Mixed-effects Maximum Likelihood Difference Scaling

Kenneth Knoblauch
Inserm U 846
Stem Cell and Brain Research Institute
Dept. Integrative Neurosciences
Bron, France

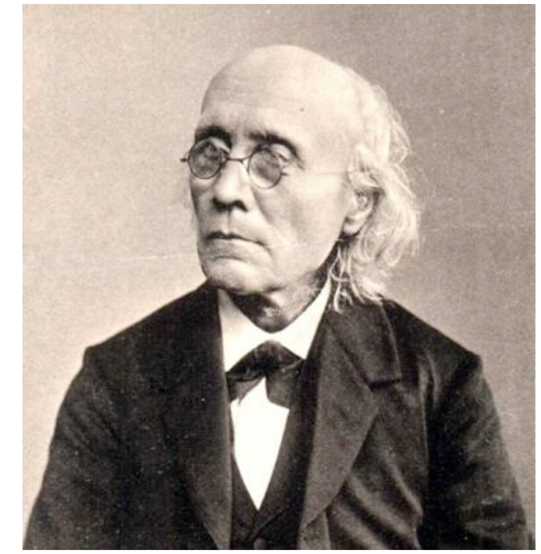
Laurence T. Maloney
Department of Psychology
Center for Neural Science
New York University
New York, NY USA

Stem-cell
and Brain Research Institute

NEW YORK UNIVERSITY

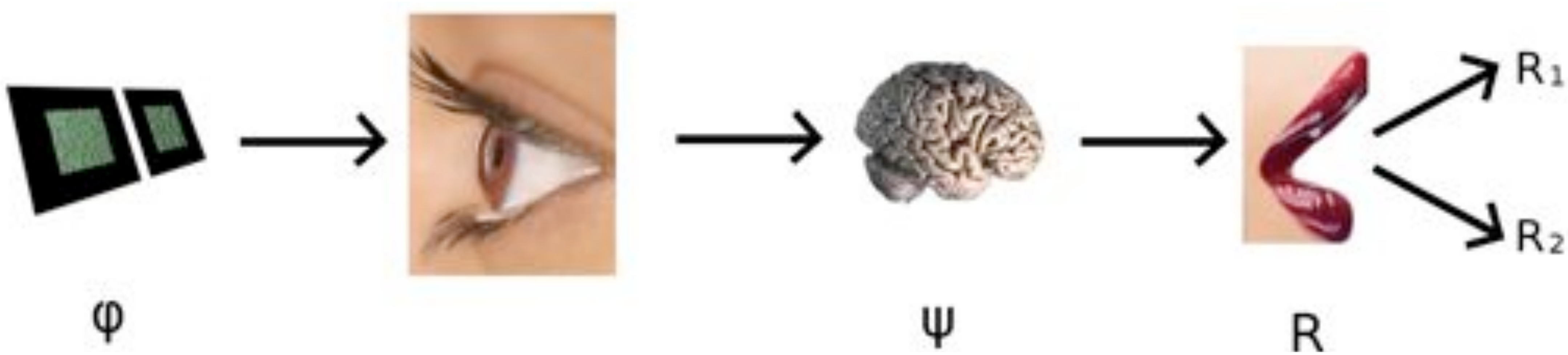
Instituts thématiques
Inserm
Institut national
de la santé et de la recherche médicale

Psychophysics,
qu'est-ce que c'est ?



Gustav Fechner (1801 - 1887)

A body of techniques and analytic methods to study the relation between physical stimuli and the organism's (classification) behavior to infer internal states of the organism or their organization.



The Design of Experiments

By

Sir Ronald A. Fisher, Sc.D., F.R.S.

II

THE PRINCIPLES OF EXPERIMENTATION, ILLUSTRATED BY A PSYCHO-PHYSICAL EXPERIMENT

Difference scaling is a psychophysical procedure used to estimate a perceptual (interval) scale for stimuli distributed along a physical continuum.

Example: VQ compressed images,

Up to what compression rate can the observer detect no loss of image quality?

Charrier, Maloney, Cherifi & Knoblauch, (2007) *J Opt Soc Am A*, **24**, 3418-3426

Difference Scaling: Experimental Procedure

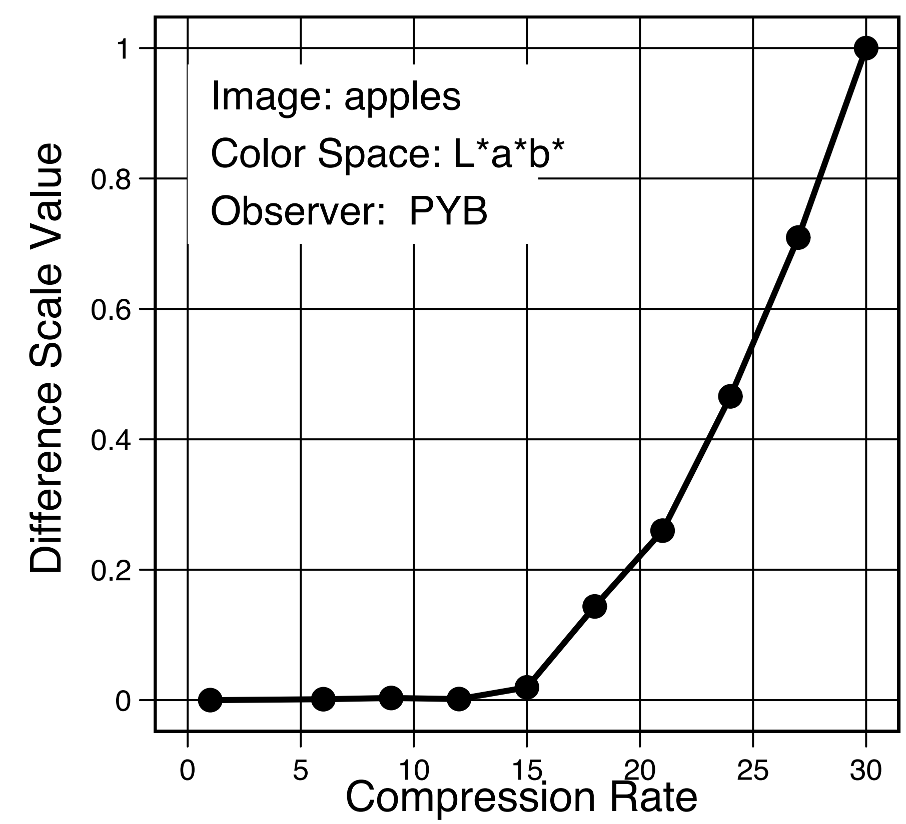
From a set of p stimuli, $\{I_1 < I_2 < \ldots < I_p\}$,

a random quadruple, $\{I_a, I_b\,;\, I_c, I_d\}$,
is chosen (w/out replacement) and presented
to the observer as in this example, on each trial:

Between which pair (upper/lower) is the perceived difference greatest?

The aim of the Maximum Likelihood Difference Scaling (MLDS) procedure is to estimate scale values, $(\psi_1, \psi_2, \ldots, \psi_p)$, that best capture the observer's judgments of the perceptual difference between the stimuli in each pair.

