Results report: Stimulus normativity and the erasure effect (preregistered secondary analysis)

Nov 1, 2026 · @Markus

1. Summary

This report gives the results of the preregistered secondary analysis registered at https://osf.io/uagy8, together with the amendment filed there before any analysis (project page: https://osf.io/e5dzg). The analysis asked whether the normativity of liking ratings places the stimulus classes with a robust erasure-with-confound effect (colors and foods) at intermediate values relative to the classes without one (Chinese characters, pictures, and tones).

The registered expectation was not met. Chinese characters and tones lie within the interval spanned by colors and foods, and only pictures lie outside it, on the more normative side. Four of the five classes share a very low normativity, with roughly 4 to 7 percent of the rating variance attributable to the stimuli, whereas pictures reach about 37 percent. The intermediate-normativity account therefore does not explain why colors and foods show the effect while Chinese characters and tones do not.

2. Data and procedure as executed

The analysis followed the registered plan as modified by the amendment; no further deviations occurred.

Data

The rating-level raw data of all 18 confounded limited-set studies of the meta-analysis were used: colors (7 studies), foods (4), tones (3), pictures (2), and Chinese characters (2). The tone study S10 entered as two studies (S10a, n = 499; S10b, n = 510), because its original data file also contained all participants of S11 and because it consisted of two successive data collections with independently drawn tone sets. The randomized studies were not used.

Normativity index

For each study, NI was computed as the one-way intraclass correlation of stimuli, $NI = \sigma^2_v / (\sigma^2_v + \sigma^2_e)$, estimated by restricted maximum likelihood in one model per study with condition (storage, erasure) as a fixed effect, so that the erasure effect itself does not enter the index. The one-way form was used for all studies because in ten studies each participant rated a single stimulus, which makes participants' general response level inseparable from the residual. In the erasure condition, the condition label served as stimulus identifier. For Maier and Dechamps (2026), Study 5, the identifiers of the erased stimuli were not recorded, so NI was estimated from the storage condition alone. Ninety-five percent intervals were obtained by a parametric bootstrap with 1,000 draws. No study had a stimulus variance at the boundary of zero, so the registered rule for that case was not needed.

Pooling and criterion

Study values were logit-transformed, pooled per class in a random-effects model with the bootstrap variances, and back-transformed. As registered, the pattern counts as consistent with the account if none of the three classes without an effect has its pooled point estimate inside the interval spanned by colors and foods, and as contradicting it if at least one does.

3. Results

The criterion was not met: two of the three classes without an effect lie inside the color-food interval.

Class-level normativity

Class	Studies	Pooled NI	95% CI	Position relative to the color-food interval [0.044, 0.073]
Food	4	0.044	[0.017, 0.106]	defines lower bound
Color	7	0.073	[0.048, 0.111]	defines upper bound
Chinese characters	2	0.054	[0.020, 0.136]	inside
Tones	3	0.072	[0.033, 0.152]	inside (just below the color value)
Pictures	2	0.366	[0.224, 0.536]	above

The tone estimate lies only marginally inside the interval and would count as fragile on its own. The result does not depend on it, because the Chinese characters lie clearly inside. The intervals of Chinese characters and tones overlap those of colors and foods almost completely; the interval of pictures does not overlap them.

Study-level values

Study	Class	Design	Stimuli	NI [95% CI]	g
MD25 S1	Color	between	6	0.088 [0.007, 0.225]	0.21
MD25 S2	Color	between	6	0.038 [0.000, 0.118]	0.50
MD25 S3	Color	between	12	0.147 [0.044, 0.273]	0.37
MD25 S4	Color	between	12	0.023 [0.003, 0.056]	0.14
MD26 S5	Color	within	18	0.084 [0.036, 0.145]	0.09
MD26 S6	Color	within	36	0.104 [0.059, 0.146]	0.18
MD26 S7	Color	within	36	0.047 [0.027, 0.073]	0.30
MVD2026a S1	Food	within	10	0.031 [0.000, 0.084]	0.65
MVD2026a S2a	Food	within	12	0.015 [0.003, 0.033]	0.50

