On the size of the model selection problem in a conjoint measurement experiment

Kenneth Knoblauch[1,2*] and Frédéric Devinck[3]

[1*]Stem Cell and Brain Research Institute, Inserm 1208, 8 avenue du Doyen Lépine, Bron, 69500, France.
[2]Université Lyon, Université Claude Bernard Lyon 1, Villeurbanne, 69100, France.
[3]Univ Rennes, Université Rennes 2, LP3C, Rennes, 35000, France.

*Corresponding author(s). E-mail(s): ken.knoblauch@inserm.fr;
Contributing authors: frederic.devinck@univ-rennes2.fr;

Abstract

We examine how the number of models respecting marginality in model selection using Maximum Likelihood Conjoint Measurement (MLCM) varies with the number of dimensions tested and explain the logic for determining this number. The results are related to Dedekind numbers, which increase super-exponentially, and, thus, suggest a limit to the number of dimensions that can be practically tested using the MLCM paradigm.

Keywords: conjoint measurement, scaling, model selection, ordinal judgements

1 Introduction

How organisms integrate features along different sensory dimensions and across modalities is a fundamental problem in perception that is often studied in the context of cue combination (Ernst & Banks, 2002; Oruç, Maloney, & Landy, 2003; Young, Landy,

& Maloney, 1993). Such experiments often require the subject to match or estimate sensory magnitudes (e.g., Badde, Navarro, & Landy, 2020). Maximum Likelihood Conjoint Measurement (MLCM) provides an alternative approach in which observers only make ordinal judgments between stimulus pairs and in which perceptual scales associated with multiple stimulus dimensions can be assessed within a single experiment. The method has seen extensive use recently in vision experiments that deal with topics as diverse as perception of surface and illuminance properties (Aguilar & Maertens, 2018, 2022; Chadwick, Cox, Smithson, & Kentridge, 2018; Emrith, Chantler, Green, Maloney, & Clarke, 2010; Hansmann-Roth & Mamassian, 2017, 2025; Harvey & Smithson, 2021; Ho, Landy, & Maloney, 2008), color perception in adults (Devinck & Knoblauch, 2023, 2025; Rogers, Knoblauch, & Franklin, 2016; Xie & Fairchild, 2023) and infants (Rogers, Franklin, & Knoblauch, 2018), texture regularity (Sun, St-Amand, Baker Jr, & Kingdom, 2021), blur perception in color vision (Kingdom, Bell, Haddad, & Bartsch, 2015), face perception in aging (Fitousi, 2021), time constancy (Lisi & Gorea, 2016), perception of numeric magnitude (Fitousi, 2022) and multimodal integration (Abbatecola, Gerardin, Beneyton, Kennedy, & Knoblauch, 2021; Abbatecola, Kennedy, & Knoblauch, 2025).

While estimation of sensory magnitudes is rapid and often generates results in agreement with those based on ordinal judgments (e.g., Devinck & Knoblauch, 2023, 2025), ordinal judgments are generally easier for subjects to perform, require minimal task demands (Firestone & Scholl, 2016) and constrain underlying perceptual scales with fewer assumptions on the observer's decision processes and on the properties of the underlying scales (Victor, Aguilar, & Waraich, 2025). MLCM is an extension of Conjoint Measurement (CM) that was orignially formulated to demonstrate that continuous scales can be derived from psychological judgments (Luce & Tukey, 1964). Specifically, MLCM extends CM to include a model of error that the original theory was lacking, to facilitate its use empirically.

The psychophysical paradigm involves ordinal comparisons between pairs of stimuli that covary independently along multiple stimulus dimensions. The underlying perceptual scales and their mutual influences are estimated to maximize the probability of the observer's choices across all experimental trials using a maximum likelihood procedure that incorporates a signal detection model of the observer's decision process (Green & Swets, 1966/1974). While the probability model depends on binary choices (paired-comparisons), the set of decision models forms a nested hierarchy equivalent to that obtained in a d–way factorial analysis of variance, including main effects for each dimension and their interactions potentially over multiple orders. Here, we describe how the number of dimensions manipulated as independent variables determines the number of nested model comparisons available to test.

2 Results

In our recent article (Devinck & Knoblauch, 2025), we presented an equation for calculating the maximum number of models that might be tested in a d–dimensional comparison in an analysis of data using MLCM that provides correct values for the small models that we analyzed in our paper, i.e., using 2 or 3 stimulus dimensions, but not for models using $d > 3$ stimulus dimensions.

