Copulas for neural and behavioral parallel systems

Hans Colonius

Department für Psychologie
Universität Oldenburg

Purdue Winer Memorial Lectures
Probability and Contextuality
November 2018

Outline

- Preliminaries
- Coupling: Definition
  - Example: Coupling Bernoulli Random Variables
  - Quantile Coupling
- Theory of Copulas
  - Defining the copula
  - Sklar's Theorem
  - Copula examples
  - Fréchet-Hoeffding bounds and copulas
  - Copulas and order statistics
- Application to parallel systems
  - The multisensory paradigm (behavioral (RT) version)
  - The multisensory paradigm (neural (spike counts) version)
- Response inhibition: stop signal paradigm
  - The independent race model
  - The paradox
  - The race model with perfect negative dependence

Intro remarks: Coupling and Copulas

Coupling and **Copula** are two important concepts of probability theory.

Their origins can be traced back to the fifties (coupling: thirties) of the last century, but interest in them waxed and waned for a long time.

Very loosely, **coupling** means the construction of the joint distribution of two or more (previously unrelated) random variables,

whereas **copulas** are functions that join multivariate distribution functions to their one-dimensional margins.

Copulas have recently become a very active area in statistics.

Preliminaries: Equality-in-distribution

- Let X be a random variable; $\hat{X}$ is a *copy* or a *representation* of X if it has the same distribution as X, denoted by

$$\hat{X} \stackrel{D}{=} X$$

First definition: Coupling

- A ***coupling*** of a collection of random variables X_i, $i \in \mathrm{I}$ (I some index set), is a family of random variables

$$(\hat{X}_i : i \in \mathrm{I}) \quad \text{such that} \quad \hat{X}_i \overset{D}{=} X_i, \quad i \in \mathrm{I}.$$

Note: The collection X_i need not be defined on a common probability space and may not have a joint distribution; the family $(\hat{X}_i : i \in \mathrm{I})$ has joint distribution such that the marginals are equal to the distributions of the X_i variables.

Example 1: Coupling two Bernoulli random variables

Let X_p be a Bernoulli random variable, i.e.,

$$P(X_p = 1) = p \text{ and } P(X_p = 0) = 1 - p.$$

Assume $p < q$; we can couple X_p and X_q as follows:
Let U be a uniform random variable on $[0, 1]$, i.e., for $0 \le a < b \le 1$,

$$P(a < U \le b) = b - a.$$

Define

$$\hat{X}_p = \begin{cases} 1 & \text{if } 0 < U \le p; \\ 0 & \text{if } p < U \le 1 \end{cases}; \quad \hat{X}_q = \begin{cases} 1 & \text{if } 0 < U \le q; \\ 0 & \text{if } q < U \le 1. \end{cases}$$

Then U serves as a common source of randomness for both $\hat{X}_p$ and $\hat{X}_q$. Moreover, $\hat{X}_p \overset{D}{=} X_p$ and $\hat{X}_q \overset{D}{=} X_q$.

Example 1: Coupling two Bernoulli random variables

Let X_p be a Bernoulli random variable, i.e.,

$$P(X_p = 1) = p \text{ and } P(X_p = 0) = 1 - p.$$

Assume $p < q$; we can couple X_p and X_q as follows:
Let U be a uniform random variable on $[0, 1]$, i.e., for $0 \leq a < b \leq 1$,

$$P(a < U \leq b) = b - a.$$

Define

$$\hat{X}_p = \begin{cases} 1 & \text{if } 0 < U \leq p; \\ 0 & \text{if } p < U \leq 1 \end{cases}; \quad \hat{X}_q = \begin{cases} 1 & \text{if } 0 < U \leq q; \\ 0 & \text{if } q < U \leq 1. \end{cases}$$

Then U serves as a common source of randomness for both $\hat{X}_p$ and $\hat{X}_q$. Moreover, $\hat{X}_p \overset{D}{=} X_p$ and $\hat{X}_q \overset{D}{=} X_q$.

Example 1: Coupling two Bernoulli random variables

Let X_p be a Bernoulli random variable, i.e.,

$$P(X_p = 1) = p \;\text{ and }\; P(X_p = 0) = 1 - p.$$

Assume $p < q$; we can couple X_p and X_q as follows:
Let U be a uniform random variable on $[0, 1]$, i.e., for $0 \leq a < b \leq 1$,

$$P(a < U \leq b) = b - a.$$

Define

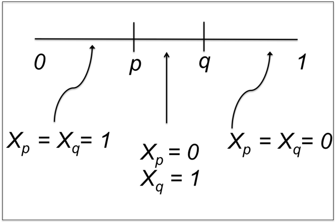
$$\hat{X}_p = \begin{cases} 1 & \text{if } 0 < U \leq p; \\ 0 & \text{if } p < U \leq 1 \end{cases}; \quad \hat{X}_q = \begin{cases} 1 & \text{if } 0 < U \leq q; \\ 0 & \text{if } q < U \leq 1. \end{cases}$$

Then U serves as a common source of randomness for both $\hat{X}_p$ and $\hat{X}_q$. Moreover, $\hat{X}_p \stackrel{D}{=} X_p$ and $\hat{X}_q \stackrel{D}{=} X_q$.

Example 1 (cont'ed): Coupling two Bernoulli random variables

- The joint distribution of $(\hat{X}_p, \hat{X}_q)$ is

$\hat{X}_p$ \ $\hat{X}_q$	0	1	
0	$1-q$	$q-p$	$1-p$
1	0	p	p
	$1-q$	q	1

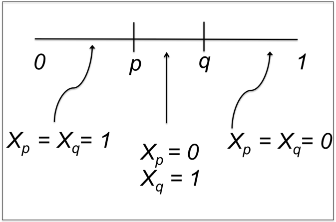


- and $\text{cov}(\hat{X}_p, \hat{X}_q) = p(1-q)$.

Example 1 (cont'ed): Coupling two Bernoulli random variables

- The joint distribution of $(\hat{X}_p, \hat{X}_q)$ is

$\hat{X}_p$ \ $\hat{X}_q$	0	1	
0	$1-q$	$q-p$	$1-p$
1	0	p	p
	$1-q$	q	1

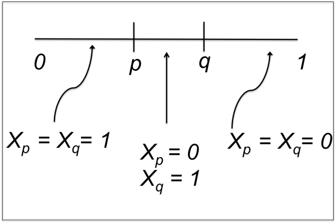
- 

- and $\text{cov}(\hat{X}_p, \hat{X}_q) = p(1-q)$.

Example 1 (cont'ed): Coupling two Bernoulli random variables

- The joint distribution of $(\hat{X}_p, \hat{X}_q)$ is

		$\hat{X}_q$ = 0	$\hat{X}_q$ = 1	
$\hat{X}_p$	0	$1-q$	$q-p$	$1-p$
	1	0	p	p
		$1-q$	q	1

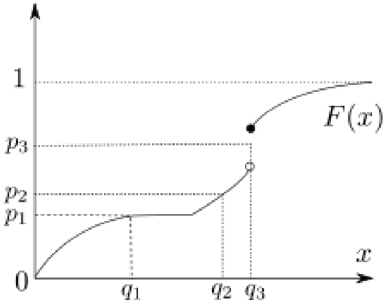
- 

- and $\mathrm{cov}(\hat{X}_p, \hat{X}_q) = p(1-q)$.

Preliminaries: The Quantile Function

Let X be a real-valued random variable with distribution function $F(x)$ which is continuous from the right. Then, the *quantile function* (or, *generalized inverse*) $Q(u)$ of X is defined as

$$Q(u) \equiv F^{-1}(u) = \inf\{x : F(x) \geq u\}, \ \ 0 \leq u \leq 1.$$



Example 2: Quantile Coupling

Let X be a random variable with distribution function F, that is,

$$P(X \leq x) = F(x), \quad x \in \Re.$$

Let U be a uniform random variable on $[0,1]$. Then, for random variable $\hat{X} = F^{-1}(U)$,

$$P(\hat{X} \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x), \quad x \in \Re,$$

that is, $\hat{X}$ is a copy of X: $\quad \hat{X} \stackrel{D}{=} X$.

Thus, letting F run over the class of all distribution functions (using the same U), yields a coupling of all differently distributed random variables, the *quantile coupling*.

Example 2: Quantile Coupling

Let X be a random variable with distribution function F, that is,

$$P(X \leq x) = F(x), \quad x \in \Re.$$

Let U be a uniform random variable on $[0, 1]$. Then, for random variable $\hat{X} = F^{-1}(U)$,

$$P(\hat{X} \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x), \quad x \in \Re,$$

that is, $\hat{X}$ is a copy of X: $\quad \hat{X} \stackrel{D}{=} X$.

Thus, letting F run over the class of all distribution functions (using the same U), yields a coupling of all differently distributed random variables, the *quantile coupling*.

Example 2: Quantile Coupling

Let X be a random variable with distribution function F, that is,

$$P(X \leq x) = F(x), \quad x \in \Re.$$

Let U be a uniform random variable on $[0, 1]$. Then, for random variable $\hat{X} = F^{-1}(U)$,

$$P(\hat{X} \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x), \quad x \in \Re,$$

that is, $\hat{X}$ is a copy of X: $\quad \hat{X} \stackrel{D}{=} X$.

Thus, letting F run over the class of all distribution functions (using the same U), yields a coupling of all differently distributed random variables, the *quantile coupling*.

Theory of Copulas

Copula: Example

Let (X, Y) be a pair of random variables with joint distribution function $F(x, y)$ and marginal distributions $F_X(x)$ and $F_Y(y)$.

To each pair of real numbers (x, y) we can associate three numbers: $F_X(x)$, $F_Y(y)$, and $F(x, y)$. Note that each of these numbers lies in the interval $[0, 1]$.

In other words, each pair (x, y) of real numbers leads to a point $(F_X(x), F_Y(y))$ in the unit square $[0, 1] \times [0, 1]$, and this ordered pair in turn corresponds to a number $F(x, y)$ in $[0, 1]$.

$$(x, y) \mapsto (F_X(x), F_Y(y)) \mapsto F(x, y) = C(F_X(x), F_Y(y))$$

$$\Re \times \Re \to [0, 1] \times [0, 1] \xrightarrow{C} [0, 1]$$

This is a copula C.

Copula: Example

Let (X, Y) be a pair of random variables with joint distribution function $F(x, y)$ and marginal distributions $F_X(x)$ and $F_Y(y)$.

To each pair of real numbers (x, y) we can associate three numbers: $F_X(x)$, $F_Y(y)$, and $F(x, y)$. Note that each of these numbers lies in the interval $[0, 1]$.

