Advances in Sensory Discrimination Testing: Thurstonian-Derived Models, Covariates, and Consumer Relevance

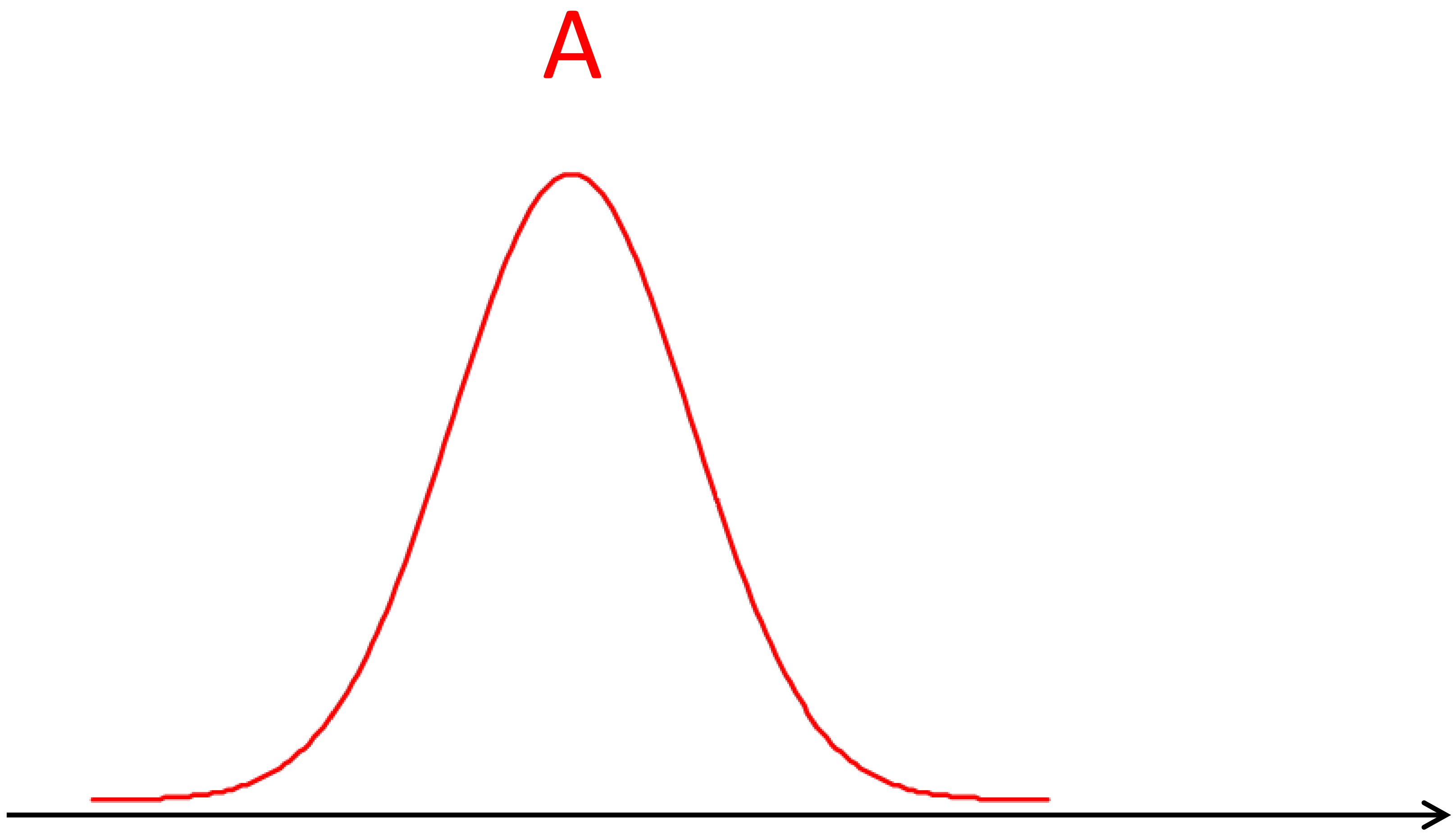
John C. Castura, Sara K. King, C. J. Findlay

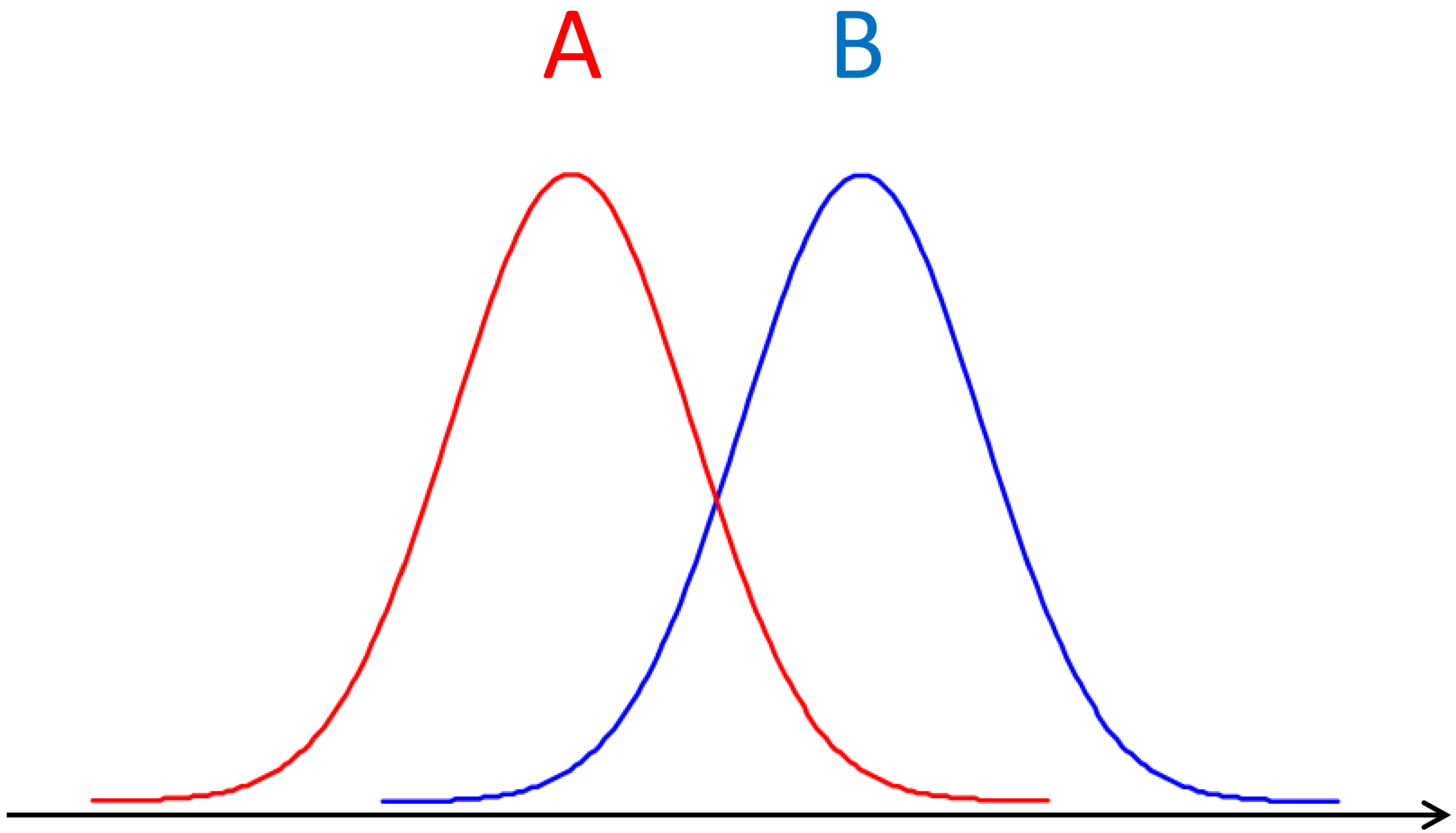
Compusense Inc.

