The Stochastic Early Reaction, Inhibition, and Late Action (SERIA) Model for Antisaccades

SERIA - A model for errors and reaction times in the antisaccade task

Eduardo A. Aponte[1,*], Dario Schoebi[1], Klaas E. Stephan[1,2], Jakob Heinzle[1*]

[1]Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zurich & Swiss Institute of Technology Zurich, Zurich, Switzerland

[2]Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom

[*]Corresponding authors

aponte@biomed.ee.ethz.ch

heinzle@biomed.ee.ethz.ch

Abstract

The antisaccade task is a classic paradigm used to study the voluntary control of eye movements. It requires participants to suppress a reactive eye movement to a visual target and to concurrently initiate a saccade in the opposite direction. Although several models have been proposed to explain error rates and reaction times in this task, no formal model comparison has yet been performed. Here, we describe a Bayesian modeling approach for the antisaccade task that allows us to formally compare different models on the basis of their model evidence. First, we provide a formal likelihood function of actions (prosaccades or antisaccades) and reactions times based on a recently published model. Second, we introduce the *Stochastic Early Reaction, Inhibition, and late Action model* (SERIA), a novel model that postulates two different types of mechanisms that interact in the antisaccade task: a race-to-threshold decision process and a binary, time-insensitive decision process. Third, we apply these models to a data set from an experiment with three mixed blocks of pro- and antisaccade trials. Bayesian model comparison demonstrates that the SERIA model explains the data better than competing models that are based only on race-to-threshold processes. Moreover, we show that the race-to-threshold decision processes postulated by the SERIA model are, to a large extent, insensitive to the cue presented on a single trial. Finally, we use the same inversion technique to infer upon model parameters and demonstrate that changes in reaction time and error rate due to the probability of a trial type (prosaccade or antisaccade) are explained mostly by faster or slower inhibition and the probability of generating late voluntary prosaccades.

Author summary

One widely replicated finding in schizophrenia research is that patients tend to make more errors in the antisaccade task, a psychometric paradigm in which participants are required to look in the opposite direction of a visual cue. This deficit has been suggested to be an endophenotype of schizophrenia, as first order relatives of patients tend to show similar but milder deficits. Currently, most statistical models applied to experimental findings in this task are limited to fit average reaction times and error rates. Here, we propose a novel statistical model that fits experimental data from the antisaccade task beyond summary statistics. For this, we suggest that antisaccades are the result of several competing decision processes that interact nonlinearly with one another. Applying this model to a relatively large experimental data set, we show that mean reaction times and error rates do not fully reflect the complexity of the processes that are likely to underlie experimental findings. In the future, our model could help to understand the nature of the deficits observed in schizophrenia by providing a statistical tool to study the biological processes from which they arise.

Introduction

In the antisaccade task ([1]; for reviews, see [2, 3]), participants are required to saccade in the contralateral direction of a visual cue. This behavior is thought to require both the inhibition of a prepotent saccadic response towards the cue and the initiation of a voluntary eye movement in the opposite direction. A failure to inhibit the reflexive response leads to an erroneous saccade towards the cue (i.e., a prosaccade), which is often followed by a corrective eye movement in the opposite direction (i.e., an antisaccade). As a probe of inhibitory capacity, the antisaccade task has been widely used to study psychiatric and neurological diseases [3]. Notably, since the initial report [4], studies have consistently found an increased number of errors in patients with schizophrenia when compared to healthy controls, independent of medication and clinical status [5,6,7,8]. Moreover, there is evidence that an increased error rate constitutes an endophenotype of schizophrenia, as antisaccade deficits are also present in non-affected first-degree relatives of diagnosed individuals (for example [5, 7]). However, not all studies have reported positive evidence for this (for example [9,10]).

Unfortunately, the exact nature of the antisaccade deficits and their biological origin in schizophrenia remain unclear. One approach to improve our understanding of experimental findings is to develop generative models of their putative computational and/or neurophysiological causes [11]. Generative models can reveal features of the data that are not apparent when only considering summary statistics such as mean error rate (ER) and reaction time (RT) [12]. Additionally, generative models can relate behavioral findings in humans to their biological substrate.

