Supplementary Material

Supplement to "Diagnosing Image-Ad Success: An Interpretable and Scalable Framework for Liking, Sharing, and Memorability."

This supplement has four parts. The first section contains the loess diagnostics for the computational measures. The second section provides more details on the synthetic consumer panel's persona construction, prompting protocol, and implementation details. The third section provides the results of a path model estimated without the enjoyment and interest constructs (M3); and the fourth section presents the three algorithms used to compute the image typicality scores.

M1 Complexity and Ease of Understanding Figures

Figures M1 and M2 show that the loess functions relating the computational measures to the image-level complexity and ease-of-understanding judgments are approximately linear.

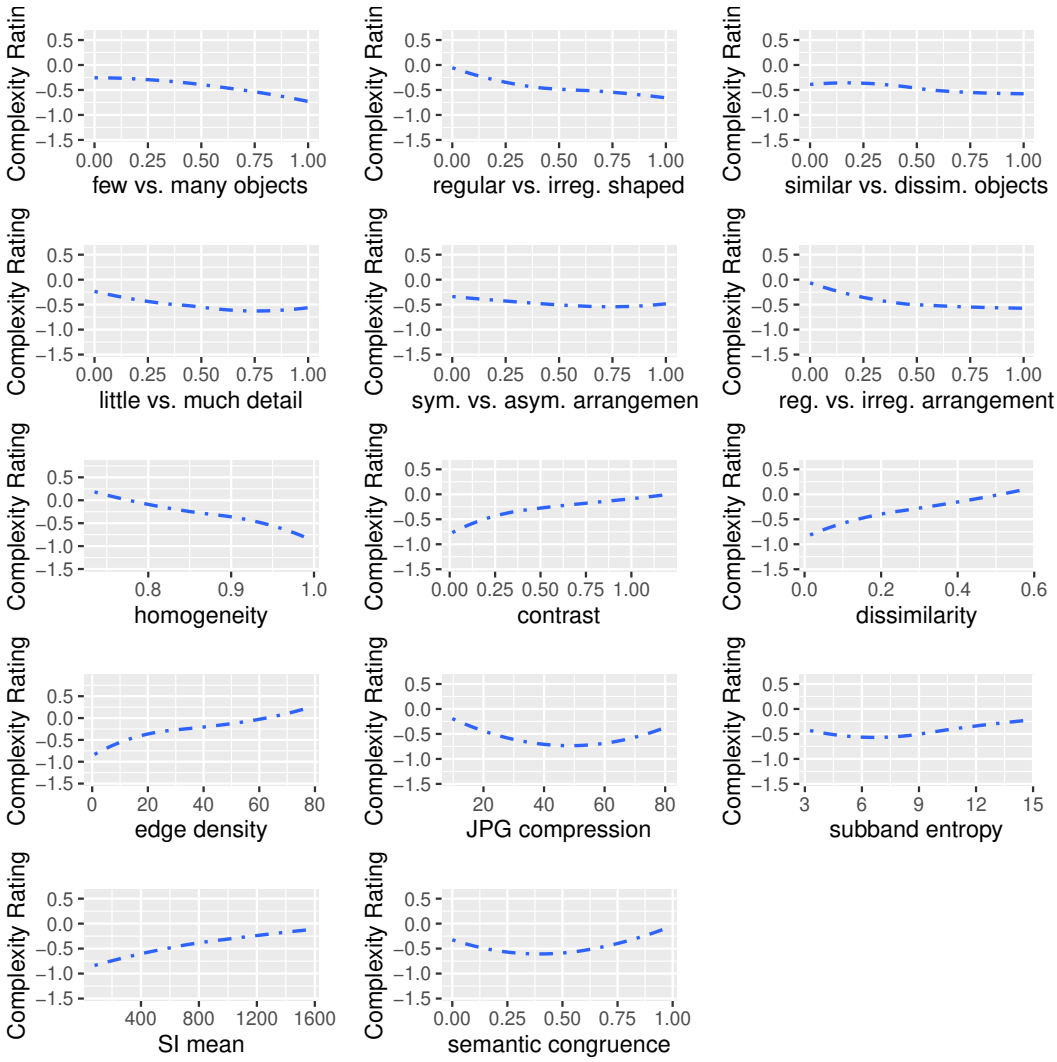

Figure M1: Loess functions relating computational complexity scores to image-level complexity judgments. Each panel plots one of the complexity measures against the standardized image-level complexity rating, pooled across the three fast-food chains. The first six panels are the design-level measures; the remaining eight comprise the seven pixel-level measures and semantic complexity.