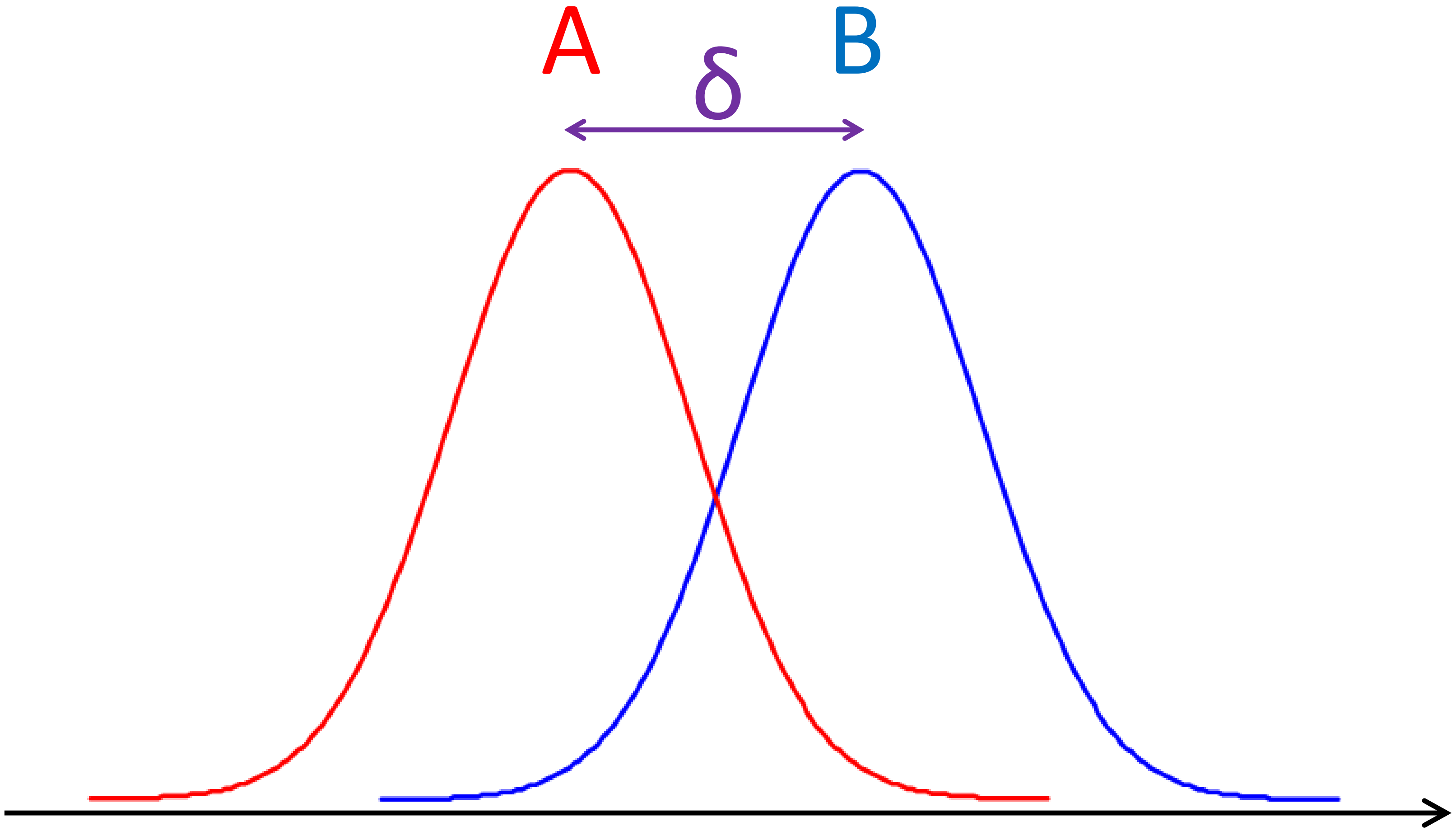
Guelph, ON, Canada

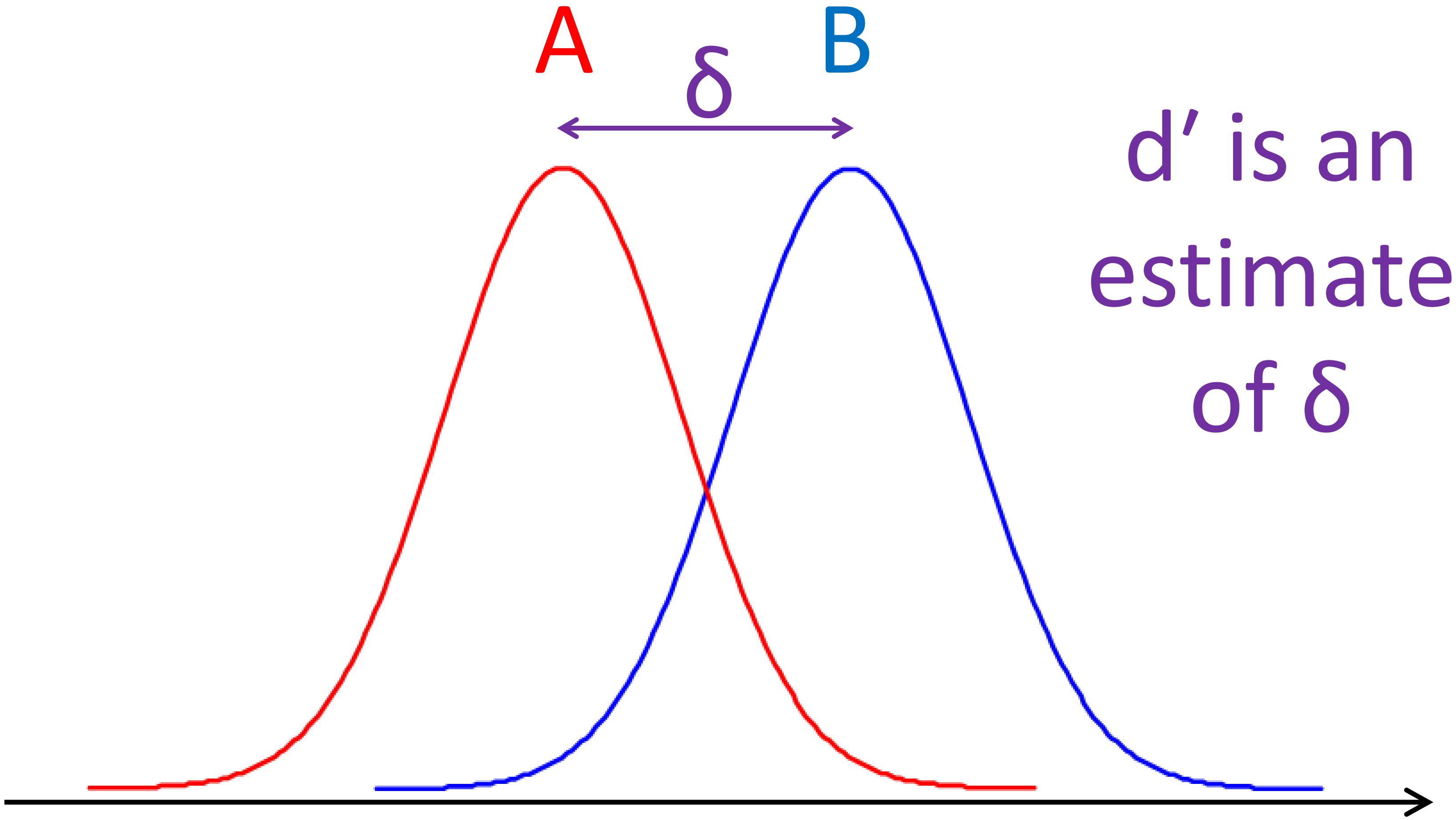
Compusense

Derivation of a psychometric function

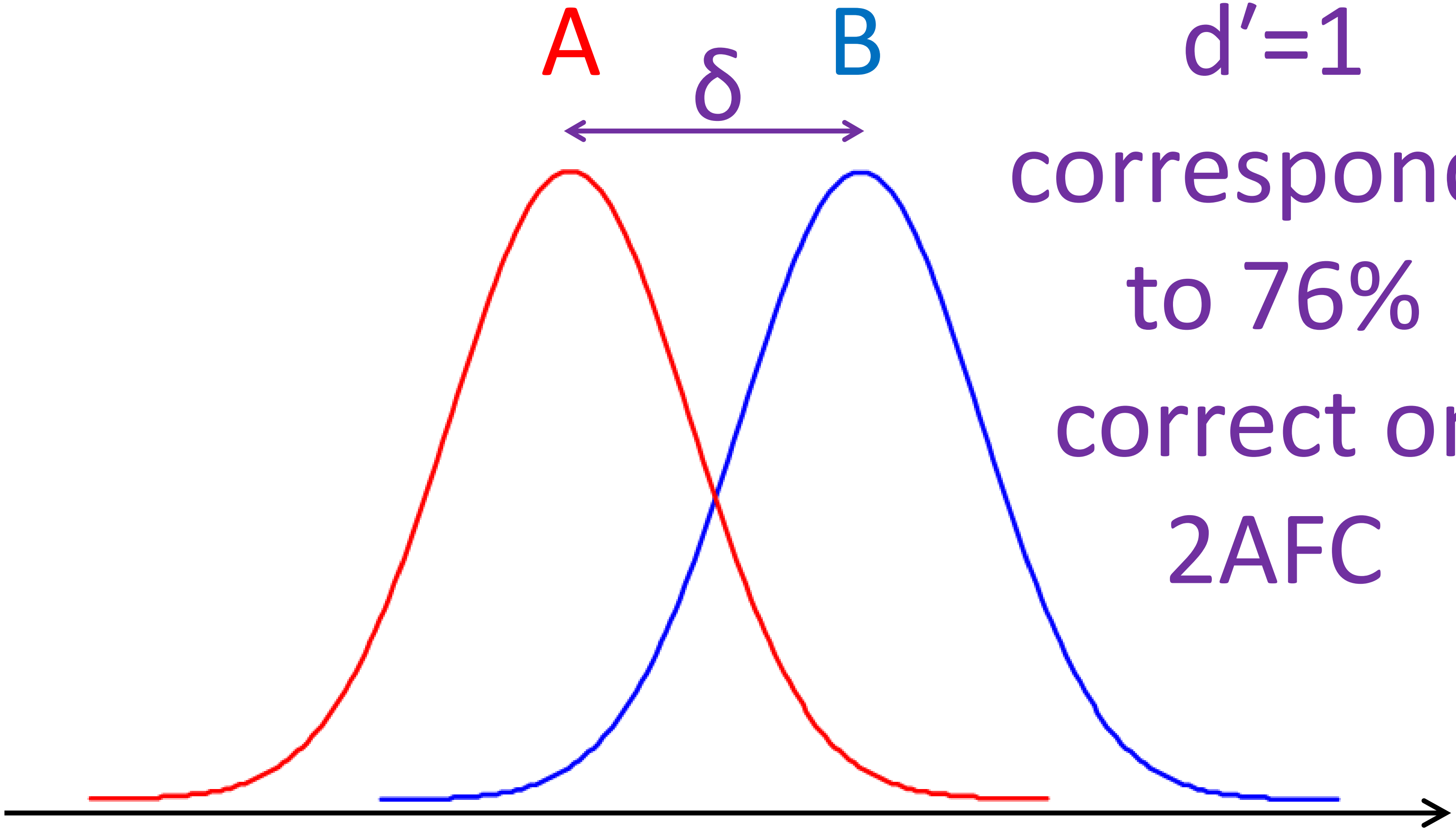

A


A
B


A
δ
B



d′ is an estimate of δ



d'=1 corresponds to 76% correct on 2AFC

Psychometric function

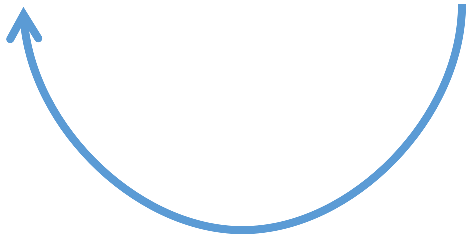
$$p_c = f_{psy}(d')$$

Psychometric function

$$p_c = f_{psy}(d')$$

Psychometric function

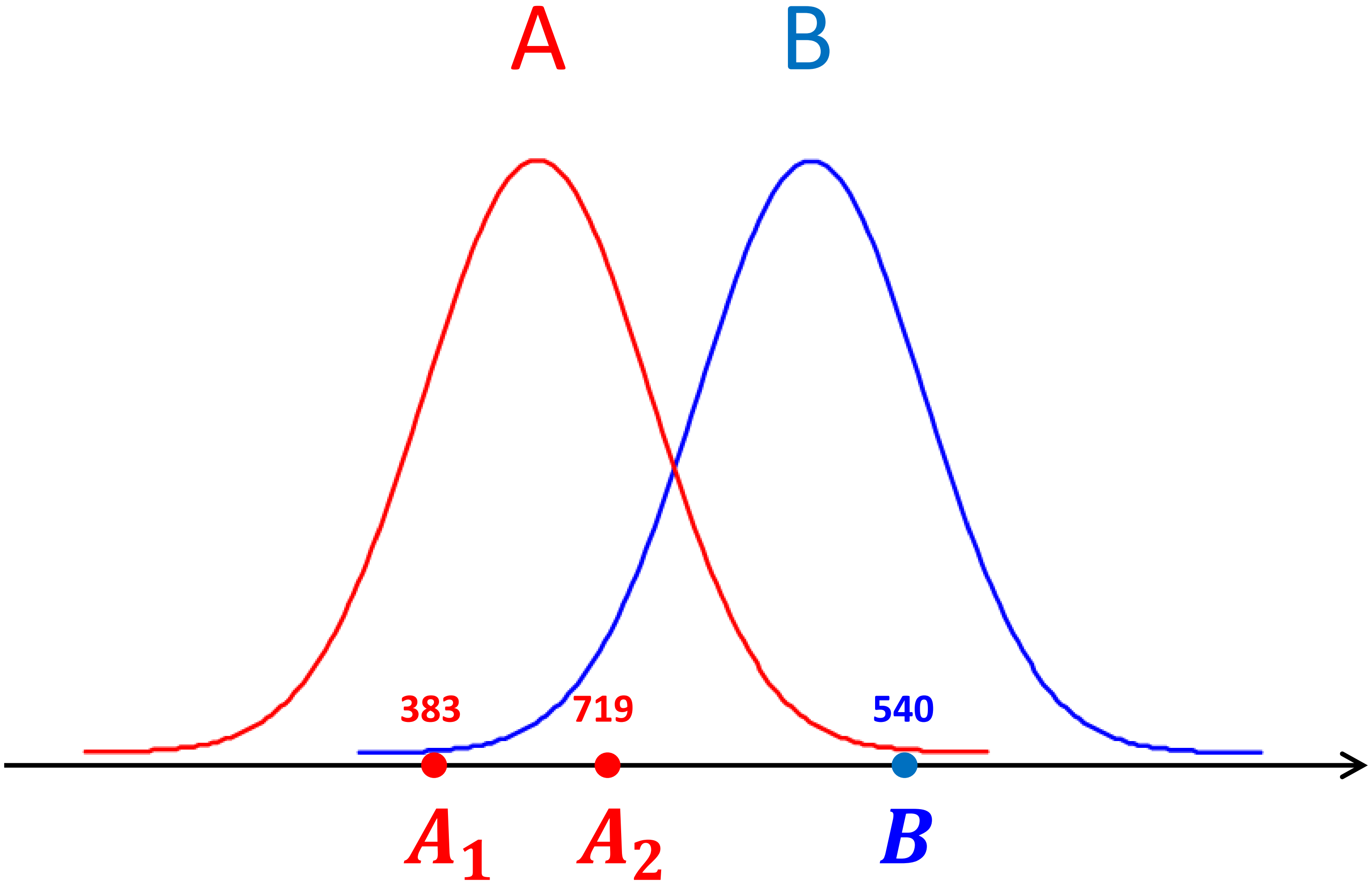
$$p_c = f_{psy}(d')$$



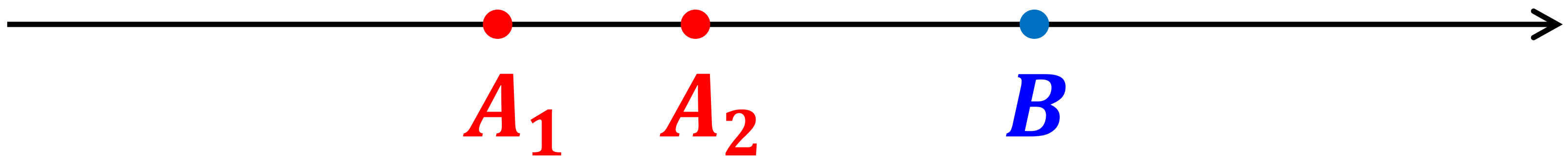
Psychometric function

$$p_c = f_{psy}(d')$$

f_{psy} encodes psychological decision-making rules into a mathematical model, and is used to create a mapping between p_c and d'

