and effect sizes vary with the measured outcome and research design [4]. Product evaluation, parent-brand evaluation and character-IP equity

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

therefore require separate definitions and analyses.

We reanalyze a public food-brand experiment to distinguish product evaluation from parent-brand evaluation change. We then specify a persona-conditioned SSR protocol for a future character-IP experiment, supported by a separate generic power sensitivity analysis. The public observations provide the measured example; the SSR workflow is a proposed extension.

The contribution is an auditable endpoint comparison and an outcome-specific planning framework linking synthetic response distributions to a subsequent human pilot.

2. Conceptual Framework

2.1. Operational definitions

Parent brand refers only to either original food brand in the public experiment. Parent-brand evaluation is the released composite of perceived quality and overall favorability, averaged over the two parent brands. Co-branded product evaluation concerns the joint product. The product gap compares these different objects; parent-brand evaluation change compares the same parent-brand composite before and after exposure. These evaluation scores are not direct measurements of brand equity.

Character IP denotes the licensed fictional character; parent brand is reserved for the public food-brand analysis. Character-IP equity denotes the incremental preference or value attributed to that character's identity. For the proposed SSR pilot, an exploratory five-point equity rating expresses that added-value judgment; it requires validation against a human equity measure. Collaboration frequency is launches within a fixed year; category diversity is category breadth conditional on launch count. Neither construct is inferred from the public purchasing-frequency fields.

2.2. Frequency and competing mechanisms

More collaborations can make a character easier to recall, while repeated licensing may also make commercial intent salient. Persuasion knowledge provides a basis for considering the latter judgments [5]. We distinguish familiarity, perceived excessive commercialization and uniqueness; their contributions to equity remain measurement questions.

A licensed beverage may be judged on taste, price and packaging, whereas character-IP equity concerns the value attributable to the character's identity. Work on asymmetric alliances [6] and superhero licensing [7] supports separating these outcomes without establishing an optimal collaboration frequency.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

2.3. Category diversity as a boundary condition

Category diversity is the number of product categories represented at a fixed collaboration count. It differs from partner fit and launch frequency. Breadth could preserve interest by reducing repetition or weaken distinctiveness by making a character ubiquitous. Creative consistency may matter independently, so both attenuation and amplification of frequency effects remain plausible.

The proposed primary interaction is the high-minus-low frequency effect under high diversity minus that under low diversity. A negative value indicates a more adverse frequency effect under broader category coverage. Existing consumer-based equity scales [8] can guide a human questionnaire; the exploratory SSR equity rating described below is not itself a validated replacement for that measurement.

3. Public Data Reanalysis

3.1. Provenance and endpoint construction

We use Experiment 1 from the release accompanying Garaus, Wolfsteiner and Florack [9], deposited by Garaus as version 2 on Mendeley Data under CC BY 4.0 [10]. Its 195 distinct cases have complete baseline, co-brand and post-exposure parent-brand evaluation composites. No observations are imputed or added. This secondary analysis concerns food brands and reproduces the original endpoint distinction.

Table 1. Public-data endpoints and the limits of their interpretation.

Quantity	Construction	Interpretation
Baseline B	Mean of four parent quality/evaluation items	Pre-exposure parent-brand evaluation
Product C	Mean of displayed co-brand quality/evaluation items	Co-branded product evaluation
Parent evaluation after A	Mean of four later parent items	Immediate parent-brand evaluation
Product gap	C − B	Different objects compared within person
Parent evaluation change	A − B	Same parent composite across phases
Endpoint difference	A − C	Difference between the two contrasts

The baseline averages quality and overall evaluation for both parent brands. The post-exposure parent score uses the corresponding four later items. The co-brand score averages quality and overall evaluation in the displayed brand-order condition. We reconstruct these scores from the released items or condition-specific composites and verify exact agreement

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

with the deposited summary fields. Each score ranges from 1 to 9. The released purchasing-frequency fields are not used as proxies for collaboration frequency.

3.2. Paired estimation

The descriptive endpoints are the mean product gap, mean parent change and their paired difference. We resample participants with replacement 10,000 times, retaining all three scores within each sampled row; seed 20261116 fixes the resampling. Percentile 95% intervals are reported alongside conventional paired t intervals, medians and 10% trimmed means. Resampling individual measurements separately would discard the pairing and answer a different precision question.

