Paradox resolved: stop signal race model with negative dependence

Hans Colonius

Department of Psychology and Cluster of Excellence “Hearing4all”,

Carl von Ossietzky Universität, Oldenburg, Germany

and

Department of Psychological Sciences, Purdue University

Adele Diederich

Focus Health: Life Sciences & Chemistry, Jacobs University Bremen

and

Department of Psychological Sciences, Purdue University

Correspondence: Hans Colonius, Department of Psychology, Carl von Ossietzky Universität, 26111 Oldenburg, Germany, Phone: +49 157 30200389, E-Mail: hans.colonius@uni-oldenburg.de

Abstract

The ability to inhibit our responses voluntarily is an important case of cognitive control. The *stop-signal paradigm* is a popular tool to study response inhibition. Participants perform a response time task (*go task*) and, occasionally, the go stimulus is followed by a stop signal after a variable delay, indicating subjects to withhold their response (*stop task*). The main interest of modeling is in estimating the unobservable stop-signal processing time, that is, the covert latency of the stopping process as a characterization of the response inhibition mechanism. In the *independent race model* the stop-signal task is represented as a race between stochastically independent go and stop processes. Without making any specific distributional assumptions about the processing times, the model allows to estimate the mean time to cancel a response. Neurophysiological studies on countermanding saccadic eye movements, however, have shown that the neural correlates of go and stop processes consist of networks of mutually *interacting* gaze-shifting and gaze-holding neurons. This poses a major challenge in formulating linking propositions between the behavioral and neural findings. Here we propose a *dependent race model* that postulates perfect negative stochastic dependence between go and stop activations. The model is consistent with the concept of interacting processes while retaining the simplicity and elegance of the distribution-free independent race model. For mean data, the dependent model's predictions remain identical to those of the independent model. The resolution of this apparent paradox advances the understanding of mechanisms of response inhibition and paves the way for modeling more complex situations.

Key words: response inhibition; stop signal paradigm; independent race model; perfect negative dependence

Paradox resolved: stop signal race model with negative dependence

A recurrent theme in cognitive modeling is the difficulty to uniquely identify the processes underlying the generation of response times and probabilities in behavioral paradigms like simple yes-no tasks or same-different judgments of stimulus pairs. A prime example is the serial vs. parallel processing issue (Algom, Eidels, Hawkins, Jefferson, & Townsend, 2015; Townsend, 1990; Townsend & Wenger, 2004) showing that even rigorous mathematical analyses of the underlying assumptions may not always resolve the difficulty completely. Recently, efforts in model-based cognitive neuroscience to constrain behavioral models by findings obtained from neuroscientific methods have intensified, from spike train analyses to EEG and fMRI recordings (Forstmann & Wagenmakers, 2015). One area where important advances have emerged from this joint endeavor is the modeling of cognitive control and, in particular, response inhibition (Logan, 2017; Verbruggen & Logan, 2009). Response inhibition refers to the ability to suppress responses that are no longer required or have become inappropriate; it is important for survival, such as stopping yourself from crossing the street when a car comes around the corner without noticing you. Deficits of response inhibition have been linked to several disorders like attention-deficit/hyperactivity, obsessive-compulsive behavior and substance abuse. Probing response inhibition has been used widely to study executive control and flexibility in behavior; for reviews, see (Logan, 1994; Verbruggen & Logan, 2008, 2009; Matzke, Verbruggen, & Logan, in press) and a recent theme issue of *Philosophical Transactions of the Royal Society B* (2017) on ‘Movement suppression: brain mechanisms for stopping and stillness’ (Carpenter & Noorani, 2017).

In the laboratory, a very useful tool for the study of inhibition is the *stop-signal paradigm* where participants perform a response time task (*go task*), such as moving their gaze to the location of a pre-defined target. Occasionally, the go stimulus is followed by a stop signal after a variable time delay, indicating subjects to withhold the response (*stop task*). Performance in the stop-signal paradigm has been modeled as a race between a “go process", triggered by the presentation of the go stimulus, and a "stop process" triggered by the presentation of the stop signal (Verbruggen & Logan, 2008, 2009). When the stop process finishes before the go process, the response is inhibited; otherwise, it is executed. The main interest of the modeler is in estimating the unobservable stop-signal reaction time (SSRT), that is, the latency of the stopping process as a characterization of the response inhibition mechanism. In the *independent race model* (IND model, for short) the stop-signal task is represented as a race between *stochastically independent* go and stop processes (Logan & Cowan, 1984). Under certain simplifying assumptions, mean SSRT can be estimated efficiently without making any specific assumptions about the distribution of the processing times (Verbruggen & Logan, 2009). In several hundreds of studies, the IND model has been applied in virtually every stop-signal experiment providing important measures of cognitive control like SSRT. Although the model makes no commitment to the underlying computational or neural processes that generate the go and stop processing times, the IND model is considered as defining constraints that any model of response inhibition must follow (Schall, Palmeri, & Logan, 2017).

However, neurophysiological studies in the frontal eye fields (FEF) and superior colliculus (SC) of macaque monkeys performing a countermanding task with saccadic eye movements have shown that the neural correlates of go and stop processes produce eye movement behavior through a network of *interacting* gaze-shifting and gaze-holding neurons (Hanes & Schall, 1995; Hanes, Patterson, & Schall, 1998; Paré & Hanes, 2003; Brown, Hanes, Schall, & Stuphorn, 2008). This discrepancy between the behavioral and neural data is widely perceived as a paradox (Boucher, Palmeri, Logan, & Schall, 2007; Schall & Godlove, 2012; Schall et al., 2017; Matzke et al., in press): how can interacting circuits of mutually inhibitory neurons instantiate stop and go processes with stochastically independent finishing times?