In other words, each pair (x, y) of real numbers leads to a point $(F_X(x), F_Y(y))$ in the unit square $[0, 1] \times [0, 1]$, and this ordered pair in turn corresponds to a number $F(x, y)$ in $[0, 1]$.

$$(x, y) \mapsto (F_X(x), F_Y(y)) \mapsto F(x, y) = C(F_X(x), F_Y(y))$$

$$\Re \times \Re \to [0, 1] \times [0, 1] \stackrel{C}{\to} [0, 1]$$

This is a copula C.

Copula: Example

Let (X, Y) be a pair of random variables with joint distribution function $F(x, y)$ and marginal distributions $F_X(x)$ and $F_Y(y)$.

To each pair of real numbers (x, y) we can associate three numbers: $F_X(x)$, $F_Y(y)$, and $F(x, y)$. Note that each of these numbers lies in the interval $[0, 1]$.

In other words, each pair (x, y) of real numbers leads to a point $(F_X(x), F_Y(y))$ in the unit square $[0, 1] \times [0, 1]$, and this ordered pair in turn corresponds to a number $F(x, y)$ in $[0, 1]$.

$$(x, y) \mapsto (F_X(x), F_Y(y)) \mapsto F(x, y) = C(F_X(x), F_Y(y))$$

$$\Re \times \Re \to [0, 1] \times [0, 1] \xrightarrow{C} [0, 1]$$

This is a copula C.

Defining the copula

Definition

An *n-copula* is an n-variate distribution function with univariate margins uniformly distributed on $[0,1]$.

- There are many equivalent definitions of copula.
- One of them, in the case $n = 2$, is the following:

Definition

A function $C : [0,1] \times [0,1] \to [0,1]$ is a $2-$*copula* if, and only if, it satisfies

1. $C(0,t) = C(t,0) = 0$, for every $t \in [0,1]$ (groundedness);
2. $C(1,t) = C(t,1) = t$, for every $t \in [0,1]$ (uniform marginals);
3. for all $a_1, a_2, b_1, b_2 \in [0,1]$, with $a_1 \leq b_1$ and $a_2 \leq b_2$,

$$(*) \quad C(a_1, a_2) - C(a_1, b_2) - C(b_1, a_2) + C(b_1, b_2) \geq 0.$$

$(*)$ is called *2-increasing, supermodularity*.

Defining the copula

Definition

An *n-copula* is an *n*-variate distribution function with univariate margins uniformly distributed on $[0, 1]$.

- There are many equivalent definitions of copula.
- One of them, in the case $n = 2$, is the following:

Definition

A function $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is a $2-$*copula* if, and only if, it satisfies

1. $C(0, t) = C(t, 0) = 0$, for every $t \in [0, 1]$ (groundedness);
2. $C(1, t) = C(t, 1) = t$, for every $t \in [0, 1]$ (uniform marginals);
3. for all $a_1, a_2, b_1, b_2 \in [0, 1]$, with $a_1 \leq b_1$ and $a_2 \leq b_2$,

$$(*) \quad C(a_1, a_2) - C(a_1, b_2) - C(b_1, a_2) + C(b_1, b_2) \geq 0.$$

$(*)$ is called *2-increasing, supermodularity.*

Defining the copula

Definition

An *n-copula* is an *n*-variate distribution function with univariate margins uniformly distributed on $[0,1]$.

- There are many equivalent definitions of copula.
- One of them, in the case $n = 2$, is the following:

Definition

A function $C : [0,1] \times [0,1] \to [0,1]$ is a $2-$*copula* if, and only if, it satisfies

1. $C(0,t) = C(t,0) = 0$, for every $t \in [0,1]$ (groundedness);
2. $C(1,t) = C(t,1) = t$, for every $t \in [0,1]$ (uniform marginals);
3. for all $a_1, a_2, b_1, b_2 \in [0,1]$, with $a_1 \leq b_1$ and $a_2 \leq b_2$,

$$(*) \quad C(a_1,a_2) - C(a_1,b_2) - C(b_1,a_2) + C(b_1,b_2) \geq 0.$$

$(*)$ is called *2-increasing, supermodularity.*

Defining the copula

Definition

An *n-copula* is an n-variate distribution function with univariate margins uniformly distributed on $[0, 1]$.

- There are many equivalent definitions of copula.
- One of them, in the case $n = 2$, is the following:

Definition

A function $C : [0, 1] \times [0, 1] \to [0, 1]$ is a $2-$*copula* if, and only if, it satisfies

1. $C(0, t) = C(t, 0) = 0$, for every $t \in [0, 1]$ (groundedness);
2. $C(1, t) = C(t, 1) = t$, for every $t \in [0, 1]$ (uniform marginals);
3. for all $a_1, a_2, b_1, b_2 \in [0, 1]$, with $a_1 \leq b_1$ and $a_2 \leq b_2$,

$$(*) \quad C(a_1, a_2) - C(a_1, b_2) - C(b_1, a_2) + C(b_1, b_2) \geq 0.$$

$(*)$ is called *2-increasing, supermodularity*.

Sklar's Theorem

Theorem (Sklar's Theorem, 1959)

Let $F(x_1, \ldots, x_n)$ be an n-variate distribution function with margins $F_1(x_1), \ldots, F_n(x_n)$; then there exists an n-copula $C : [0,1]^n \longrightarrow [0,1]$ that satisfies

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)), \quad (x_1, \ldots x_n) \in \Re^n.$$

If all univariate margins $F_1, \ldots, F_n$ are continuous, then the copula is unique.

If $F_1^{-1}, \ldots, F_n^{-1}$ are the quantile functions of the margins, then for any $(u_1, \ldots, u_n) \in [0,1]^n$

$$C(u_1, \ldots, u_n) = F(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)).$$

Note: The copula can be considered 'independent' of the univariate margins.

Sklar's Theorem

Theorem (Sklar's Theorem, 1959)

Let $F(x_1, \ldots, x_n)$ be an n-variate distribution function with margins $F_1(x_1), \ldots, F_n(x_n)$; then there exists an n-copula $C : [0,1]^n \longrightarrow [0,1]$ that satisfies

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)), \quad (x_1, \ldots x_n) \in \Re^n.$$

If all univariate margins $F_1, \ldots, F_n$ are continuous, then the copula is unique.

If $F_1^{-1}, \ldots, F_n^{-1}$ are the quantile functions of the margins, then for any $(u_1, \ldots, u_n) \in [0,1]^n$

$$C(u_1, \ldots, u_n) = F(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)).$$

Note: The copula can be considered 'independent' of the univariate margins.

How to prove that a function is a copula

- Find a suitable probabilistic model (i.e., random vector) whose distribution function is concentrated on $[0, 1]^n$ and has uniform marginals; or,
- prove that properties (1) groundedness, (2) uniform marginals, and (3) n-increasing (supermodularity) are satisfied.

How to prove that a function is a copula

- Find a suitable probabilistic model (i.e., random vector) whose distribution function is concentrated on $[0, 1]^n$ and has uniform marginals; or,
- prove that properties (1) groundedness, (2) uniform marginals, and (3) n-increasing (supermodularity) are satisfied.

Properties of copulas

Sklar's theorem shows that copulas remain invariant under strictly increasing transformations of the underlying random variables.

It is possible to construct a wide range of multivariate distributions by choosing the marginal distributions and a suitable copula.

Example 1: Gumbel's bivariate exponential copula

Let H_θ be the joint distribution function given by

$$H_\theta(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y+\theta xy)} & \text{if } x \geq 0, y \geq 0; \\ 0, & \text{otherwise}; \end{cases}$$

where θ is a parameter in $[0,1]$.

Then the marginals are exponentials, with quantile functions $F^{-1}(u) = -\ln(1-u)$ and $G^{-1}(v) = -\ln(1-v)$ for $u, v \in [0,1]$.
The corresponding copula is

$$C_\theta(u,v) = u + v - 1 + (1-u)(1-v)e^{-\theta \ln(1-u)\ln(1-v)}.$$

Example 1: Gumbel's bivariate exponential copula

Let H_θ be the joint distribution function given by

$$H_\theta(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y+\theta xy)} & \text{if } x \geq 0, y \geq 0; \\ 0, & \text{otherwise;} \end{cases}$$

where θ is a parameter in $[0,1]$.
Then the marginals are exponentials, with quantile functions $F^{-1}(u) = -\ln(1-u)$ and $G^{-1}(v) = -\ln(1-v)$ for $u, v \in [0,1]$.
The corresponding copula is

$$C_\theta(u,v) = u + v - 1 + (1-u)(1-v)e^{-\theta \ln(1-u)\ln(1-v)}.$$

Example 1: Gumbel's bivariate exponential copula

Let H_θ be the joint distribution function given by

$$H_\theta(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y+\theta xy)} & \text{if } x \geq 0, y \geq 0; \\ 0, & \text{otherwise;} \end{cases}$$

where θ is a parameter in $[0,1]$.
Then the marginals are exponentials, with quantile functions $F^{-1}(u) = -\ln(1-u)$ and $G^{-1}(v) = -\ln(1-v)$ for $u, v \in [0,1]$.
The corresponding copula is

$$C_\theta(u,v) = u + v - 1 + (1-u)(1-v)e^{-\theta \ln(1-u)\ln(1-v)}.$$

Example 2: Bivariate extreme value distribution

Let X and Y be random variables with a joint distribution given by

$$H_\theta(x, y) = \exp[-(e^{-\theta x} + e^{-\theta y})^{1/\theta}]$$

for all $x, y \in \bar{\Re}$, where $\theta \geq 1$.

The corresponding *Gumbel-Hougaard* copula is given by

$$C_\theta(u, v) = \exp\left(-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right).$$

Example 2: Bivariate extreme value distribution

Let X and Y be random variables with a joint distribution given by

$$H_\theta(x, y) = \exp[-(e^{-\theta x} + e^{-\theta y})^{1/\theta}]$$

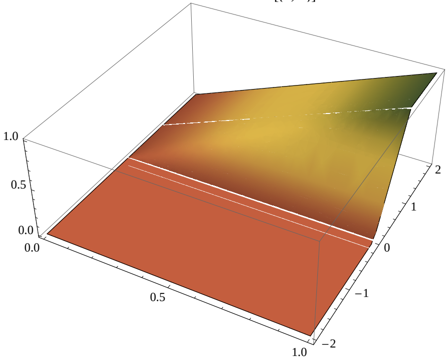
for all $x, y \in \bar{\Re}$, where $\theta \geq 1$.