Direction counts separate negative, zero and positive changes. A two-by-two table compares lower product evaluation with lower parent-brand evaluation. The cross-object product gap is descriptive; the within-parent contrast measures immediate evaluation change. Without an untreated counterfactual, neither establishes a causal or lasting change in parent-brand equity.

3.3. Public-data results

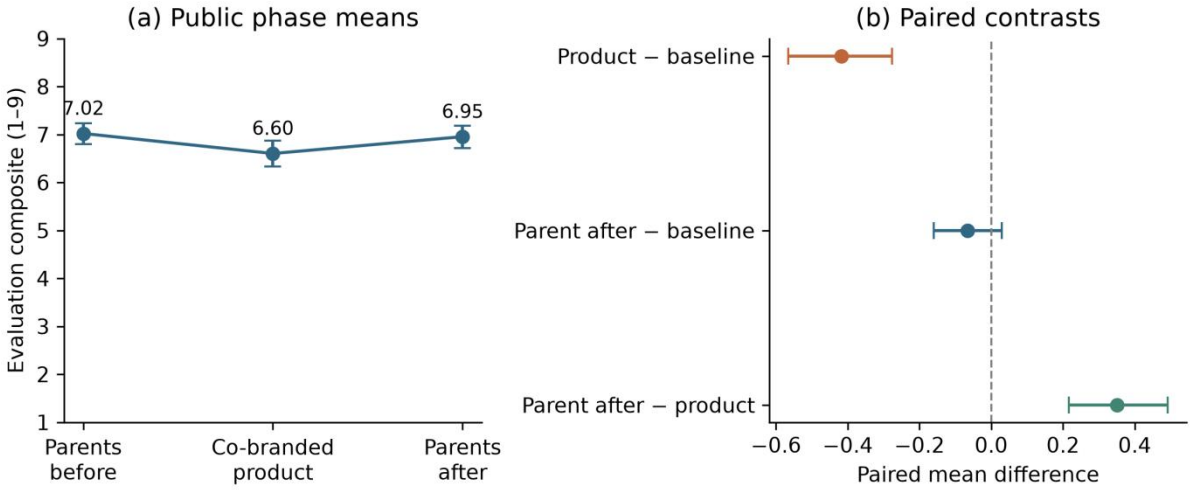


Figure 1. Public observations only. Evaluation means and paired contrasts with 95% participant-bootstrap intervals. The product gap compares different objects; parent-brand evaluation change compares the same parent-brand composite across phases.

Table 2. Recomputed public-data estimates; n = 195, nine-point scale.

Endpoint	Mean	95% paired-bootstrap CI	10% trimmed mean
Product gap C − B	-0.418	[-0.565, -0.277]	-0.347
Parent evaluation change A − B	-0.067	[-0.160, 0.029]	-0.056
Endpoint difference A − C	0.351	[0.217, 0.492]	0.275

The baseline parent-brand evaluation mean is 7.021, the co-branded product mean 6.603 and the later parent-brand mean 6.954. The cross-object product gap is −0.418; the separate within-

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

parent evaluation change is −0.067. Their paired difference is 0.351 (95% bootstrap CI 0.217 to 0.492). Conventional t intervals are −0.564 to −0.272 for the product gap and −0.163 to 0.030 for the within-parent change, consistent with the bootstrap comparison.

The direction reproduces the original report [9]. The within-parent interval includes zero, but is not an equivalence test. Both change distributions have a median of zero. The product-gap trimmed mean is −0.347, so its negative mean is not solely carried by the most extreme tenth of observations at each tail.

Ninety-seven participants evaluate the co-branded product below their baseline parent-brand average. Of these, 53 also report lower later parent-brand evaluation, while 44 report the same or higher evaluation. Across all 195 cases, the within-parent contrast is negative for 73, zero for 65 and positive for 57. Thus lower product evaluation does not imply a within-parent decline for the same individual.