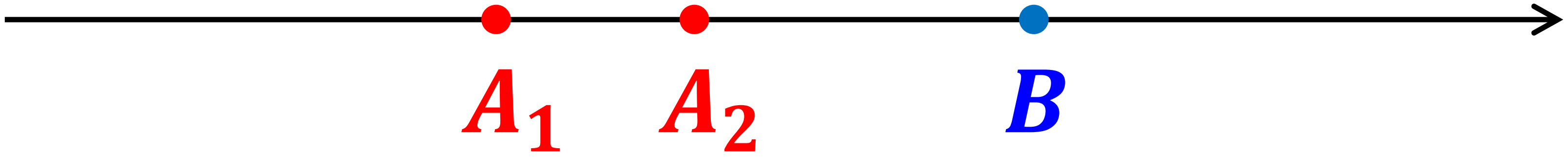

A
B
383
719
540
A_1
A_2
B

Psychometric function: 3AFC



Psychometric function: 3AFC

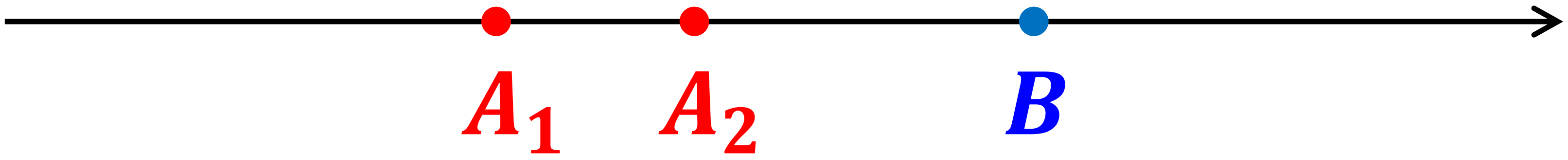
$$A_1 < B$$



Psychometric function: 3AFC

$$A_1 < B$$

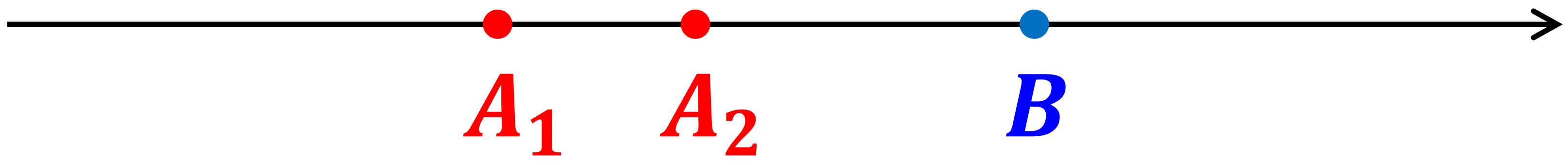
$$A_2 < B$$



Psychometric function: 3AFC

$$A_1 < B$$

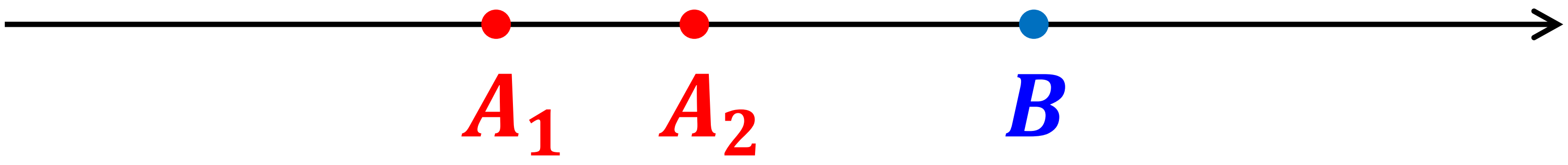
$$A_2 < B$$



Psychometric function: 3AFC

$A_1 < B$

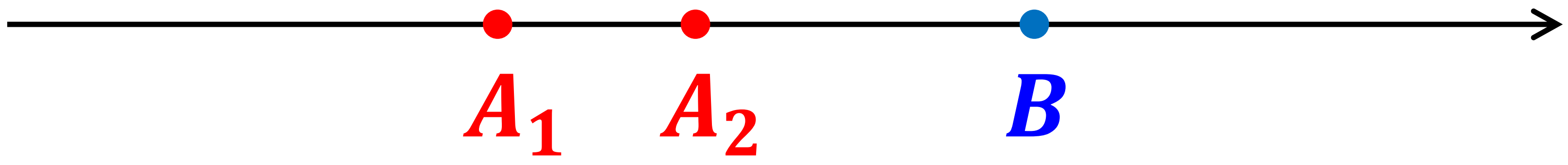
AND $A_2 < B$



Psychometric function: 3AFC

$A_1 < B$

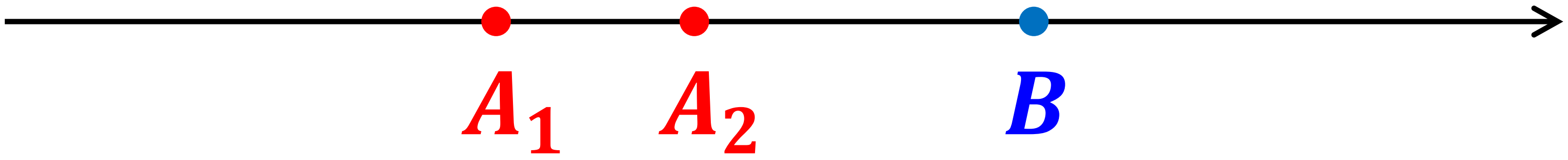
<u>*AND*</u> $A_2 < B$



Psychometric function: 3AFC

$$A_1 < B$$

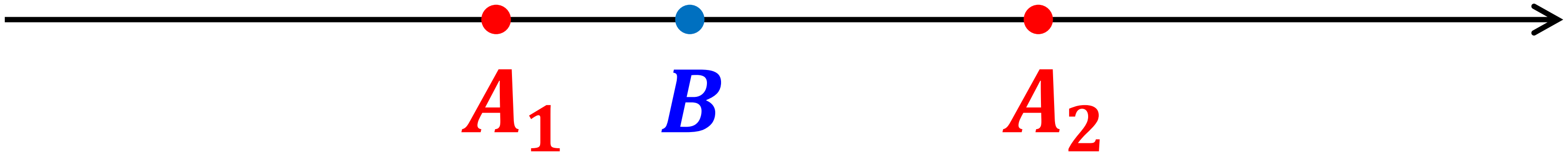
$$A_2 < B$$



Psychometric function: 3AFC

$$A_1 < B$$

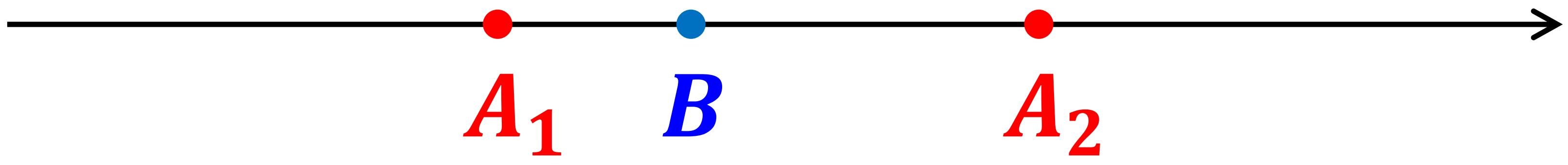
$$B < A_2$$



Psychometric function: 3AFC

$$A_1 < B$$

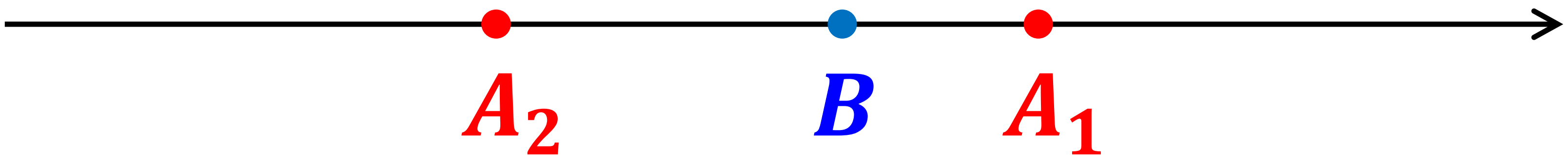
$$B < A_2$$



Psychometric function: 3AFC

$$B < A_1$$

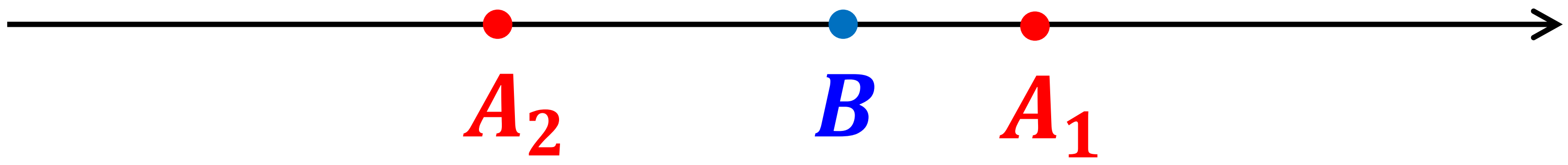
$$A_2 < B$$



Psychometric function: 3AFC

$$B < A_1$$

$$A_2 < B$$



Psychometric function: 3AFC

$$B < A_1$$

$$B < A_2$$

Psychometric function: 3AFC

$$B < A_1$$

<u>***OR***</u> $$B < A_2$$

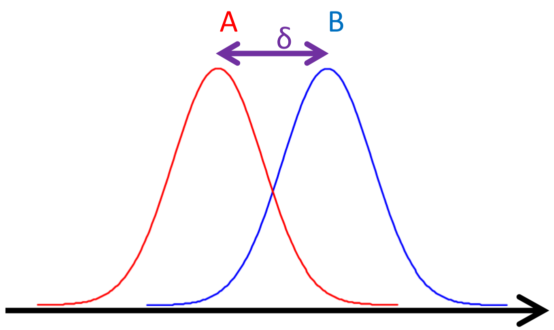
Psychometric function for 3AFC

$$\boldsymbol{p_c} \quad = \quad P(A_1 < B, A_2 < B)$$

Adapted from Christensen, R. H. B. (2012). *Sensometrics: Thurstonian and Statistical Models*. Technical University of Denmark, Kongens Lyngby, Denmark. PhD Thesis, p. 15.

Psychometric function for 3AFC

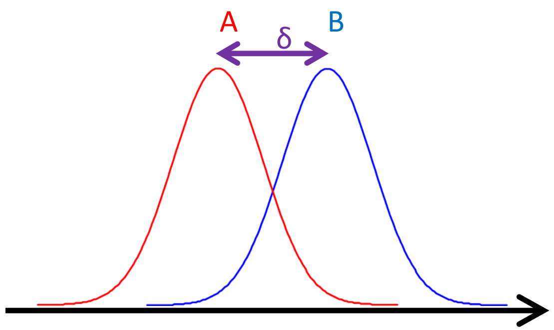
$$\begin{aligned} \boldsymbol{p_c} &= P(A_1 < B, A_2 < B) \\ &= \int_{-\infty}^{\infty} P(A_1 < B, A_2 < B, B = z)\mathrm{d}z \end{aligned}$$



Adapted from Christensen, R. H. B. (2012). *Sensometrics: Thurstonian and Statistical Models*. Technical University of Denmark, Kongens Lyngby, Denmark. PhD Thesis, p. 15.

Psychometric function for 3AFC

$$\begin{aligned} p_c &= P(A_1 < B, A_2 < B) \\ &= \int_{-\infty}^{\infty} P(A_1 < B, A_2 < B, B = z)\mathrm{d}z \\ &= \int_{-\infty}^{\infty} P(A_1 < z)P(A_2 < z)P(A_1 + \boldsymbol{\delta} = z)\mathrm{d}z \end{aligned}$$



Adapted from Christensen, R. H. B. (2012). *Sensometrics: Thurstonian and Statistical Models*. Technical University of Denmark, Kongens Lyngby, Denmark. PhD Thesis, p. 15.

Psychometric function for 3AFC

$$
\begin{aligned}
\boldsymbol{p_c} &= P(A_1 < B, A_2 < B) \\
&= \int_{-\infty}^{\infty} P(A_1 < B, A_2 < B, B = z)\mathrm{d}z \\
&= \int_{-\infty}^{\infty} P(A_1 < z)P(A_2 < z)P(A_1 + \boldsymbol{\delta} = z)\mathrm{d}z \\
&= \int_{-\infty}^{\infty} P(A_1 < z)^2 P(A_1 = z - \boldsymbol{\delta})\mathrm{d}z
\end{aligned}
$$

Adapted from Christensen, R. H. B. (2012). *Sensometrics: Thurstonian and Statistical Models*. Technical University of Denmark, Kongens Lyngby, Denmark. PhD Thesis, p. 15.

Psychometric function for 3AFC

$$
\begin{aligned}
p_c &= P(A_1 < B, A_2 < B) \\
&= \int_{-\infty}^{\infty} P(A_1 < B, A_2 < B, B = z)\mathrm{d}z \\
&= \int_{-\infty}^{\infty} P(A_1 < z)P(A_2 < z)P(A_1 + \boldsymbol{\delta} = z)\mathrm{d}z \\
&= \int_{-\infty}^{\infty} P(A_1 < z)^2 P(A_1 = z - \boldsymbol{\delta})\mathrm{d}z \\
&= \int_{-\infty}^{\infty} \phi(z)^2 \varphi(z - \boldsymbol{\delta})\mathrm{d}z
\end{aligned}
$$

Adapted from Christensen, R. H. B. (2012). *Sensometrics: Thurstonian and Statistical Models*. Technical University of Denmark, Kongens Lyngby, Denmark. PhD Thesis, p. 15.

Psychometric function for 3AFC

$$
\begin{aligned}
p_c &= P(A_1 < B, A_2 < B) \\
&= \int_{-\infty}^{\infty} P(A_1 < B, A_2 < B, B = z)\mathrm{d}z \\
&= \int_{-\infty}^{\infty} P(A_1 < z)P(A_2 < z)P(A_1 + \boldsymbol{\delta} = z)\mathrm{d}z \\
&= \int_{-\infty}^{\infty} P(A_1 < z)^2 P(A_1 = z - \boldsymbol{\delta})\mathrm{d}z \\
&= \int_{-\infty}^{\infty} \underbrace{\phi(z)}_{\text{cumulative distribution function}}{}^2 \varphi(z - \boldsymbol{\delta})\mathrm{d}z
\end{aligned}
$$

cumulative distribution function

Adapted from Christensen, R. H. B. (2012). *Sensometrics: Thurstonian and Statistical Models*. Technical University of Denmark, Kongens Lyngby, Denmark. PhD Thesis, p. 15.

Psychometric function for 3AFC

$$
\begin{aligned}
\boldsymbol{p_c} &= P(A_1 < B, A_2 < B) \\
&= \int_{-\infty}^{\infty} P(A_1 < B, A_2 < B, B = z)\mathrm{d}z \\
&= \int_{-\infty}^{\infty} P(A_1 < z)P(A_2 < z)P(A_1 + \boldsymbol{\delta} = z)\mathrm{d}z \\
&= \int_{-\infty}^{\infty} P(A_1 < z)^2 P(A_1 = z - \boldsymbol{\delta})\mathrm{d}z \\
&= \int_{-\infty}^{\infty} \phi(z)^2 \varphi(z - \boldsymbol{\delta})\mathrm{d}z
\end{aligned}
$$

probability density function

Adapted from Christensen, R. H. B. (2012). *Sensometrics: Thurstonian and Statistical Models*. Technical University of Denmark, Kongens Lyngby, Denmark. PhD Thesis, p. 15.

Psychometric function: 3AFC

$$p_c = \int_{-\infty}^{\infty} \phi(z)^2 \varphi(z - \delta) \mathrm{d}z$$

It is possible to set $\boldsymbol{\delta}$, then solve for p_c

if $\delta = 0.5$ then $p_c = 0.48$

if $\delta = 1.0$ then $p_c = 0.63$

if $\delta = 1.5$ then $p_c = 0.77$

Adapted from Christensen, R. H. B. (2012). *Sensometrics: Thurstonian and Statistical Models*. Technical University of Denmark, Kongens Lyngby, Denmark. PhD Thesis, p. 15.

Psychometric function

$$p_c = f_{psy}(d')$$

The R package **sensR** can be used to obtain **d′** estimates

Christensen, R. H. B. & P. B. Brockhoff (2015). sensR - An R-package for sensory discrimination. R package version 1.4-5. http://www.cran.r-project.org/package=sensR/.

Generalized linear models with a psychometric link function

GLM with Thurstonian Link Function

Tetrad

$$\boldsymbol{X}\beta = \boldsymbol{f}_{\boldsymbol{psy}}{}^{-1}(\boldsymbol{\delta}) = \boldsymbol{g}(\boldsymbol{f}_{\boldsymbol{psy}})$$

E.g.

$$\boldsymbol{g}_{\text{tetrad}}(p_{ij}) = \beta_0 + \beta_j \boldsymbol{X}_i + \varepsilon_{ij}$$

GLM with Thurstonian Link Function

Original formulation vs. four prototypes

	vs.P1	vs.P2	vs.P3	vs.P4
A	6/19	8/19	11/19	13/19
B	12/19	13/19	16/19	17/19

where A and B are consumer segments

Adapted from: Brockhoff, P. B. & Christensen, R. H. B. (2010). Thurstonian models for sensory discrimination tests as generalized linear models. *Food Quality and Preference* **21**, 330-338.

GLM with Thurstonian Link Function

```
require(sensR)

# tetrad data
data <- expand.grid(  conc = 1:4,
                      segment = c("A", "B") )
data$correct <- c(6, 8, 11, 13, 12, 13, 16, 17)
data$total   <- rep(19, 8)

# glm with appropriate thurstonian link function specified
model <- glm( cbind(correct, total - correct) ~ segment + conc,
       data, family = tetrad )
```

Adapted from: Brockhoff, P. B. & Christensen, R. H. B. (2010). Thurstonian models for sensory discrimination tests as generalized linear models. *Food Quality and Preference* **21**, 330-338.

Analysis of Deviance (ANODE) table

```
round(summary(model)$coefficients, 3)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.105	0.509	-0.206	0.837
segmentB	1.015	0.323	3.139	0.002
conc	0.437	0.151	2.899	0.004

Adapted from: Brockhoff, P. B. & Christensen, R. H. B. (2010). Thurstonian models for sensory discrimination tests as generalized linear models. *Food Quality and Preference* **21**, 330-338.

Analysis of Deviance (ANODE) table

```
round(summary(model)$coefficients, 3)

            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.105      0.509  -0.206    0.837
segmentB       1.015      0.323   3.139    0.002
conc           0.437      0.151   2.899    0.004
```

Adapted from: Brockhoff, P. B. & Christensen, R. H. B. (2010). Thurstonian models for sensory discrimination tests as generalized linear models. *Food Quality and Preference* **21**, 330-338.

Analysis of Deviance (ANODE) table

```
round(summary(model)$coefficients, 3)

             Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.105      0.509  -0.206    0.837
segmentB        1.015      0.323   3.139    0.002
conc            0.437      0.151   2.899    0.004
```

Adapted from: Brockhoff, P. B. & Christensen, R. H. B. (2010). Thurstonian models for sensory discrimination tests as generalized linear models. *Food Quality and Preference* **21**, 330-338.

Analysis of Deviance (ANODE) table

```
round(summary(model)$coefficients, 3)

            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.105      0.509  -0.206    0.837
segmentB       1.015      0.323   3.139    0.002
conc           0.437      0.151   2.899    0.004
```

Adapted from: Brockhoff, P. B. & Christensen, R. H. B. (2010). Thurstonian models for sensory discrimination tests as generalized linear models. *Food Quality and Preference* **21**, 330-338.

Confidence Intervals

Brockhoff & Christensen (2010) also propose using likelihood intervals instead of Wald intervals (based on a normal approximation).

This recommendation stands for all results, but seems especially important if $\boldsymbol{p_c} \cong 1$.

d-prime calculation

```
# compare with d-prime calculation from discrim method
discrim(34, 55, method="threeAFC", statistic="likelihood")

Estimates for the threeAFC discrimination protocol with 34 correct
answers in 55 trials. One-sided p-value and 95 % two-sided confidence
intervals are based on the likelihood root statistic.

        Estimate Std. Error  Lower  Upper
pc        0.6182    0.06551 0.4865 0.7390
pd        0.4273    0.09826 0.2298 0.6086
d-prime   0.9467    0.22335 0.5128 1.3893

Result of difference test:
Likelihood Root statistic = 4.311726, p-value: 8.099e-06
Alternative hypothesis: d-prime is greater than 0
```

d-prime calculation

```
# compare with d-prime calculation from discrim method
discrim(34, 55, method="threeAFC", statistic="likelihood")

Estimates for the threeAFC discrimination protocol with 34 correct
answers in 55 trials. One-sided p-value and 95 % two-sided confidence
intervals are based on the likelihood root statistic.