The corresponding *Gumbel-Hougaard* copula is given by

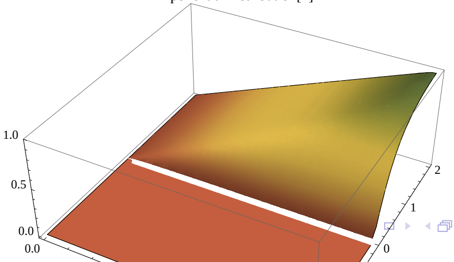
$$C_\theta(u, v) = \exp\left(-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right).$$

Plotting Gumbel-Hougaard copula with different marginals

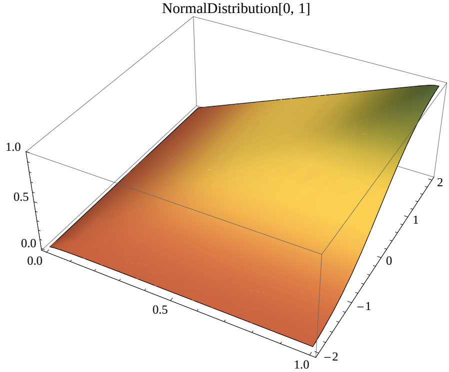
UniformDistribution[{0, 1}]

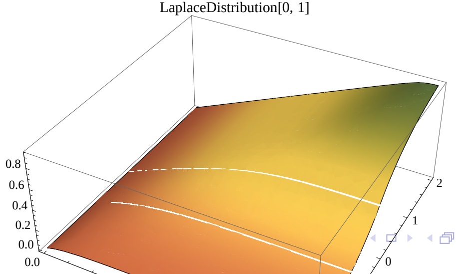


ExponentialDistribution[2]

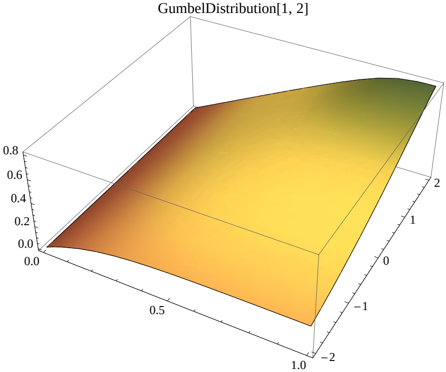


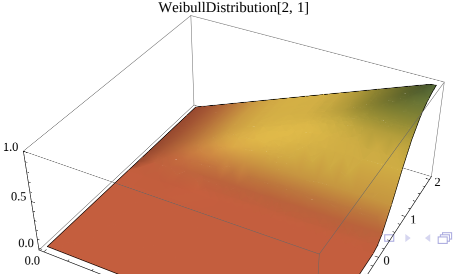
Plotting Gumbel-Hougaard copula with different marginals





Plotting Gumbel-Hougaard copula with different marginals





Fréchet-Hoeffding bounds and copulas

These will be central topics in the applications presented here.

Fréchet-Hoeffding copulas

Let $C(u, v)$ be a 2-copula; then, for $u, v \in [0, 1]$,

$$W(u, v) \equiv \max\{u + v - 1, 0\} \leq C(u, v) \leq \min\{u, v\} \equiv M(u, v),$$

and M and W are also copulas, the *upper* and *lower* *Fréchet-Hoeffding copula*.

The dependence properties translate into:
For W, $P(U + V = 1) = 1$,
and for M, $P(U = V) = 1$.

Fréchet-Hoeffding copulas

Let $C(u, v)$ be a 2-copula; then, for $u, v \in [0, 1]$,

$$W(u, v) \equiv \max\{u + v - 1, 0\} \leq C(u, v) \leq \min\{u, v\} \equiv M(u, v),$$

and M and W are also copulas, the *upper* and *lower* *Fréchet-Hoeffding copula*.

The dependence properties translate into:
For W, $P(U + V = 1) = 1$,
and for M, $P(U = V) = 1$.

Copulas and order statistics: diagonal section

Maximum

Let $U_1, U_2, \ldots, U_n$ be random variables defined on the same prob. space., having uniform distribution on $(0,1)$ and C as their distribution function; then, for every $t \in [0,1]$

$$P(\max\{U_1, U_2, \ldots, U_n\}) \leq t) = P\left(\bigcap_{j=1}^{n}\{U_j \leq t\}\right)$$
$$= C(t, t, \ldots, t)$$

$\delta_C(t) := C(t, t, \ldots, t)$ is called the *diagonal section of the copula C*.

Copulas and order statistics: diagonal section

Maximum

Let $U_1, U_2, \dots, U_n$ be random variables defined on the same prob. space., having uniform distribution on $(0,1)$ and C as their distribution function; then, for every $t \in [0,1]$

$$\begin{aligned} P(\max\{U_1, U_2, \dots, U_n\}) \leq t) &= P\left(\bigcap_{j=1}^{n} \{U_j \leq t\}\right) \\ &= C(t, t, \dots, t) \end{aligned}$$

$\delta_C(t) := C(t, t, \dots, t)$ is called the *diagonal section of the copula* C.

Copulas and order statistics: diagonal section

Minimum

For $n = 2$, it follows

$$P(\min\{U, V\}) \leq t) = P(U \leq t) + P(V \leq t) - P(U \leq t \cap V \leq t)$$
$$= 2t - \delta_C(t).$$

Thus, determining a bivariate copula C with prescribed diagonal section δ is equivalent to determining a random vector (U, V) such that $(U, V) \sim C$ and the marginal distribution functions of the order statistics of (U, V) are known.

Note: the diagonal of a copula does not uniquely determine the underlying copula.

Copulas and order statistics: diagonal section

Minimum

For $n = 2$, it follows

$$P(\min\{U, V\}) \leq t) = P(U \leq t) + P(V \leq t) - P(U \leq t \cap V \leq t)$$
$$= 2t - \delta_C(t).$$

Thus, determining a bivariate copula C with prescribed diagonal section δ is equivalent to determining a random vector (U, V) such that $(U, V) \sim C$ and the marginal distribution functions of the order statistics of (U, V) are known.

Note: the diagonal of a copula does not uniquely determine the underlying copula.

Copulas and order statistics: diagonal section

Minimum

For $n = 2$, it follows

$$
\begin{aligned}
P(\min\{U, V\}) \le t) &= P(U \le t) + P(V \le t) - P(U \le t \cap V \le t) \\
&= 2t - \delta_C(t).
\end{aligned}
$$

Thus, determining a bivariate copula C with prescribed diagonal section δ is equivalent to determining a random vector (U, V) such that $(U, V) \sim C$ and the marginal distribution functions of the order statistics of (U, V) are known.

Note: the diagonal of a copula does not uniquely determine the underlying copula.

Application to parallel systems

We loosely define

> Parallel system: "Existence of one or more multivariate distributions on possibly different probability spaces, where random variables relate to certain behavioral or neural processes".

Copulas for neural and behavioral parallel systems

(1) Multisensory integration: reaction times
(2) Multisensory integration: spike numbers (impulse numbers)
(3) Response inhibition: stop signal paradigm

In (1) and (2), we present a new measure of multisensory integration that unifies behavioral and neural data.

In (3), we modify a classic model of inhibitory behavior in a way to solve a paradox between behavioral and neural data.

Copulas for neural and behavioral parallel systems

(1) Multisensory integration: reaction times

(2) Multisensory integration: spike numbers (impulse numbers)

(3) Response inhibition: stop signal paradigm

In (1) and (2), we present a new measure of multisensory integration that unifies behavioral and neural data.

In (3), we modify a classic model of inhibitory behavior in a way to solve a paradox between behavioral and neural data.

Copulas for neural and behavioral parallel systems

(1) Multisensory integration: reaction times
(2) Multisensory integration: spike numbers (impulse numbers)
(3) Response inhibition: stop signal paradigm

In (1) and (2), we present a new measure of multisensory integration that unifies behavioral and neural data.

In (3), we modify a classic model of inhibitory behavior in a way to solve a paradox between behavioral and neural data.

The multisensory paradigm (behavioral version)

- *Unimodal* condition: a stimulus of a single modality (visual, auditory, tactile) is presented and the participant is asked to respond (by button press or eye movement) as quickly as possible upon detecting the stimulus (*reaction time, RT, task*).
- *Bi- or trimodal* condition: stimuli from two or three modalities are presented (nearly) simultaneously and the participant is asked to respond as quickly as possible upon detecting a stimulus of any modality (*redundant signals task*)

The multisensory paradigm (behavioral version)

- We refer to $\mathcal{V}, \mathcal{A}, \mathcal{T}$ as the unimodal context where visual, auditory, or tactile stimuli are presented, resp. Simlarly, $\mathcal{VA}$ denotes a bimodal (visual-auditory) context, etc.
- For each stimulus, or stimulus combination, we observe samples from a random variable representing the reaction time measured in any given trial. Let $F_V(t), F_A(t), F_{VA}(t), \ldots$ denote the distribution functions of reaction time in a unimodal visual, auditory, or a visual-auditory context, etc. when a specific stimulus (combination) is presented.

The multisensory paradigm (behavioral version)

- We refer to $\mathcal{V}, \mathcal{A}, \mathcal{T}$ as the unimodal context where visual, auditory, or tactile stimuli are presented, resp. Simlarly, $\mathcal{VA}$ denotes a bimodal (visual-auditory) context, etc.
- For each stimulus, or stimulus combination, we observe samples from a random variable representing the reaction time measured in any given trial. Let $F_V(t), F_A(t), F_{VA}(t), \ldots$ denote the distribution functions of reaction time in a unimodal visual, auditory, or a visual-auditory context, etc. when a specific stimulus (combination) is presented.

Existence of couplings

- Each context $\mathcal{V}, \mathcal{A}, \mathcal{VA}$ refers to a different event space (σ-algebra), so, from an empirical point of view, no couplings between the reaction time random variables in these different conditions necessarily exist.
- A common assumption, often not stated explicitly, is that there exists a coupling between visual and auditory RT, for example, such that the margins of the bivariate distribution H_{VA} are equal to F_V and F_A.
- Given a coupling exists, an assumption on how H_{VA} is related to the margins of F_V and F_A, is required ($\Rightarrow$ copula ?).
- In principle, this could be tested empirically. However, the margins of H_{VA} are not observable, only the distribution function of RTs in the bimodal context, F_{VA}, is.

Existence of couplings

- Each context $\mathcal{V}, \mathcal{A}, \mathcal{VA}$ refers to a different event space (σ-algebra), so, from an empirical point of view, no couplings between the reaction time random variables in these different conditions necessarily exist.
- A common assumption, often not stated explicitly, is that there exists a coupling between visual and auditory RT, for example, such that the margins of the bivariate distribution H_{VA} are equal to F_V and F_A.
- Given a coupling exists, an assumption on how H_{VA} is related to the margins of F_V and F_A, is required ($\Rightarrow$ copula ?).
- In principle, this could be tested empirically. However, the margins of H_{VA} are not observable, only the distribution function of RTs in the bimodal context, F_{VA}, is.

Existence of couplings

- Each context $\mathcal{V}, \mathcal{A}, \mathcal{VA}$ refers to a different event space (σ-algebra), so, from an empirical point of view, no couplings between the reaction time random variables in these different conditions necessarily exist.
- A common assumption, often not stated explicitly, is that there exists a coupling between visual and auditory RT, for example, such that the margins of the bivariate distribution H_{VA} are equal to F_V and F_A.
- Given a coupling exists, an assumption on how H_{VA} is related to the margins of F_V and F_A, is required ($\Rightarrow$ copula ?).
- In principle, this could be tested empirically. However, the margins of H_{VA} are not observable, only the distribution function of RTs in the bimodal context, F_{VA}, is.