Here we propose a variant of the race model which, instead of independence, assumes *perfect negative stochastic dependence* between go and stop processes (PND model, for short). It resolves the apparent paradox and nonetheless retains the distribution-free property of the independent race model. Moreover, the PND model's predictions, considered at the level of mean SSRT, are shown to be necessarily identical to those of the independent model. Notably, we argue that it is very difficult to empirically distinguish between the two race model versions without introducing further, parametric assumptions. Thus, there is no reason to uphold the stochastic independence assumption of the race model whenever a distribution-free version is to be considered.

Neurally inspired modeling

Investigating the neural underpinnings of response inhibition in saccades, Hanes and Schall (Hanes & Schall, 1995) first showed that macaque monkey behavior in saccade countermanding corresponded in detail to that of human performance in manual stop-signal tasks. Then, recording from the frontal eye fields they isolated neurons involved in gaze-shifting and gaze-holding that represent a larger circuit of such neurons that extends from cortex through basal ganglia and superior colliculus to brainstem (Hanes et al., 1998). Importantly, that result was based on their postulate that for neurons to participate in controlling movement initiation two criteria must be met: first, neurons must discharge differently when movements are initiated or withheld; if neurons still discharge when movements are canceled, their activity was not affected by the stop process. Second, the differential modulation on canceled trials must occur before SSRT; otherwise, the neural modulation happens after the movement has already been canceled (see (Schall & Godlove, 2012), p. 1012). From these findings, Boucher and colleagues (Boucher et al., 2007) developed an *interactive race model* linking the interacting circuits of mutually inhibitory gaze-holding and gaze-shifting neurons with stochastic accumulation in the go and stop processes of the race model (see also, Hanes & Schall, 1995). A response is observed only if the go process reaches a certain threshold of activation. Stopping occurs if the stop process interferes with the go process by inhibiting activation in the go accumulator to prevent it from reaching the threshold. Alternative models assume that response inhibition results from blocking the input to the go unit ("blocked-input models") (Logan, Yamaguchi, Schall, & Palmeri, 2015) or postulate a spiking neural network of hundreds of units representing populations of movement neurons, fixation neurons and inhibitory interneurons, and a control unit that turns the fixation neurons on and off (Lo, Boucher, Paré, Schall, & Wang, 2009).

These neural models are computationally explicit: they are based on systems of stochastic differential equations; estimating parameters to best fit the behavioral data, they are able to predict the distribution of cancel times, i.e., the times at which neural activity modulates on trials on which subjects stop successfully, relative to SSRT (Schall et al., 2017). The models fit the behavioral data just as well as the independent race model. Note that this is just another instantiation of the paradox. An attempt to resolve the contradiction between stochastic independence for behavioral data and interdependence at the neural level has been to postulate that the stop process is independent of the go process for much of its duration, followed by a late and potent interaction between stopping and going that reverses the trajectory of go activation (Boucher et al., 2007; Schall et al., 2017). Strictly speaking, this would require introducing two serial processing stages in the IND model, independence followed by strong interaction. No formal modeling of this extension has been undertaken, however, to the best of our knowledge.

Background: General race model

Next, we present a formal framework for the general race model making no assumption at all about the dependency between stop and go processes. One should distinguish between two different experimental conditions termed context $\mathcal{GO}$ where only a go signal is presented, and context $\mathcal{STOP}$ where a stop signal is presented in addition. A race between processes triggered by the go and stop signal is represented by two random variables: T_{go} and T_{stop} (referred to as SSRT above) denote the random processing times for the go and stop signal, respectively, in context $\mathcal{STOP}$ with a bivariate distribution function denoted H,

$$H(s,t) = \mathrm{P}\,[T_{go} \leq s, T_{stop} \leq t], \tag{1}$$

defined for all real numbers s and t, with $s, t \geq 0$. The marginal distributions of $H(s,t)$ are

$$F_{go}(s) = \mathrm{P}\,[T_{go} \leq s, T_{stop} < \infty] \text{ and}$$
$$F_{stop}(t) = \mathrm{P}\,[T_{go} < \infty, T_{stop} \leq t].$$

In context $\mathcal{STOP}$ the go signal triggers a realization of random variable T_{go} and the stop signal triggers a realization of random variable T_{stop}. In context $\mathcal{GO}$, however, only processing of the go signal occurs. In general, the latter one may be different from the marginal distribution $F_{go}(s)$ in context $\mathcal{STOP}$. However, the *general race model* rules this out by adding the important assumption of "context invariance", also know as "context independence" (e.g., Matzke et al., in press):

Context invariance (CI). In context $\mathcal{GO}$, the distribution of go signal processing time is assumed to be

$$F_{go}(s) = \mathrm{P}\,[T_{go} \leq s, T_{stop} < \infty], \tag{2}$$

i.e., it is identical to the marginal distribution $F_{go}(s)$ in context $\mathcal{STOP}$.

From these assumptions, the probability of observing a response to the go signal, given a stop signal is presented with delay t_d [ms] ($t_d \geq 0$) after the go signal, is defined by

$$p_r(t_d) = \mathrm{P}\,[T_{go} < T_{stop} + t_d]. \tag{3}$$

According to the model, the probability of observing a response to the go signal no later than time t, given the stop signal was presented with delay t_d, is given by (conditional) distribution function

$$F_{sr}(t \,|\, t_d) = \mathrm{P}\,[T_{go} \leq t \,|\, T_{go} < T_{stop} + t_d], \tag{4}$$

known as *signal-response RT* distribution.