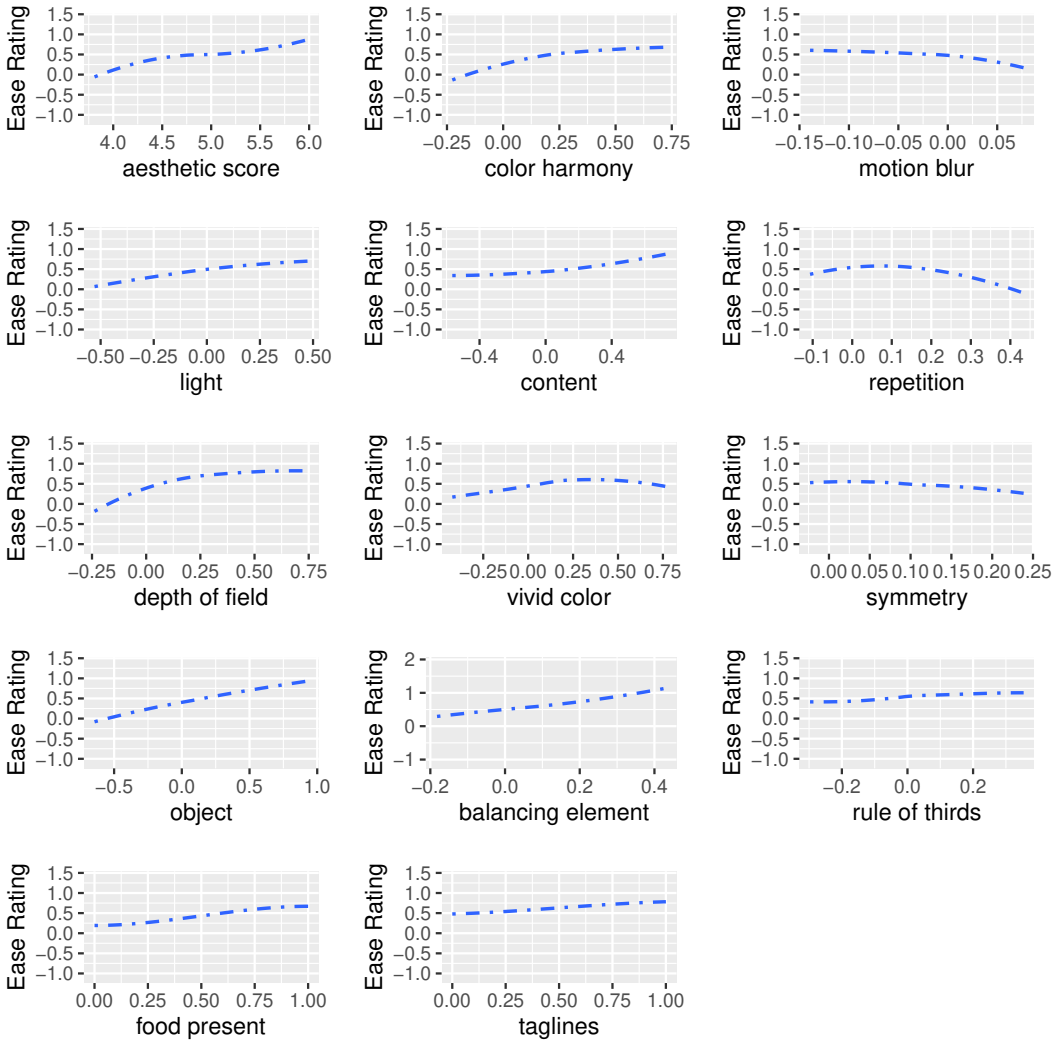

Figure M2: Loess relationships between computational measures and image-level ease-of-understanding judgments. Each panel relates one computational measure to the standardized image-level ease rating pooled across the three fast-food chains. Twelve panels show the eleven AADB aesthetic attributes and the overall aesthetic score. The other two show the food-present and taglines (brand-name indicator); their fitted lines summarize differences between the two groups.

M2 Synthetic consumer panel: methodology and supplementary path model

This section describes the persona-conditioned synthetic consumer panel and the supplementary path model estimated from its ratings.

Overview The study simulated a consumer survey by assigning a synthetic persona to an AI agent and asking the agent to evaluate brand-related image ads. Each persona was constructed from observed demographic and transaction variables rather than generated ad hoc. The design introduces heterogeneity across synthetic respondents while standardizing the evaluation task across images.

The panel evaluated the same substantive items used in Study I: interest, complexity, pleasantness, typicality, ease of understanding, warmth, and engagingness. As in the main text, pleasantness and warmth are treated as indicators of enjoyment, and interestingness and engagingness are treated as indicators of interest.

Persona construction For each evaluation, the model received a detailed system prompt specifying the persona that it was asked to embody. The prompt contained two types of information. First, it included demographic descriptors such as gender, income bracket, ethnicity, age, marital status, parental status, and education. Second, it included brand-patronage variables for McDonald's, Wendy's, and Chipotle, such as the number of historical visits, average basket size, and the recency of the most recent visit. These brand-history variables were intended to anchor the image evaluations in prior brand experience rather than let each image be judged in isolation.