Existence of couplings

- Each context $\mathcal{V}, \mathcal{A}, \mathcal{VA}$ refers to a different event space (σ-algebra), so, from an empirical point of view, no couplings between the reaction time random variables in these different conditions necessarily exist.
- A common assumption, often not stated explicitly, is that there exists a coupling between visual and auditory RT, for example, such that the margins of the bivariate distribution H_{VA} are equal to F_V and F_A.
- Given a coupling exists, an assumption on how H_{VA} is related to the margins of F_V and F_A, is required ($\Rightarrow$ copula ?).
- In principle, this could be tested empirically. However, the margins of H_{VA} are not observable, only the distribution function of RTs in the bimodal context, F_{VA}, is.

The (possibly non-independent) race model

The model studied most often is the *(possibly non-independent) race model*: Let V and A be the random reaction times in unimodal conditions $\mathcal{V}$ and $\mathcal{A}$, with distribution functions $F_V(t)$ and $F_A(t)$, resp.

- Assume bimodal RT is determined by the "winner" of the race between the modalities:

$$F_{VA}(t) = P(V \leq t \text{ or } A \leq t)$$

- Then

$$F_{VA}(t) = F_V(t) + F_A(t) - H_{VA}(t,t),$$

with $H_{VA}(t,t) = P(V \leq t \text{ and } A \leq t)$.

The (possibly non-independent) race model

The model studied most often is the *(possibly non-independent) race model*: Let V and A be the random reaction times in unimodal conditions $\mathcal{V}$ and $\mathcal{A}$, with distribution functions $F_V(t)$ and $F_A(t)$, resp.

- Assume bimodal RT is determined by the "winner" of the race between the modalities:

$$F_{VA}(t) = P(V \leq t \text{ or } A \leq t)$$

- Then

$$F_{VA}(t) = F_V(t) + F_A(t) - H_{VA}(t,t),$$

with $H_{VA}(t,t) = P(V \leq t \text{ and } A \leq t)$.

The (possibly non-independent) race model

The model studied most often is the *(possibly non-independent) race model*: Let V and A be the random reaction times in unimodal conditions $\mathcal{V}$ and $\mathcal{A}$, with distribution functions $F_V(t)$ and $F_A(t)$, resp.

- Assume bimodal RT is determined by the "winner" of the race between the modalities:

$$F_{VA}(t) = P(V \leq t \text{ or } A \leq t)$$

- Then

$$F_{VA}(t) = F_V(t) + F_A(t) - H_{VA}(t,t),$$

with $H_{VA}(t,t) = P(V \leq t \text{ and } A \leq t)$.

The (possibly non-independent) race model

- from the Fréchet bounds,

$$\max\{F_V(t) + F_A(t) - 1, 0\} \leq H_{VA}(t, t) \leq \min\{F_V(t), F_A(t)\}.$$

- Inserting and rearranging yields

$$\max\{F_V(t), F_A(t)\} \leq F_{VA}(t) \leq \min\{F_V(t) + F_A(t), 1\}, \ t \geq 0,$$

('race model inequality', Miller 1982)

- The upper bound corresponds to maximal negative dependence between V and A, the lower bound to maximal positive dependence.
- Empirical violation of the upper bound (occurring only for small enough t) is interpreted as evidence against the race mechanism ("bimodal RT faster than predictable from unimodal conditions").

The (possibly non-independent) race model

- from the Fréchet bounds,

$$\max\{F_V(t) + F_A(t) - 1, 0\} \leq H_{VA}(t,t) \leq \min\{F_V(t), F_A(t)\}.$$

- Inserting and rearranging yields

$$\max\{F_V(t), F_A(t)\} \leq F_{VA}(t) \leq \min\{F_V(t) + F_A(t), 1\}, \; t \geq 0,$$

('race model inequality', Miller 1982)

- The upper bound corresponds to maximal negative dependence between V and A, the lower bound to maximal positive dependence.
- Empirical violation of the upper bound (occurring only for small enough t) is interpreted as evidence against the race mechanism ("bimodal RT faster than predictable from unimodal conditions").

The (possibly non-independent) race model

- from the Fréchet bounds,

  $$\max\{F_V(t) + F_A(t) - 1, 0\} \le H_{VA}(t,t) \le \min\{F_V(t), F_A(t)\}.$$

- Inserting and rearranging yields

  $$\max\{F_V(t), F_A(t)\} \le F_{VA}(t) \le \min\{F_V(t) + F_A(t), 1\}, \; t \ge 0,$$

  ('race model inequality', Miller 1982)
- The upper bound corresponds to maximal negative dependence between V and A, the lower bound to maximal positive dependence.
- Empirical violation of the upper bound (occurring only for small enough t) is interpreted as evidence against the race mechanism ("bimodal RT faster than predictable from unimodal conditions").

The (possibly non-independent) race model

- from the Fréchet bounds,

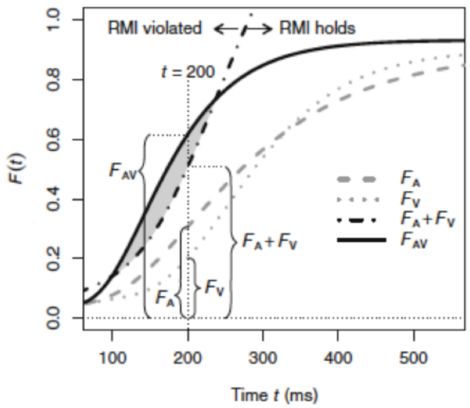
  $$\max\{F_V(t) + F_A(t) - 1, 0\} \leq H_{VA}(t, t) \leq \min\{F_V(t), F_A(t)\}.$$

- Inserting and rearranging yields

  $$\max\{F_V(t), F_A(t)\} \leq F_{VA}(t) \leq \min\{F_V(t) + F_A(t), 1\}, \ t \geq 0,$$

  ('race model inequality', Miller 1982)
- The upper bound corresponds to maximal negative dependence between V and A, the lower bound to maximal positive dependence.
- Empirical violation of the upper bound (occurring only for small enough t) is interpreted as evidence against the race mechanism ("bimodal RT faster than predictable from unimodal conditions").

Race Model Inequality Test



Gondan & Minakata 2016

The grey area between the upper bound $F_A + F_V$ and the bimodal RT distribution F_{VA} is taken as a **measure** of the amount of violation of the race model inequality. **This area is the (sample estimate of the) expected value of** $\min(A, V)$:

$$\mathrm{E}^{(-)}[\min(V, A)],$$

under **maximal negative** **dependence between** V **and** A (Colonius & Diederich *PsyRev* 2006). Estimation is straightforward.

Crossmodal response enhancement (CRE)

- We will use maximal negative probability summation to define a new measure of crossmodal response enhancemen for RT.
- Response enhancement in RT means "faster average responses":

$$\mathrm{CRE}_{\mathrm{RT}} = \frac{\min\{\mathrm{E}[RT_V], \mathrm{E}[RT_A]\} - \mathrm{E}[RT_{VA}]}{\min\{\mathrm{E}[RT_V], \mathrm{E}[RT_A]\}} \times 100$$

Crossmodal response enhancement (CRE)

- We will use maximal negative probability summation to define a new measure of crossmodal response enhancemen for RT.
- Response enhancement in RT means "faster average responses":

$$\mathrm{CRE}_{\mathrm{RT}} = \frac{\min\{\mathrm{E}[RT_V], \mathrm{E}[RT_A]\} - \mathrm{E}[RT_{VA}]}{\min\{\mathrm{E}[RT_V], \mathrm{E}[RT_A]\}} \times 100$$

The new measure of CRE in RT

1. Replace the traditional

$$\mathrm{CRE_{RT}} = \frac{\min\{\mathrm{E}[RT_V], \mathrm{E}[RT_A]\} - \mathrm{E}[RT_{VA}]}{\min\{\mathrm{E}[RT_V], \mathrm{E}[RT_A]\}} \times 100$$

   by

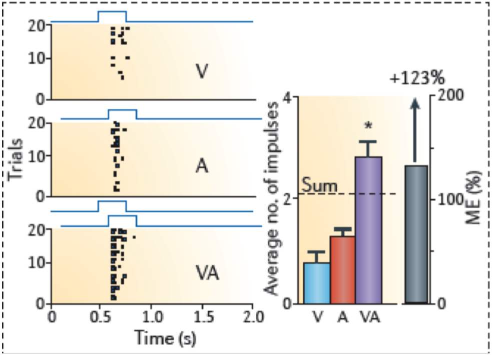
$$\mathrm{CRE_{RT}}^{(-)} = \frac{\mathrm{E}^{(-)}[\min(V, A)] - \mathrm{E}[RT_{VA}]}{\mathrm{E}^{(-)}[\min(V, A)]} \times 100$$

2. $\mathrm{CRE_{RT}}^{(-)} \leq \mathrm{CRE_{RT}}$

The multisensory paradigm (neural version)

CRE at the level of a single neuron

Response strength: the absolute number of impulses (spikes) registered within in a fixed time interval after stimulus presentation



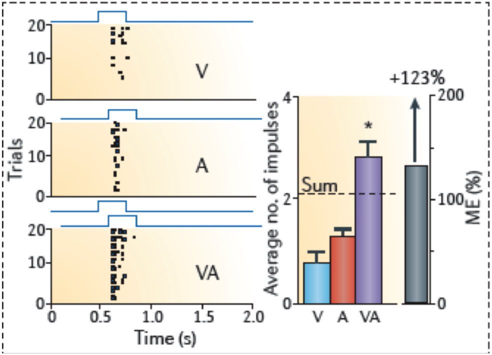
Stein et al., Nat Rev Neurosci 2014

$$\mathrm{CRE} = \frac{\mathrm{CM}_{VA} - \max\{\mathrm{UM}_V, \mathrm{UM}_A\}}{\max\{\mathrm{UM}_V, \mathrm{UM}_A\}} \times 100,$$

$\mathrm{CRE} = 123[\%]$

CRE at the level of a single neuron

Response strength: the absolute number of impulses (spikes) registered within in a fixed time interval after stimulus presentation



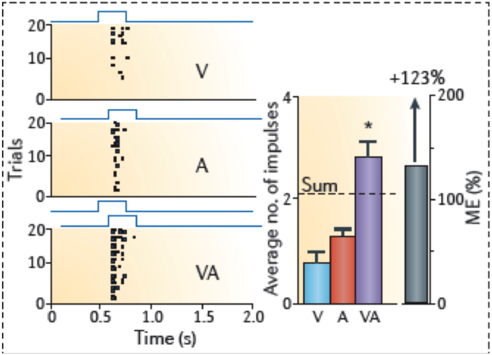
Stein et al., Nat Rev Neurosci 2014

$$\mathrm{CRE} = \frac{\mathrm{CM}_{VA} - \max\{\mathrm{UM}_V, \mathrm{UM}_A\}}{\max\{\mathrm{UM}_V, \mathrm{UM}_A\}} \times 100,$$