        Estimate Std. Error  Lower  Upper
pc        0.6182    0.06551 0.4865 0.7390
pd        0.4273    0.09826 0.2298 0.6086
d-prime   0.9467    0.22335 0.5128 1.3893

Result of difference test:
Likelihood Root statistic = 4.311726, p-value: 8.099e-06
Alternative hypothesis: d-prime is greater than 0
```

Same-different test – Thurstonian analysis

```
Coefficients
      Estimate Std. Error  Lower  Upper P-value
tau     1.7919     0.1279 1.5502 2.0516  <2e-16 ***
delta   2.7760     0.2115 2.3643 3.1953  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1

Log Likelihood: -154.5183       AIC: 313.0366
```

Same-different test – Thurstonian analysis

```
Coefficients
        Estimate Std. Error  Lower  Upper P-value
tau       1.7919     0.1279 1.5502 2.0516  <2e-16 ***
delta     2.7760     0.2115 2.3643 3.1953  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 
0.1 ‘ ’ 1

Log Likelihood: -154.5183       AIC: 313.0366
```

d′

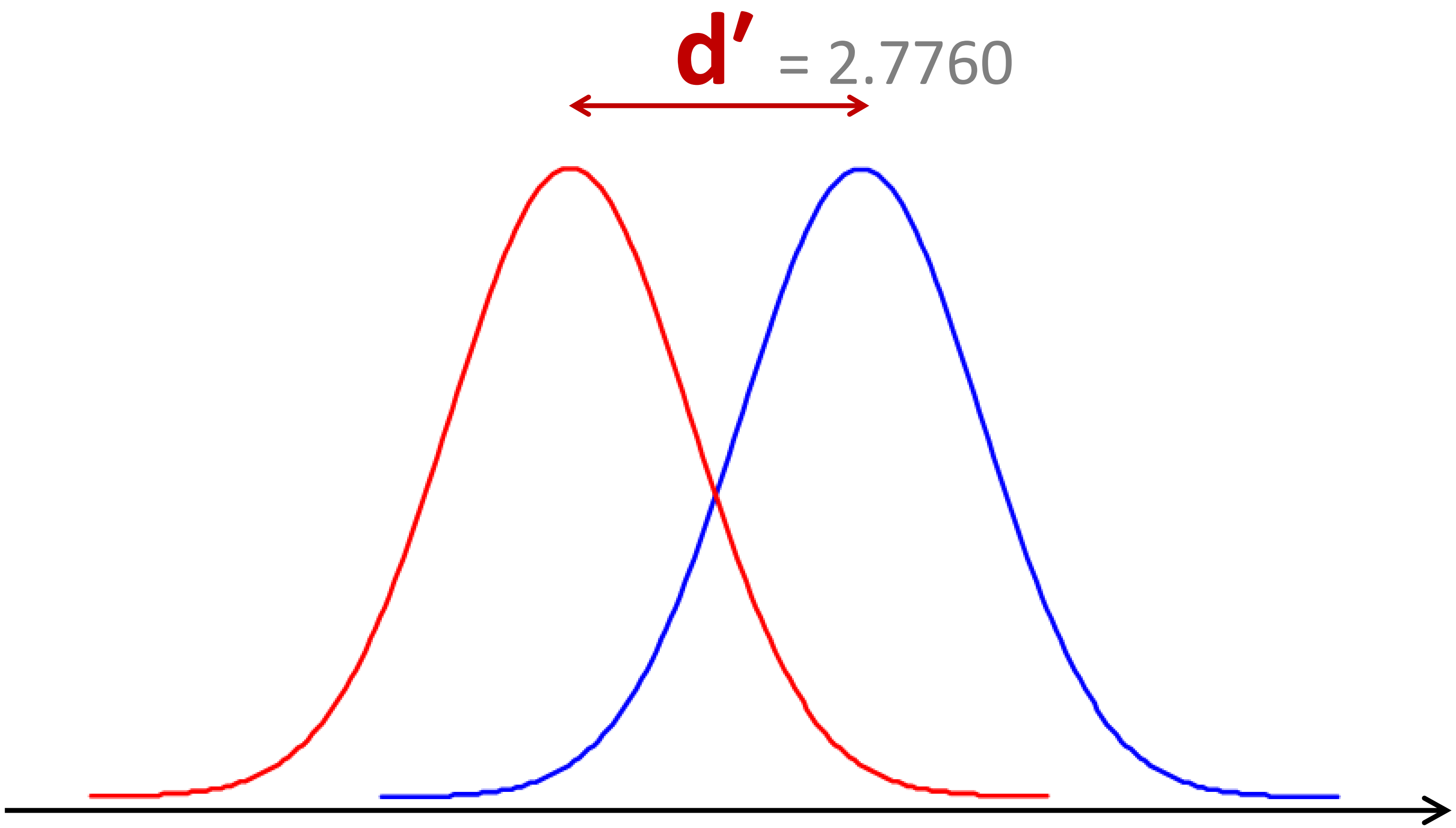
Same-different test – Thurstonian analysis

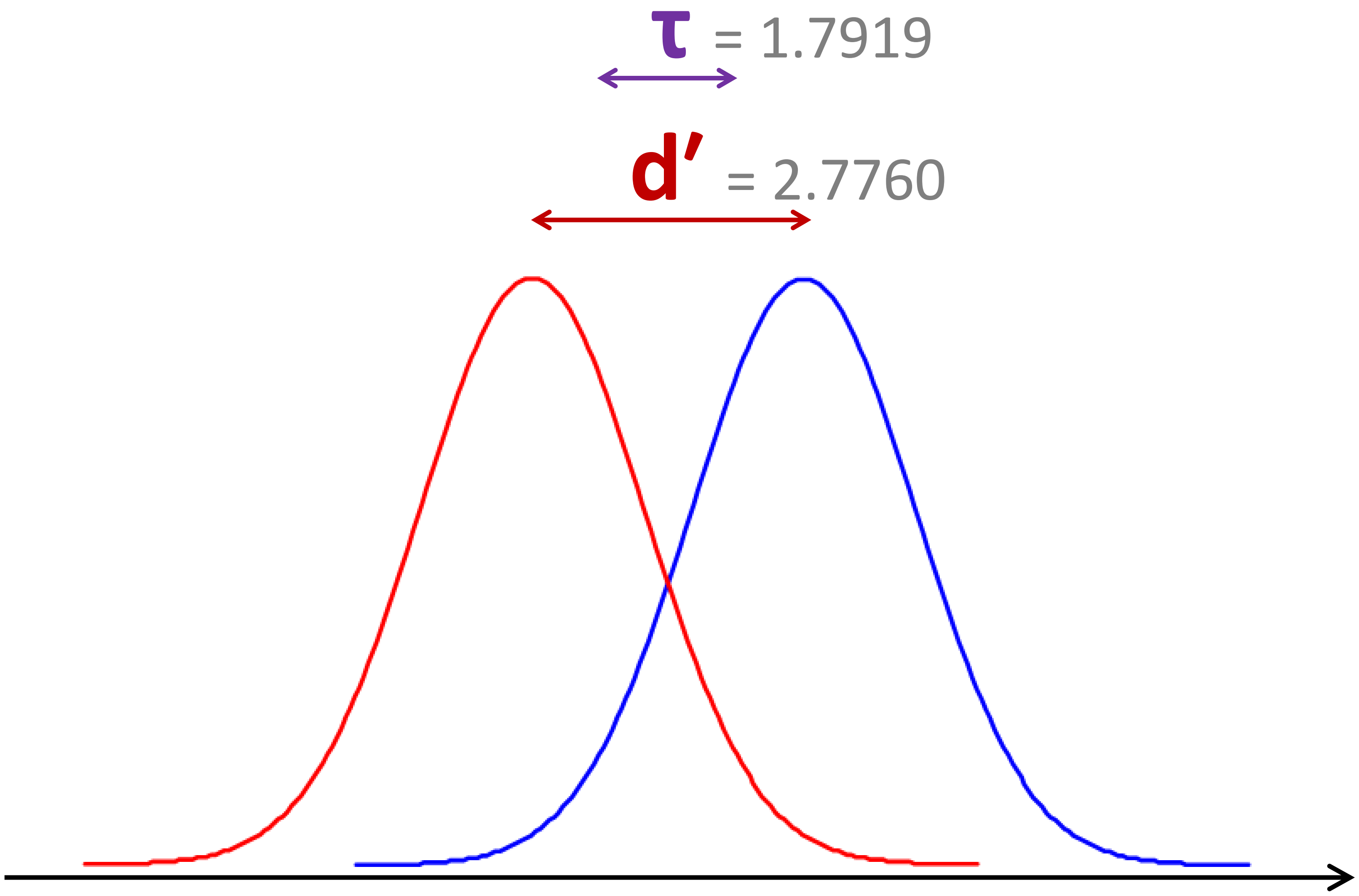
```
Coefficients
       Estimate Std. Error  Lower  Upper P-value
tau      1.7919     0.1279 1.5502 2.0516  <2e-16 ***
delta    2.7760     0.2115 2.3643 3.1953  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1