The main interest in race modeling is to obtain information about the distribution of the unobservable stop signal processing time T_{stop}, or about some of its parameters, given sample estimates of $F_{go}(t)$, $F_{sr}(t \mid t_d)$, and $p_r(t_d)$. As observed in (Logan & Cowan, 1984), letting stop signal delay t_d vary, the *inhibition function*[1] $p_r(t_d)$ can be formally considered as the distribution function of a random variable, $T_d \equiv T_{go} - T_{stop}$, say. Then,

$$\mathrm{E}[T_d] = \mathrm{E}[T_{go}] - \mathrm{E}[T_{stop}], \tag{5}$$

with E[ ] denoting expected value (mean) of a random variable. Solving for the estimate of $\mathrm{E}[T_{stop}]$ immediately yields an estimate of the mean of the unobservable distribution of T_{stop}. It is well known that the reliability of this estimation method, known as the *mean method*, depends on how precise the inhibition function $p_r(t_d)$ and the mean of T_{go} are estimated (Matzke et al., in press; Verbruggen & Logan, 2009). Obtaining estimates of higher moments, or the entire distribution of T_{stop}, requires further non-parametric assumptions about the bivariate distribution $H(s, t)$.

Independent vs. negatively dependent race model

The most common version of the race model, as introduced by Logan & Cowan (Logan & Cowan, 1984), postulates stochastic independence between T_{go} and T_{stop}:

Stochastic independence:

$$H(s,t) = \mathrm{P}\left[T_{go} \leq s\right] \times \mathrm{P}\left[T_{stop} \leq t\right] = F_{go}(s) \times F_{stop}(t),$$

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

In addition to estimating the mean via Equation [5], an estimate of the variance of stop signal processing time is obtainable in the IND model from

$$\mathrm{Var}[T_{stop}] = \mathrm{Var}[T_d] - \mathrm{Var}[T_{go}], \tag{6}$$

due to assuming stochastic independence. Finally, it can be shown that the unobservable stop signal distribution function $F_{stop}(t)$ is expressible as a function of the observables

[1] Strictly speaking, $p_r(t_d)$ is the probability of *no-inhibition*, but the terminology used here is common.

$p_r(t_d), f_{go}(t)$ and $F_{sr}(t \,|\, t_d)$ (see appendix). In order to resolve the paradox described above, a race model with negative dependency between go and stop signal processing times is proposed next. We define a bivariate distribution function for T_{go} and T_{stop} exhibiting *perfect negative dependence* (see, e.g., Nelsen, 2006; Joe, 2015) as follows:

Perfect negative stochastic dependence:

$$H^{-}(s,t) = \max\{F_{go}(s) + F_{stop}(t) - 1, 0\}. \tag{7}$$

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 $F_{stop}(t)$. It can be shown (see appendix) that Equation 7 implies that

$$F_{stop}(T_{stop}) = 1 - F_{go}(T_{go}) \tag{8}$$

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 [8] is also interpretable as “T_{stop} *is (almost surely) a decreasing function of* T_{go}”.

The PND model arguably 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. 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.

Predictions from the PND race model

We list some predictions from the PND race model (for further details, see appendix):

Probability of no-stopping. Under perfect negative dependence between T_{go} and T_{stop}, the probability of no-stopping $p_r(t_d)$ is an increasing function of t_d, as it should be.

Signal-response RT distribution. Signal-response RT distribution

$$F_{sr}(t \,|\, t_d) = \mathrm{P}\,[T_{go} \leq t \,|\, T_{go} < T_{stop} + t_d], \tag{4}$$

approaches the go signal distribution $F_{go}(t)$ with increasing stop signal delay t_d. Moreover, for varying values of t_d, $F_{sr}(t \mid t_d)$ exhibits the "fan effect" typically found in empirical data and also predicted by the IND race model (Logan & Cowan, 1984; Verbruggen & Logan, 2009): all signal-response RT distributions $F_{sr}(t \mid t_d)$ are strictly ordered by delay (t_d) from left to right, converging toward the right-most curve, the no-stop signal distribution $F_{go}(t)$ for $t_d \to \infty$.

[Figure 1 about here]

To illustrate, Figure 1 presents the go signal distribution and signal-response RT distributions with different delay values t_d for both the IND (dashed curves) and the PND race model (solid lines) assuming exponential distributions for T_{go} and T_{stop}. While the exponential distribution lacks empirical support, it was chosen here just to illustrate quantitative differences between the model versions.

Figure 1 suggests an important feature of the PND model: each signal-response distribution $F_{sr}(t \mid t_d)$ crosses 1 at a certain finite point t. It is shown below that this point is predictable from the observables $p_r(t_d)$ and $F_{go}(t)$. While this "crossing" property is not shared by the IND model (because the "coupling" of T_{go} and T_{stop} as expressed in Equation (8) is absent), the strength of this test will of course depend on the precision of estimates of $F_{sr}(t \mid t_d)$.

Estimating moments of the T_{stop} ***distribution.*** Given that the marginal distributions F_{go} and F_{stop} are the same under both models, any estimates for $\mathrm{E}[T_{stop}]$ based on the mean method (Eq. [5]) are necessarily identical for the IND and PND model.

However, for the variance we obtain

$$\mathrm{Var}[T_{stop}] = \mathrm{Var}[T_d] - \mathrm{Var}[T_{go}] + 2\,\mathrm{Cov}[T_{go}, T_{stop}]. \tag{9}$$

As the covariance in the above equation is unknown, an estimate for the stop signal variance is no longer available under the PND model. Nevertheless, given that $\mathrm{Cov}[T_{go}, T_{stop}]$ is the most extreme negative covariance under any bivariate distribution for

T_{go} and T_{stop} (Denuit & Dhaene, 2003), the PND model stop signal variance can never be larger than the one under independence.