$\mathrm{CRE} = 123[\%]$

CRE at the level of a single neuron

Response strength: the absolute number of impulses (spikes) registered within in a fixed time interval after stimulus presentation



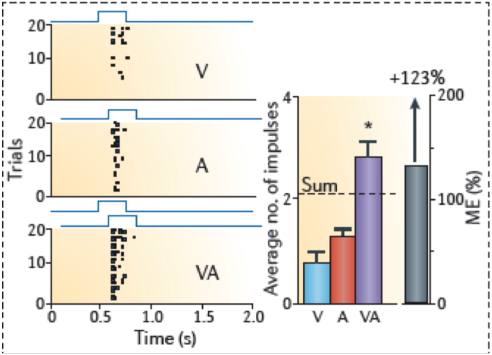
Stein et al., Nat Rev Neurosci 2014

$$\text{CRE} = \frac{\text{CM}_{VA} - \max\{\text{UM}_V, \text{UM}_A\}}{\max\{\text{UM}_V, \text{UM}_A\}} \times 100,$$

$\text{CRE} = 123[\%]$

CRE at the level of a single neuron

Response strength: the absolute number of impulses (spikes) registered within in a fixed time interval after stimulus presentation



Stein et al., Nat Rev Neurosci 2014

$$\mathrm{CRE} = \frac{\mathrm{CM}_{VA} - \max\{\mathrm{UM}_V, \mathrm{UM}_A\}}{\max\{\mathrm{UM}_V, \mathrm{UM}_A\}} \times 100,$$

$\mathrm{CRE} = 123[\%]$

CRE at the level of a single neuron

- this measure is a very useful tool, but it is purely descriptive
- no theoretical foundation in terms of the possible operations SC neurons may perform

  $\Longrightarrow$ Being responsive to multiple sensory modalities does not guarantee that a neuron has actually engaged in integrating its multiple sensory inputs rather than simply responding to the most salient stimulus.

  Multisensory computations performed by SC neurons still not fully understood

CRE at the level of a single neuron

- this measure is a very useful tool, but it is purely descriptive
- no theoretical foundation in terms of the possible operations SC neurons may perform

$\Longrightarrow$ Being responsive to multiple sensory modalities does not guarantee that a neuron has actually engaged in integrating its multiple sensory inputs rather than simply responding to the most salient stimulus.

Multisensory computations performed by SC neurons still not fully understood

CRE at the level of a single neuron

- this measure is a very useful tool, but it is purely descriptive
- no theoretical foundation in terms of the possible operations SC neurons may perform

  $\Longrightarrow$ Being responsive to multiple sensory modalities does not guarantee that a neuron has actually engaged in integrating its multiple sensory inputs rather than simply responding to the most salient stimulus.

  Multisensory computations performed by SC neurons still not fully understood

CRE at the level of a single neuron

- this measure is a very useful tool, but it is purely descriptive
- no theoretical foundation in terms of the possible operations SC neurons may perform

  $\Longrightarrow$ Being responsive to multiple sensory modalities does not guarantee that a neuron has actually engaged in integrating its multiple sensory inputs rather than simply responding to the most salient stimulus.

  Multisensory computations performed by SC neurons still not fully understood

CRE for single neuron data

- **Task:** Develop a measure of crossmodal enhancement for single neuron data, in analogy to $\mathrm{CRE_{RT}}^{(-)}$.
- N_V, N_A, and N_{VA} denote random number of impulses emitted by a neuron, following unisensory (visual, auditory) and cross-modal stimulation. The traditional index is:

$$\mathrm{CRE_{MAX}} = \frac{\mathrm{E}[N_{VA}] - \max\{\mathrm{E}[N_V], \mathrm{E}[N_A]\}}{\max\{\mathrm{E}[N_V], \mathrm{E}[N_A]\}} \times 100.$$

Towards a new index of CRE in neural data

$G_V(m) = P_V(N_V > m)$ and $G_A(m) = P_A(N_A > m)$, $m = 0, 1, \ldots$ the (survivor) distribution functions of N_V and N_A.

In analogy to the race model inequality, from the *Fréchet-Hoeffding bounds*,

$$\begin{aligned}\min\{G_V(m), G_A(m)\} &\leq P(\max\{N_V, N_A\} > m) \\ &\leq \max\{0, G_V(m) + G_A(m) - 1\}\end{aligned}$$

for $m = 0, 1, \ldots$

Again, upper and lower bounds are distribution functions (survival fcts) for random variable $\max\{N_V, N_A\}$!

Summing over m and applying *Jensen' s inequality*, we obtain

Proposition on Coupling N_V and N_A

Proposition

For any coupling of the univariate response random variables N_V and N_A with expected value $\mathrm{E}[\max\{N_V, N_A\}]$,

$$\max\{\mathrm{E}[N_V], \mathrm{E}[N_A]\} \leq \mathrm{E}[\max\{N_V, N_A\}] \leq \mathrm{E}^{(-)}[\max\{N_V, N_A\}],$$

where $\mathrm{E}^{(-)}[\max\{N_V, N_A\}]$ is the expected value under maximal negative dependence between N_V and N_A.

A a new index of CRE in neural data

Replace the traditional index

$$\mathrm{CRE}_{\mathrm{MAX}} = \frac{\mathrm{E}[N_{VA}] - \max\{\mathrm{E}[N_V], \mathrm{E}[N_A]\}}{\max\{\mathrm{E}[N_V], \mathrm{E}[N_A]\}} \times 100.$$

by

$$\mathrm{CRE}^{(-)}_{\mathrm{MAX}} = \frac{\mathrm{E}[N_{VA}] - \mathrm{E}^{(-)}[\max\{N_V, N_A\}]}{\mathrm{E}^{(-)}[\max\{N_V, N_A\}]} \times 100.$$

$\mathrm{CRE}^{(-)}_{\mathrm{MAX}}$ compares the observed bimodal response $\mathrm{E}N_{VA}$ with the largest bimodal response achievable by coupling the unisensory responses via negative stochastic dependence.

An Important Consequence

- $\mathrm{CRE}^{(-)}_{\mathrm{MAX}} \leq \mathrm{CRE}_{\mathrm{MAX}} \Rightarrow$
  Thus, a neuron labeled as "multisensory" under $\mathrm{CRE}_{\mathrm{MAX}}$ may lose that property under $\mathrm{CRE}^{(-)}_{\mathrm{MAX}}$.

Illustrative Example: Single SC Neuron Data

* Data set from Mark Wallace (Vanderbilt U.):
  84 sessions, with 15 trials each, from 3 superior colliculus neurons (cat)

  (Results in Colonius & Diederich, *Scientific Reports*, 2017.)

Intermediate summary

- We suggest to replace the traditional multisensory index $\mathrm{CRE}_{\mathrm{MAX}}$ by the new one, $\mathrm{CRE}^{(-)}_{\mathrm{MAX}}$.
- The new one has a theoretical foundation: it measures the degree by which a neuron's observed multisensory response surpasses the level obtainable by optimally combining the unisensory responses (assuming that the neuron simply reacts to the more salient modality in any given cross-modal trial).
- $\mathrm{CRE}^{(-)}_{\mathrm{MAX}}$ is easy to compute and does not require any specific assumption about the distribution of the spikes.
- It is in straightforward analogy to a measure well-established in the domain of reaction times ("race model inequality").
- It is sensitive to the variability of the data set.

Intermediate summary

- We suggest to replace the traditional multisensory index $\mathrm{CRE}_{\mathrm{MAX}}$ by the new one, $\mathrm{CRE}_{\mathrm{MAX}}^{(-)}$.
- The new one has a theoretical foundation: it measures the degree by which a neuron's observed multisensory response surpasses the level obtainable by optimally combining the unisensory responses (assuming that the neuron simply reacts to the more salient modality in any given cross-modal trial).
- $\mathrm{CRE}_{\mathrm{MAX}}^{(-)}$ is easy to compute and does not require any specific assumption about the distribution of the spikes.
- It is in straightforward analogy to a measure well-established in the domain of reaction times ("race model inequality").
- It is sensitive to the variability of the data set.

Intermediate summary

- We suggest to replace the traditional multisensory index $\mathrm{CRE}_{\mathrm{MAX}}$ by the new one, $\mathrm{CRE}_{\mathrm{MAX}}^{(-)}$.
- The new one has a theoretical foundation: it measures the degree by which a neuron's observed multisensory response surpasses the level obtainable by optimally combining the unisensory responses (assuming that the neuron simply reacts to the more salient modality in any given cross-modal trial).
- $\mathrm{CRE}_{\mathrm{MAX}}^{(-)}$ is easy to compute and does not require any specific assumption about the distribution of the spikes.
- It is in straightforward analogy to a measure well-established in the domain of reaction times ("race model inequality").
- It is sensitive to the variability of the data set.

Intermediate summary

- We suggest to replace the traditional multisensory index $\mathrm{CRE}_{\mathrm{MAX}}$ by the new one, $\mathrm{CRE}_{\mathrm{MAX}}^{(-)}$.
- The new one has a theoretical foundation: it measures the degree by which a neuron's observed multisensory response surpasses the level obtainable by optimally combining the unisensory responses (assuming that the neuron simply reacts to the more salient modality in any given cross-modal trial).
- $\mathrm{CRE}_{\mathrm{MAX}}^{(-)}$ is easy to compute and does not require any specific assumption about the distribution of the spikes.
- It is in straightforward analogy to a measure well-established in the domain of reaction times ("race model inequality").
- It is sensitive to the variability of the data set.

Intermediate summary

- We suggest to replace the traditional multisensory index $\mathrm{CRE}_{\mathrm{MAX}}$ by the new one, $\mathrm{CRE}_{\mathrm{MAX}}^{(-)}$.
- The new one has a theoretical foundation: it measures the degree by which a neuron's observed multisensory response surpasses the level obtainable by optimally combining the unisensory responses (assuming that the neuron simply reacts to the more salient modality in any given cross-modal trial).
- $\mathrm{CRE}_{\mathrm{MAX}}^{(-)}$ is easy to compute and does not require any specific assumption about the distribution of the spikes.
- It is in straightforward analogy to a measure well-established in the domain of reaction times ("race model inequality").
- It is sensitive to the variability of the data set.

Copulas for neural and behavioral parallel systems

(1) Multisensory integration: reaction times

(2) Multisensory integration: spike numbers (impulse numbers)

(3) Response inhibition: stop signal paradigm

In (1) and (2), we present a new measure of multisensory integration that unifies behavioral and neural data.