Study	Class	Design	Stimuli	NI [95% CI]	g
MVD2026a S2b	Food	within	12	0.057 [0.013, 0.112]	0.29
MVD2026a S2c	Food	within	12	0.096 [0.026, 0.188]	0.26
Storch S5	Chinese characters	between	12	0.031 [0.000, 0.085]	0.29
Storch S6	Chinese characters	between	12	0.061 [0.015, 0.125]	−0.13
Storch S7	Pictures	between	12	0.394 [0.146, 0.577]	0.29
Maier & Vogel (2026)	Pictures	within	12	0.343 [0.148, 0.535]	−0.53
Storch S10a	Tones	between	12	0.049 [0.002, 0.117]	0.15
Storch S10b	Tones	between	12	0.082 [0.013, 0.174]	0.17
Storch S11	Tones	between	11	0.078 [0.016, 0.161]	−0.05

For MD26 S5 the stimuli are the 18 stored colors only. In S11 two condition labels used the same tone file, which was treated as one stimulus.

Robustness

- Alternative normativity measures. In the eight within-subjects studies, the crossed-model version that removes participants' response level and the model-free mean Spearman correlation between participants' rating profiles both correlate at r = .999 with NI and leave the ordering unchanged.
- Storage versus erasure condition. NI estimated separately within each condition agreed across the 17 studies with both conditions at r = .83 (Spearman ρ = .35; mean absolute difference 0.057). The agreement is driven by the clear separation of the picture studies; among the low values, the ordering varies considerably, as expected with three to six stimuli per condition.

Exploratory analyses

- Side of the null classes. Pictures lie on the normative side; Chinese characters and tones lie within the color-food range rather than on either side.
- NI and effect size. Within colors the rank correlation between study NI and Hedges' g is ρ = .11 (k = 7), within foods ρ = −.80 (k = 4), and across both classes ρ = −.36 (k = 11). There is no indication that intermediate normativity goes along with larger effects.
- Stimulus range. For the color studies, the spread of hue and lightness of the stored colors showed no discernible relation to NI. The food data contain no physical parameter, so this check was not possible for foods.
- Idiosyncratic share. Apart from three repeated presentations in one food study (Section 5), no study contained repeated ratings of the same stimulus by the same participant, so the idiosyncratic share could not be separated from measurement error, as anticipated in the registration.

4. Interpretation and limitations

The result speaks against the idea that the erasure effect requires an intermediate level of normativity in the sense measured here.

Normativity separates pictures from all other classes but does not separate the classes with an effect (colors, foods) from Chinese characters and tones. For pictures, whose liking was largely shared across participants, a position on the normative side remains possible, but this explains only one of the three null classes. As stated in the registration, the analysis is agnostic about whether the color and food effects reflect erasure itself or the chance-dependent stimulus selection that the confound permits, and the result does not bear on that question. Nor does it bear on the competing physical-parameter account, which distinguishes colors and foods by a property of the objective record rather than by the normativity of liking.

Three limitations qualify the result. First, with five classes and two to seven studies per class, the class estimates are imprecise, and the intervals of four classes overlap almost completely. Second, NI depends on how heterogeneous the chosen stimulus sets were, not only on how shared liking is, so class differences cannot be attributed to normativity alone. Third, in the between-subjects studies the index includes participants' general response level, which the within-subjects versions show to change the values only slightly. For these reasons the outcome is best read as a lack of support for the intermediate-normativity account rather than as its definitive refutation.

The prediction band registered for future work (the NI range of colors and foods, 0.044 to 0.073) is retained as reported, but its diagnostic value is limited, since Chinese characters and tones fall inside it without showing an effect.

5. Files and reproducibility

All results can be regenerated from the raw data with the scripts listed below.

- Raw data: the 18 confounded study files, named as in the meta-analysis, in the folder "confounded" of the raw-data archive, with a README listing participants and MD5 checksums per file.
- NI_load.py reads the raw files into a common long format (participant, stimulus, condition, rating).
- NI_estimate.py computes NI per study by restricted maximum likelihood, the parametric bootstrap intervals, the within-condition estimates, and, for within-subjects studies, the crossed-model and Spearman measures.
- NI_pool.py pools the study values per class and applies the registered criterion.
- NI_results_by_study.csv, NI_results_by_class.csv, and NI_results_figure.png contain the reported numbers and the figure.

One data detail is not covered above. In Maier, Vogel and Dechamps (2026a), Study 1, stimuli were drawn with replacement, so that three participants were shown one food image twice and another not at all; the two ratings of the repeated image were averaged. The bootstrap uses a fixed random seed (20261101), so the reported intervals are reproducible.