Here, we apply a generative modelling approach to the antisaccade task. First, we introduce a novel model of this paradigm based on previous race-to-threshold models [13-16]. For this, we formalize the model introduced by Noorani and Carperter [15] and extend it into what we refer to as the *Stochastic Early Response, Inhibition and late Action* (SERIA) model. We then apply both models to an experimental data set of three mixed blocks of pro- and antisaccades trials with different trial type probability using formal Bayesian inference. More specifically, we compare several models using Bayesian model comparison. Thirdly, we use the parameter estimates from the best model to investigate the effects of our experimental manipulation. We found that there was positive evidence in favor of the SERIA model when compared to our formalization of the model proposed

in [15]. Moreover, the parameters estimated through model inversion revealed a complex picture of the decision processes underlying the antisaccade task that is not obvious from mean RT and ER.

This paper is organized as follows. First, we formalize the model developed in [15] and introduce the SERIA model. Second, we present our experimental setup. Third, in the results section, we present our behavioral findings in terms of summary statistics (mean RT and ER), the comparison between different models, and the parameter estimates. Finally, we review our findings, discuss other recent models, and potential future developments and translational applications.

Materials and methods

Ethics statement

All participants gave written informed consent before the study. All experimental procedures were approved by the local ethics board (Kantonale Ethikkomission Zürich, KEK-ZH-Nr.2014-0246).

Race to threshold models for antisaccades

In this section, we derive a formal description of the models evaluated in this paper. We start with a formalized version of the model in [15] and proceed to extend this approach.

The pro, stop, and antisaccade model (PROSA)

Following [15], we assume that the RT and the type of saccade generated in a given trial is caused by the interaction of three competing race-to-threshold units. The first unit u_p represents a command to perform a prosaccade, the second unit u_s represents an inhibitory command to stop a prosaccade, and the third unit u_a represents a command to perform an antisaccade. The time required for each unit to arrive to threshold is given by:

$$s_i = r_i t, \tag{1}$$

$$\frac{s_i}{r_i} = t, \tag{2}$$

where r_i represents the slope or increase rate of unit i, s_i represents the height of the threshold, and t represents time. We also assume that, on each trial, the increase rates are stochastic and independent from each other.

The time and order in which the units reach their thresholds s_i determines the action and RT in a trial. If the prosaccade unit u_p reaches threshold before any other unit at time t, a prosaccade is elicited at t. If the antisaccade unit arrives first, an antisaccade is elicited at t. Finally, if the stop unit arrives before the prosaccade unit, an antisaccade is elicited at the time when the antisaccade unit reaches threshold.

Formally (but in a slight abuse of language), the two random variables of interest, the reaction time $T \in [0, \infty[$ and the type of action performed $A \in \{pro, anti\}$, depend on three further random variables: the arrival times $U_p, U_s, U_a \in [0, \infty[$ of each of the units.

The probability of performing a prosaccade at time t is given by the probability of the prosaccade unit arriving at time t, and the stop and antisaccade unit arriving afterwards:

$$p(A = pro, T = t) = p(U_p = t)p(U_a > t)p(U_s > t). \tag{3}$$

The probability of performing an antisaccade at time t is given by

$$p(A = anti, T = t) = \tag{4}$$

$$p(U_a = t)p\big(U_p > t\big)p(U_s > t) + p(U_a = t)\int_0^t p\,(U_s = \tau)p\big(U_p > \tau\big)d\tau.$$

The first term on the right side of Eq. 4 corresponds to the unlikely case that the antisaccade unit arrives before the prosaccade and the stop unit. The second term describes trials in which the stop unit arrives before the prosaccade unit. It can be decomposed into two terms:

$$p(U_a = t)\int_0^t p\,(U_s = \tau)p(U_p > \tau)d\tau \tag{5}$$

$$= p(U_a = t)\left(p(U_s < t)p\big(U_p > t\big) + \int_0^t p\,(U_s = \tau)p\big(\tau < U_p < t\big)d\tau\right)$$

$$= p(U_a = t)\left(p(U_s < t)p\big(U_p > t\big) + \int_0^t p(U_s < \tau)p\big(U_p = \tau\big)d\tau\right). \tag{6}$$