In (3), we modify a classic model of inhibitory behavior in a way to solve a paradox between behavioral and neural data.

Copulas for neural and behavioral parallel systems

(1) Multisensory integration: reaction times

(2) Multisensory integration: spike numbers (impulse numbers)

(3) Response inhibition: stop signal paradigm

In (1) and (2), we present a new measure of multisensory integration that unifies behavioral and neural data.

In (3), we modify a classic model of inhibitory behavior in a way to solve a paradox between behavioral and neural data.

Copulas for neural and behavioral parallel systems

(1) Multisensory integration: reaction times

(2) Multisensory integration: spike numbers (impulse numbers)

(3) Response inhibition: stop signal paradigm

In (1) and (2), we present a new measure of multisensory integration that unifies behavioral and neural data.

In (3), we modify a classic model of inhibitory behavior in a way to solve a paradox between behavioral and neural data.

Copulas for neural and behavioral parallel systems

(1) Multisensory integration: reaction times
(2) Multisensory integration: spike numbers (impulse numbers)
(3) Response inhibition: stop signal paradigm

In (1) and (2), we present a new measure of multisensory integration that unifies behavioral and neural data.

In (3), we modify a classic model of inhibitory behavior in a way to solve a paradox between behavioral and neural data.

Copulas for neural and behavioral parallel systems

(1) Multisensory integration: reaction times

(2) Multisensory integration: spike numbers (impulse numbers)

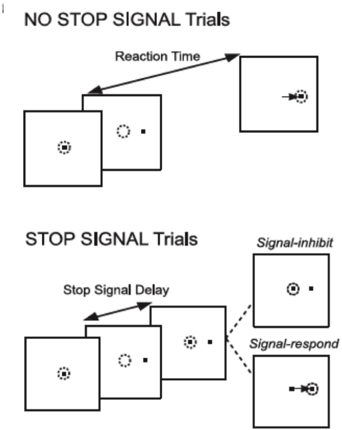
(3) Response inhibition: stop signal paradigm

In (1) and (2), we present a new measure of multisensory integration that unifies behavioral and neural data.

In (3), we modify a classic model of inhibitory behavior in a way to solve a paradox between behavioral and neural data.

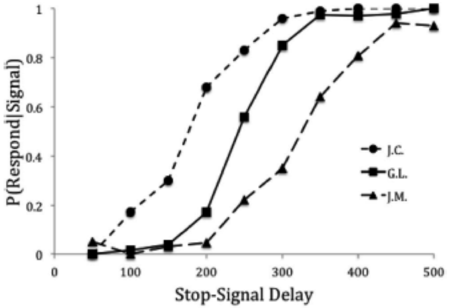
Response inhibition: stop signal paradigm

Stop signal paradigm



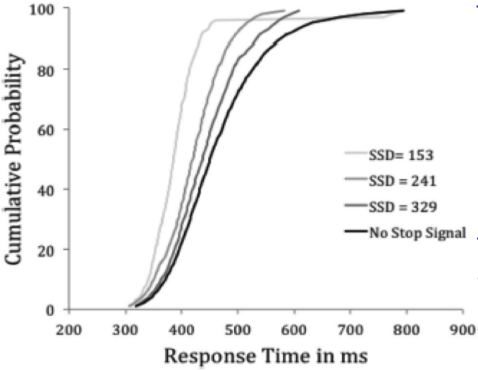
- Subjects are instructed to make a response as quickly as possible to a go signal (no-stop-signal trial)
- On a minority of trials, a stop signal is presented and subjects have to inhibit the previously planned response (stop-signal trial)

Stop signal paradigm: inhibition functions



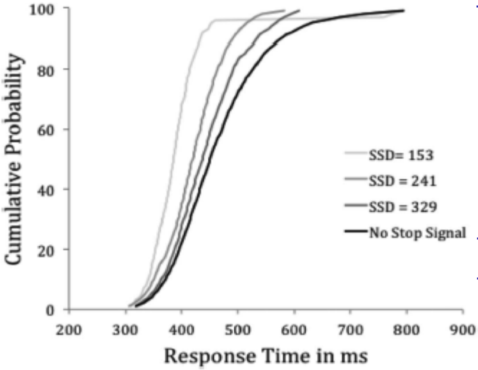
- Inhibition functions of three subjects (Logan & Cowan, 1984)
- The inhibition function is determined by stop-signal delay, but it also depends strongly on RT in the go task; the probability of responding given a stop signal is lower the longer the go RT

RT distributions with and without stop signal



- Observed response time distributions for no-stop -signal trials and signal-respond trials with stop-signal delays (SSDs) of 153, 241, and 329 ms (from Logan et al 2014):
- $\mathrm{RT}_{\text{signal-respond}} < \mathrm{RT}_{\text{go}}$
- faster for shorter stop-signal delays than for longer ones.

RT distributions with and without stop signal



- Observed response time distributions for no-stop -signal trials and signal-respond trials with stop-signal delays (SSDs) of 153, 241, and 329 ms (from Logan et al 2014):
- $\mathrm{RT}_{\text{signal-respond}} < \mathrm{RT}_{\text{go}}$
- faster for shorter stop-signal delays than for longer ones.

The general race model

The general race model (1)

- Distinguish two different experimental conditions termed context GO, where only a go signal is presented, and context STOP, where a stop signal is presented in addition.
- In STOP, let T_{go} and T_{stop} denote the random processing time for the go and the stop signal, respectively, with (unobservable !) bivariate distribution function

$$H(s,t) = \Pr(T_{go} \leq s, T_{stop} \leq t),$$

for all $s, t \geq 0$.
- The marginal distributions of $H(s,t)$ are denoted as

$$F_{go}(s) = \Pr(T_{go} \leq s, T_{stop} < \infty)$$
$$F_{stop}(t) = \Pr(T_{go} < \infty, T_{stop} \leq t).$$

The general race model (1)

- Distinguish two different experimental conditions termed context GO, where only a go signal is presented, and context STOP, where a stop signal is presented in addition.
- In STOP, let T_{go} and T_{stop} denote the random processing time for the go and the stop signal, respectively, with (unobservable !) bivariate distribution function

$$H(s,t) = \Pr(T_{go} \leq s, T_{stop} \leq t),$$

for all $s, t \geq 0$.

- The marginal distributions of $H(s,t)$ are denoted as

$$F_{go}(s) = \Pr(T_{go} \leq s, T_{stop} < \infty)$$
$$F_{stop}(t) = \Pr(T_{go} < \infty, T_{stop} \leq t).$$

The general race model (1)

- Distinguish two different experimental conditions termed context GO, where only a go signal is presented, and context STOP, where a stop signal is presented in addition.
- In STOP, let T_{go} and T_{stop} denote the random processing time for the go and the stop signal, respectively, with (unobservable !) bivariate distribution function

$$H(s,t) = \Pr(T_{go} \leq s, T_{stop} \leq t),$$

  for all $s, t \geq 0$.
- The marginal distributions of $H(s,t)$ are denoted as

$$F_{go}(s) = \Pr(T_{go} \leq s, T_{stop} < \infty)$$
$$F_{stop}(t) = \Pr(T_{go} < \infty, T_{stop} \leq t).$$

The general race model (2)

NOTE: The distribution of T_{go} could be different in context GO and in context STOP. However, the general race model rules this out by adding the important

Context invariance assumption *The distribution of go signal processing time T_{go} is the same in context GO and context STOP.*

Race assumption Probability of a response despite stop signal at delay t_d:

$$p_r(t_d) = \Pr(T_{go} < T_{stop} + t_d) \tag{1}$$

Assume $H(s, t) = \Pr(T_{go} \leq s, T_{stop} \leq t)$ is absolutely continuous, so that density functions for the marginals exist, denoted as $f_{go}(s)$ and $f_{stop}(t)$.

The general race model (2)

NOTE: The distribution of T_{go} could be different in context GO and in context STOP. However, the general race model rules this out by adding the important

Context invariance assumption *The distribution of go signal processing time T_{go} is the same in context GO and context STOP.*

Race assumption Probability of a response despite stop signal at delay t_d:

$$p_r(t_d) = \Pr(T_{go} < T_{stop} + t_d) \tag{1}$$

Assume $H(s,t) = \Pr(T_{go} \le s, T_{stop} \le t)$ is absolutely continuous, so that density functions for the marginals exist, denoted as $f_{go}(s)$ and $f_{stop}(t)$.

The general race model (2)

NOTE: The distribution of T_{go} could be different in context GO and in context STOP. However, the general race model rules this out by adding the important

Context invariance assumption *The distribution of go signal processing time T_{go} is the same in context GO and context STOP.*

Race assumption Probability of a response despite stop signal at delay t_d:

$$p_r(t_d) = \Pr(T_{go} < T_{stop} + t_d) \tag{1}$$

Assume $H(s,t) = \Pr(T_{go} \leq s, T_{stop} \leq t)$ is absolutely continuous, so that density functions for the marginals exist, denoted as $f_{go}(s)$ and $f_{stop}(t)$.

The general race model (3)

The RT distribution of responses given a stop signal at delay t_d (signal-response distribution) is

$$F_{sr}(t \mid t_d) = \Pr(T_{go} \leq t \mid T_{go} < T_{stop} + t_d)$$

Goal: Estimate stop-signal processing distribution $F_{stop}(t)$

The independent race model

The **independent** race model (1)

Logan & Cown (1984) suggested the *independent race model* assuming stochastic independence between T_{go} and T_{stop}:

Stochastic independence: for all s, t

$$H(s,t) = \Pr(T_{go} \leq s) \Pr(T_{stop} \leq t) = F_{go}(s)\, F_{stop}(t)$$

Then

$$\begin{aligned} p_r(t_d) &= \Pr(T_{go} < T_{stop} + t_d) \\ &= \int_0^\infty f_{go}(t)\,[1 - F_{stop}(t - t_d)]\,\mathrm{d}t. \end{aligned} \tag{2}$$

The **independent** race model (1)

Logan & Cown (1984) suggested the *independent race model* assuming stochastic independence between T_{go} and T_{stop}:

Stochastic independence: for all s, t

$$H(s,t) = \Pr(T_{go} \le s)\Pr(T_{stop} \le t) = F_{go}(s)\, F_{stop}(t)$$

Then

$$\begin{aligned} p_r(t_d) &= \Pr(T_{go} < T_{stop} + t_d) \\ &= \int_0^\infty f_{go}(t)\,[1 - F_{stop}(t - t_d)]\,\mathrm{d}t. \end{aligned} \tag{2}$$

The **independent** race model (2)

Density of the signal-response time distribution $F_{sr}(t|t_d)$, for $t > t_d$

$$f_{sr}(t \mid t_d) = f_{go}(t)\,[1 - F_{stop}(t - t_d)]/p_r(t_d).$$

Solving for $F_{stop}(t - t_d)$,

$$F_{stop}(t - t_d) = 1 - \frac{f_{sr}(t \mid t_d)p_r(t_d)}{f_{go}(t)}, \qquad (3)$$

known as the "Colonius" method (Colonius 1990). Unfortunately obtaining reliable estimates for the stop signal distribution using Equation (3) requires unrealistically large numbers of observations in practice (Band et al. 2003; Matzke et al. 2013).