Log Likelihood: -154.5183       AIC: 313.0366
```

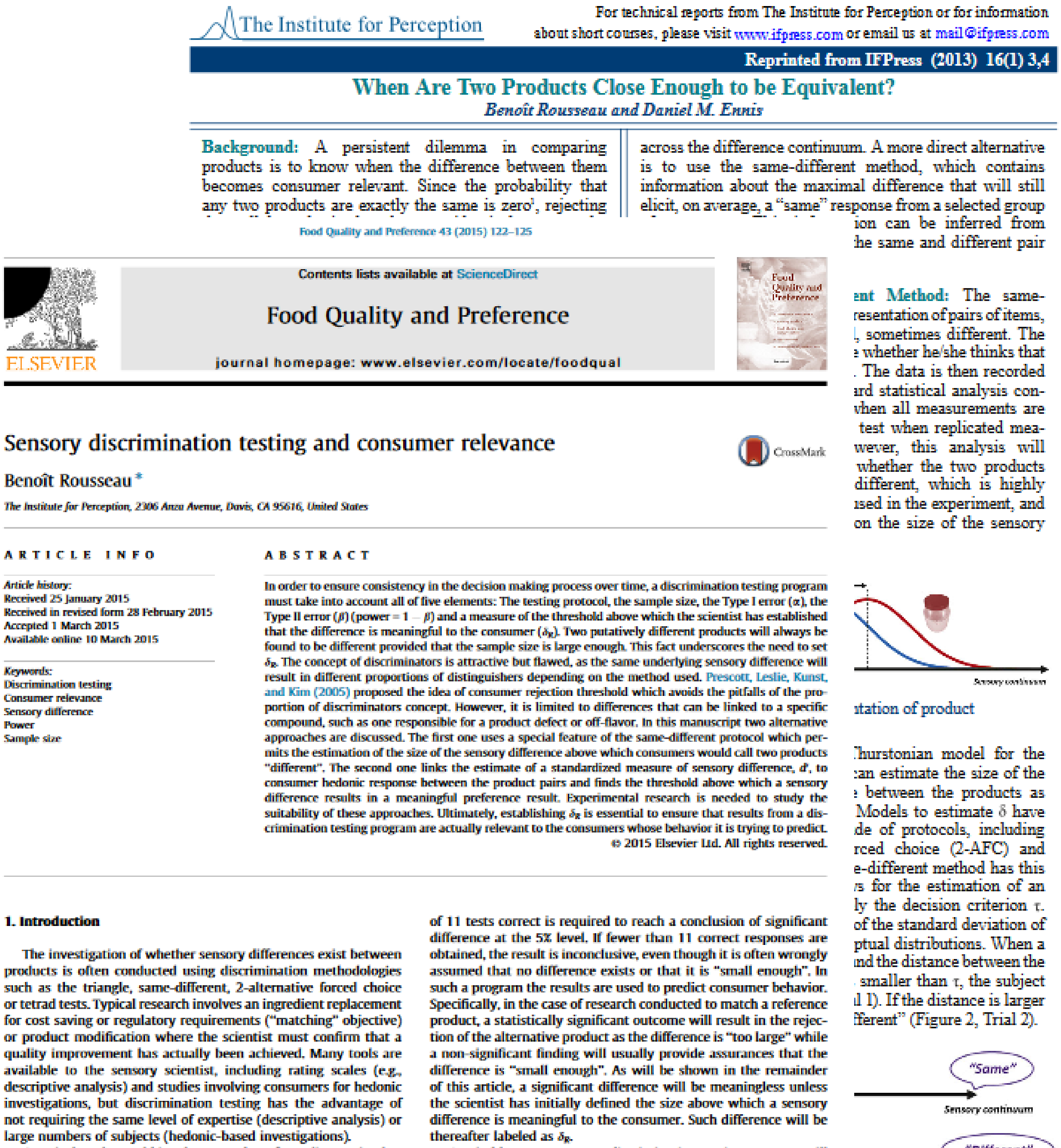
τ

d′


d' = 2.7760


τ = 1.7919
d′ = 2.7760

Rousseau and Ennis (2013) and Rousseau (2015) propose a strategy for determining consumer-relevant differences



The Institute for Perception

For technical reports from The Institute for Perception or for information about short courses, please visit www.ifpress.com or email us at mail@ifpress.com

Reprinted from IFPress (2013) 16(1) 3,4

When Are Two Products Close Enough to be Equivalent?

Benoît Rousseau and Daniel M. Ennis

Background: A persistent dilemma in comparing products is to know when the difference between them becomes consumer relevant. Since the probability that any two products are exactly the same is zero[1], rejecting

across the difference continuum. A more direct alternative is to use the same-different method, which contains information about the maximal difference that will still elicit, on average, a "same" response from a selected group

ion can be inferred from
he same and different pair

ent Method: The same-
resentation of pairs of items,
, sometimes different. The
whether he/she thinks that
. The data is then recorded
ard statistical analysis con-
vhen all measurements are
test when replicated mea-
wever, this analysis will
whether the two products
different, which is highly
ised in the experiment, and
on the size of the sensory

Sensory continuum

itation of product

Thurstonian model for the
can estimate the size of the
e between the products as
Models to estimate δ have
de of protocols, including
rced choice (2-AFC) and
e-different method has this
s for the estimation of an
ly the decision criterion τ.
of the standard deviation of
ptual distributions. When a
nd the distance between the
smaller than τ, the subject
l 1). If the distance is larger
fferent" (Figure 2, Trial 2).

"Same"

Sensory continuum

Food Quality and Preference 43 (2015) 122–125

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Sensory discrimination testing and consumer relevance

Benoît Rousseau *

The Institute for Perception, 2306 Anza Avenue, Davis, CA 95616, United States

ARTICLE INFO

Article history:
Received 25 January 2015
Received in revised form 28 February 2015
Accepted 1 March 2015
Available online 10 March 2015

Keywords:
Discrimination testing
Consumer relevance
Sensory difference
Power
Sample size

ABSTRACT