Discussion

Testing IND vs. PND race models

Because stop signal processing times are not observable, empirical testing of non-parametric stop signal race models is severely limited in general. In particular, both the IND and PND version of the race model can only be tested in conjunction with context invariance (CI). Presuming CI is valid, the following predictions can be tested: (1) mean signal-response RT should be faster than mean go signal RT; (2) mean signal-response RT should increase with stop signal delay t_d; and (3) both these tests are implied by the "fan" structure of the distribution functions permitting an additional qualitative test of IND and PND race models. Because data not consistent with this "fan" structure would be evidence against both the IND and PND model, obviously, none of these tests would allow us to distinguish between IND and PND.

By construction, the PND model has the same (marginal) distribution for stop signal processing as the IND model; therefore, estimates for mean SSRT, i.e., the expected value of stop signal processing, will be identical for either model as long as estimates are based on the mean method. In a sense, this is good news because it implies that adopting perfect negative dependency between go and stop signal processing does not invalidate previous SSRT estimates of all empirical studies employing the mean method. Note that this holds as well for estimating SSRT via the *integration method*, another way to estimate SSRT. However, because the integration method presumes stop signal processing time to be constant, the distinction between IND and PND model becomes meaningless.

This leaves us with the shape of the family of signal-response distributions, that is, $F_{sr}(t \mid t_d)$ for varying values of t_d, as the only potential means of distinguishing between the two model versions. Whether or not this "crossing" test that -as outlined above- consists of checking the way in which the signal-response distributions approach their upper bound of 1, proves to be as visible in real data compared to the exponential toy example (Figure 1) will have to be seen in suitable data (with large enough samples) as it depends on the specific, but unobservable, stop signal distribution.

Conclusion

Several authors have stressed that the level of description provided by the race model is quite different from that of neural models (Boucher et al., 2007; Logan et al., 2015): the independent race model is mute about the possible underlying mechanisms of response inhibition, and its primary purpose is to provide a measure of stop signal processing time, which the model allows without having to assume a specific distribution or estimate parameters. Nevertheless, it has been claimed that the independent race model "captures the essence of computation and ...that it formulates the constraints that any model of response inhibition must follow" (cf. Logan et al., 2015, p.3). The race model with perfect negative dependence suggested here operates at the same level of generality and yields the same measure of (mean) stop signal processing time but, importantly, resolves the paradox the independent model is facing due to the neurophysiological findings. We conclude that the race model with perfect negative stochastic dependence is a natural way to unify the observation of interacting circuits of mutually inhibitory gaze-holding and gaze-shifting neurons with data on the behavioral level.

Beyond the simple stop signal task, several other forms of acts of control have been studied that pose an even greater challenge for efforts to identify the underlying cognitive processes, such as switching tasks or shifting of attention (Logan, 2017). One promising area is *selective stopping* where response inhibition is required for some stimuli (e.g. red light) but not others (green light) (e.g., Bissett & Logan, 2014). Representing cognitive processes for such more complex tasks may call for types of graded, instead of perfect, dependency. Appropriate race models can, in principle, be introduced by appealing to more general forms of stochastic dependency via copulas (Nelsen, 2006; Joe, 2015).

Appendix

The appendix presents some computational details for both the IND and PND model, in general and for the special case of exponential distributions (for producing Figure 1).

We assume that distribution functions F_{go} and F_{stop} possess densities, f_{go} and f_{stop}, and are increasing, where "increasing" always refers to "strictly increasing" here. Note that, although it is usually taken for granted in the race model that densities exist, strictly speaking, this is not required neither by the independent nor the dependent race model. [2]

Independent race model

Under stochastic independence of T_{go} and T_{stop},

$$\begin{aligned} p_r(t_d) &= \mathrm{P}\left[T_{go} < T_{stop} + t_d\right] \\ &= \int_0^\infty f_{go}(t)[1 - F_{stop}(t - t_d)]\,\mathrm{d}t. \end{aligned} \tag{10}$$

Moreover,

$$\mathrm{Var}[T_d] = \mathrm{Var}[T_{stop}] + \mathrm{Var}[T_{go}], \tag{11}$$

implying Equation [6].

Writing $f_{sr}(t \,|\, t_d)$ for the density function of signal-response time distribution $F_{sr}(t|t_d)$, it has been shown in (Colonius, 1990) that

$$f_{sr}(t \,|\, t_d) = f_{go}(t)\,[1 - F_{stop}(t - t_d)]/p_r(t_d). \tag{12}$$

From that, an explicit expression of the distribution of unobservable stop signal processing time T_{stop} is given by:

$$F_{stop}(t - t_d) = 1 - \frac{f_{sr}(t \,|\, t_d) p_r(t_d)}{f_{go}(t)}. \tag{13}$$

[2] Without this assumption, one can rewrite, e.g., Equation (10) for $p_r(t_d)$ using the *Lebesgue-Stieltjes* integral:

$$p_r(t_d) = \int_0^\infty [1 - F_{stop}(t - t_d)]\,\mathrm{d}F_{go}(t).$$

Unfortunately, as investigated in (Band, van der Molen, & Logan, 2003; Matzke, Dolan, Logan, Brown, & Wagenmakers, 2013), gaining reliable estimates for the stop signal distribution using Equation [13] requires unrealistically large numbers of observations.