The MLDS package, available on CRAN, provides tools for performing this analysis in R. An example scale obtained from an observer for the "apples" sequence of VQ compressed images is shown on the right:



The decision model

Given a quadruple, $\mathbf{q} = (a, b\,;\, c, d)$,
from a single trial, we assume that the observer chooses the upper pair to be further apart when

$$\Delta(a, b\,;\, c, d) = |\psi_d - \psi_c| - |\,\psi_b - \psi_a| + \epsilon > 0,$$

where ψ_i are estimated scale values, $\epsilon \sim \mathcal{N}(0, \sigma^2)$
and σ a scale factor.

Estimation of Scale Values

Maloney and Yang (2003) used a direct method for estimating the maximum likelihood scale values,

$$L(\mathbf{\Psi}, \sigma) = \prod_{k=1}^{n} \Phi\left(\frac{\delta\left(\mathbf{q}^k\right)}{\sigma}\right)^{1-R_k} \left(1 - \Phi\left(\frac{\delta\left(\mathbf{q}^k\right)}{\sigma}\right)\right)^{R_k}$$

where

$$\Psi = (\psi_2, \psi_3, \ldots, \psi_{p-1})$$

$$\delta(\mathbf{q}^k) = |\psi_d - \psi_c| - |\psi_b - \psi_a|$$

Φ is the cumulative standard Gaussian (a probit analysis)

R_k is 0/1 if the judgment is lower/upper

$\psi_1 = 0, \psi_p = 1$ for identifiability,

leaving $p - 1$ parameters to estimate

Maloney LT, Yang JN (2003). "Maximum Likelihood Difference Scaling." *Journal of Vision*, **3**(8), 573–585. URL http://www.journalofvision.org/3/8/5.

Estimation of Scale Values

The problem can also be conceptualized as a GLM.

Each level of the stimulus is treated as a covariate in the model matrix, taking on values of 0 or ± 1 in the design matrix,

depending on the presence of the stimulus in a trial and its weight in the decision variable, with absolute value signs removed.

	resp	S1	S2	S3	S4
1	0	4	8	2	3
2	1	2	3	6	11
3	1	2	6	7	10
4	0	4	11	1	2
5	0	9	11	7	8
6	0	7	10	1	3

$$\begin{matrix} p_1 & p_2 & p_3 & p_4 & p_5 & p_6 & p_7 & p_8 & p_9 & p_{10} & p_{11} \end{matrix}$$

$$\begin{pmatrix} 0 & 1 & -1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

For model identifiability, we drop the first column (fixing $\psi_1 = 0$ and $\sigma = 1$).

Knoblauch & Maloney (2008) *J. Stat. Software*, **25**, 1-25

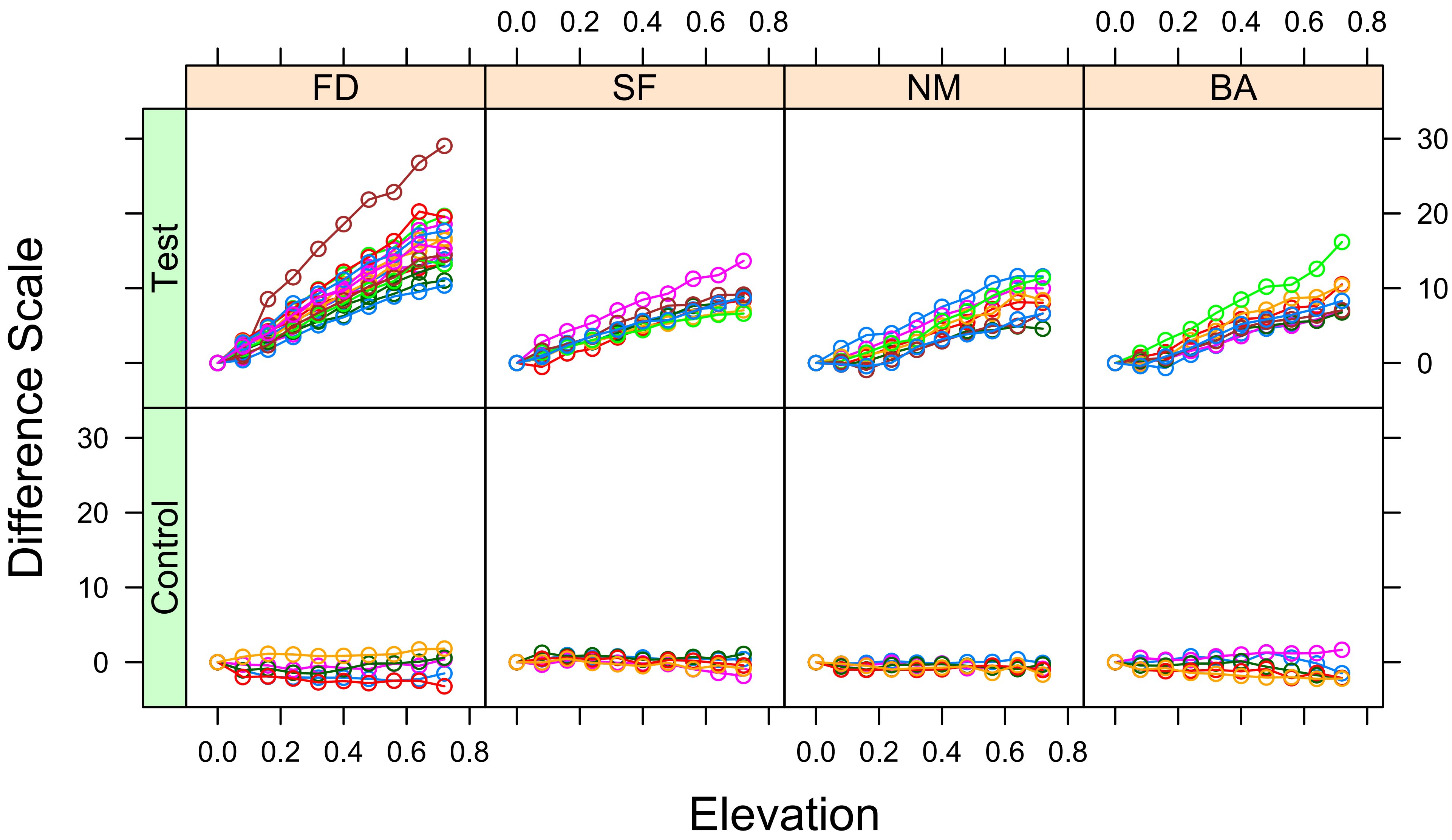
Estimation of Scale Values

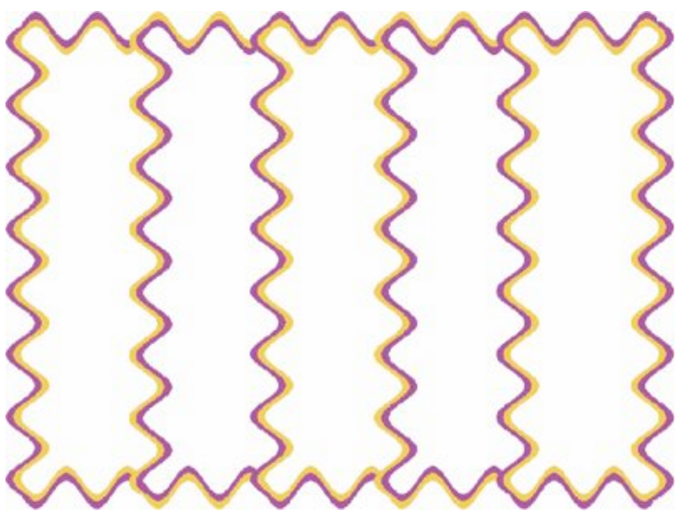
```
> kk.ix <- make.ix.mat(kk)
> head(kk.ix)
  resp stim.2 stim.3 stim.4 stim.5 stim.6 stim.7 stim.8 stim.9 stim.10 stim.11
1    1      1      0     -1      0     -1      0      1      0       0       0
2    1      0      0     -1      0     -1      0      0      1       0       0
3    1      1     -1      0      0      0     -1      0      1       0       0
4    1      1      0      0     -1     -1      1      0      0       0       0
5    0      0     -1      0      0     -1      1      0      0       0       0
6    0      0      0      0     -1     -1      0      0      1       0       0
```

$$\eta\left(\mathrm{E}\left[Y\right]\right) = X\beta$$

```
> glm(resp ~ . - 1, family = binomial( "probit" ), data = kk.ix)
```