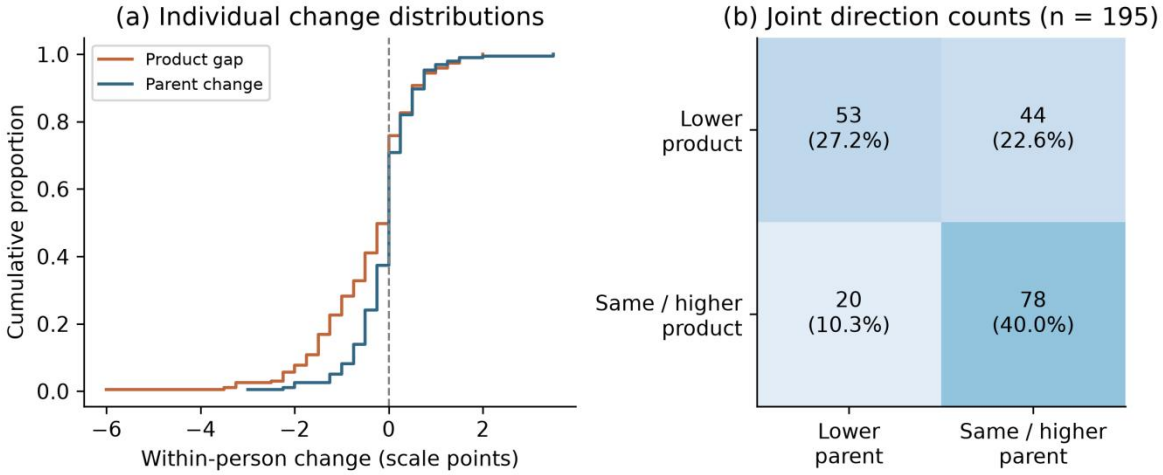


Figure 2. Public observations only. Change distributions and joint direction counts. Product gaps compare different objects; parent-brand evaluation changes compare the same parent-brand composite. Same/higher groups include zeros.

4. SSR Protocol for Character IP Study Planning

4.1. Methodological basis and execution status

Maier et al. [11] describe SSR, which maps LLM-generated free text to Likert probability distributions using similarity to rating anchors. Their purchase-intention benchmark reports roughly 90% correlation attainment relative to a human test–retest baseline; this is not 90% individual-level accuracy or validation of character-IP equity. We adapt the elicitation idea as the prospective protocol below. No LLM responses, embeddings or SSR-derived equity estimates have been generated for this manuscript.

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

The previous 400-record parametric illustration is removed from the main analysis and retained in the companion archive as a legacy sensitivity exercise. Its assumed coefficients cannot validate the proposed SSR workflow. The completed public-data reanalysis and generic power calculations remain distinct from the planned SSR stage (Figure 3).

4.2. Factorial stimuli and persona conditioning

The proposed design crosses four versus twelve collaborations within one year with one versus four product categories. Every high-diversity portfolio uses the same four category labels; product appeal, price, artwork, launch timing and creative coherence should be matched. Each portfolio presents one character IP. Several character versions and independently authored stimulus variants would be required before generalizing beyond a single example.

An exploratory grid crosses three adult age bands (18–29, 30–44 and 45+), two fandom levels, two discretionary-budget levels and two commercialization-sensitivity levels, giving 24 hypothetical personas. These are designed scenarios, not representative population segments. Each persona is evaluated in all four cells in separate generation contexts, with ten planned repeated generations per persona and cell. The resulting 960 planned responses are repeated model outputs, not 960 independent consumers. Equal persona weights are exploratory; any population weighting would require an external sampling basis.

The generation prompt should request a short free-text judgment of the character's added value after viewing the portfolio, without requesting a numerical rating or prescribing whether diversity helps. Product purchase intention is a separate secondary prompt. Fandom and commercialization sensitivity must describe prior orientation without instructing the direction of the response. Context order should be randomized and sessions isolated. The protocol should archive exact model versions, prompts, stimulus files, generation parameters, timestamps, repeated-run identifiers and unedited responses.

4.3. Equity anchors and distributional scoring

Before generation, the exploratory character-IP equity judgment should be anchored from negative added value to strong positive added value (Table 3). Purchase-intention anchors from SSR research must not simply be renamed as equity anchors. The proposed statements require cognitive testing with humans, and should be compared with a validated equity battery before supporting recruitment decisions. Alternate wording sets should be fixed in advance and reported separately, not selected for producing a desired interaction.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

Table 3. Proposed character-IP equity anchors for piloting; not validated scale items.

Rating	Proposed semantic anchor
1	The character identity makes the offering less valuable to me.
2	The character identity adds no value to an otherwise comparable offering.
3	The character identity adds a small amount of value to an otherwise comparable offering.
4	The character identity adds clear value to an otherwise comparable offering.
5	The character identity adds substantial value to an otherwise comparable offering.

One fixed embedding model should encode each response and every anchor. The proposed baseline mapping subtracts the minimum of the five cosine similarities and normalizes the nonnegative remainder to sum to one, without temperature reshaping. Equal similarities are flagged as uninformative rather than silently converted into confident ratings. Full five-category probability vectors and mapping diagnostics must be retained. This mapping is sensitive to wording and model choice, so a second embedding model and prespecified anchor variants are robustness checks, not interchangeable replications of human responses.