Perfect negative dependent race model

Defining the bivariate distribution by

$$H^{-}(s,t) = \max\{F_{go}(s) + F_{stop}(t) - 1, 0\} \tag{7}$$

for all s, t $(s, t \geq 0)$, the marginal distributions of $H^{-}(s,t)$ are again $F_{go}(s)$ and $F_{stop}(t)$. A classic result from the *theory of copulas* (see appendix) asserts that

(i) H^{-} is the lower bound of all bivariate distributions H of (T_{go}, T_{stop}), i.e., $H^{-}(s,t) \leq H(s,t)$ for all (s,t);

(ii) $\mathrm{P}\left[F_{go}(T_{go}) + F_{stop}(T_{stop}) = 1\right] = 1$.

From (i) the covariance between the random variables can be shown to be the smallest possible across all bivariate distributions $H(s,t)$ (Denuit & Dhaene, 2003). Statement (ii) is equivalent to

$$F_{stop}(T_{stop}) = 1 - F_{go}(T_{go}). \tag{8}$$

holding almost surely.

Probability of stopping. Under perfect negative dependence between T_{go} and T_{stop},

$$\begin{aligned} p_r(t_d) &= \mathrm{P}\left[T_{go} - T_{stop} < t_d\right] \\ &= \mathrm{P}\left[T_{go} - F_{stop}^{-1}[1 - F_{go}(T_{go})] < t_d\right] \\ &= \mathrm{P}\left[g(T_{go}) < t_d\right] \\ &= \mathrm{P}\left[T_{go} < g^{-1}(t_d)\right], \end{aligned} \tag{14}$$

where function g, defined as $g(t) = t - F_{stop}^{-1}[1 - F_{go}(t)]$, is increasing and so its inverse g^{-1} is increasing as well. Thus, $p_r(t_d)$ is increasing in t_d.

Signal-response RT distribution. For stop signal delay t_d,

$$\begin{aligned} F_{sr}(t \mid t_d) &= \mathrm{P}\left[T_{go} \le t \mid T_{go} < T_{stop} + t_d\right] \\ &= \mathrm{P}\left[T_{go} \le t \cap T_{go} < T_{stop} + t_d\right]/p_r(t_d) \\ &= \mathrm{P}\left[T_{go} \le t \cap T_{go} < g^{-1}(t_d)\right]/\mathrm{P}\left[T_{go} < g^{-1}(t_d)\right] \\ &= F_{go}(\min\{t, g^{-1}(t_d)\})/F_{go}(g^{-1}(t_d)). \end{aligned}$$

To see that $F_{sr}(t \mid t_d)$ obeys the "fan effect" consider two different delays, t_d and t_d^*, say, with $t_d < t_d^*$; for a fixed value of t, we first assume $t < g^{-1}(t_d) < g^{-1}(t_d^*)$. Then

$$\begin{aligned} F_{sr}(t \mid t_d^*) &= F_{go}(t)/F_{go}(g^{-1}(t_d^*)) \\ &< F_{go}(t)/F_{go}(g^{-1}(t_d)) \\ &= F_{sr}(t \mid t_d). \end{aligned}$$

The cases $g^{-1}(t_d) < t < g^{-1}(t_d^*)$ and $g^{-1}(t_d) < g^{-1}(t_d^*) < t$ can be shown similarly.

Expected stop signal processing time. A formal expression for mean (expected) stop signal processing time is obtained as follows. We have

$$\begin{aligned} \mathrm{E}[T_{stop}] &= \int_0^\infty [1 - F_{stop}(s)]\, \mathrm{d}s \\ &= \int_\infty^0 F_{go}(t) \frac{\mathrm{d}s}{\mathrm{d}t}\, \mathrm{d}t \quad \text{(a.s.)}, \end{aligned} \tag{15}$$

using Equation [8] and the fact that $1 - F_{stop}(s)$ and $F_{go}(t)$ have opposite limit values. With

$$s = F_{stop}^{-1}[1 - F_{go}(t)]$$

we have

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -f_{go}(t)/f_{stop}\left[F_{stop}^{-1}[1 - F_{go}(t)]\right].$$

Inserting into [15] yields

$$\mathrm{E}[T_{stop}] = \int_0^\infty \frac{F_{go}(t)\, f_{go}(t)}{f_{stop}\left[F_{stop}^{-1}[1 - F_{go}(t)]\right]}\, \mathrm{d}t. \tag{16}$$

IND model: exponential case

Define independent, exponential distributions for T_{go} and T_{stop} with parameters $\lambda_{go} > 0$ and $\lambda_{stop} > 0$ for context $\mathcal{STOP}$ by

$$\begin{aligned} H(s,t) &= \mathrm{P}\left[T_{go} \leq s\right] \times \mathrm{P}\left[T_{stop} \leq t\right] \\ &= (1 - \exp[-\lambda_{go}\, s]) \times (1 - \exp[-\lambda_{stop}\, t]), \end{aligned}$$

for all $s, t \geq 0$. Then

$$\begin{aligned} p_r(t_d) &= \int_0^{t_d} f_{go}(t)\, \mathrm{d}t + \int_{t_d}^{\infty} f_{go}(t)\left[1 - F_{stop}(t - t_d)\right] \\ &= 1 - \frac{\lambda_{stop}}{\lambda_{stop} + \lambda_{go}} \exp[-\lambda_{go} t_d]. \end{aligned} \tag{17}$$