Can the MLDS analysis be extended within a mixed-effects modeling framework to account for random sensitivity variations across sessions and across observers?





Difference Scaling of the Watercolor Illusion
Devinck & Knoblauch (in preparation)

Mixed-effects models with MLDS

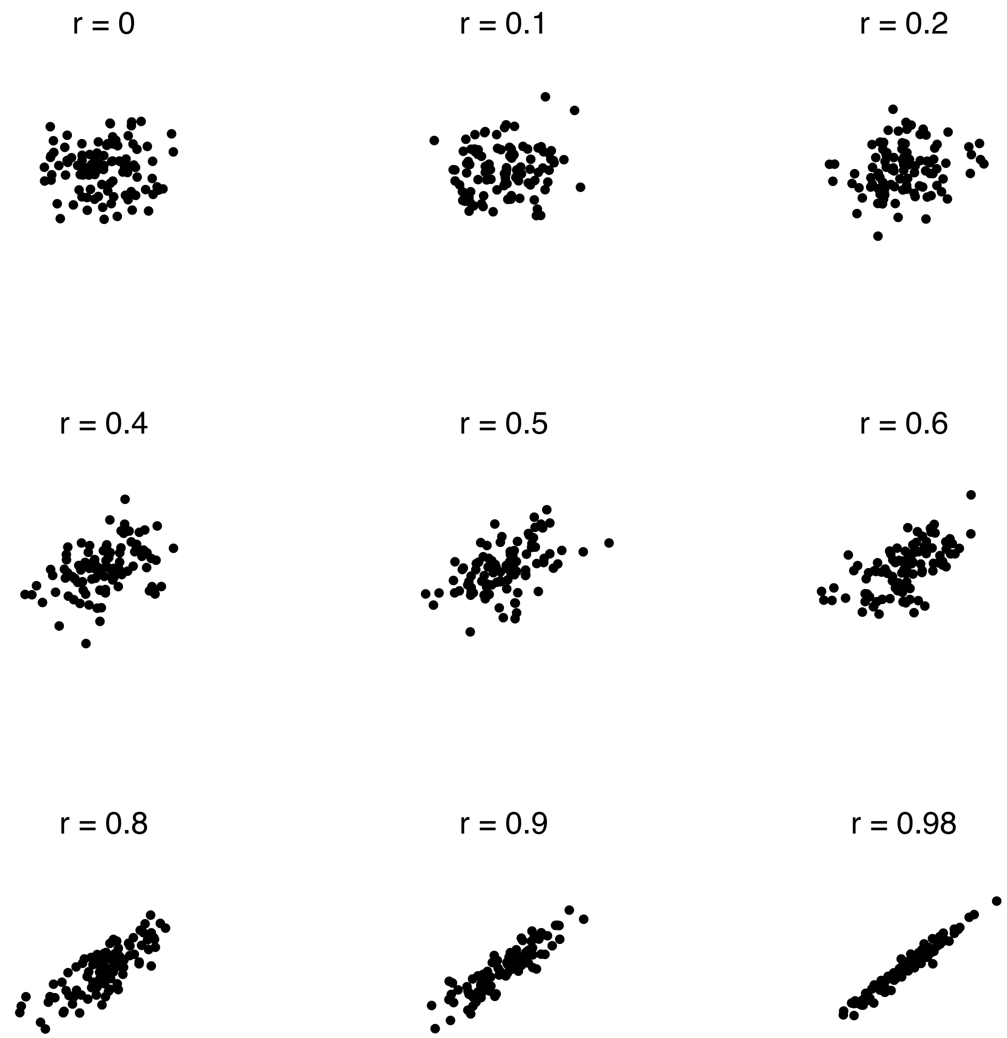
Three Strategies

1. Re-parameterize in terms of parametric decision variable

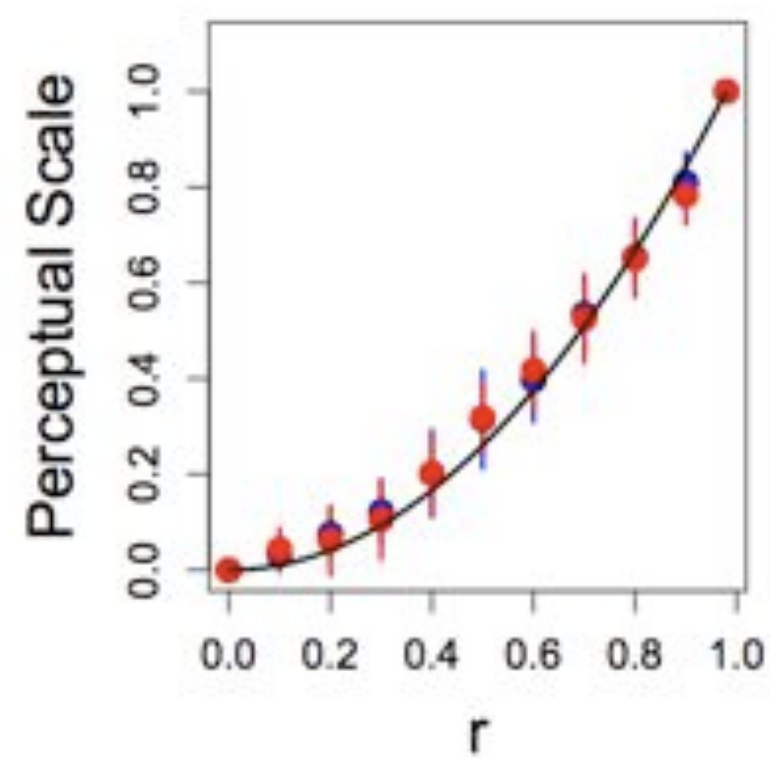