In order to ensure consistency in the decision making process over time, a discrimination testing program must take into account all of five elements: The testing protocol, the sample size, the Type I error (α), the Type II error (β) (power = 1 – β) and a measure of the threshold above which the scientist has established that the difference is meaningful to the consumer (δ_R). Two putatively different products will always be found to be different provided that the sample size is large enough. This fact underscores the need to set δ_R. The concept of discriminators is attractive but flawed, as the same underlying sensory difference will result in different proportions of distinguishers depending on the method used. Prescott, Leslie, Kunst, and Kim (2005) proposed the idea of consumer rejection threshold which avoids the pitfalls of the proportion of discriminators concept. However, it is limited to differences that can be linked to a specific compound, such as one responsible for a product defect or off-flavor. In this manuscript two alternative approaches are discussed. The first one uses a special feature of the same-different protocol which permits the estimation of the size of the sensory difference above which consumers would call two products "different". The second one links the estimate of a standardized measure of sensory difference, d', to consumer hedonic response between the product pairs and finds the threshold above which a sensory difference results in a meaningful preference result. Experimental research is needed to study the suitability of these approaches. Ultimately, establishing δ_R is essential to ensure that results from a discrimination testing program are actually relevant to the consumers whose behavior it is trying to predict.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The investigation of whether sensory differences exist between products is often conducted using discrimination methodologies such as the triangle, same-different, 2-alternative forced choice or tetrad tests. Typical research involves an ingredient replacement for cost saving or regulatory requirements ("matching" objective) or product modification where the scientist must confirm that a quality improvement has actually been achieved. Many tools are available to the sensory scientist, including rating scales (e.g., descriptive analysis) and studies involving consumers for hedonic investigations, but discrimination testing has the advantage of not requiring the same level of expertise (descriptive analysis) or large numbers of subjects (hedonic-based investigations).