For $t > t_d$, the density of the signal-response distribution is given by,

$$\begin{aligned} f_{sr}(t \,|\, t_d) &= f_{go}(t)\left[1 - F_{stop}(t - t_d)\right] / p_r(t_d) \\ &= \frac{\lambda_{go} \exp[-\lambda_{go} t]\, \exp[-\lambda_{stop}(t - t_d)]}{\left(1 - \dfrac{\lambda_{stop}}{\lambda_{stop} + \lambda_{go}} \exp[-\lambda_{go} t_d]\right)} \\ &= \frac{1}{K}(\lambda_{go} + \lambda_{stop})\, \exp[-(\lambda_{go} + \lambda_{stop})(t - t_d)], \end{aligned} \tag{18}$$

with $K = \exp[\lambda_{go}\, t_d](1 + \lambda_{stop}/\lambda_{go}) - \lambda_{stop}/\lambda_{go}$. For $t_d = 0$, we have $K = 1$ and the signal-respond density is identical to an exponential density for an independent race between T_{stop} and T_{go}, with parameter $\lambda_{go} + \lambda_{stop}$ and $p_r(t_d) = \lambda_{go}/(\lambda_{go} + \lambda_{stop})$. For $t \leq t_d$, the density simplifies to

$$\begin{aligned} f_{sr}(t \,|\, t_d) &= f_{go}(t)] / p_r(t_d) \\ &= \frac{1}{\lambda_{stop} + \lambda_{go}} \exp[-\lambda_{go}(t - t_d)]. \end{aligned} \tag{19}$$

Computation of the expected value of signal-response RTs yields:

$$\begin{aligned} \mathrm{E}[T_{go} \,|\, T_{go} < T_{stop} + t_d] &= \int\limits_0^{\infty} t\, f_{sr}(t \,|\, t_d) \mathrm{d}t \\ &= \frac{\lambda_{go}\left[1 + (\lambda_{go} + \lambda_{stop}) t_d\right]}{(\lambda_{go} + \lambda_{stop})\{\exp[\lambda_{go}\, t_d](\lambda_{go} + \lambda_{stop}) - \lambda_{stop}\}}. \end{aligned} \tag{20}$$

In particular, for $t_d = 0$, we obtain $\mathrm{E}[T_{go} \,|\, T_{go} < T_{stop} + t_d] = 1/(\lambda_{go} + \lambda_{stop})$, consistent with the density we mentioned above for this value of the stop signal delay.

PND model: exponential example

Inserting exponential margins into the bivariate distribution,

$$\begin{aligned} H^{-}(s,t) &= \max\{F_{go}(s) + F_{stop}(t) - 1, 0\} \\ &= \max\{1 - \exp[-\lambda_{go}\, s] - \exp[-\lambda_{stop}\, t], 0\}, \end{aligned} \tag{21}$$

for all $s, t \geq 0$. From [14],

$$\begin{aligned} p_r(t_d) &= \mathrm{P}\left[T_{go} - T_{stop} < t_d\right] \\ &= \mathrm{P}\left[T_{go} - F_{stop}^{-1}[1 - F_{go}(T_{go})] < t_d\right] \\ &= \mathrm{P}\left[T_{go} - F_{stop}^{-1}[\exp(-\lambda_{go}\, T_{go})] < t_d\right] \\ &= \mathrm{P}\left[T_{go} + 1/\lambda_{stop} \log[1 - \exp(-\lambda_{go}\, T_{go})] < t_d)\right] \\ &= \mathrm{P}\left[g(T_{go}) < t_d)\right] \\ &= \mathrm{P}\left[T_{go} < g^{-1}(t_d)\right]. \end{aligned} \tag{22}$$

Note that function

$$g(T_{go}) \equiv T_{go} + 1/\lambda_{stop} \log[1 - \exp(-\lambda_{go}\, T_{go})]$$

cannot be solved explicitly for T_{go}. Therefore, in order to compute $p_r(t_d)$ and plot signal-response time distributions

$$F_{sr}(t \,|\, t_d) = F_{go}(\min\{t, g^{-1}(t_d)\})/F_{go}(g^{-1}(t_d)),$$

we sampled ($n = 100,000$) from the bivariate distribution function $H^{-}(s,t)$ using function simCop (based on the conditional simulation method, see (Nelsen, 2006)) from the **copBasic** package of the open source software R (http://www.r-project.org).

Fréchet-Hoeffding bounds and perfect dependency

The dependence between two random variables of a random vector (X, Y) is completely described by its probability distribution. Let $G(x, y)$ be the bivariate distribution function of some random vector (X, Y):

$$G(x, y) = \mathrm{P}\left(X \leq x, Y \leq y\right)$$

with marginal distributions F_X and F_Y. Then, it always holds that

$$\begin{aligned} G^-(x, y) =& \max\{F_X(x) + F_Y(y) - 1, 0\} \leq G(x, y) \\ & \leq \min\{F_X(x), F_Y(y)\} = G^+(x, y), \end{aligned}$$

for all x, y for which G is defined (the support of G). Both $G^-(x, y)$ and $G^+(x, y)$ are known as *Fréchet-Hoeffding bounds* and are themselves distribution functions for (X, Y): G^- corresponds to "perfect" negative dependence between X and Y, while G^+ corresponds to "perfect" positive dependence (see (Nelsen, 2006; Joe, 2015) for proofs of this and other claims in this section). "Perfect" dependency can be formulated in various ways. For simplicity, we assume X and Y have continuous distribution functions, the only case we will need below. Here are three equivalent characterizations of *perfect negative dependence* (analogous results exist for perfect positive dependence, but they are not of concern here):

(i) Either $\mathrm{P}\left(X > x, Y > y\right) = 0$ for all x, y or
$\mathrm{P}\left(X \leq x, Y \leq y\right) = 0$ for all x, y;

(ii) $\mathrm{P}\left[F_X(X) + F_Y(Y) = 1\right] = 1$;

(iii) X is almost surely a decreasing function of Y.