The **independent** race model (2)

Density of the signal-response time distribution $F_{sr}(t|t_d)$, for $t > t_d$

$$f_{sr}(t \mid t_d) = f_{go}(t)\,[1 - F_{stop}(t - t_d)]/p_r(t_d).$$

Solving for $F_{stop}(t - t_d)$,

$$F_{stop}(t - t_d) = 1 - \frac{f_{sr}(t \mid t_d)p_r(t_d)}{f_{go}(t)}, \tag{3}$$

known as the "Colonius" method (Colonius 1990). Unfortunately obtaining reliable estimates for the stop signal distribution using Equation (3) requires unrealistically large numbers of observations in practice (Band et al. 2003; Matzke et al. 2013).

The **independent** race model (3): Integration method

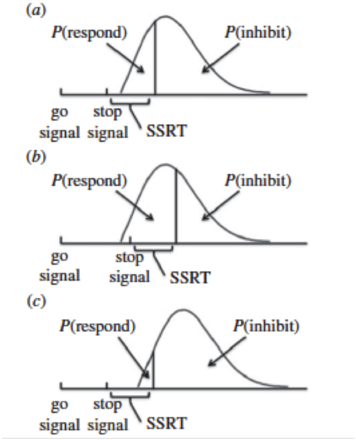


Figure
from Schall et al. 2017

Assume constant stop signal processing time: $T_{stop} = \mathrm{SSRT}$

$$p_r(t_d) = \int_0^{\mathrm{SSRT}+t_d} f_{go}(t)\mathrm{d}t$$

- Because estimates of both $p_r(t_d)$ and $f_{go}(t)$ are available, this allows estimation of stop signal processing mean SSRT.

The mean method

$$
\begin{aligned}
p_r(t_d) &= \Pr(T_{go} < T_{stop} + t_d) \\
&= \Pr(T_{go} - T_{stop} < t_d) \\
&= \Pr(T_{go} - T_{stop} < t_d) \equiv \Pr(T_d < t_d) \\
\Rightarrow E[T_d] &= E[T_{go}] - E[T_{stop}]
\end{aligned}
$$

Independent race model:

$$\mathrm{Var}[T_d] = \mathrm{Var}[T_{go}] + \mathrm{Var}[T_{stop}]$$

$$\Rightarrow \mathrm{Var}[T_{stop}] = \mathrm{Var}[T_{go}] - \mathrm{Var}[T_d]$$

The mean method

$$p_r(t_d) = \Pr(T_{go} < T_{stop} + t_d)$$
$$= \Pr(T_{go} - T_{stop} < t_d)$$
$$= \Pr(T_{go} - T_{stop} < t_d) \equiv \Pr(T_d < t_d)$$
$$\Rightarrow E[T_d] = E[T_{go}] - E[T_{stop}]$$

Independent race model:

$$\mathrm{Var}[T_d] = \mathrm{Var}[T_{go}] + \mathrm{Var}[T_{stop}]$$

$$\Rightarrow \mathrm{Var}[T_{stop}] = \mathrm{Var}[T_{go}] - \mathrm{Var}[T_d]$$

The mean method

$$
\begin{aligned}
p_r(t_d) &= \Pr(T_{go} < T_{stop} + t_d) \\
&= \Pr(T_{go} - T_{stop} < t_d) \\
&= \Pr(T_{go} - T_{stop} < t_d) \equiv \Pr(T_d < t_d)
\end{aligned}
$$

$$\Rightarrow E[T_d] = E[T_{go}] - E[T_{stop}]$$

Independent race model:

$$\mathrm{Var}[T_d] = \mathrm{Var}[T_{go}] + \mathrm{Var}[T_{stop}]$$

$$\Rightarrow \mathrm{Var}[T_{stop}] = \mathrm{Var}[T_{go}] - \mathrm{Var}[T_d]$$

The mean method

$$
\begin{aligned}
p_r(t_d) &= \Pr(T_{go} < T_{stop} + t_d) \\
&= \Pr(T_{go} - T_{stop} < t_d) \\
&= \Pr(T_{go} - T_{stop} < t_d) \equiv \Pr(T_d < t_d) \\
\Rightarrow \mathrm{E}[T_d] &= \mathrm{E}[T_{go}] - \mathrm{E}[T_{stop}]
\end{aligned}
$$

Independent race model:

$$\mathrm{Var}[T_d] = \mathrm{Var}[T_{go}] + \mathrm{Var}[T_{stop}]$$

$$\Rightarrow \mathrm{Var}[T_{stop}] = \mathrm{Var}[T_{go}] - \mathrm{Var}[T_d]$$

The mean method

$$
\begin{aligned}
p_r(t_d) &= \Pr(T_{go} < T_{stop} + t_d) \\
&= \Pr(T_{go} - T_{stop} < t_d) \\
&= \Pr(T_{go} - T_{stop} < t_d) \equiv \Pr(T_d < t_d) \\
\Rightarrow E[T_d] &= E[T_{go}] - E[T_{stop}]
\end{aligned}
$$

Independent race model:

$$\mathrm{Var}[T_d] = \mathrm{Var}[T_{go}] + \mathrm{Var}[T_{stop}]$$

$$\Rightarrow \mathrm{Var}[T_{stop}] = \mathrm{Var}[T_{go}] - \mathrm{Var}[T_d]$$

The mean method

$$
\begin{aligned}
p_r(t_d) &= \Pr(T_{go} < T_{stop} + t_d) \\
&= \Pr(T_{go} - T_{stop} < t_d) \\
&= \Pr(T_{go} - T_{stop} < t_d) \equiv \Pr(T_d < t_d) \\
\Rightarrow \mathrm{E}[T_d] &= \mathrm{E}[T_{go}] - \mathrm{E}[T_{stop}]
\end{aligned}
$$

Independent race model:

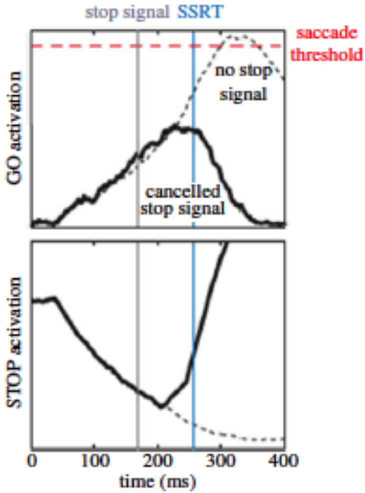
$$\mathrm{Var}[T_d] = \mathrm{Var}[T_{go}] + \mathrm{Var}[T_{stop}]$$

$$\Rightarrow \mathrm{Var}[T_{stop}] = \mathrm{Var}[T_{go}] - \mathrm{Var}[T_d]$$

The paradox

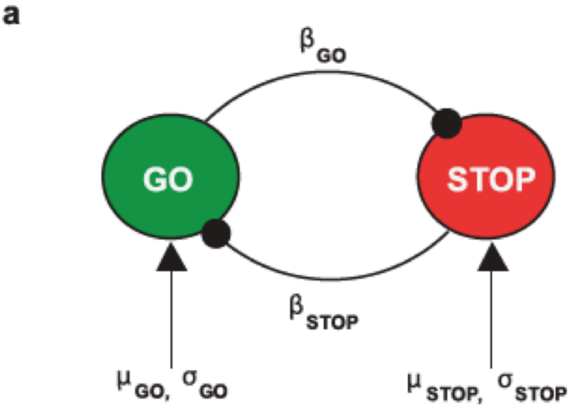
The paradox

Studying saccade countermanding in monkeys, Hanes and colleagues (Hanes & Schall 1995, Hanes et al. 1998) recorded from frontal and supplem. eye fields. They found neurons involved in gaze-shifting and gaze-holding that modulate on stop-signal trials, just before SSRT when the monkey stopped successfully. (Figure from Schall & Logan 2017)



The paradox

- The paradox: How can interacting circuits of mutually inhibitory gaze-holding and gaze-shifting neurons instantiate STOP and GO processes with independent finishing times?
- Proposed solution: interactive race model (Boucher et al. 2007):



The paradox

- The paradox: How can interacting circuits of mutually inhibitory gaze-holding and gaze-shifting neurons instantiate STOP and GO processes with independent finishing times?
- Proposed solution: interactive race model (Boucher et al. 2007), blocked-input model (Logan et al. 2015)
- stochastic differential equations; no longer non-parametric.

The race model with perfect negative dependence

The Fréchet-Hoeffding bounds

- For any bivariate distribution function

  $$H(s,t) = \Pr(T_{go} \leq s, T_{stop} \leq t)$$

  the following inequality holds:

  $$H^{-}(s,t) \leq H(s,t)$$

  with $H^{-}(s,t) = \max\{F_{go}(s) + F_{stop}(t) - 1, 0\}$

- $H^{-}(s,t)$ is a distribution function. Specifically, it correspond to perfect negative dependence between T_{go} and T_{stop}.

Perfect negative dependence

What does it mean?

$$H^{-}(s,t) = \max\{F_{go}(s) + F_{stop}(t) - 1, 0\}. \qquad (4)$$

for all s, t $(s, t \geq 0)$.

The marginal distributions of $H^{-}(s,t)$ are the same as before, that is, $F_{go}(s)$ and $F_{stop}(t)$.

Note that this perfect negative stochastic dependence (PND) model is parameter-free just like the IND race model, that is, we do not assume any specific parametric distribution.

Perfect negative dependence

What does it mean?

$$H^{-}(s,t) = \max\{F_{go}(s) + F_{stop}(t) - 1, 0\}. \tag{4}$$

for all s, t $(s, t \geq 0)$.

The marginal distributions of $H^{-}(s,t)$ are the same as before, that is, $F_{go}(s)$ and $F_{stop}(t)$.

Note that this perfect negative stochastic dependence (PND) model is parameter-free just like the IND race model, that is, we do not assume any specific parametric distribution.

Perfect negative dependence

What does it mean?

$$H^{-}(s,t) = \max\{F_{go}(s) + F_{stop}(t) - 1, 0\}. \qquad (4)$$

for all s, t ($s, t \geq 0$).

The marginal distributions of $H^{-}(s,t)$ are the same as before, that is, $F_{go}(s)$ and $F_{stop}(t)$.

Note that this perfect negative stochastic dependence (PND) model is parameter-free just like the IND race model, that is, we do not assume any specific parametric distribution.

Perfect negative dependence: the key property

$$F_{stop}(T_{stop}) = 1 - F_{go}(T_{go}) \tag{5}$$

holds "almost surely", that is, with probability 1.