The equation that we gave for the number, N, of possible models respecting marginality conditions is

$$N(d) = \sum_{k=0}^{d-1} \binom{d}{k} (2^{d-k} - 1) \tag{1}$$

where d is the number of dimensions tested. For values of d equal to 2 and 3, the equation yields the correct values of 5 and 19, respectively, but not for higher values. For example, the equation gives the values of 65, 211 and 665 for $d = 4$, 5 and 6 respectively, whereas the correct values are 167, 7580 and 7828353. We verified the

correct value for $d = 4$ by brute force using the dredge function from the package **MuMIn** (Bartoń, 2025) in the openSource software R (R Core Team, 2025) that generates all models respecting marginality for automatic model selection. This function only works, however, for models having up to 4 main effects. To understand the rapid growth of the correct values requires additional explanation.

The question is, given a d–factor experiment, how many nested model sequences can be evaluated (see also Nardi & Rinaldo, 2012, for a discussion of this issue in the context of log-linear models.)? Specifically, in nested likelihood ratio testing, a model is evaluated against a similar model with one term from the higher order model removed or set to 0 with all other marginal terms retained (Knoblauch & Maloney, 2012; Wilks, 1938). For example, the last column of Table 1 lists all possible models respecting marginality for a 2–factor experiment with factors a and b. The 4 model terms are indicated in the heading with the 1 indicating an intercept and the term a:b a 2–way interaction. The significance of the 2–way interaction can be tested against the additive model by comparing the models in rows 5 and 6. The additive model contains all of the terms in the column heading except the interaction. Similarly, the additive model can be compared with either of the 1–factor models (row 3 or 4), and each of the 1–factor models with the intercept (row 1). The model relations might also be depicted in a hierarchical graph or Hasse diagram with the two main effects models (rows 3 and 4) on the same level (Figure 1), illustrating the marginality relations more clearly. An example of a model comparison not respecting marginality would be comparing the model 1 + a + b + a:b with 1 + b + a:b. The a term marginal to the interaction is missing. Note that there are 5 such nested models that can be described. In a 3-factor model there would be 19, as we illustrated previously (Devinck & Knoblauch, 2025). With so many model comparisons to perform, the issue of conducting multiple hypothesis tests also arises. We treat that problem elsewhere (Devinck & Knoblauch, 2025).

Table 1 Boolean sequences for inclusion/exclusion of terms for all models of a 2–way factorial experiment. The 0/1 in columns 2–5 indicate the absence/presence of a model term and the Model column indicates the model specification for that row using R formula notation. All pairs are monotone except rows 3 and 4. However, these are monotone with the other rows above and below them.

	1	a	b	a:b	Model
1	0	0	0	0	~ 0
2	1	0	0	0	~ 1
3	1	1	0	0	$\sim 1 + a$
4	1	0	1	0	$\sim 1 + b$
5	1	1	1	0	$\sim 1 + a + b$
6	1	1	1	1	$\sim 1 + a + b + a{:}b$

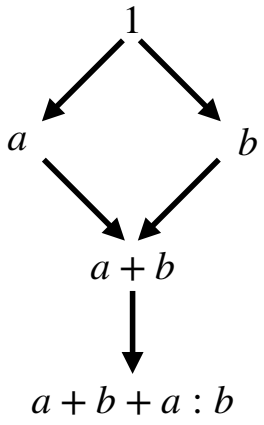


Fig. 1 Hasse diagram for the hierarchical arrangement of models from a 2–d MLCM experiment

The 0's and 1's in Table 1 are Boolean variables indicating the absence or presence, respectively, of a model term. For models to be nested, the columns must follow a monotone Boolean sequence. This means that within a column, passing from row to row, the sequence can be 0 to 0, 0 to 1 or 1 to 1 but not 1 to 0. For a pair of models

to be nested, all columns must be monotone between the two rows. Thus, rows 3 and 4 do *not* form a monotone sequence, but 3 and 5 and 4 and 5 do. Such sequences are counted by Dedekind numbers (Dedekind, 1897), and they grow super-exponentially with the number of factors. Interestingly, the highest Dedekind number computed until now is 9, contains 42 digits and has been determined only recently (Jäkel, 2023; Van Hirtum et al., 2024).

The monotone Boolean sequences include the model in which all terms equal 0, row 1, but such a model is only exceptionally of interest in statistical comparisons. With MLCM, the null model corresponds to the situation for which the probability of a choice is independent of all of the dimensions, i.e., row 2 in Table 1. Thus, the number of models in which we are interested is one less than the Dedekind number or 5 in this case.