2. Normalize to common scale

3. Regression on estimated coefficients

Difference Scaling: Correlation in scatterplots



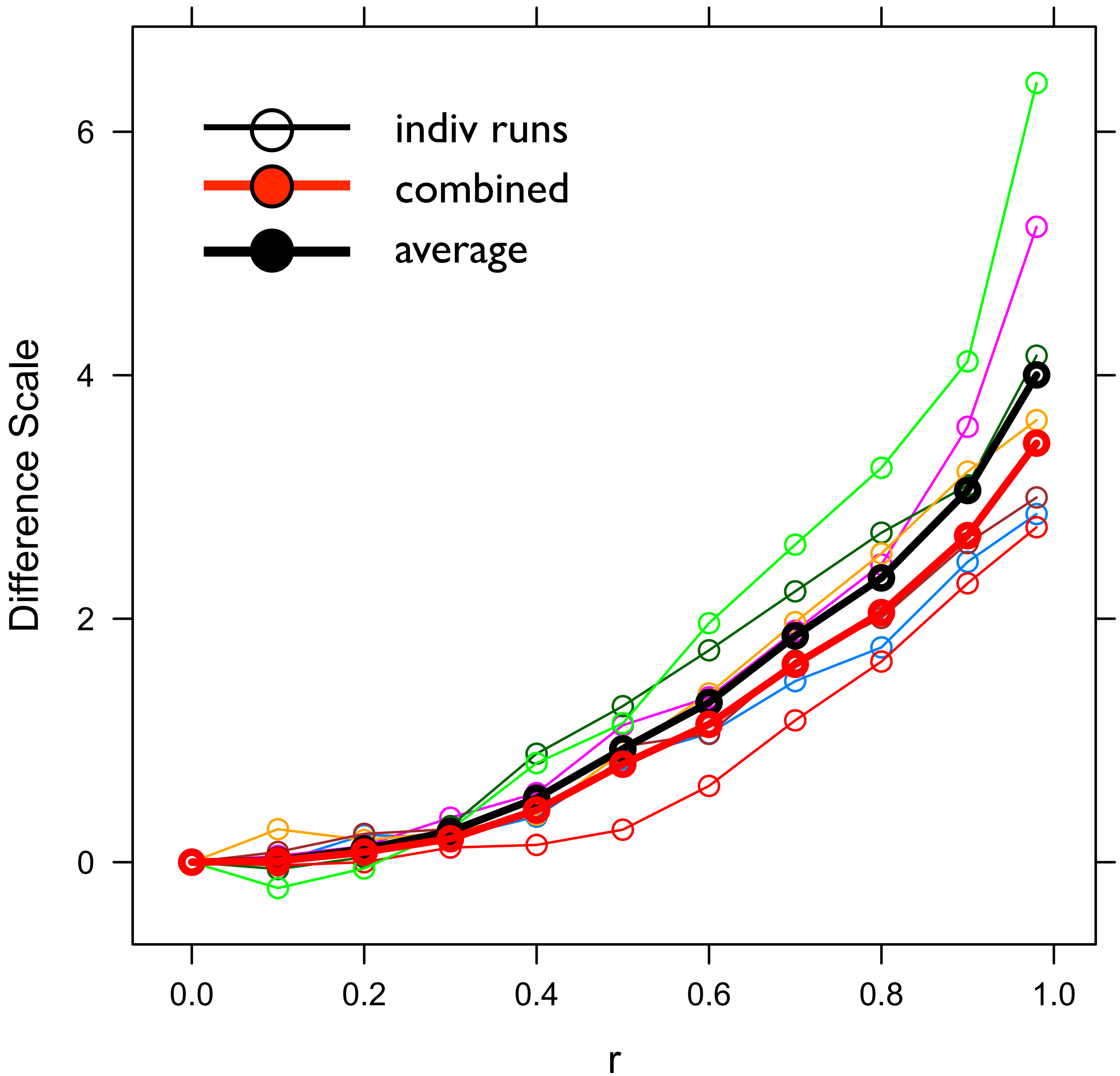
$$\psi(r) \approx r^2$$



Knoblauch & Maloney, *J. Stat. Soft.*, **25**, 1 - 26.

Mixed-effects models with MLDS:

Re-parameterize in terms of decision variable



$$\Delta = \psi_d - \psi_c - \psi_b + \psi_a$$

re-parameterized as empirical decision variable:

$$\mathcal{DV} = \rho_d^2 - \rho_c^2 - \rho_b^2 + \rho_a^2$$

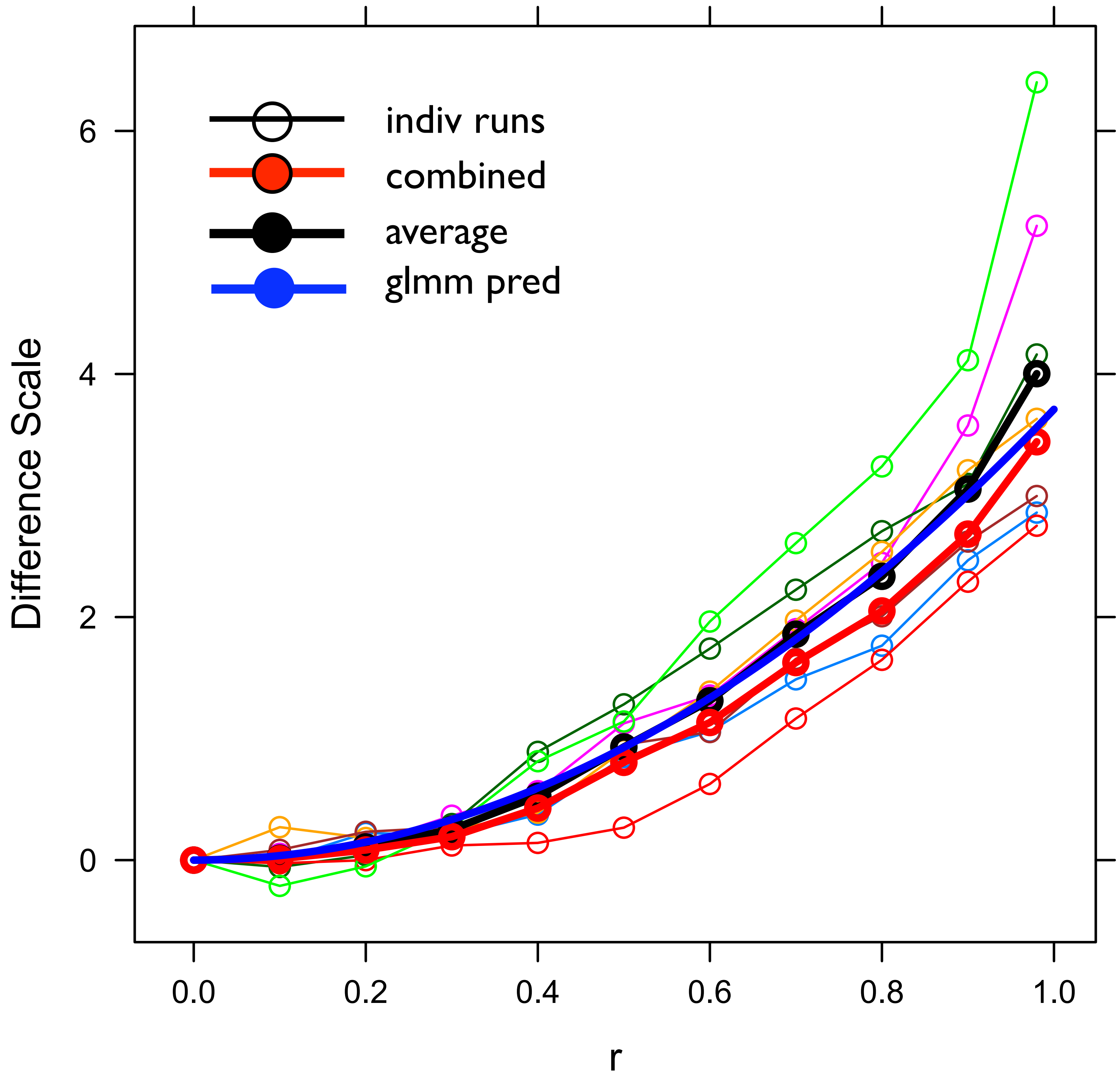
then, fit GLMM

$$\Phi^{-1}(\mathsf{E}[Y]) = (\beta + b_i)\mathcal{DV},$$
$$b \sim \mathcal{N}(0, \sigma^2)$$

```
  resp S1 S2 S3 S4 Obs    DV OL
1    0  8 11  4  7  S1 -0.20  1
2    1  1  3  5 10  S1  0.61  2
3    1  9 10  1  8  S1  0.32  3
4    0  6 11  3  4  S1 -0.66  4
5    1  9 10  5  7  S1  0.03  5
6    1  8 10  3  6  S1 -0.11  6

> library(lme4)
    . . .
> gm1 <- glmer( resp ~ DV + (DV + 0 | Obs) + 0,
       allraw.df, binomial(probit) )

> summary( gm1 )

Generalized linear mixed model fit by the Laplace approximation
Formula: resp ~ DV + (DV + 0 | Obs) + 0
   Data: allraw.df
  AIC  BIC logLik deviance
 4304 4317  -2150     4300
Random effects:
 Groups Name Variance Std.Dev.
 Obs    DV   0.30811  0.55507
Number of obs: 4620, groups: Obs, 7

Fixed effects:
   Estimate Std. Error z value Pr(>|z|)
DV   3.7092     0.2345   15.82   <2e-16 ***
```

Mixed-effects models with MLDS:

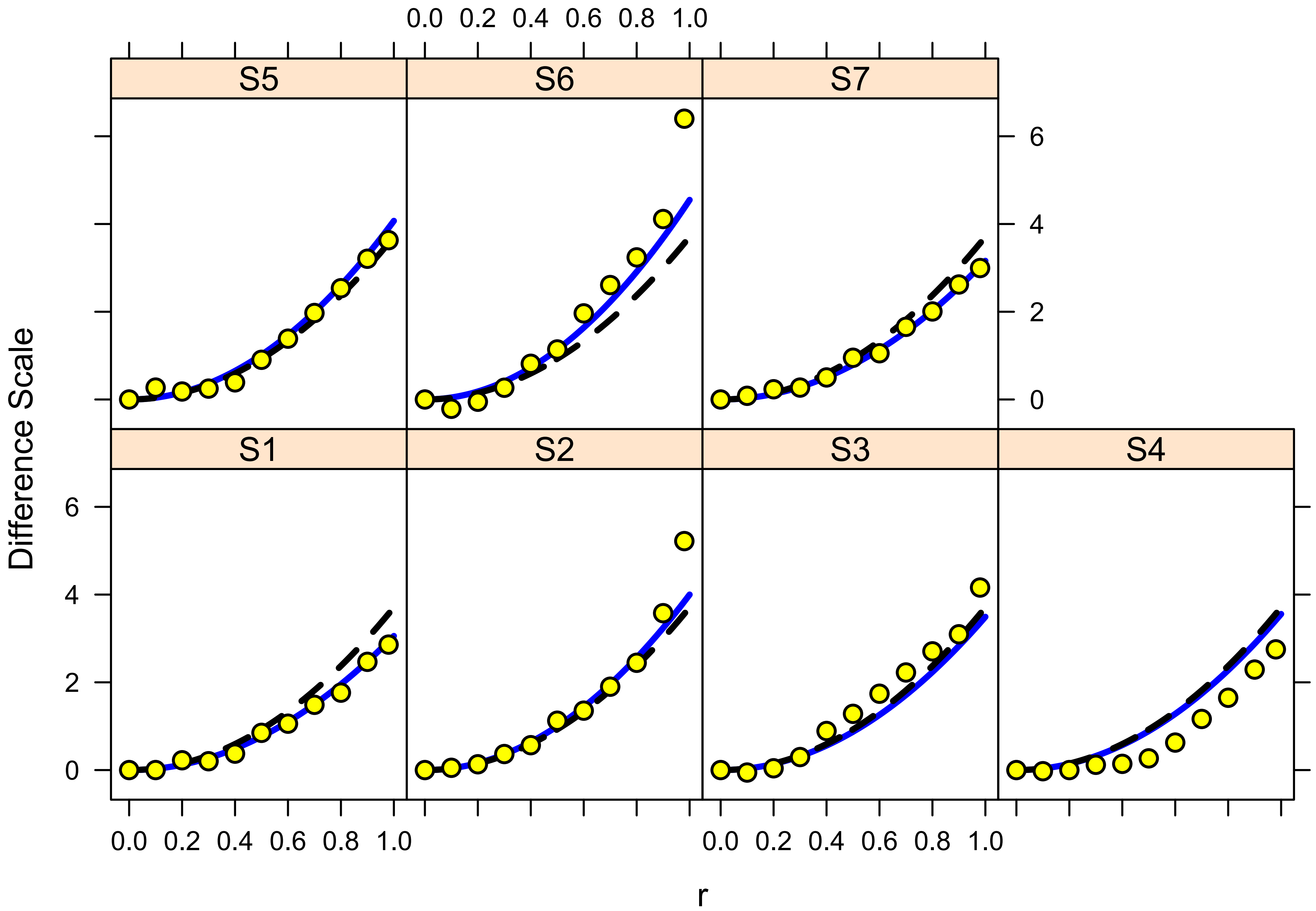
Re-parameterize in terms of decision variable