Consider the equation inside the "P " expression in **(ii)**: Solving for X by taking the inverse function of $F_X(X)$ we get

$$X = F_X^{-1}[1 - F_Y(Y)],$$

from which (iii) follows. Random variables with perfect negative dependence are also known as *antithetic variates* in simulation studies.

Acknowledgments

We thank Harry Joe (UBC) for some help with the statistical concepts. Authors take full responsibility for any remaining faults and expressions of opinion. HC was supported by a grant from German Science Foundation/DFG (SFB/TRR31) and and the Cluster of Excellence (Hearing4all) of Oldenburg University. AD was supported by grant DI 506/12-1 (German Science Foundation/DFG). First results of this paper were presented at the 50th Annual Meeting of the Society for Mathematical Psychology, University of Warwick, UK from July 22–25, 2017. A preliminary version of this paper has been posted on bioRxiv February 18, 2018 (https://doi.org/10.1101/158485).

References

Algom, D., Eidels, A., Hawkins, R., Jefferson, B., & Townsend, J. (2015). Features of response times: identification of cognitive mechanisms through mathematical modeling. In J. Busemeyer, Z. Wang, J. Townsend, & A. Eidels (Eds.), The Oxford Handbook of Mathematical and Computational Psychology (pp. 63–98). New York, NY 10016: Oxford University Press.

Band, G., van der Molen, M., & Logan, G. (2003). Horse-race model simulations of the stop-signal procedure. Acta Psychologica, 112, 105–142.

Bissett, P., & Logan, G. (2014). Selective stopping ? Maybe not. Journal of Experimental Psychology: General, 143(1), 455–472.

Boucher, L., Palmeri, T., Logan, G., & Schall, J. (2007). Inhibitory control in mind and brain: an interactive race model of countermanding saccades. Psychological Review, 114(2), 376–397.

Brown, J., Hanes, D., Schall, J., & Stuphorn, V. (2008). Relation of frontal eye field activity to saccade initiation during a countermanding task. Experimental Brain Research, 190, 135–151. (DOI 10.1007/s00221-008-1455-0)

Carpenter, R., & Noorani, I. (2017). Preface: Movement suppression: brain mechanisms for stopping and stillness. In (Vol. 372). The Royal Society Publishing. (10.1098/rstb.2016.0542)

Colonius, H. (1990). A note on the stop-signal paradigm, or how to observe the unobservable. Psychological Review, 97(2), 309–312.

Denuit, M., & Dhaene, J. (2003). Simple characterizations of comonotonicity and countermonotonicity by extremal correlations. Belgian Actuarial Bulletin, 3(1), 22–27.

Forstmann, B., & Wagenmakers, E. (Eds.). (2015). An introduction to model-based cognitive neuroscience (1st ed.). NewYork Heidelberg Dordrecht London: Springer-Verlag GmbH.

Hanes, D., Patterson, W., & Schall, J. (1998). Role of frontal eye field in countermanding saccades: visual, movement and fixation activity. Journal of Neurophysiology, 79, 817–834.

Hanes, D., & Schall, J. (1995). Countermanding saccades in macaque. Visual Neuroscience, 12, 929–937.

Joe, H. (2015). Dependence modeling with copulas (Vol. 134). Boca Raton, FL 33487-2742: CRC Press Chapman & Hall.

Lo, C., Boucher, L., Paré, M., Schall, J., & Wang, X. (2009). Proactive inhibitory control and attractor dynamics in countermanding action: a spiking neural circuit model. Journal of Neuroscience, 29, 9059—9071.

Logan, G. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In D. Dagenbach & T. Carr (Eds.), Inhibitory processes in attention, memory, and language (pp. 189–239). San Diego: Academic Press.

Logan, G. (2017). Takong control of cognition: an instance perspective on acts of control. American Psychologist, 72(9), 875–884.

Logan, G., & Cowan, W. (1984). On the ability to inhibit thought and action: a theory of an act of control. Psychological Review, 91(3), 295–327.

Logan, G., Yamaguchi, M., Schall, J., & Palmeri, T. (2015). Inhibitory control in mind and brain 2.0: Blocked-input models of saccadic countermandingon the ability to inhibit thought and action: General and special theories of an act of control. Psychological Review, 122, 115–147.

Matzke, D., Dolan, C., Logan, G., Brown, S., & Wagenmakers, E.-J. (2013). Bayesian parametric estimation of stop-signal reaction time distributions. Journal of Experimental Psychology: General, 142(4), 1047–1073.

Matzke, D., Verbruggen, F., & Logan, G. (in press). The stop-signal paradigm. In E. Wagenmakers (Ed.), Stevens' handbook of experimental psychology and cognitive neuroscience: Methodology (4th ed., Vol. 4). John Wiley & Sons.

Nelsen, R. B. (2006). An introduction to copulas (2nd ed., Vol. 139). New York, NY: Springer-Verlag.

Paré, M., & Hanes, D. (2003). Controlled movement processing: superior colliculus activity associated with countermanded saccades. Journal of Neuroscience, 23, 6480–6489.

Schall, J., & Godlove, D. (2012). Current advances and pressing problems in studies of stopping. Current Opinion in Neurobiology, 22, 1012–1021.