Kisielewicz (1988) derived the following equation to calculate Dedekind numbers

$$M(d) = \sum_{k=1}^{2^{2^d}} \prod_{j=1}^{2^d-1} \prod_{i=0}^{j-1} \left(1 - b_i^k b_j^k \prod_{m=0}^{\log_2 i} (1 - b_m^i + b_m^i b_m^j) \right), \tag{2}$$

where b_i^k is the i^{th} bit of the number k in base 2 and the $\log_2$ value of the upper limit of the innermost product is truncated to the nearest integer. However, the formula is only useful for calculating low values. For example, using an inline C++ script embedded in R via the package **Rcpp** (Eddelbuettel, 2013) (shown in the code below), the formula is accurate for values up to $d = 4$, afterwhich the upper limit of the summation overflows the size of the default 32-bit integer and yields incorrect values.

```
library(Rcpp)
cppFunction(
'double DdkndNum(int n) {
    double ksum = 0.0;
    int k_max = 1 << (1 << n);
```

```
    int j_max = (1 << n) - 1;

    for (int k = 1; k <= k_max; ++k) {
        double jprod = 1.0;
        for (int j = 1; j <= j_max; ++j) {
            double iprod = 1.0;
            for (int i = 0; i < j; ++i) {
                int ii = (i == 0) ? 1 : i;
                double mprod = 1.0;
                for (int m = 0; m <= floor(log2(ii)); ++m) {
                    mprod *= (1 - bik(m, i) + bik(m, i) * bik(m, j));
                }
                iprod *= (1 - bik(i, k) * bik(j, k) * mprod);
            }
            jprod *= iprod;
        }
        ksum += jprod;
    }
    return ksum;
}
', includes = 'int bik(int i, int k) {
    return (k / (1 << i)) % 2;
}')
```

3 Discussion

Considering the combinatorial complexity demonstrated here, MLCM experiments have mostly employed just 2 dimensions, but a few have used 3 (Devinck & Knoblauch, 2025; Hansmann-Roth, 2026; Sun et al., 2021). Thus, the issues considered above do not affect the overall results or interpretations in our paper. Given the number of

possible models to evaluate, however, it is unlikely that 4 or more dimensions would be attempted (the exception being, perhaps, if there are only a small number of *a priori* hypotheses of interest), but in the spirit of general understanding and giving pause to researchers planning large MLCM experiments, we thought that it would be of interest and value to describe the logic of the more general solution to this problem. The full list of known Dedekind numbers is available on the Wikipedia page for Dedekind number (https://en.wikipedia.org/wiki/Dedekind_number).

Acknowledgements. K. K. was supported by LABEX CORTEX (ANR- 11-LABX-0042) of Université de Lyon (ANR-11-IDEX-0007) operated by the French National Research Agency. We thank J. S. Werner and L. T. Maloney for feedback on an earlier draft of this text.

Declarations

- Funding: LABEX CORTEX (ANR- 11-LABX-0042) of Université de Lyon (ANR-11-IDEX-0007) operated by the French National Research Agency.
- Conflict of interest/Competing interests: None
- Ethics approval and consent to participate: Not applicable
- Consent for publication: Not applicable
- Data availability: Not applicable
- Materials availability: Not applicable
- Code availability: Not applicable
- Author contribution: Both authors contributed to the conception and editing of the manuscript

References

Abbatecola, C., Gerardin, P., Beneyton, K., Kennedy, H., Knoblauch, K. (2021). The role of unimodal feedback pathways in gender perception during activation of

voice and face areas. *Frontiers in Systems Neuroscience*, *15*, 669256.

Abbatecola, C., Kennedy, H., Knoblauch, K. (2025). Directed attention influences optimality of top-down and bottom-up multi-modal perceptual integration. *Scientific Reports*, *15*(1), 25742.

Aguilar, G., & Maertens, M. (2018). Using conjoint measurement to estimate scales of perceived surface lightness. *Journal of Vision*, *18*(10), 177–177.

Aguilar, G., & Maertens, M. (2022). Conjoint measurement of perceived transparency and perceived contrast in variegated checkerboards. *Journal of Vision*, *22*(2), 2–2.

Badde, S., Navarro, K.T., Landy, M.S. (2020). Modality-specific attention attenuates visual-tactile integration and recalibration effects by reducing prior expectations of a common source for vision and touch. *Cognition*, *197*, 104170.

Bartoń, K. (2025). MuMIn: Multi-model inference [Computer software manual]. (R package version 1.48.11, https://CRAN.R-project.org/package=MuMIn)

Chadwick, A.C., Cox, G., Smithson, H.E., Kentridge, R.W. (2018). Beyond scattering and absorption: Perceptual unmixing of translucent liquids. *Journal of Vision*, *18*(11), 18–18.

Dedekind, R. (1897). Über zerlegung von zahlen durch ihre grössten gemeinsamen theiler. *Gesammelte Werke*, *2*, .

Devinck, F., & Knoblauch, K. (2023). Color appearance of spatial patterns compared by direct estimation and conjoint measurement. *Journal of the Optical Society of America A*, *40*(3), A99–A106.

Devinck, F., & Knoblauch, K. (2025). Warm/cool judgments as a function of hue, value, and chroma. *Journal of the Optical Society of America A*, *42*(5), B68–B75.