For each cell, the mean is the weighted sum of rating values under the persona-averaged probability distribution. The full rating variance is its second moment minus the squared mean. Using only the variance of expected persona ratings discards within-response probability mass and can understate planning variance. Conversely, SSR probability spread is semantic uncertainty, not established human response variation. Persona weights, repeated-generation variability and within-response probability spread should therefore be reported separately.

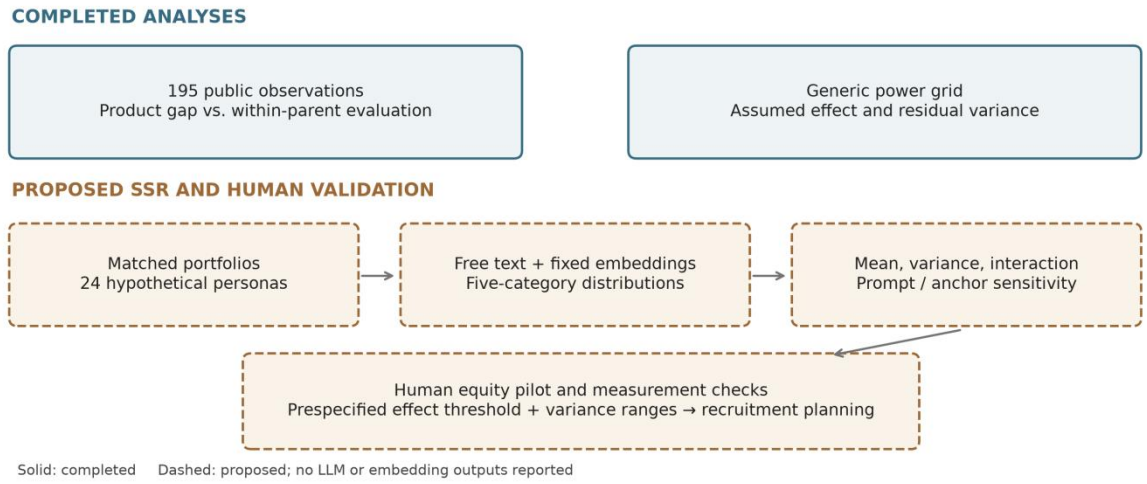


Figure 3. Evidence and planning workflow. Public-data analysis and generic power calculations are completed. The persona, text-generation, embedding and human-pilot stages are proposed; arrows do not represent completed SSR results.

4.4. Interaction, validation and power inputs

The prospective interaction is the difference in frequency contrasts between high- and low-diversity cells. Its sensitivity should be assessed across persona weights, prompt variants,

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

character versions, anchor sets and models. Resampling should retain the four-cell structure of each base persona, with repeated generations nested within persona; model repetitions must not be counted as independent human participants. Any resulting interval characterizes that model workflow, not population sampling uncertainty.

A human pilot should administer the same stimuli and an independently validated character-IP equity instrument. Distribution shape, mean contrasts, ceiling effects and residual variation should be compared between human and SSR outputs. A purchase-intention benchmark cannot supply equity validity. SSR may suggest plausible planning ranges, but the smallest meaningful interaction must be chosen on substantive grounds in advance; it should not be set equal to the largest synthetic effect. Null, attenuated and opposite-sign interactions, and inflated residual variances, should be included in sensitivity analyses [12,13].

4.5. Completed generic power sensitivity analysis

For a numerical precision benchmark independent of SSR, the completed generic simulation uses balanced four-cell normal outcomes with residual standard deviation one. Total sample sizes are 100, 200, 400, 800 and 1,600, with interactions −0.30, 0 and 0.30 residual SD. Cell means in low-frequency/low-diversity, low-frequency/high-diversity, high-frequency/low-diversity and high-frequency/high-diversity order are 0, 0.10, −0.10 and the specified interaction. These are assumed sensitivity inputs, not LLM-derived estimates.

For each of the 15 conditions, 3,000 independent datasets are represented by their exact normal-theory sufficient statistics. Four sample means are drawn with standard error equal to one divided by the square root of cell size; the pooled residual sum of squares follows a chi-square distribution with N − 4 degrees of freedom. We then compute the ordinary factorial t interval. This is equivalent to drawing all individual normal responses for these performance measures. Seed 20261117 fixes this analysis. It evaluates a continuous homoscedastic planning model, not the rounded seven-point composites or the public data.