Schall, J., Palmeri, T., & Logan, G. (2017). Models of inhibitory control. Philosophical Transactions of the Royal Society of London B, 372(20160193). (http://dx.doi.org/10.1098/rstb.2016.0193)

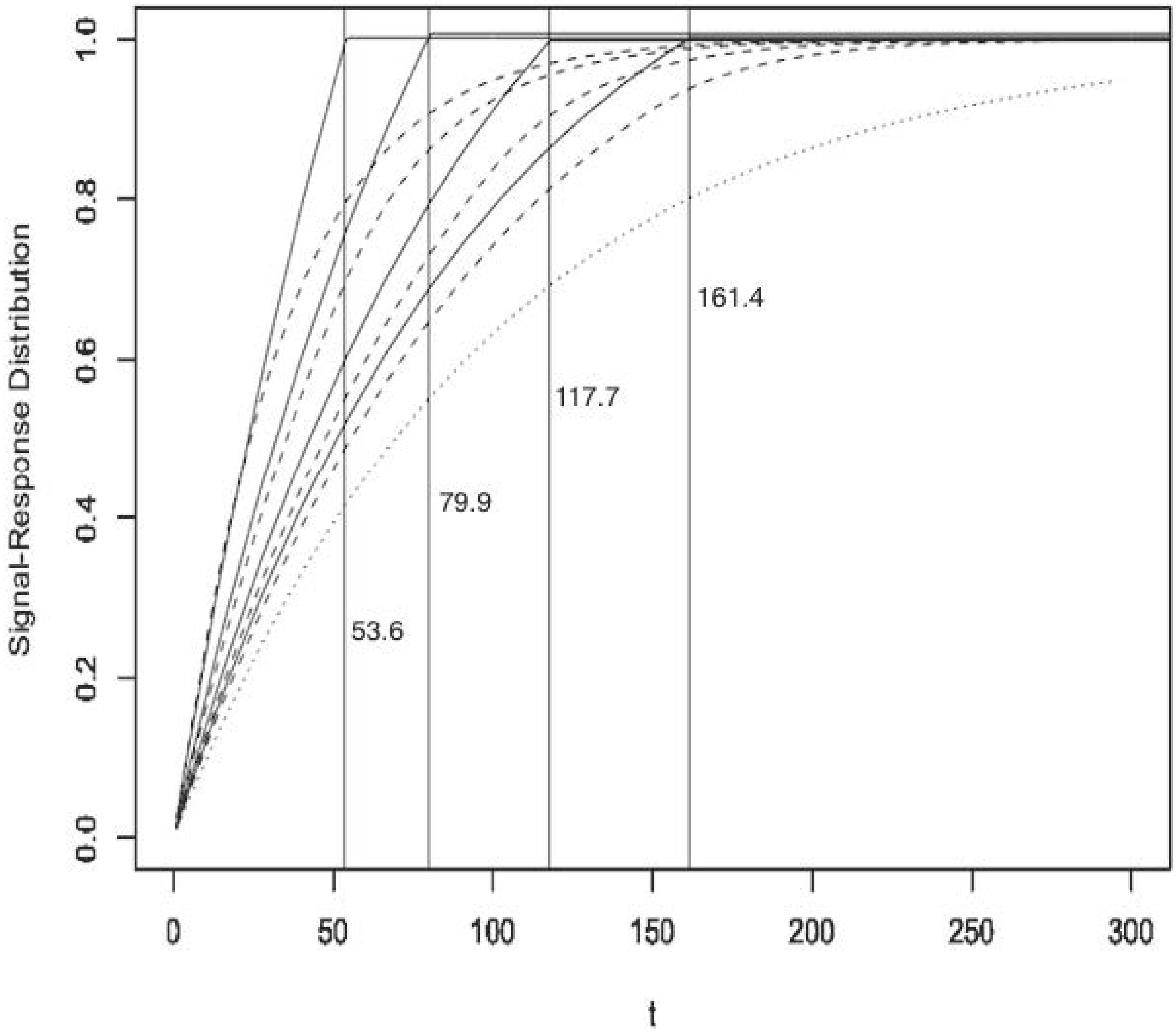
Townsend, J. (1990). Serial vs. parallel processing: Sometimes theylook like tweedledum and tweedledee but they can (and should be)distinguished. Psychological Science, 1, 46–54.

Townsend, J., & Wenger, M. (2004). A theory of interactive parallel processing: new capacity measures and predictions for a response time inequality series. Psychological Review, 111(4), 1003–1035.

Verbruggen, F., & Logan, G. (2008). Response inhibition in the stop-signal paradigm. Trends in Cognitive Sciences, 12, 418–424. (http://doi.org/10.1016/j.tics.2008.07.005)

Verbruggen, F., & Logan, G. (2009). Models of response inhibition in the stop-signal and stop-change paradigms. Neuroscience & Biobehavioral Reviews,, 33, 647–661. (http://doi.org/10.1016/j.neubiorev.2008.08.014)

Figure caption for Figure 1. The “fan” effect: all signal-response RT distributions $F_{sr}(t \,|\, t_d)$ are strictly ordered by delay (t_d) from left to right, for $t_d = 10, 50, 100, 150$ [ms], converging toward no-stop signal distribution $F_{go}(t)$ (rightmost curve) for $t_d \to \infty$, for exponential distributions with rate parameters $\lambda_{go} = .01$ and $\lambda_{stop} = .02$. The vertical lines and their corresponding $g^{-1}(t_d)$ values indicate the crossing points where $F_{sr}(t \,|\, t_d) = 1$ for the PND model. The PND model curves were obtained by simulation (see appendix); dashed: IND model; solid: PND model.


Signal-Response Distribution
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
250
300
t
53.6
79.9
117.7
161.4