of 11 tests correct is required to reach a conclusion of significant difference at the 5% level. If fewer than 11 correct responses are obtained, the result is inconclusive, even though it is often wrongly assumed that no difference exists or that it is "small enough". In such a program the results are used to predict consumer behavior. Specifically, in the case of research conducted to match a reference product, a statistically significant outcome will result in the rejection of the alternative product as the difference is "too large" while a non-significant finding will usually provide assurances that the difference is "small enough". As will be shown in the remainder of this article, a significant difference will be meaningless unless the scientist has initially defined the size above which a sensory difference is meaningful to the consumer. Such difference will be thereafter labeled as δ_R.

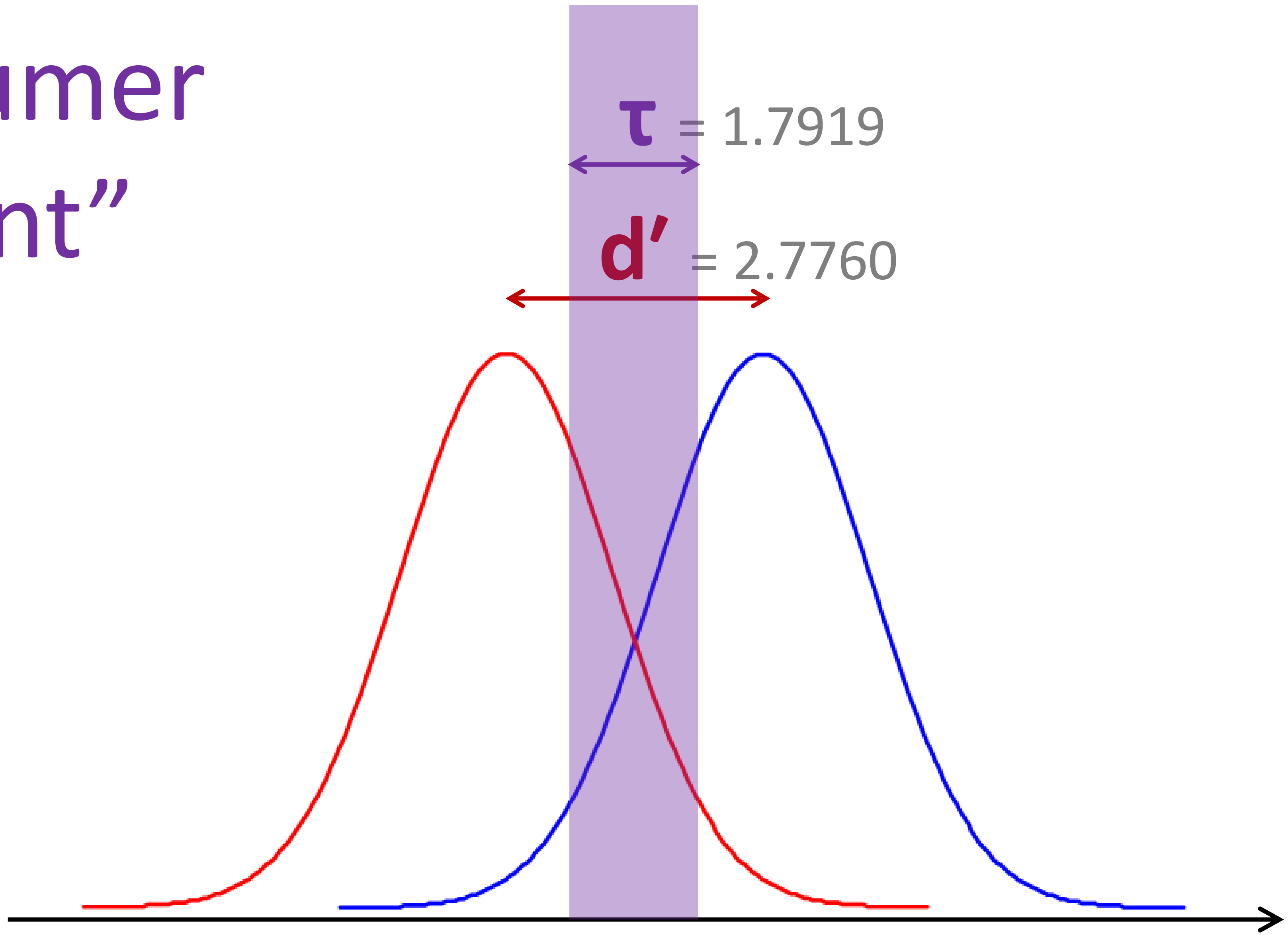
Rousseau's strategy

1. Collect same-different data and obtain a good estimate of **τ**.

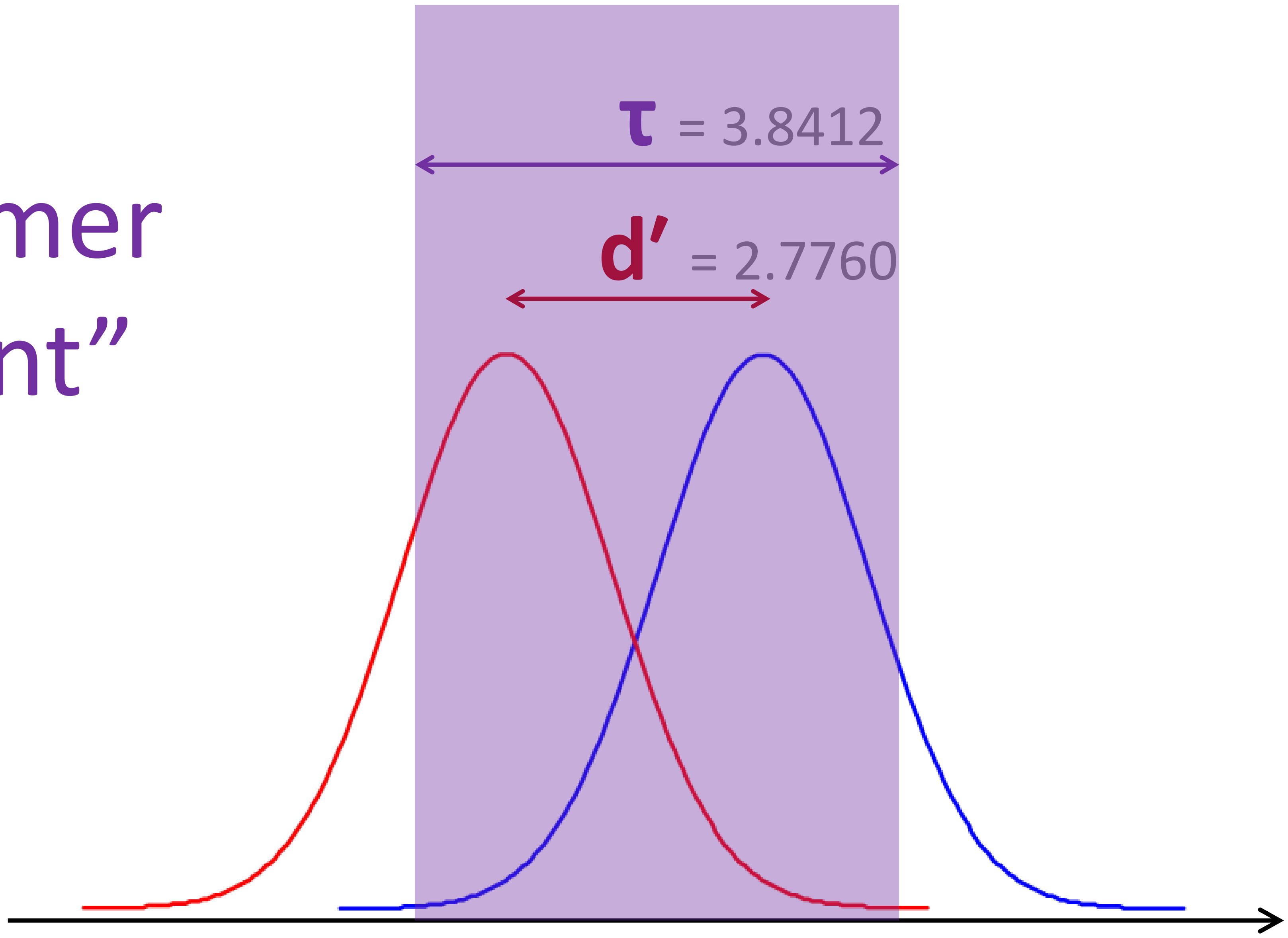
2. The reference for future decision-making is **τ**.

3. Consider the difference to be consumer-relevant if **d′ > τ**. Otherwise, consider the difference to be non-consumer-relevant.

“consumer relevant”



“not consumer relevant”

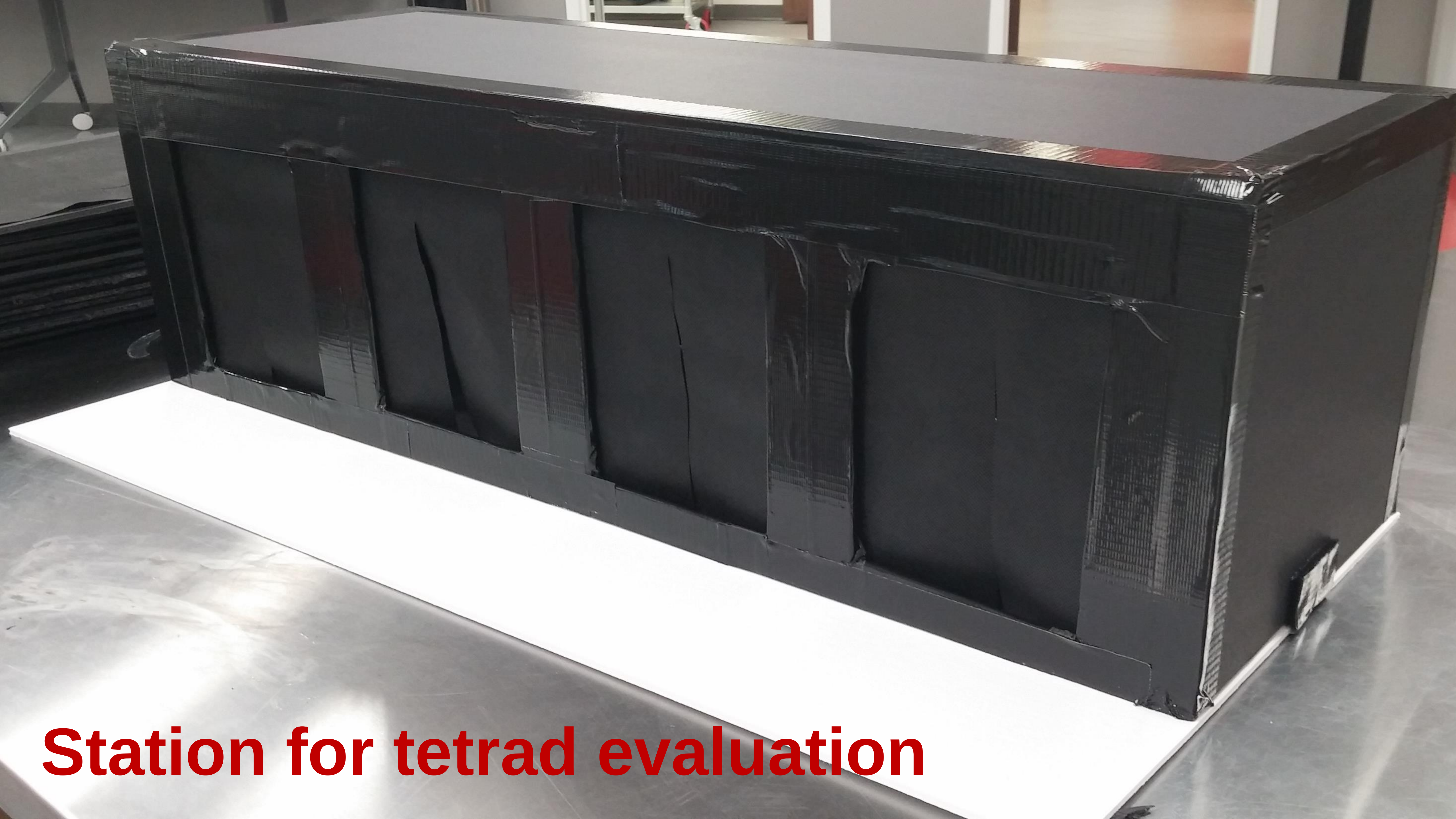


So far the published evidence used to justify this strategy has been based on simulated data.

Equivalence study

Materials and Methods


A
B
C
D
psi=27
psi=35
psi=45
psi=65



Station for tetrad evaluation

Same-Different test

In front of you are two samples.

Firmly press down on the samples in the order indicated below and indicate if the samples are

SAME **or** ***DIFFERENT***.

Same

Dif

Pair	Possible Presentations			
AB	AB	AA	BA	BB
AC	AC	AA	CA	CC
AD	AD	AA	DA	DD
BC	BC	BB	CB	CC
BD	BD	BB	DB	DD
CD	CD	CC	DC	DD

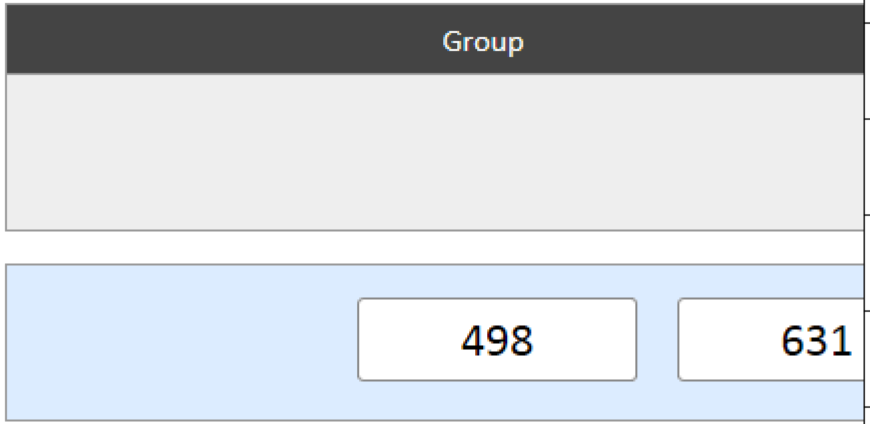
Tetrad test

In front of you are **four samples**.

Firmly press down on each of the samples in the order indicated below from left to right.

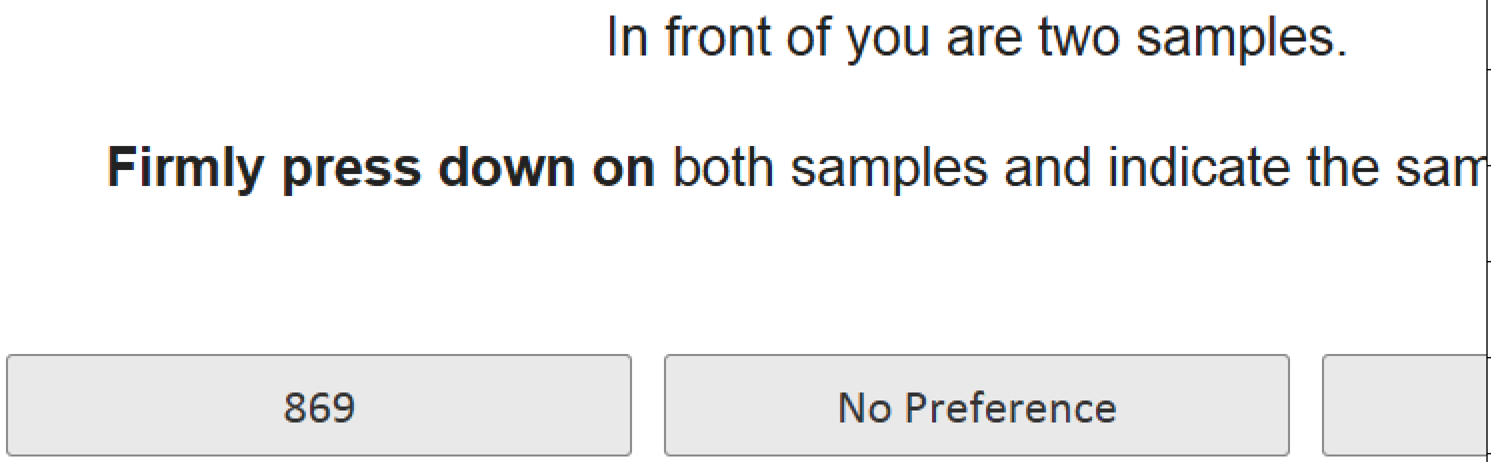
Decide which samples are similar.

Drag the codes on the screen

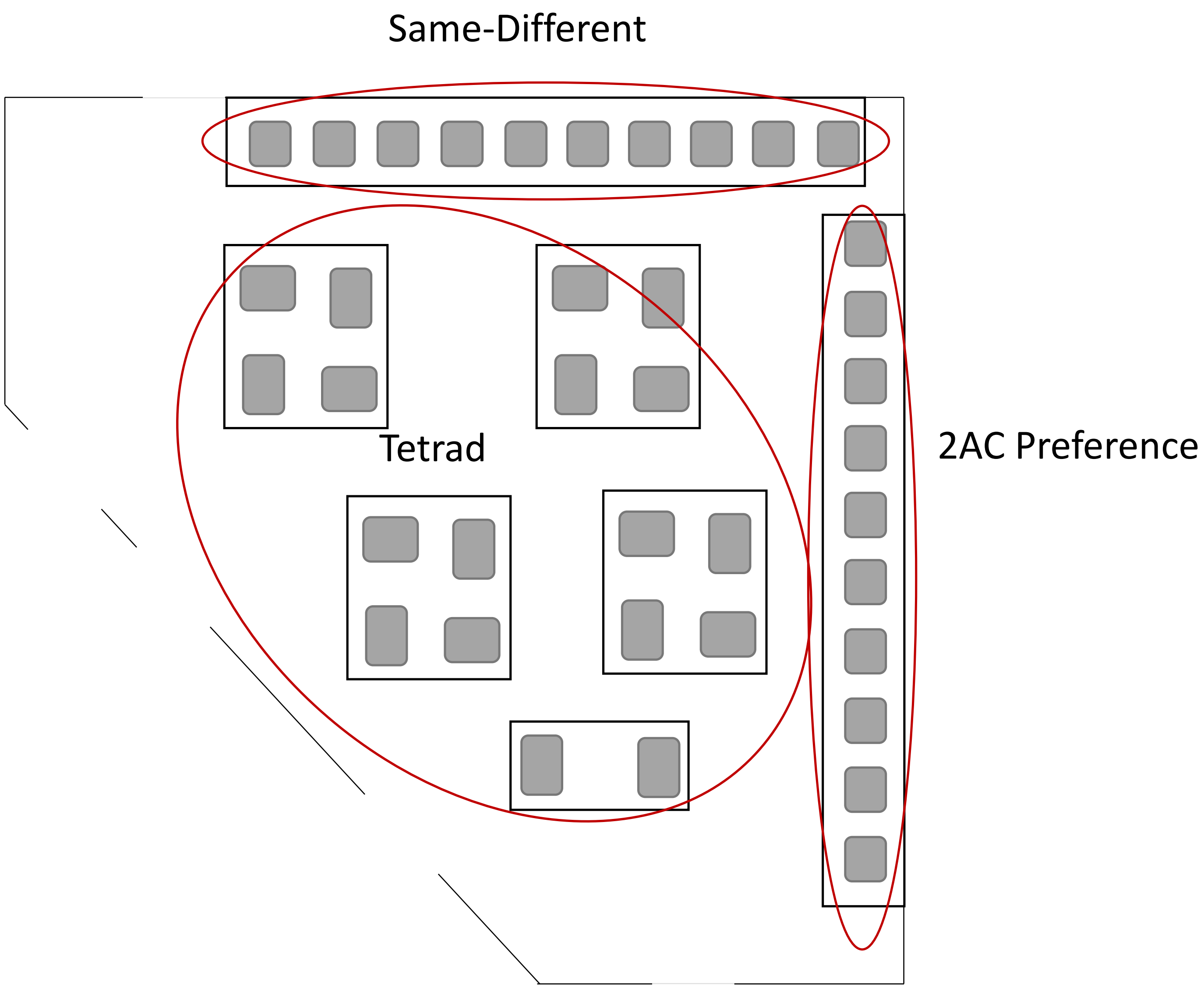


Pair	Possible Presentations					
AB	ABAB	AABB	BABA	BBAA	ABBA	BAAB
AC	ACAC	AACC	CACA	CCAA	ACCA	CAAC
AD	ADAD	AADD	DADA	DDAA	ADDA	DAAD
BC	BCBC	BBCC	CBCB	CCBB	BCCB	CBBC
BD	BDBD	BBDD	DBDB	DDBB	BDDB	DBBD
CD	CDCD	CCDD	DCDC	DDCC	CDDC	DCCD

2AC Preference test

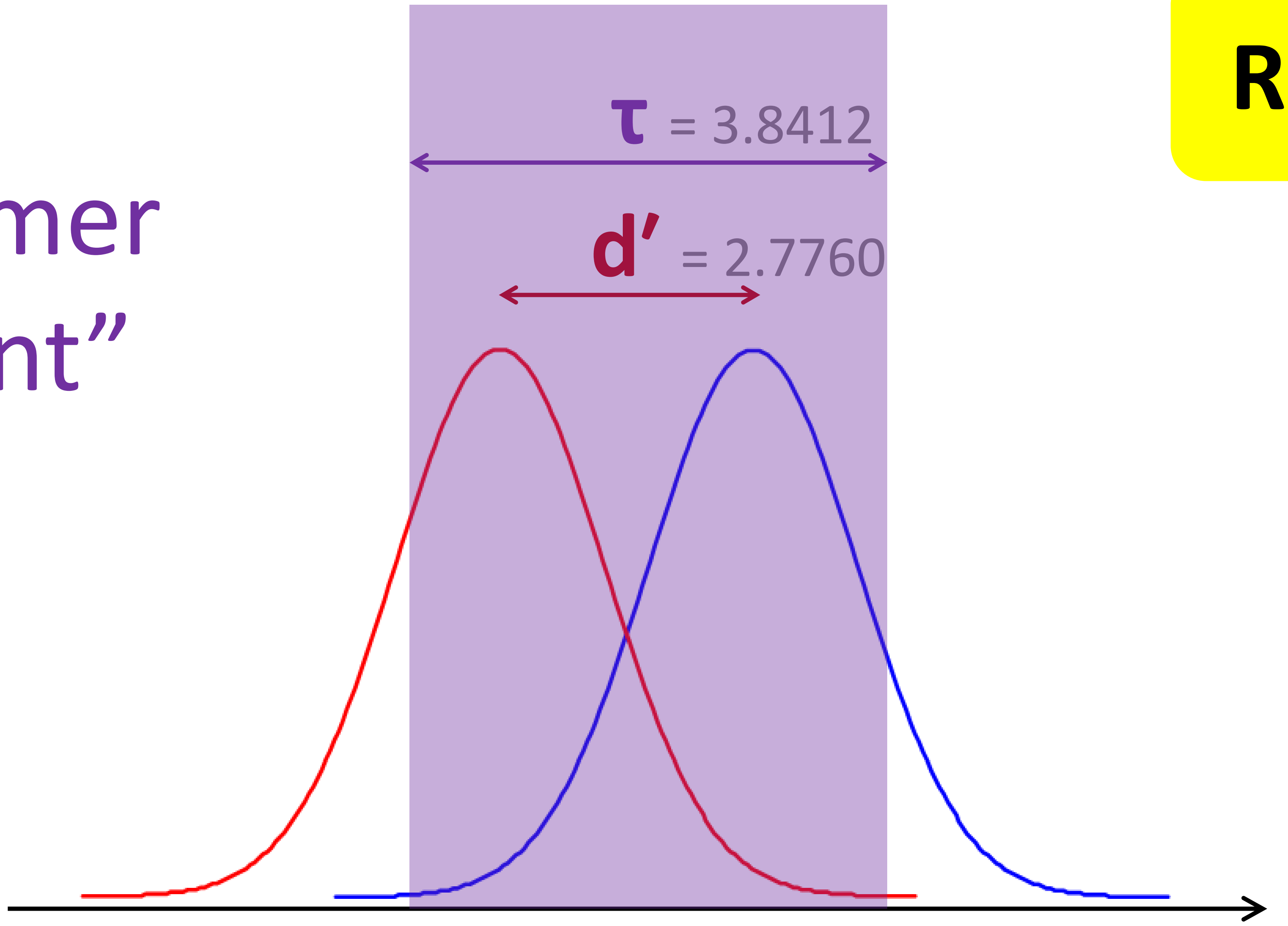


Pair	Possible Presentations	
AB	AB	BA
AC	AC	CA
AD	AD	DA
BC	BC	CB
BD	BD	DB
CD	CD	DC
AA	AA	
BB	BB	
CC	CC	
DD	DD	

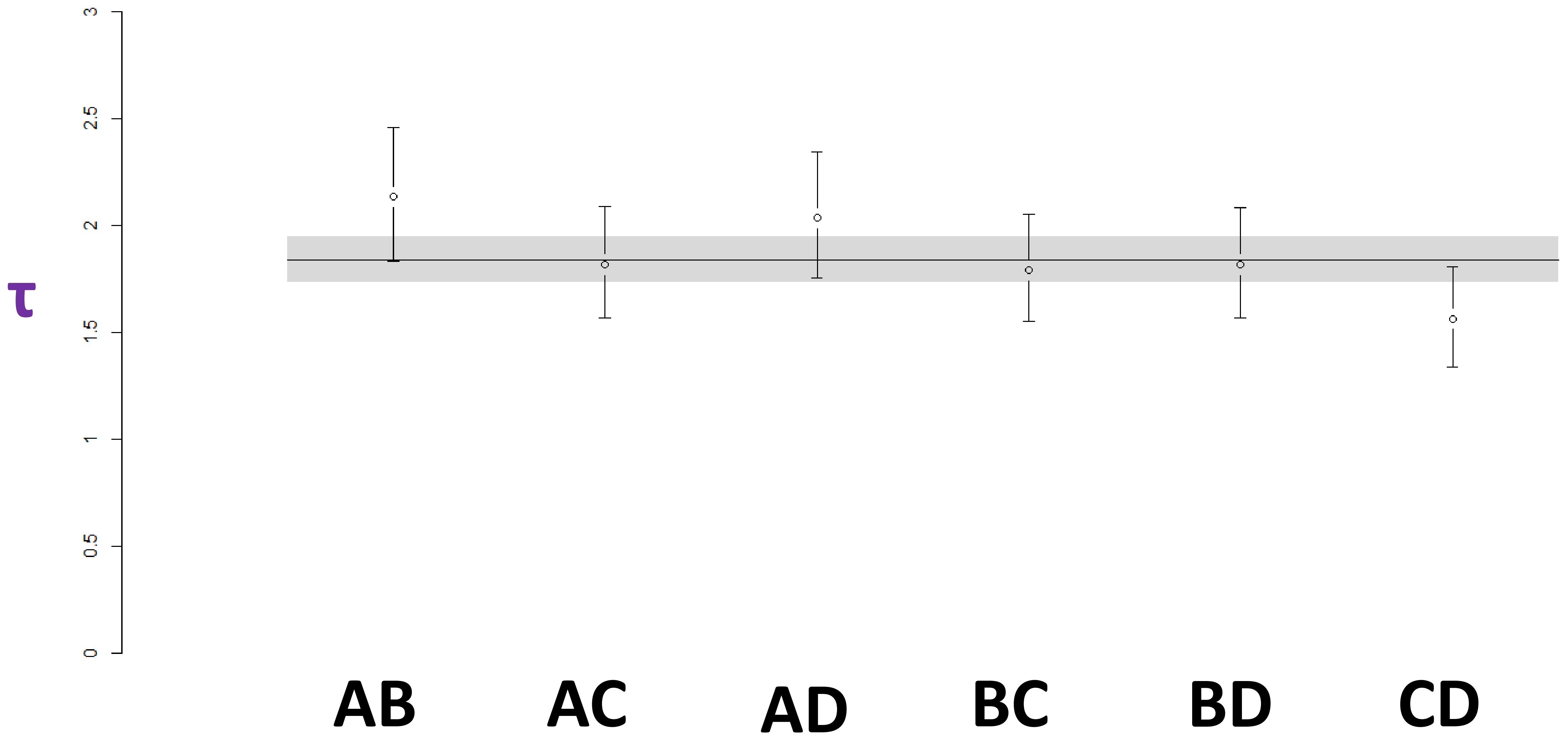

Same-Different
Tetrad
2AC Preference

“not consumer relevant”