Table 4. Simulated interaction detection and precision; 3,000 repetitions per condition.

Total N	Detect −0.30	Reject at zero	Detect +0.30	Mean CI width at zero
100	0.110	0.049	0.122	1.583
200	0.187	0.044	0.184	1.114
400	0.309	0.049	0.340	0.787
800	0.580	0.044	0.551	0.555
1600	0.853	0.056	0.846	0.392

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

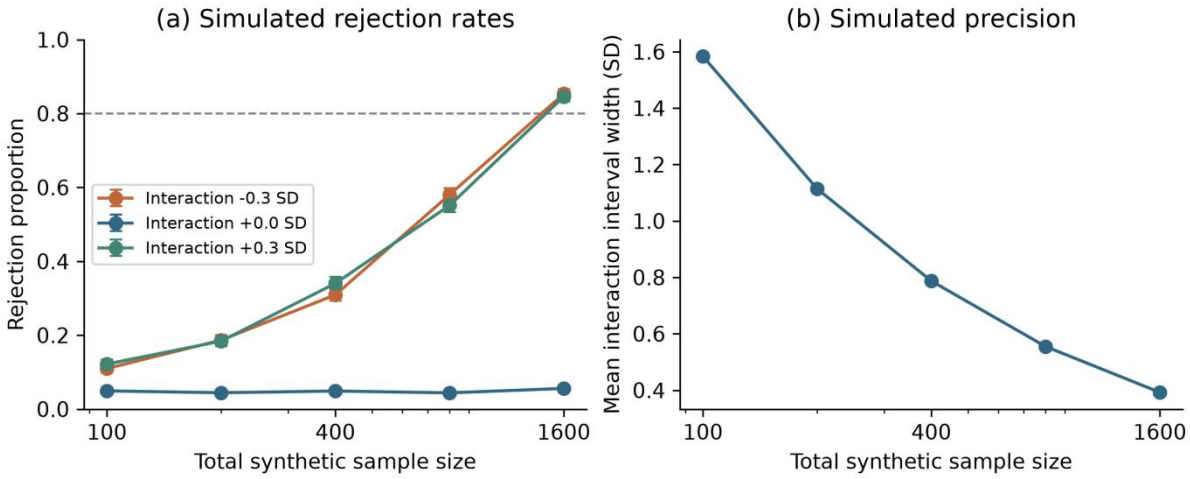


Figure 4. Independent continuous-outcome simulations only. Detection/null rejection and interval width across total sample sizes. Error bars show ±1.96 Monte Carlo standard errors; the dashed line marks 80% detection.

At N = 400, detection is 30.9% for −0.30 SD and 34.0% for +0.30 SD; the difference reflects Monte Carlo variation, not an assumed directional advantage. At N = 1,600, the corresponding rates are 85.3% and 84.6%. Across conditions, coverage ranges from 94.4% to 95.6%. Null rejection ranges from 4.4% to 5.6%. The maximum Monte Carlo standard error of a detection proportion is approximately 0.0091. Mean interval width at zero shrinks from 1.583 SD at N = 100 to 0.392 SD at N = 1,600.

The generic grid illustrates how power depends on assumed interaction and residual variation. It neither validates SSR nor fixes a human recruitment target. Final planning should combine a substantively chosen effect threshold, uncertainty from the human pilot and prespecified SSR sensitivity ranges [13].

5. Discussion

5.1. Endpoint separation and study planning

The public data show a cross-object product-evaluation gap alongside a much smaller within-parent evaluation change. This supports measuring the character IP separately from its licensed product. The SSR protocol extends the measurement plan; the completed power grid quantifies precision under stated assumptions.

Outcome-dependent alliance estimates [4] and licensing models [14] caution against equating a product's reception with its parent brand's reputation. Synthetic persona outputs could help refine stimuli and identify sensitivity ranges, while human observations remain necessary to establish character-IP equity effects.

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

5.2. Implications for a future consumer experiment

The next empirical step should randomize participants to matched character portfolios. At each frequency level, category breadth must vary without inadvertently changing product attractiveness, average price, artistic coherence or launch timing. Otherwise the interaction may reflect an unbalanced stimulus. Several character versions would help assess dependence on a single design. Familiarity and fandom should be measured before exposure, with subgroup analyses specified in advance.