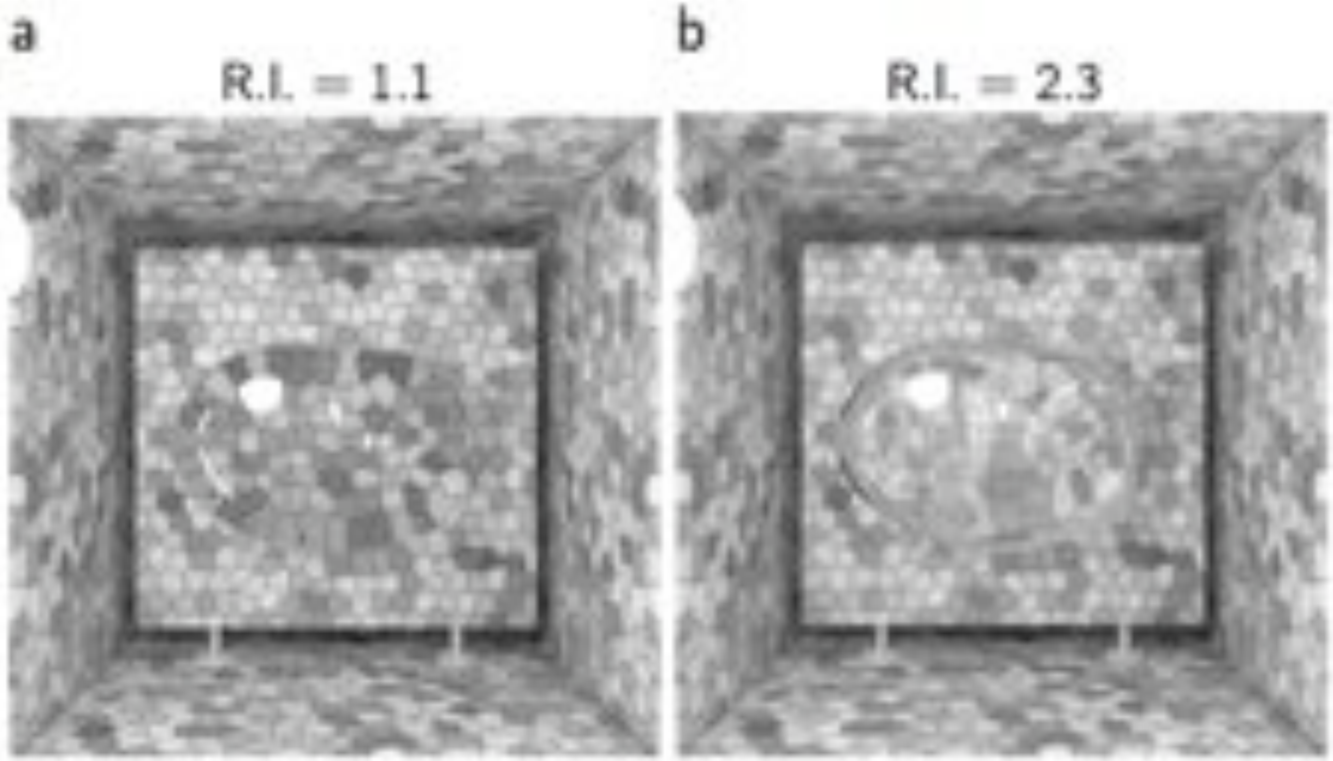

```
> coef(gm1)
$Obs
         DV
S1 3.056327
S2 4.000103
S3 3.493026
S4 3.557268
S5 4.071183
S6 4.549447
S7 3.160450
```

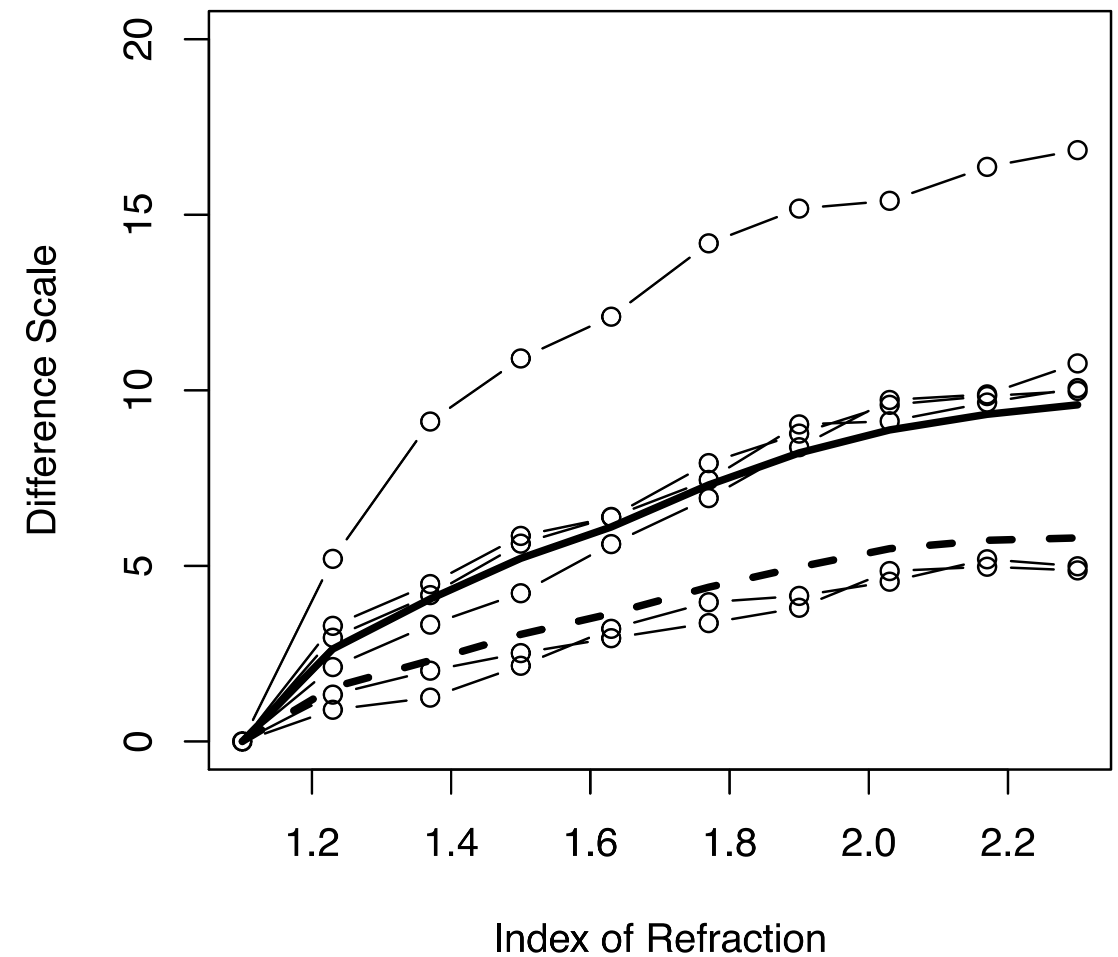
Mixed-effects models with MLDS:

Re-parameterize in terms of decision variable



Mixed-effects models with MLDS:
Normalize to common scale





Experiment of Fleming, Jäkel and Maloney (2011).
Perception of transparency as a function of rendered index of refraction

No simple functional description of relation because of kink in curve

Use each individual's scale value to compute decision variables and fit GLMM to these value; normalizes out individual shape differences.