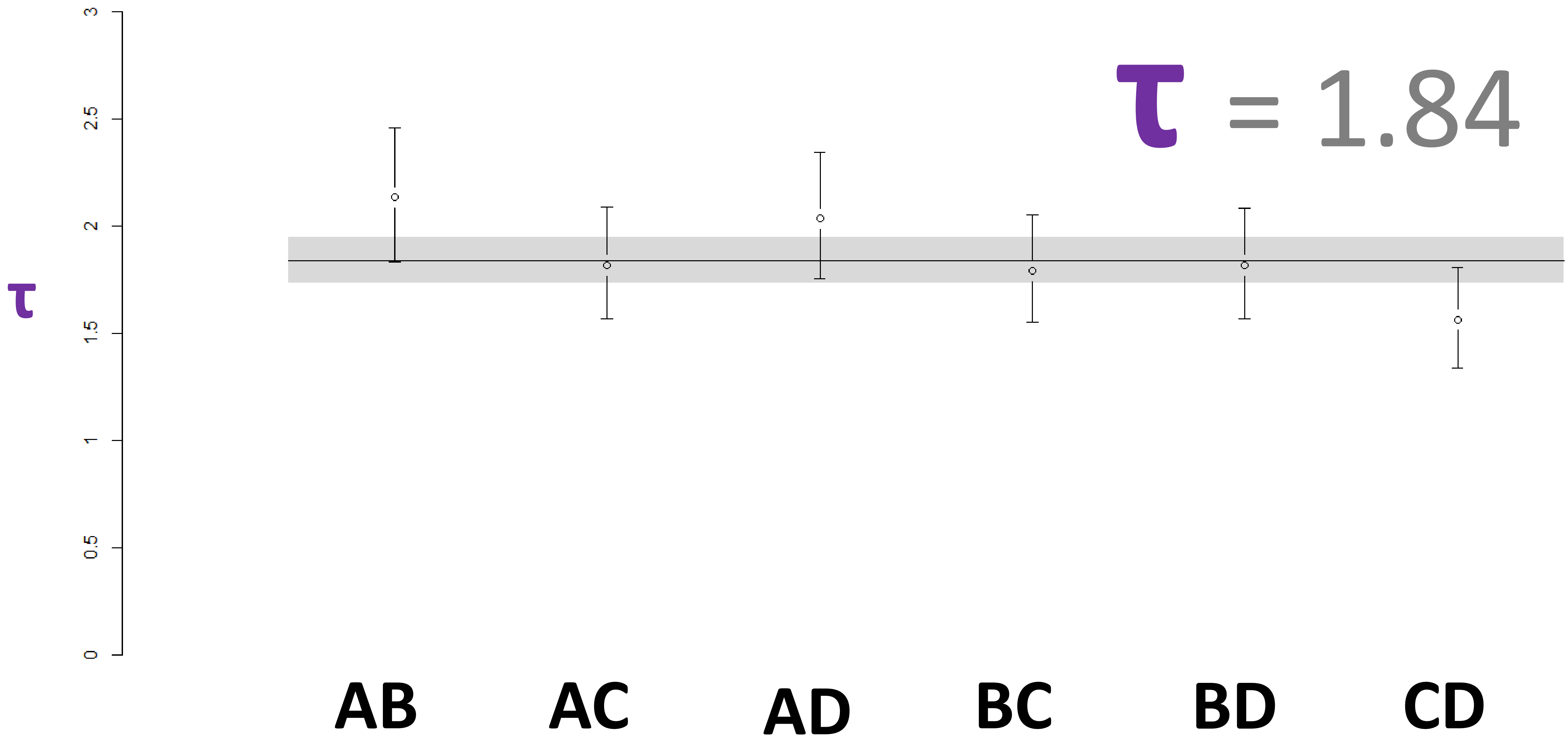
RECALL

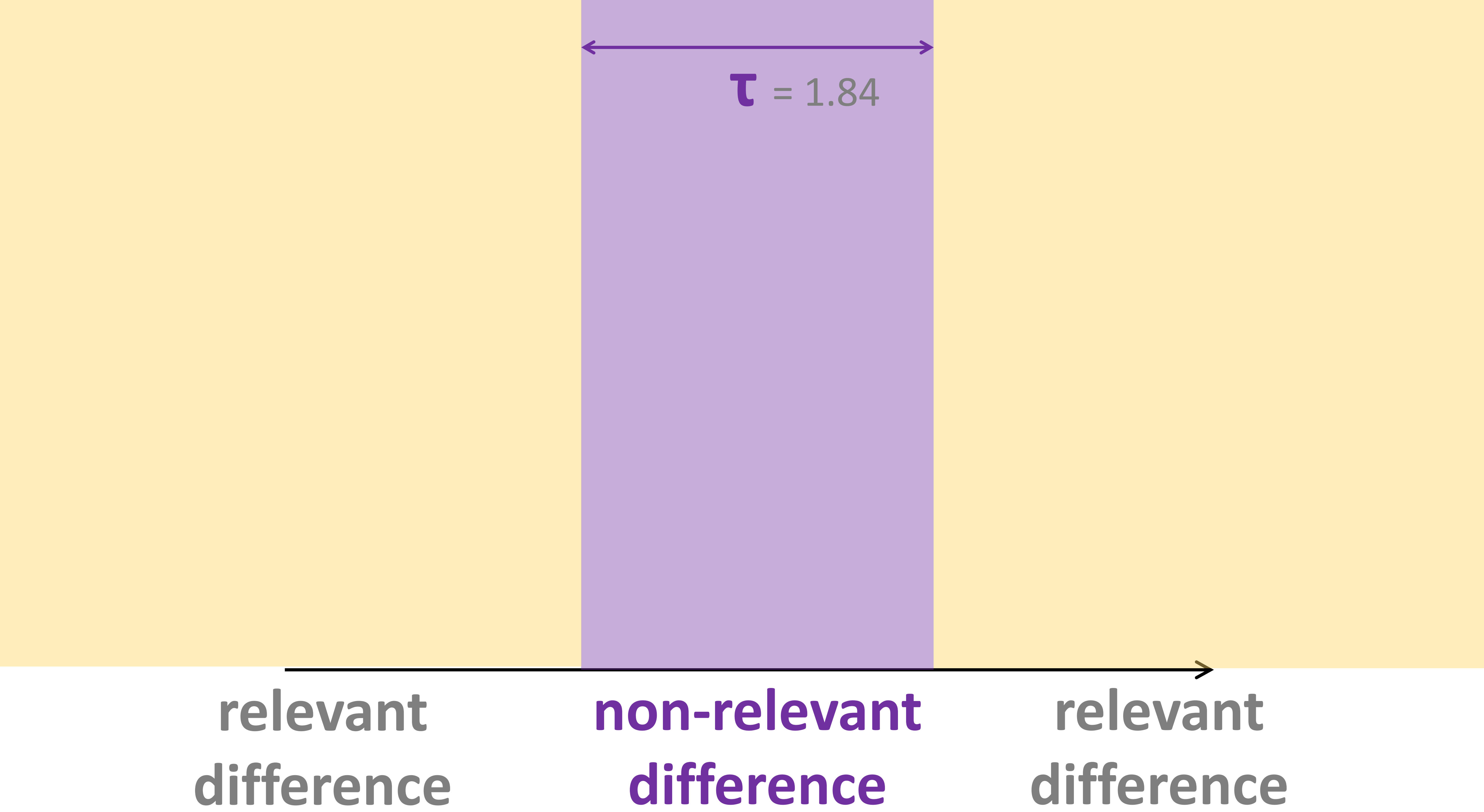


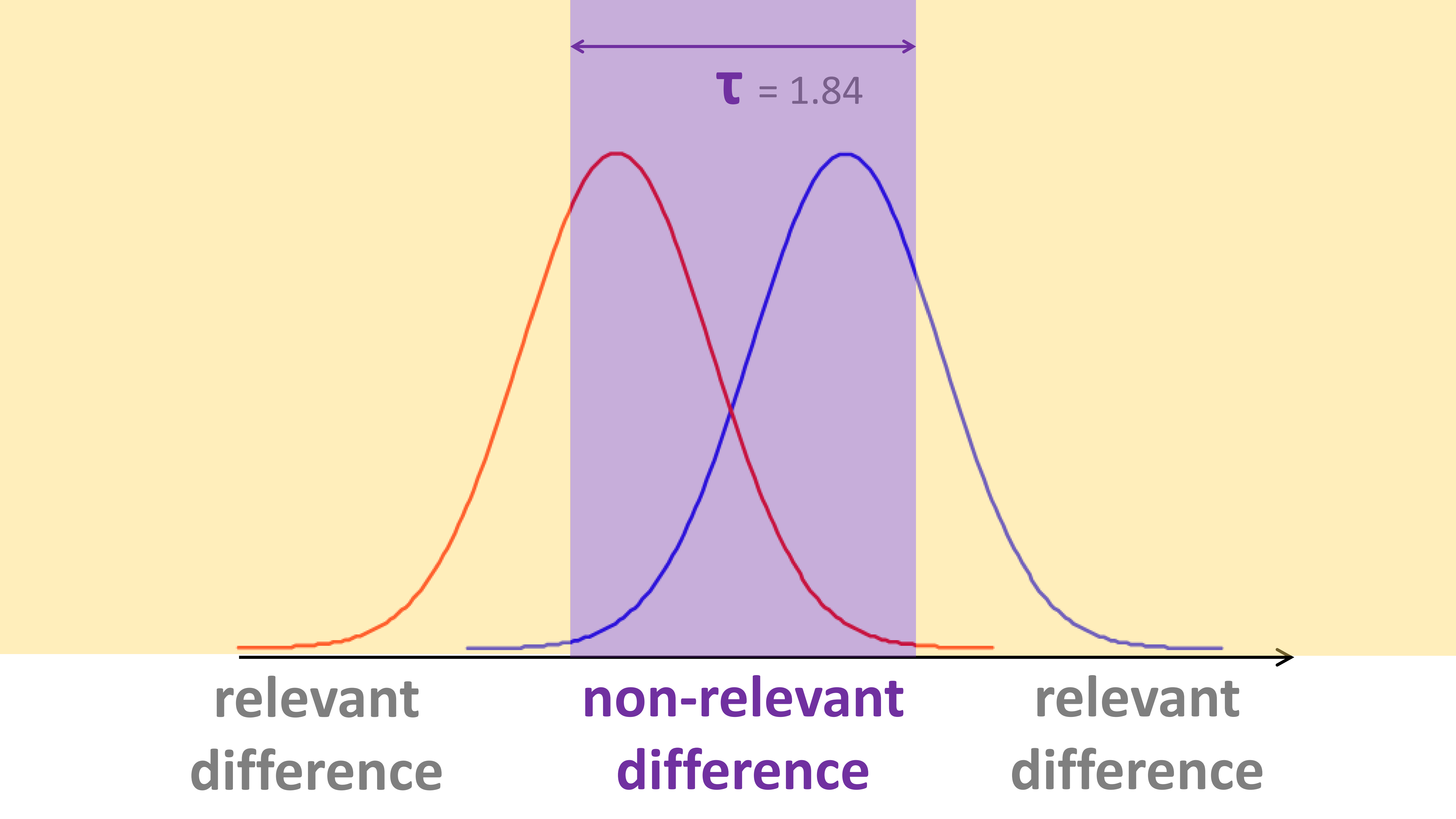
τ from Same–Different data

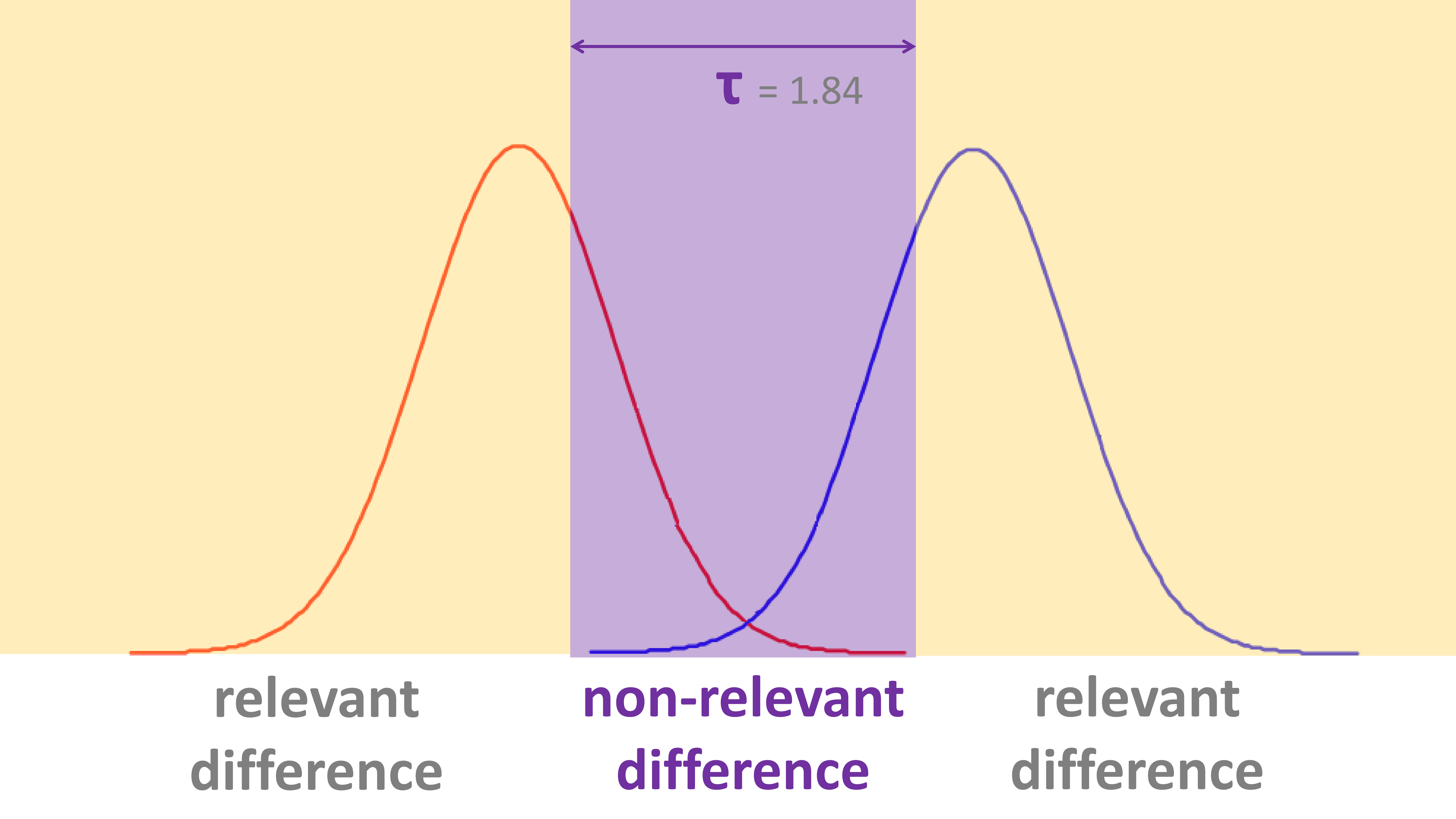


τ from Same–Different data




τ = 1.84
relevant difference
non-relevant difference
relevant difference

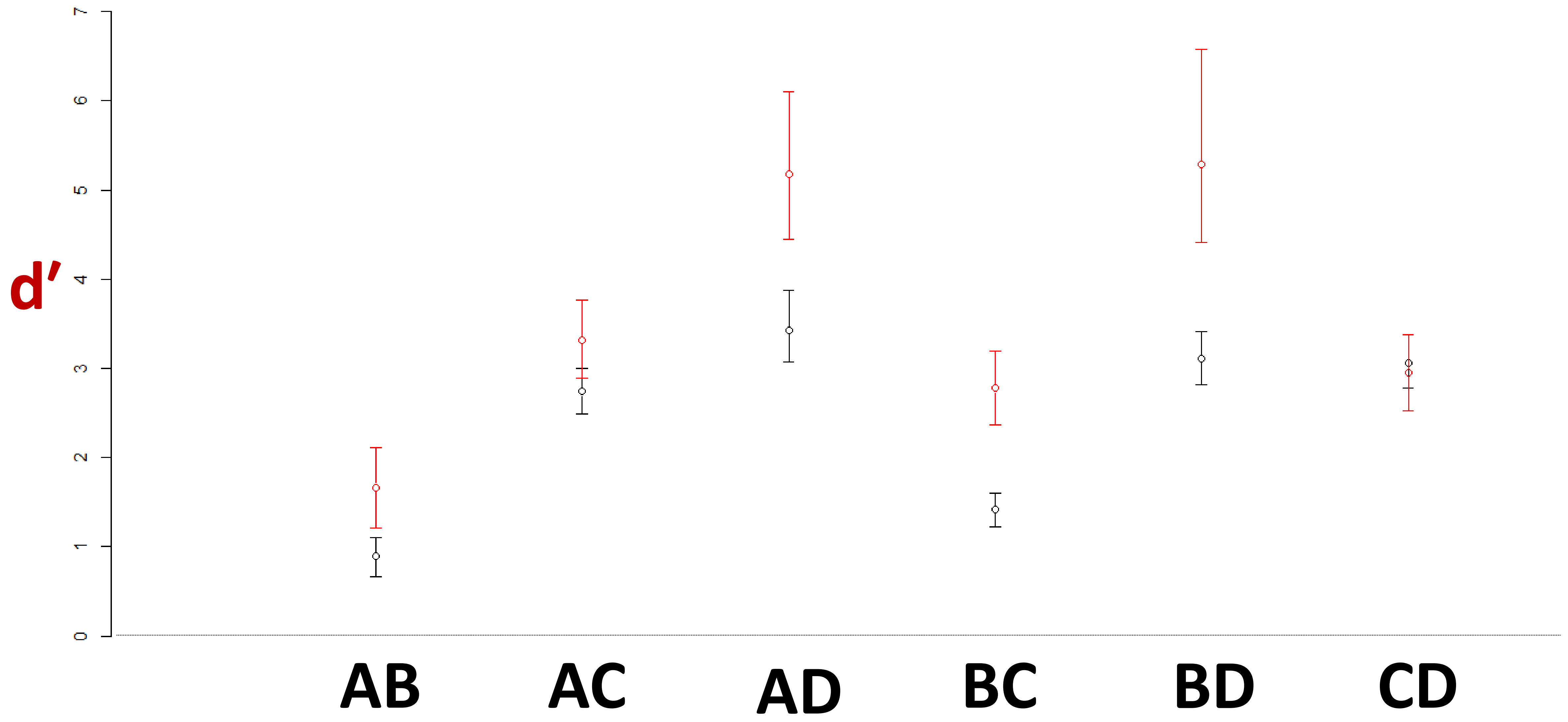

τ = 1.84
relevant difference
non-relevant difference
relevant difference


τ = 1.84
relevant difference
non-relevant difference
relevant difference

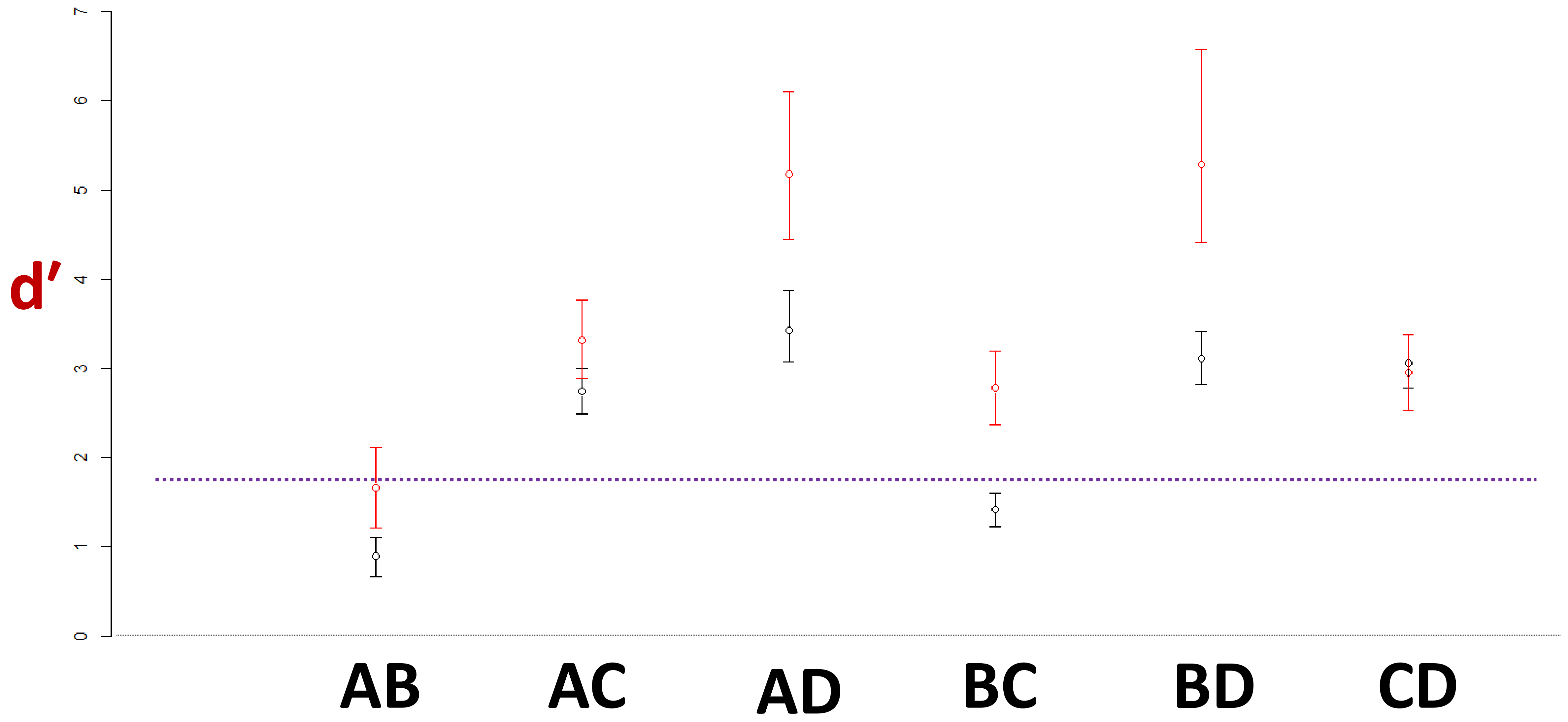
Pair	Same-different d’	Consumer relevant?
AB	1.67	No
AC	3.32	Yes
AD	5.18	Yes
BC	2.77	Yes
BD	5.29	Yes
CD	2.95	Yes

Pair	Same-different d’	Consumer relevant?	Tetrad d’	Consumer relevant?
AB	1.67	No	0.90	No
AC	3.32	Yes	2.74	Yes
AD	5.18	Yes	3.42	Yes
BC	2.77	Yes	1.42	No
BD	5.29	Yes	3.10	Yes
CD	2.95	Yes	3.06	Yes

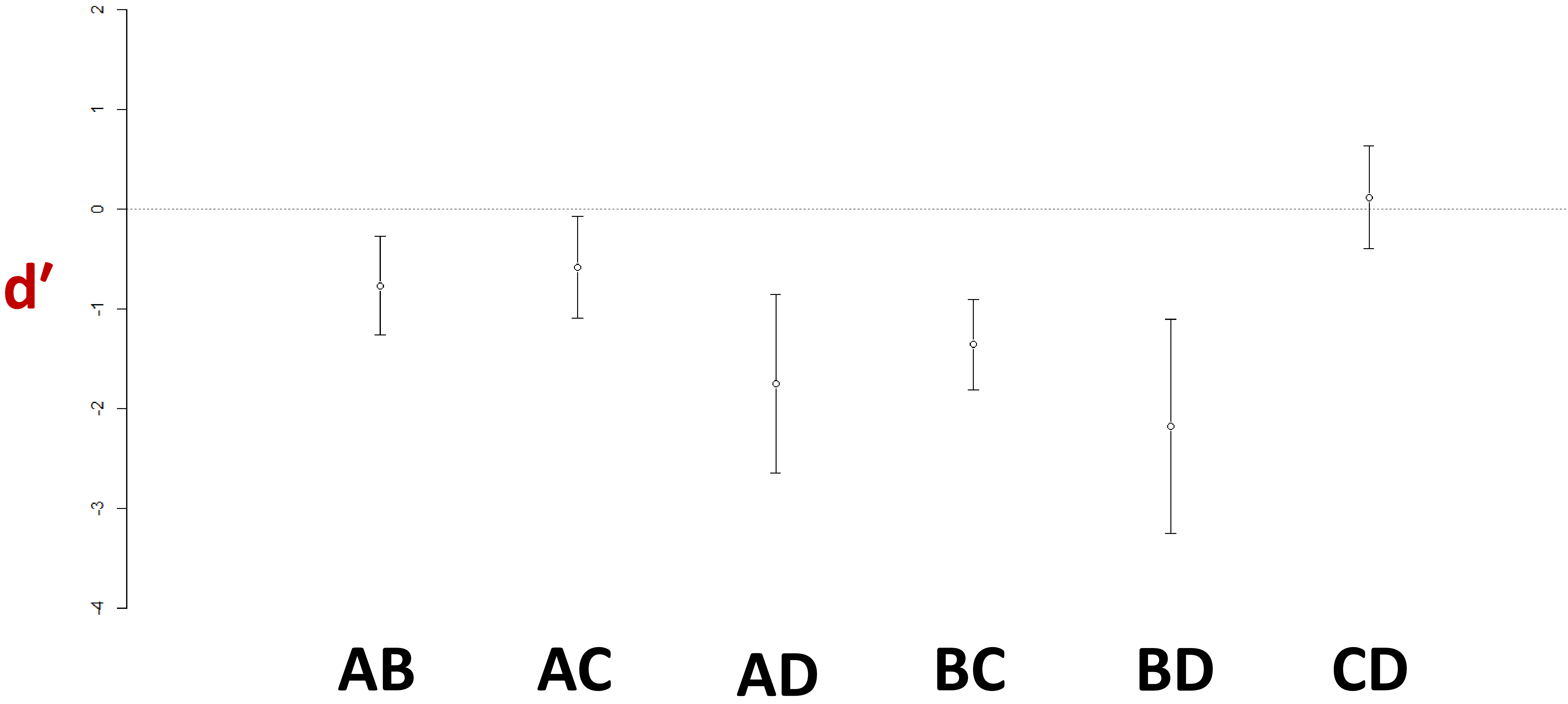
Same–Different d' ≠ Tetrad d'



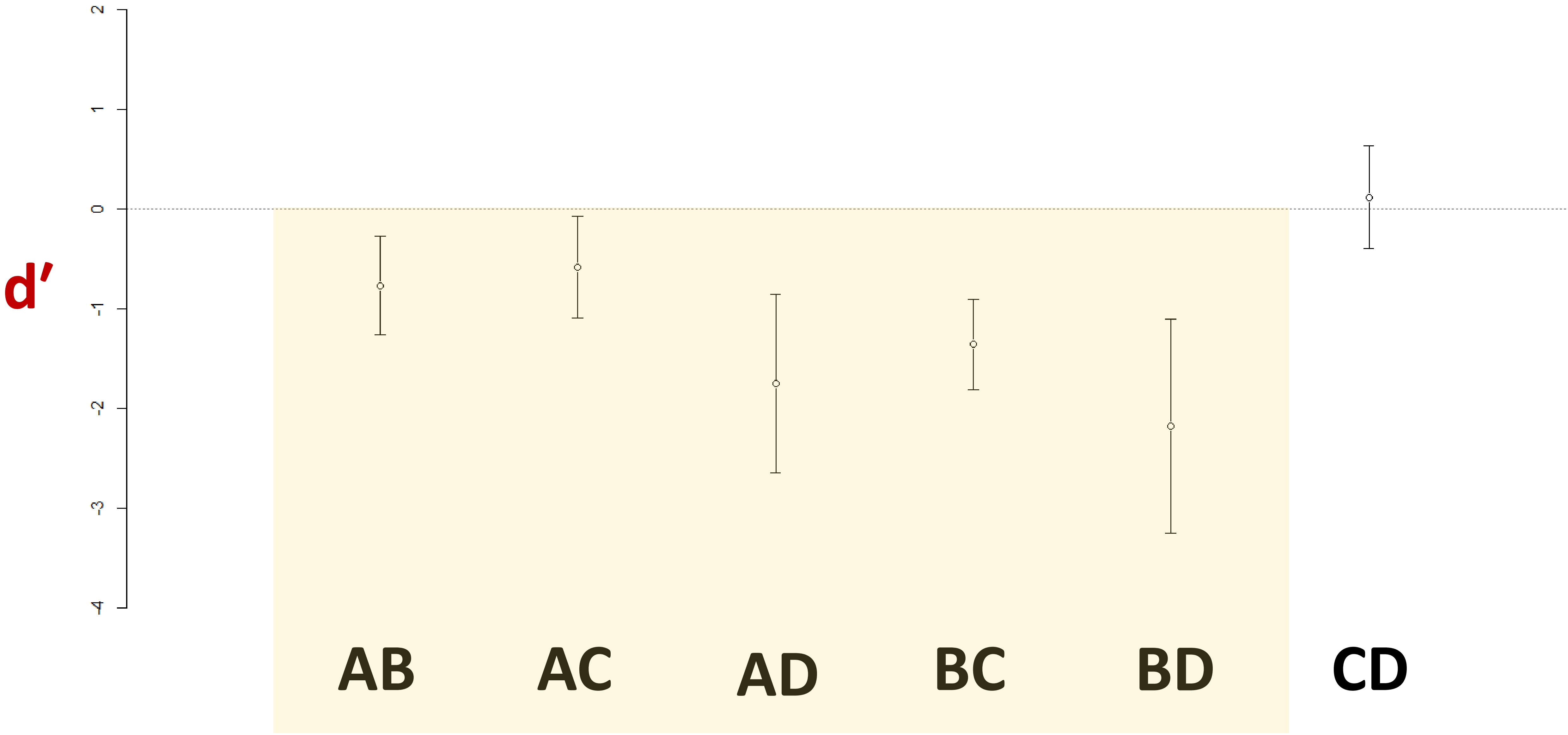
Same–Different d′ ≠ Tetrad d′



Tetrad d' – Same–Different d' ≠ 0



Tetrad d' – Same–Different d' ≠ 0



Conclusions

consumer data ≠ simulated data

Same–Different d′ ≠ Tetrad d′

Conclusions

Tetrad d' = discriminal distance

Reference: Castura & Franczak, 2017

Conclusions

Tetrad d' = discriminal distance

Same–Different d' = conceptual distance

Reference: Castura & Franczak, 2017

Conclusions

τ from **Same–Different** might not be method–independent!

Conclusions

For now, findings suggest that Rousseau's strategy for determining consumer relevance **cannot** be used across sensory method types

Conclusions

Follow up studies are needed to confirm or disconfirm these findings.

Selected References

Castura, J. C., & Franczak, B. C. (2017). Statistics for use in difference testing. In: L. Rogers (ed.), *Discrimination Testing in Sensory Science - A Practical Handbook*. Boca Raton, FL: Woodhead Publishing. (Forthcoming.)

Christensen, R. H. B. (2012). *Sensometrics: Thurstonian and Statistical Models*. Technical University of Denmark, Kongens Lyngby, Denmark. PhD Thesis.

Christensen, R. H. B., & Brockhoff, P. B. (2016). sensR - An R-package for sensory discrimination. R package version 1.4-7. http://www.cran.r-project.org/package=sensR/.

Rousseau, B. (2015). Sensory discrimination testing and consumer relevance. *Food Quality and Preference*, 43, 122-125.

Rousseau, B., & Ennis, D. M. (2013). When Are Two Products Close Enough to be Equivalent? *IFPress*, 16(1), 3-4.

Thank you for your kind attention

John C. Castura
Compusense Inc.
Guelph, ON, Canada

jcastura@compusense.com

Compusense