Character-IP equity should be the primary human outcome; product evaluation and purchase intention should be secondary outcomes. Familiarity, commercialization and uniqueness require distinct measures. A five-category SSR mapping is an exploratory proxy, and must not be pooled with the public nine-point evaluation composite. Randomizing frequency and diversity also does not randomize mediators, so causal process claims require additional design and assumptions.

Managers should monitor character-IP equity alongside product performance. The present evidence does not justify a licensing-frequency threshold or category-expansion rule; those decisions require character-specific human and longitudinal evidence.

5.3. Limitations

The public reanalysis covers one food-brand experiment, immediate evaluations and a bounded nine-point composite. Procedural effects may influence before–after differences. Participant bootstrap intervals do not represent variation across character IPs or licensing portfolios, and the near-zero within-parent evaluation mean is not an equivalence result.

The proposed SSR stage has not been executed or validated for character-IP equity. Persona construction, prompt wording, shared model biases and anchor choice may change the inferred distribution. Repeated model outputs cannot establish human heterogeneity or causal mediation. The completed generic power grid assumes independent normal outcomes and does not resolve these measurement limitations. A preregistered human pilot and subsequent experiment are required.

6. Conclusion

The public reanalysis distinguishes product evaluation from within-parent evaluation change. An operationally defined character-IP equity outcome, a proposed persona-conditioned SSR protocol and a separate generic power grid provide a transparent basis for planning a human

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

experiment. Frequency and diversity effects on character-IP equity remain untested empirically.

Data availability. Public data: https://doi.org/10.17632/t876rkzyf4.2 (CC BY 4.0). The revised companion archive separates public-data analyses, legacy parametric simulations and the proposed SSR protocol. It contains no LLM responses or embeddings. No new human participants were recruited.

References

[1] Keller, K. L. (1993). Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing*, *57*(1), 1–22. https://doi.org/10.1177/002224299305700101

[2] Loken, B., & John, D. R. (1993). Diluting brand beliefs: When do brand extensions have a negative impact? *Journal of Marketing*, *57*(3), 71–84. https://doi.org/10.1177/002224299305700305

[3] Paydas Turan, C. (2021). Success drivers of co-branding: A meta-analysis. *International Journal of Consumer Studies*, *45*(4), 911–936. https://doi.org/10.1111/ijcs.12682

[4] Voss, K. E., Mohan, M., Jung, J. H., & Jiménez, F. R. (2025). The effect of brand alliances across dependent variables and research designs: A meta-analysis. *Journal of Business Research*, *191*, Article 115248. https://doi.org/10.1016/j.jbusres.2025.115248

[5] Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research*, *21*(1), 1–31. https://doi.org/10.1086/209380

[6] Vaidyanathan, R., & Aggarwal, P. (2022). Asymmetric brand alliances: When joint promotions with strong brands hurt. *Journal of Business Research*, *141*, 213–228. https://doi.org/10.1016/j.jbusres.2021.12.022

[7] Monika, M., & Antonio, F. (2022). How superhero characters shape brand alliances and leverage the local brand: The evidence from Indonesia. *Businesses*, *2*(1), 33–53. https://doi.org/10.3390/businesses2010003

[8] Yoo, B., & Donthu, N. (2001). Developing and validating a multidimensional consumer-based brand equity scale. *Journal of Business Research*, *52*(1), 1–14. https://doi.org/10.1016/S0148-2963(99)00098-3

[9] Garaus, M., Wolfsteiner, E., & Florack, A. (2022). When food co-branding backfires: The overexpectation effect. *Foods*, *11*(14), Article 2136. https://doi.org/10.3390/foods11142136

[10] Garaus, M. (2022). *Datasets* (Version 2) [Data set]. Mendeley Data. https://doi.org/10.17632/t876rkzyf4.2

[11] Maier, B. F., Aslak, U., Fiaschi, L., Rismal, N., Fletcher, K., Luhmann, C. C., Dow, R., Pappas, K., & Wiecki, T. V. (2025). *LLMs reproduce human purchase intent via semantic similarity elicitation of Likert ratings* (Version 3) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2510.08338

[12] Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102. https://doi.org/10.1002/sim.8086

[13] Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), Article 33267. https://doi.org/10.1525/collabra.33267

[14] Bacchiega, E., Colucci, M., Denicolò, V., & Magnani, M. (2024). Only the ugly face? A theoretical model of brand dilution. *Management Science*, *70*(5), 3182–3199. https://doi.org/10.1287/mnsc.2022.00852

© 2026 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.