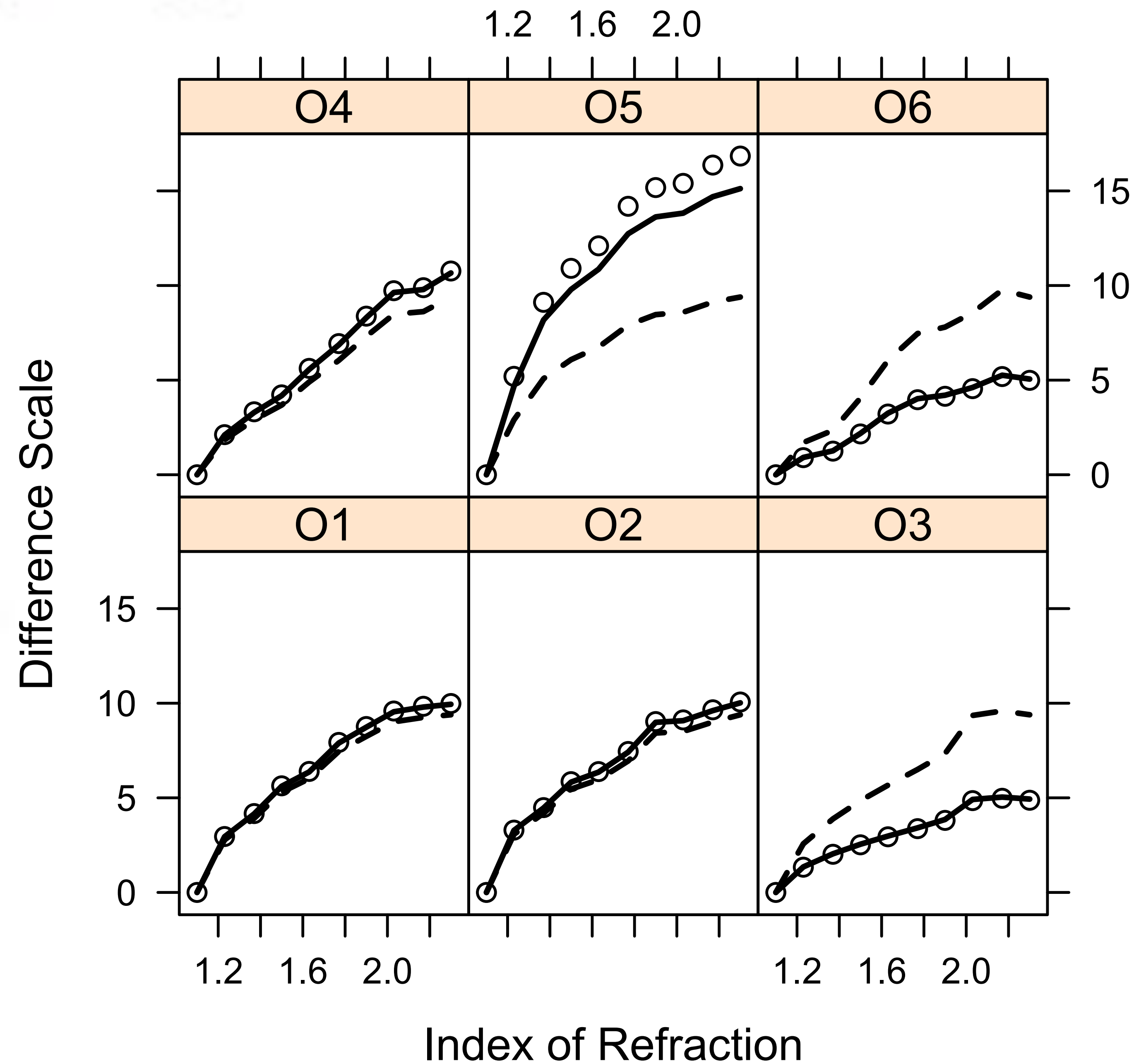
$$\mathrm{DV}_o = \hat{\psi}_{d,o} - \hat{\psi}_{c,o} - \hat{\psi}_{b,o} + \hat{\psi}_{a,o}$$

Fleming, RW, Jäkel, F, Maloney, LT **(2011)**. Visual perception of thick transparent materials. *Psychol Sci*, 22, 6:812-20.

Mixed-effects models with MLDS:

Normalize to common scale

```
Generalized linear mixed model fit by the Laplace approximation
Formula: resp ~ DV + (DV + 0 | Obs) - 1
   Data: Transparency
  AIC  BIC logLik deviance
 1485 1497   -741     1481
Random effects:
 Groups Name Variance Std.Dev.
 Obs    DV   13.7     3.69
Number of obs: 2520, groups: Obs, 6

Fixed effects:
   Estimate Std. Error z value Pr(>|z|)
DV     9.39       1.57       6    2e-09 ***
```



Mixed-effects models with MLDS:

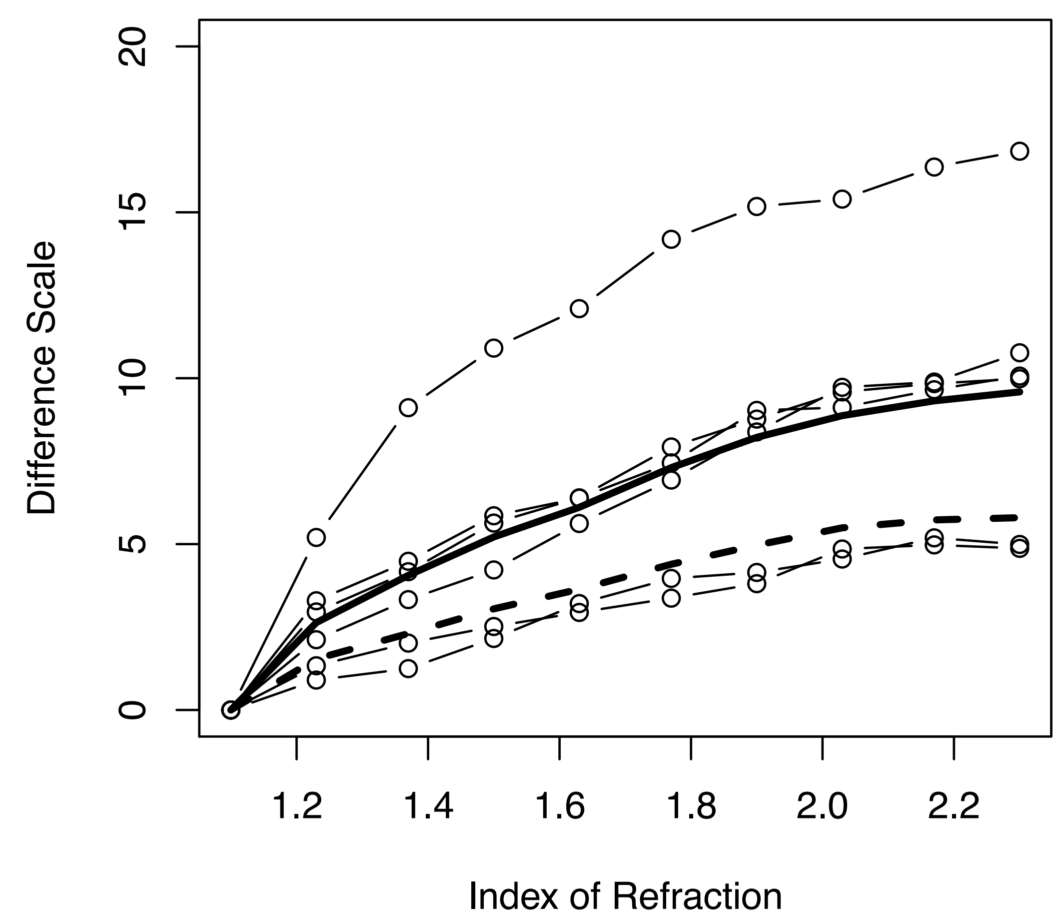
Regression on estimated coefficients

For this approach we use lmer and fit the coefficients as a function of the stimulus level using MLDS directly.

$$\hat{\psi}(S) \sim (\beta_1 + b_1)S + (\beta_2 + b_2)S^2 + \cdots + \epsilon$$

By taking the log of the coefficients, we transform the multiplicative effect to additive. We use polynomials to fit the fixed effect but also to model random differences in the shapes of the function across observers

$$\log(\hat{\psi}(S)) \sim (\beta'_0 + b'_0) + (\beta'_1 + b'_1)S + (\beta'_2 + b'_2)S^2 + \cdots + \epsilon'$$



Mixed-effects models with MLDS:
Regression on estimated coefficients

$$\log(\hat{\psi}(S)) \sim (\beta_0' + b_0') + (\beta_1' + b_1')S + (\beta_2' + b_2')S^2 + \cdots + \epsilon'$$

First, test random effects:

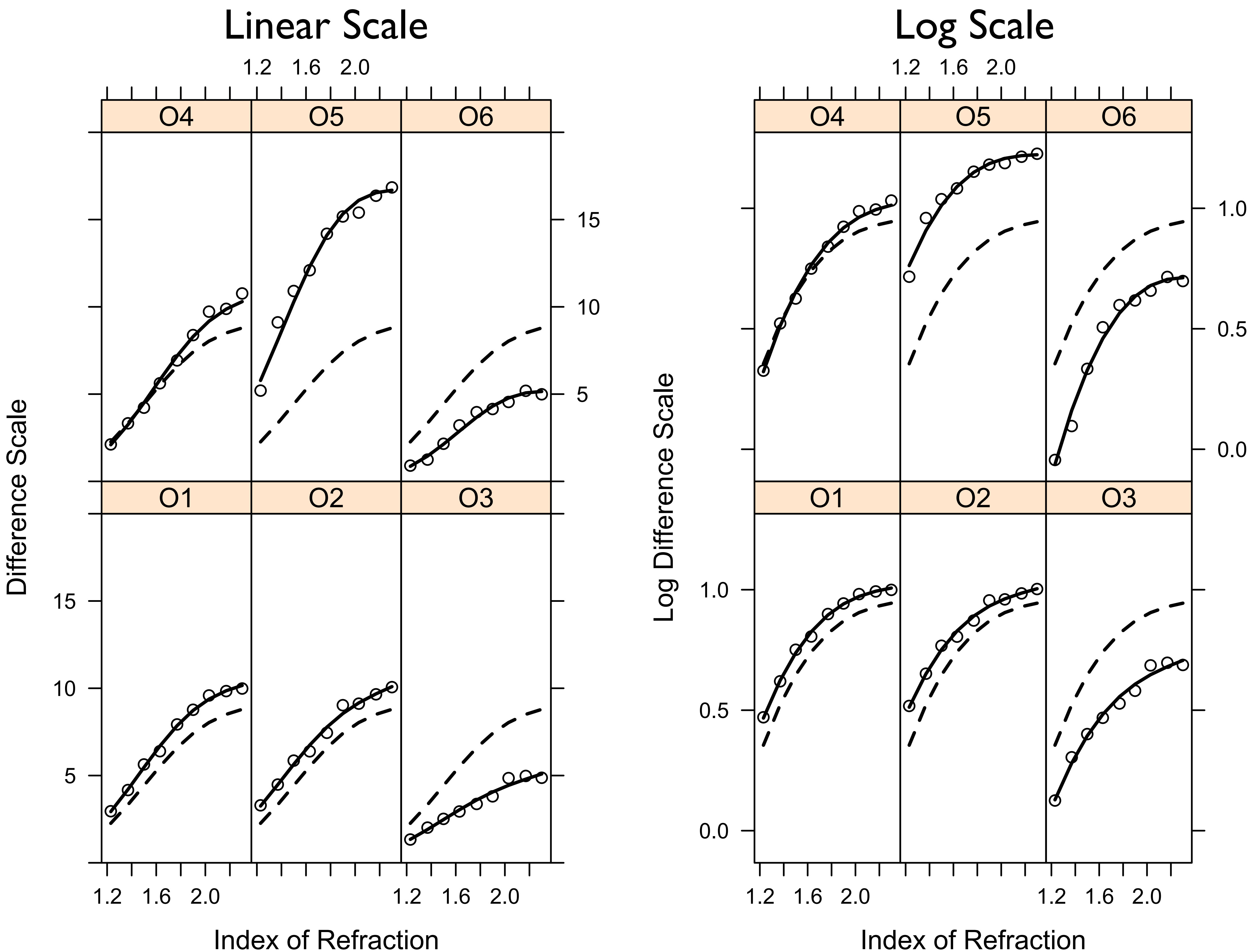
```
> P3 <- lmer(logDS ~ poly(Stim, degree = 6) +
+          (Stim + I(Stim^2) + I(Stim^3) | Obs), Tr.df)
> P2 <- lmer(logDS ~ poly(Stim, degree = 6) +
+          (Stim + I(Stim^2) | Obs), Tr.df)
> P1 <- lmer(logDS ~ poly(Stim, degree = 6) +
+          (Stim | Obs), Tr.df)
> P0 <- lmer(logDS ~ poly(Stim, degree = 6) +
+          (1 | Obs), Tr.df)
> anova(P0, P1, P2, P3)

Data: Tr.df
Models:
P0: logDS ~ poly(Stim, degree = 6) + (1 | Obs)
P1: logDS ~ poly(Stim, degree = 6) + (Stim | Obs)
P2: logDS ~ poly(Stim, degree = 6) + (Stim + I(Stim^2) |
P2:     Obs)
P3: logDS ~ poly(Stim, degree = 6) + (Stim + I(Stim^2) +
P3:     I(Stim^3) | Obs)
   Df  AIC  BIC logLik Chisq Chi Df Pr(>Chisq)
P0  9 -126 -108   71.9
P1 11 -157 -135   89.7 35.45      2      2e-08 ***
P2 14 -162 -135   95.2 11.19      3      0.011 *
P3 18 -158 -123   97.2  3.94      4      0.414
```

Mixed-effects models with MLDS:
Regression on estimated coefficients

Then, test fixed effects:

```
> P2.2 <- lmer(logDS ~ poly(Stim, degree = 2) +
+          (Stim + I(Stim^2) | Obs), Tr.df)
> P2.3 <- lmer(logDS ~ poly(Stim, degree = 3) +
+          (Stim + I(Stim^2) | Obs), Tr.df)
> P2.4 <- lmer(logDS ~ poly(Stim, degree = 4) +
+          (Stim + I(Stim^2) | Obs), Tr.df)
> P2.5 <- lmer(logDS ~ poly(Stim, degree = 5) +
+          (Stim + I(Stim^2) | Obs), Tr.df)
> anova(P2, P2.5, P2.4, P2.3, P2.2)

Data: Tr.df
Models:
P2.2: logDS ~ poly(Stim, degree = 2) + (Stim + I(Stim^2) |
P2.2:     Obs)
P2.3: logDS ~ poly(Stim, degree = 3) + (Stim + I(Stim^2) |
P2.3:     Obs)
P2.4: logDS ~ poly(Stim, degree = 4) + (Stim + I(Stim^2) |
P2.4:     Obs)
P2.5: logDS ~ poly(Stim, degree = 5) + (Stim + I(Stim^2) |
P2.5:     Obs)
P2: logDS ~ poly(Stim, degree = 6) + (Stim + I(Stim^2) |
P2:     Obs)
     Df  AIC  BIC logLik Chisq Chi Df Pr(>Chisq)
P2.2 10 -163 -143   91.5
P2.3 11 -167 -145   94.5  6.03      1      0.014 *
P2.4 12 -166 -142   95.1  1.24      1      0.266
P2.5 13 -164 -138   95.1  0.04      1      0.838
P2   14 -162 -136   95.2  0.22      1      0.637
```

Mixed-effects models with MLDS:
Regression on estimated coefficients



- *Difference Scaling* is a psychophysical technique that permits estimation of interval perceptual scales by maximum likelihood
- The approach is implemented in the R package MLDS on CRAN.
- We can introduce mixed-effects into MLDS models using the lme4 package (and perhaps others) via a number of strategies.

Thank you.