The persona descriptions were written in natural language so that they could be incorporated directly into the prompt. Each panel member therefore represented a synthetic consumer profile linking demographic characteristics to prior quick-service-restaurant behavior.

Hierarchical prompting The evaluation protocol followed a two-stage design. In Phase 1, the synthetic respondent answered one or more text-based questions about the focal brand. These baseline questions established a general brand attitude before any image was shown. In Phase 2, the same synthetic respondent evaluated the brand images. Each image prompt instructed the model to respond as the assigned persona would and to keep its answers consistent with the baseline sentiment established in Phase 1. The image questions covered the same items as Study I.

Brand-level sentiment was elicited first, and image-level evaluations were generated conditionally on it. Thus, the model was not asked to treat each image as a de novo judgment task.

Rolling context and scale anchoring To promote response-scale consistency across repeated image evaluations, the script implemented a rolling-context mechanism. Beginning with the second image in a sequence, the prompt included the synthetic respondent's response to the immediately preceding image. This feature provided a local comparison standard across the image sequence.

Question format and outputs The image-level prompts asked the model to report how interesting the image is, how complex it feels, how pleasant it feels, how typical it is for the focal brand, how easy it is to understand, how warm the overall feeling is, and how engaging the image is. The prompts were brand-specific, so each image was evaluated with respect to McDonald's, Wendy's, or Chipotle, depending on its source.

The model was required to return a structured JSON object. For each item, the output schema specified two elements: a quantitative response on a 1–7 scale and a short textual justification, typically one or two sentences, written in a manner consistent with the assigned persona.

Implementation details The study used GPT-5.1 with a temperature setting of 0.7. The model was instructed explicitly to answer as a survey respondent rather than as an analyst, and the system prompt emphasized that ratings and justifications should reflect the assigned demographic profile and brand-patronage history.

Aggregation and supplementary path analysis The item-level outputs were aggregated to form image-level synthetic measures. Enjoyment was formed by averaging the warm and pleasant items, and interest by averaging the interesting and engaging items; complexity, typicality, and ease of understanding were represented by their corresponding items. We did not apply measurement-error corrections to the image-level synthetic means.

As an additional robustness exercise, we used these synthetic measures in place of the human construct ratings to predict the observed ad-success measures. A model that constrained all path coefficients to be equal across the three fast-food chains fit less well than the corresponding human-judgment and computational-measure models ($\chi^2 = 155.9$, $df = 46$), indicating substantial brand-level heterogeneity. The estimates are reported in Table M1.

Table M1: Supplementary path coefficients (standard errors) using synthetic-panel measures

Predictor	Memorability	Liking	Sharing	Enjoyment	Interest	Ease
Enjoyment	**-.24** (.063)	**.07** (.017)	**.01** (.005)	x	x	x
Interest	**.29** (.127)	**.07** (.033)	**.03** (.011)	x	x	x
Ease	.19 (.109)	-.07 (.030)	-.02 (.010)	**.77** (.148)	**.16** (.068)	x
Typicality	**-.20** (.040)	**.03** (.010)	.01 (.003)	.06 (.061)	-.01 (.027)	**.13** (.023)
Complexity	-.07 (.077)	-.03 (.024)	-.01 (.008)	.07 (.124)	**.32** (.048)	-.10 (.052)
$R^{2(a)}$	.26, .36, .10	.37, .23, .28	.13, .12, .20	.17, .13, .09	.26, .14, .13	.13, .21, .22

$^{(a)}$ The estimated observed-variable R^2 values are reported in the order Wendy's, McDonald's, and Chipotle.

The supplementary path results show some qualitative similarities to the human and computational analyses, but the fit and explained variance are weaker and less stable across brands.

M3 Path analysis without the enjoyment and interest constructs

This section reports the results of a path model that estimates how ease of understanding, typicality, and complexity relate to the three success measures, as these constructs have been used predominantly in the literature to predict liking responses. The path-analytic model including only these three constructs yields a satisfactory model fit ($\chi^2 = 18.9$; $df = 18$). Ease of understanding has a positive association with liking but a negative association with sharing, suggesting that images that are easier to understand are shared less often. Complexity predicts all three success measures. Specifically, memorability declines as image complexity increases, whereas liking and sharing are positvely associated with complexity.