- Thus, for any F_{go} percentile we immediately obtain the corresponding F_{stop} percentile as complementary probability and vice versa, which expresses perfect negative dependence between T_{go} and T_{stop}.
- The relation in Equation [5] is also interpretable as "T_{stop} *is (almost surely) a decreasing function of* T_{go}".
- It constitutes the most direct implementation of the notion of "mutual inhibition" observed in neural data: any increase of inhibitory activity (speed-up of T_{stop}) elicits a corresponding decrease in "go" activity (slow-down of T_{go}) and vice versa.

Perfect negative dependence: the key property

$$F_{stop}(T_{stop}) = 1 - F_{go}(T_{go}) \quad (5)$$

holds "almost surely", that is, with probability 1.

- Thus, for any F_{go} percentile we immediately obtain the corresponding F_{stop} percentile as complementary probability and vice versa, which expresses perfect negative dependence between T_{go} and T_{stop}.
- The relation in Equation [5] is also interpretable as "T_{stop} *is (almost surely) a decreasing function of* T_{go}".
- It constitutes the most direct implementation of the notion of "mutual inhibition" observed in neural data: any increase of inhibitory activity (speed-up of T_{stop}) elicits a corresponding decrease in "go" activity (slow-down of T_{go}) and vice versa.

Perfect negative dependence: the key property

$$F_{stop}(T_{stop}) = 1 - F_{go}(T_{go}) \quad (5)$$

holds "almost surely", that is, with probability 1.

- Thus, for any F_{go} percentile we immediately obtain the corresponding F_{stop} percentile as complementary probability and vice versa, which expresses perfect negative dependence between T_{go} and T_{stop}.
- The relation in Equation [5] is also interpretable as "*T_{stop} is (almost surely) a decreasing function of T_{go}*".
- It constitutes the most direct implementation of the notion of "mutual inhibition" observed in neural data: any increase of inhibitory activity (speed-up of T_{stop}) elicits a corresponding decrease in "go" activity (slow-down of T_{go}) and vice versa.

Predictions from perfect negative dependence

Do we have to throw away all measures obtained using the independent model, like estimates of mean $T_{stop} \equiv$ SSRT?

No ! Because the (marginal) distribution of T_{stop} are the same under independence and perfect negative dependence. Thus

$$\mathrm{E}[T_{stop} \,|\, \mathrm{IND}] = \mathrm{E}[T_{stop} \,|\, \mathrm{PND}]$$

- But for the variance,

$$\mathrm{Var}[T_{stop}] = \mathrm{Var}[T_d] - \mathrm{Var}[T_{go}] + 2\,\mathrm{Cov}[T_{go}, T_{stop}]. \qquad (6)$$

Predictions from perfect negative dependence

Do we have to throw away all measures obtained using the independent model, like estimates of mean $T_{stop} \equiv \text{SSRT}$?

No ! Because the (marginal) distribution of T_{stop} are the same under independence and perfect negative dependence. Thus

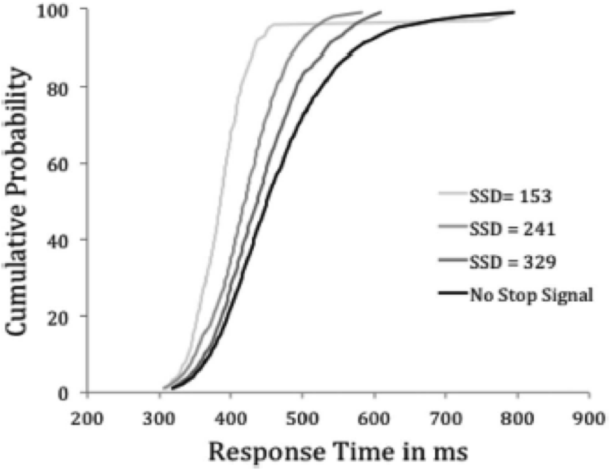
$$\mathsf{E}[T_{stop} \,|\, \mathrm{IND}] = \mathsf{E}[T_{stop} \,|\, \mathrm{PND}]$$

- But for the variance,

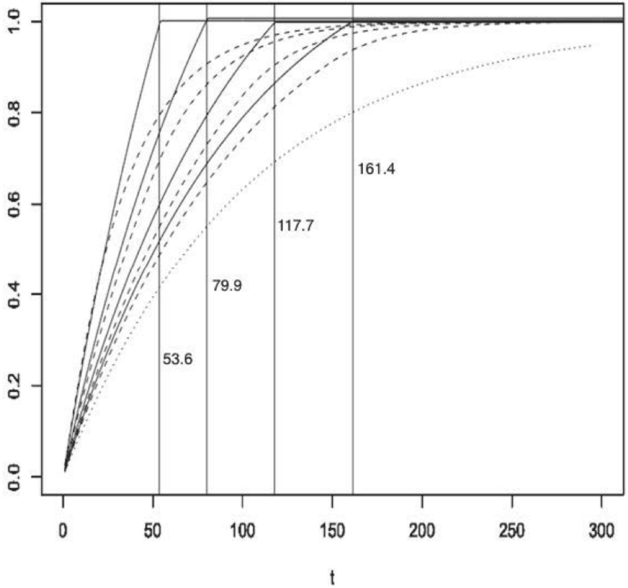
$$\mathsf{Var}[T_{stop}] = \mathsf{Var}[T_d] - \mathsf{Var}[T_{go}] + 2\,\mathsf{Cov}[T_{go}, T_{stop}]. \qquad (6)$$

Can we test for PND ?

"Fan effect":



Can we test for PND ?



dashed = IND
line = PND
* T_{go}, T_{stop}: exponential distribution
*simulation: **copBasic** package in **R**

Is there a paradox?

Is there a paradox?

- We need to distinguish different levels of description, behavioral vs. neural. The race model (whether independent or dependent) does not describe the neural processes underlying stopping behavior.
- "Linking propositions" are theories about how the specific, observable aspects of the neuroscientific data should be related to specific, but often latent, aspects of the formal models.
- However: "Linking propositions" are affected by the behavioral model **and can go astray**: "*In short, the interaction of the STOP with the GO unit must be late and potent – late to preserve the independence of the GO and STOP processes through SSRT and potent because it must be late.*" (Schall et al. 2017)
- Therefore, it DOES matter what kind of dependency is assumed in the race model.

Is there a paradox?

- We need to distinguish different levels of description, behavioral vs. neural. The race model (whether independent or dependent) does not describe the neural processes underlying stopping behavior.
- "Linking propositions" are theories about how the specific, observable aspects of the neuroscientific data should be related to specific, but often latent, aspects of the formal models.
- However: "Linking propositions" are affected by the behavioral model **and can go astray**: "*In short, the interaction of the STOP with the GO unit must be late and potent – late to preserve the independence of the GO and STOP processes through SSRT and potent because it must be late.*" (Schall et al. 2017)
- Therefore, it DOES matter what kind of dependency is assumed in the race model.

Is there a paradox?

- We need to distinguish different levels of description, behavioral vs. neural. The race model (whether independent or dependent) does not describe the neural processes underlying stopping behavior.
- "Linking propositions" are theories about how the specific, observable aspects of the neuroscientific data should be related to specific, but often latent, aspects of the formal models.
- However: "Linking propositions" are affected by the behavioral model **and can go astray**: "*In short, the interaction of the STOP with the GO unit must be late and potent – late to preserve the independence of the GO and STOP processes through SSRT and potent because it must be late.*" (Schall et al. 2017)
- Therefore, it DOES matter what kind of dependency is assumed in the race model.

Is there a paradox?

- We need to distinguish different levels of description, behavioral vs. neural. The race model (whether independent or dependent) does not describe the neural processes underlying stopping behavior.
- "Linking propositions" are theories about how the specific, observable aspects of the neuroscientific data should be related to specific, but often latent, aspects of the formal models.
- However: "Linking propositions" are affected by the behavioral model **and can go astray**: "*In short, the interaction of the STOP with the GO unit must be late and potent – late to preserve the independence of the GO and STOP processes through SSRT and potent because it must be late.*" (Schall et al. 2017)
- Therefore, it DOES matter what kind of dependency is assumed in the race model.

Summary and Conclusions

- Behavioral and neural models have distinct levels of description based on different types of data, and either level can lead to insights into the cognitive process.
- The goal of "closing the gap" may be approached by "mutual inspiration".
- Linking hypotheses are an important part of closing the gap, but their usefulness depends on how good the models are (in both areas).

Summary and Conclusions

- Behavioral and neural models have distinct levels of description based on different types of data, and either level can lead to insights into the cognitive process.
- The goal of “closing the gap” may be approached by “mutual inspiration”.
- Linking hypotheses are an important part of closing the gap, but their usefulness depends on how good the models are (in both areas).

Summary and Conclusions

- Behavioral and neural models have distinct levels of description based on different types of data, and either level can lead to insights into the cognitive process.
- The goal of "closing the gap" may be approached by "mutual inspiration".
- Linking hypotheses are an important part of closing the gap, but their usefulness depends on how good the models are (in both areas).

Summary and Conclusions

- Behavioral and neural models have distinct levels of description based on different types of data, and either level can lead to insights into the cognitive process.
- The goal of "closing the gap" may be approached by "mutual inspiration".
- Linking hypotheses are an important part of closing the gap, but their usefulness depends on how good the models are (in both areas).

Acknowledgment

- Supported by DFG (German Science Foundation) SFB/TRR-31 (Project B4, HC), DFG Cluster of Excellence EXC 1077/1 Hearing4all (HC) and DFG Grant DI 506/12-1 (AD)

Acknowledgment

- *Collaborator*: Prof. Adele Diederich, Jacobs University Bremen
- *Student assistants*: Lisa Wergen, Sarah Blum, Felix Wolff
- *Funding*: DFG (German Science Foundation)

Bibliography

- Colonius H, Diederich A (2018). Paradox resolved: stop signal race model with negative dependence. *Psychological Review*, 10.1037/rev0000127.
- Colonius H (2017) Selected concepts from probability. Chpt. 1 in WH Batchelder etal, *New Handbook of Mathematical Psychology*, Cambridge U. Press.
- Colonius H, Diederich A (2017) Measuring multisensory integration: from reaction times to spike counts. *Scientific Reports*, 7:3023. DOI:10.1038/s41598-017-03219-5. rdcu.be/tjo1.
- Colonius H (2016) An invitation to coupling and copulas: with applications to multisensory modeling. *Journal of Mathematical Psychology*, 74, 2–10.
- Colonius H, Diederich A (2006). Race model inequality: Interpreting a geometric measure of the amount of violation. *Psychological Review*, 113 (1), 148–54.
- Colonius, H. (1990). A note on the stop-signal paradigm, or how to observe the unobservable. *Psychological Review*, 97 (2), 309–312.