The term $p(U_a = t)\int_0^t p(U_s < \tau)p(U_p = \tau)d\tau$ describes the condition in which the prosaccade unit is inhibited by the stop unit allowing for an antisaccade. Note that if the prosaccade unit arrives later than the antisaccade unit, the arrival time of the stop unit is irrelevant. That means that we can simplify Eq. 4 to

$$p(A = anti, T = t) = p(U_a = t)\left(p(U_p > t) + \int_0^t p(U_s < \tau)p(U_p = \tau)d\tau\right). \tag{7}$$

Eq. 3 and 7 constitute the likelihood function of a single trial, defining the joint probability of an action and the corresponding RT. We refer to this likelihood function as the PRO-Stop-Antisaccade (PROSA) model. This model shares the central assumptions of [15]: (i) the time to reach threshold of each of the units is assumed to depend linearly on the rate r, (ii) it includes a stop unit whose function is to inhibit prosaccades and (iii) both models

assume no lateral inhibition between the different units. Finally, (iv) the reaction times are assumed to be equal to the reach-to-threshold times. Note that the RT distributions are different from the arrival-time distributions because of the interactions between the units described above. The main difference of this model compared to [15] is that we do not exclude *a priori* the possibility of the antisaccade unit arriving earlier than the other units. Otherwise, both models are conceptually equivalent.

The Stochastic Early Reaction, Inhibition, and Late Action Model (SERIA)

The PROSA model is characterized by a strict association between units and action types. In other words, the unit u_p leads unequivocally to a prosaccade, whereas the unit u_a always triggers an antisaccade. This implies that if the distribution of the arrival times of the units is unimodal and strictly positive, the PROSA model cannot predict voluntary slow prosaccades with a late peak. Hence, the PROSA model cannot account for slow, voluntary prosaccades that have been postulated in the antisaccade task [17]. Similarly, it has been argued that prosaccade RT can be described by the mixture of two distributions [18]. To account for this, we introduce the Stochastic Early Reaction, Inhibition and Late Action model (SERIA).

According to this model, and in analogy to the PROSA model, an early reaction takes place at time t if the early unit u_e arrives before the late and inhibitory units, u_l and u_i, respectively. If the inhibitory or late unit arrives before the early unit, a late response is triggered at the time the late unit reaches threshold. Crucially, both early and late responses can trigger pro- and antisaccades with a certain probability. Thus, in parallel to the race-to-threshold processes which determines RTs, an independent, secondary decision process is responsible for which reaction is generated. Fig. 1 shows the structure of the SERIA model.

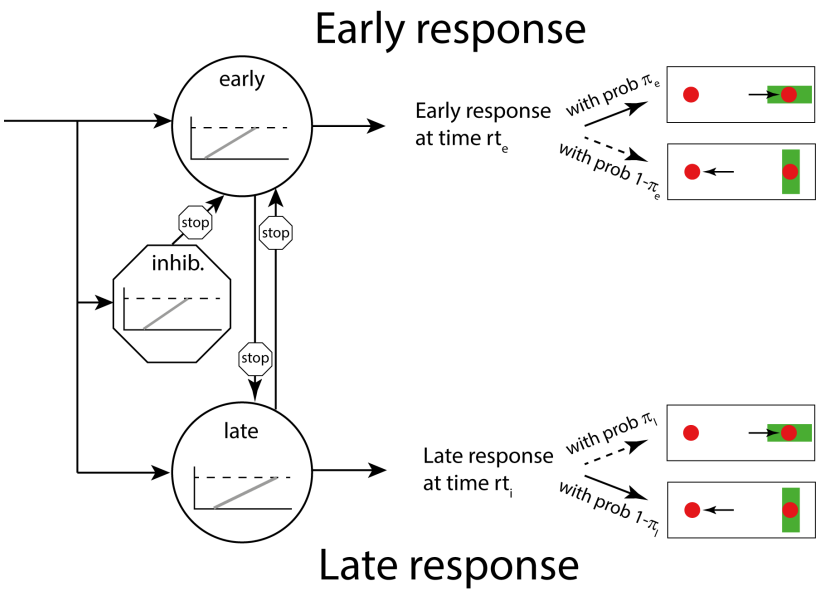

Fig 1. Layout of the SERIA model.

The presentation of a visual cue (a green bar) triggers the race of three independent units. The inhibitory unit can stop an early response. Importantly, both early and late responses can trigger pro- and antisaccades. Note that the PROSA model is a special case of the SERIA model in which $\pi_e = 1$ and $\pi_l = 0$, i.e. all early responses are prosaccades, whereas all late responses are antisaccades.