Table M2 also reports the estimated latent-variable R^2 statistics values for the three fast-food chains. For memorability, liking, and sharing, these values are lower than the corresponding values in Table 4, which includes enjoyment and interest. This descriptive comparison supports the usefulness of including these evaluative constructs in modeling ad success. [1]

Table M2: Estimated path coefficients (standard errors) of the three FF chains without Enjoyment and Interest constructs

Predictor	Memorability	Liking	Sharing	Ease
Ease	.13 (.119)	**.28 (.033)**	**-.06 (.015)**	x
Typicality[(a)]	**-1.00 (.120)**	-.01 (.028)	.00 (.010)	**.60 (.031)**
Complexity	**-.21 (.100)**	**.20 (.025)**	**.08 (.009)**	**-.12 (.059)**
$R^{2(b)}$	.57, .73, .21	.47, .68, .55	.44, .32, .43	.88, .68, .74

[(a)] The table reports the typicality effects for Wendy's. The estimated McDonald's coefficient for typicality and liking is -.08 (.024) and for typicality and d′ = -.52 (.099). The corresponding path coefficients for Chipotle are -.02 (.029) and -.42 (.109), respectively.

[(b)] The estimated latent-variable R^2 are reported in this order: Wendy's, McDonald's, and Chipotle.

[1] The two path models are estimated separately and are not nested. We therefore interpret their R^2 differences as descriptive comparisons between the fitted specifications, rather than as a formal decomposition of the variance explained in Table 4 or a quantification of the unique contributions of enjoyment and interest.

M4 Algorithms for image typicality

This section reports the three algorithms behind the computational typicality measures. The manuscript refers to them as Algorithms S1, S2 and S3.

Algorithm S1 Object (LSA)-based Typicality

Input: A set of images $I = \{i_1, i_2, \cdots, i_n\}$
for each image i in fast-food chain **do**
    Identify top k objects: $\{f_1^i, f_2^i, \cdots, f_k^i\}$ ▷ Google Cloud Vision API
Formulate an image-object occurrence matrix $M = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & v_{ij} & \vdots \\ v_{n1} & \cdots & v_{nm} \end{bmatrix}$
▷ m is the number of all unique detected objects.
▷ $v_{ij} = tfidf(f_j, i)$ if the j^{th} object occurs in the i^{th} image, 0 otherwise.
$TS_{LSA} = \{TS^1, TS^2, \cdots, TS^c\} \leftarrow \mathbf{LSA}(M, c), S \in \mathbb{R}^{n \times c}$
▷ c is the reduced dimensionality
Output: typicality scores= TS^1_{LSA}

Algorithm S2 Design-based Typicality

Input: $I = \{i_1, i_2, \cdots, i_n\}$ and human judgment ratings on 6 design features
for each image i in fast-food chain **do**
    for each design feature f **do**
        generate the true label $l_f^i \leftarrow OneHotEncoding$, where $l_f^i \in \{0, 1\}^4$
▷ Based on the quartiles of the human response proportions
Split I into *training* and *testing*
Fine tune DenseNet using *training* ▷ 4-class classification tasks
Test using *testing* ▷ prediction accuracy
obtain the predicted score $s_f^i = max(p_1, p_2, p_3, p_4)$ of each f for each image i
▷ where $p_k \in (0, 1), \sum_{k=1}^{4} p_k = 1$
each image is represented as $s^i = (s_{f1}^i, s_{f2}^i, \cdots, s_{f6}^i)$
Formulate a pairwise design-based dissimilarity matrix $M = \begin{bmatrix} s^{11} & \cdots & s^{1n} \\ \vdots & s^{ij} & \vdots \\ s^{n1} & \cdots & s^{nn} \end{bmatrix}$
▷ where $s^{ij} = Euclidean(s^i, s^j)$ for images i and j
$TS_{MDS} = \{TS^1, TS^2, \cdots, TS^c\} \leftarrow \mathbf{MDS}(M, c), S \in \mathbb{R}^{n \times c}$
▷ c is the reduced dimensionality of non-metric MDS solution
Output: typicality scores= TS^1_{MDS}

Algorithm S3 Pixel (SSIM)-based Typicality

Input: A set of images $I = \{i_1, i_2, \cdots, i_n\}$ in fast-food chain
structural similarity score matrix $S = \{s_{ij}\} \in 0^{n \times n}$
for each image $i \in I$ **do**
    for each image $j \in I$ **do**
        $s_{ij} \leftarrow SSIM(i, j)$ ▷ Structural similarity between images i and j via **skimage**
$(\mathbf{v}, \boldsymbol{\lambda}) \leftarrow eig(S)$
▷ computing the eigenvectors $\mathbf{v} = \{v^1, v^2, \cdots\}$ and the eigenvalues $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \cdots\}$
Output: typicality scores= v^1

References

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74, 829–836.

Kong, S., Shen, X., Lin, Z., Mech, R., Fowlkes, C., 2016. Photo aesthetics ranking network with attributes and content adaptation. *ECCV* 9905, 662–679.

Pieters, R., Wedel, M., Batra, R., 2010. The stopping power of advertising: Measures and effects of visual complexity. *Journal of Marketing* 74, 48–60.