Eddelbuettel, D. (2013). *Seamless R and C++ integration with Rcpp*. New York: Springer. (ISBN 978-1-4614-6867-7)

Emrith, K., Chantler, M., Green, P., Maloney, L., Clarke, A. (2010). Measuring perceived differences in surface texture due to changes in higher order statistics. *Journal of the Optical Society of America A*, *27*(5), 1232–1244.

Ernst, M.O., & Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.

Firestone, C., & Scholl, B.J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, *39*, e229.

Fitousi, D. (2021). How facial aging affects perceived gender: Insights from maximum likelihood conjoint measurement. *Journal of Vision*, *21*(12), 12–12.

Fitousi, D. (2022). Conjoint measurement of physical size and numerical magnitude: Numerals do not automatically activate their semantic meaning. *Psychonomic Bulletin & Review*, *29*(1), 134–144.

Green, D.M., & Swets, J.A. (1966/1974). *Signal Detection Theory and Psychophysics*. Huntington: Robert E. Krieger Publishing Company. (ISBN 978-0932146236)

Hansmann-Roth, S. (2026). The influence of specular reflectance, albedo, and shape on perceived gloss: A case for three-dimensional maximum likelihood conjoint measurement (mlcm). *i-Perception*, *17*(2), 20416695261435363.

Hansmann-Roth, S., & Mamassian, P. (2017). A glossy simultaneous contrast: Conjoint measurements of gloss and lightness. *i-Perception*, *8*(1), 2041669516687770.

Hansmann-Roth, S., & Mamassian, P. (2025). Perceiving gloss through transparency. *i-Perception*, *16*(4), 20416695251355381.

Harvey, J.S., & Smithson, H.E. (2021). Low level visual features support robust material perception in the judgement of metallicity. *Scientific Reports*, *11*(1), 16396.

Ho, Y.-X., Landy, M.S., Maloney, L.T. (2008). Conjoint measurement of gloss and surface texture. *Psychological Science*, *19*(2), 196–204.

Jäkel, C. (2023). A computation of the ninth Dedekind number. *Journal of Computational Algebra*, *6*, 100006.

Kingdom, F.A.A., Bell, J., Haddad, C., Bartsch, A. (2015). Perceptual scales for chromatic and luminance blur in noise textures. *Journal of Vision*, *15*(9), 6-6, https://doi.org/10.1167/15.9.6

Kisielewicz, A. (1988). A solution of Dedekind's problem on the number of isotone boolean functions. *Journal für die Reine und Angewandte Mathematik*, *386*, 139–144.

Knoblauch, K., & Maloney, L.T. (2012). *Modeling Psychophysical Data in R* (Vol. 32). New York: Springer Science & Business Media. (ISBN 978-1-4614-4474-9)

Lisi, M., & Gorea, A. (2016). Time constancy in human perception. *Journal of Vision*, *16*(14), 3–3.

Luce, R.D., & Tukey, J.W. (1964). Simultaneous conjoint measurement: a new scale type of fundamental measurement. *Journal of Mathematical Psychology*, *32*, 466–473.

Nardi, Y., & Rinaldo, A. (2012). The log-linear group-lasso estimator and its asymptotic properties. *Bernoulli*, *18*, 945–974, https://doi.org/10.3150/11-BEJ364

Oruç, I., Maloney, L.T., Landy, M.S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*(23), 2451–2468.

R Core Team (2025). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. (https://www.R-project.org/)

Rogers, M., Franklin, A., Knoblauch, K. (2018). A novel method to investigate how dimensions interact to inform perceptual salience in infancy. *Infancy*, *23*(6), 833–856.

Rogers, M., Knoblauch, K., Franklin, A. (2016). Maximum likelihood conjoint measurement of lightness and chroma. *Journal of the Optical Society of America A*, *33*(3), A184–A193.

Sun, H.-C., St-Amand, D., Baker Jr, C.L., Kingdom, F.A. (2021). Visual perception of texture regularity: Conjoint measurements and a wavelet response-distribution model. *PLoS Computational Biology*, *17*(10), e1008802.

Van Hirtum, L., De Causmaecker, P., Goemaere, J., Kenter, T., Riebler, H., Lass, M., Plessl, C. (2024). A computation of the ninth Dedekind number using FPGA supercomputing. *ACM Transactions on Reconfigurable Technology and Systems*,

17(3), 1–28.

Victor, J.D., Aguilar, G., Waraich, S.A. (2025). Ordinal characterization of similarity judgments. *Mathematical Neuroscience and Applications*, *5*, 1–64, https://doi.org/10.46298/mna.12457

Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, *9*(1), 60–62.

Xie, H., & Fairchild, M.D. (2023). Representing color as multiple independent scales: brightness versus saturation. *Journal of the Optical Society of America A*, *40*(3), 452–461.

Young, M.J., Landy, M.S., Maloney, L.T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research*, *33*(18), 2685–2696.