To formalize the concept of early and late responses, we introduce a new unobservable random variable that represents the type of response $R \in \{early, late\}$. The distribution of the RTs is analogous to the PROSA-model, such that e.g. the probability of an early response at time t is given by

$$p(R = early, T = t) = p(U_e = t)p(U_i > t)p(U_l > t) \tag{8}$$

where U_e, U_i and U_l represent the arrival times of the early, inhibitory, and late units respectively. The fundamental assumption of the SERIA model is that the action type (pro- or antisaccade) is conditionally independent of the RT given the response type (early or late). Hence, the distribution of RTs is not *a priori* coupled to the saccade type anymore; RT distributions for both pro- and antisaccades could in principle be bimodal, consisting of both fast reactive and slow voluntary saccades. Formally, the conditional independency assumption can be written down as

$$p(A, T|R) = p(A|R)p(T|R), \tag{9}$$

$$p(A, T|R)p(R) = p(A|R)p(T|R)p(R), \tag{10}$$

$$p(A,T,R) = p(A|R)p(T,R). \tag{11}$$

The term $p(A|R)$ is simply the probability of an action, given a response type. We denote it as

$$p(A = pro|R = early) = \pi_e \in [0,1], \tag{12}$$

$$p(A = anti|R = early) = 1 - \pi_e, \tag{13}$$

$$p(A = pro|R = late) = \pi_l \in [0,1], \tag{14}$$

$$p(A = anti|R = late) = 1 - \pi_l. \tag{15}$$

Since the type of response R is not observable, it is necessary to marginalize it out in Eq. [11] to obtain the likelihood of the SERIA model:

$$p(A,T) = p(A,T,R = early) + p(A,T,R = late). \tag{16}$$

The complete likelihood of the model is given by substituting the terms in Eq. [16]:

$$p(A = pro, T = t) = \pi_e p(U_e = t)p(U_i > t)p(U_l > t) + \tag{17}$$

$$\pi_l p(U_l = t)\left(p(U_e > t) + \int_0^t p(U_e = \tau)p(U_i < \tau)d\tau\right),$$

$$p(A = anti, T = t) = (1 - \pi_e)p(U_e = t)p(U_i > t)p(U_l > t) + \tag{18}$$

$$(1 - \pi_l)p(U_l = t)\left(p(U_e > t) + \int_0^t p(U_e = \tau)p(U_i < \tau)d\tau\right).$$

It is worth noting here that the PROSA model is a special case of the SERIA model, namely, it corresponds to the assumption that $\pi_e = 1$ and $\pi_l = 0$. The SERIA model allows for bimodal distributions, as both early and late responses can be pro- and antisaccades. Importantly, one prediction of the model is that late prosaccades have the same distribution as late antisaccades.

Non-decision time

The models above can be further finessed to account for non-decision times δ by transforming the reaction times t to $t_\delta = t - \delta$. The delay δ might be caused, for example, by conductance delays from the retina to the cortex. In addition, the antisaccade (or

"late") unit might include a constant delay δ_a, which is often referred to as the antisaccade cost [1]. Note that the model is highly sensitive to δ since any RT lower than δ has zero probability. In order to relax this condition and to account for early outliers, we assumed that saccades could be generated before δ at a rate $\eta \in [0,1]$ such that the marginal likelihood of an outlier is

$$p(T < \delta) = p(T_\delta < 0) = \eta. \tag{19}$$

For simplicity, we assume that outliers are generated with uniform probability in the interval $[0, \delta]$:

$$p(T = t) = \frac{\eta}{\delta} \text{ if } t < \delta. \tag{20}$$

Furthermore, we assume that the probability of an early outlier being a pro- or antisaccade is equal. Because of the new parameter η, the distribution of saccades with RT larger than δ needs to be renormalized by the factor $1 - \eta$. In the case of the PROSA model for example this means that now the joint distribution of action and reaction time is given by the conditional probability

$$p(A = pro, T = t_\delta | t_\delta > 0) = \quad p(U_p = t_\delta) p(U_a > t_\delta - \delta_a) p(U_s > t_\delta), \tag{21}$$

$$p(U_a < 0) = 0, \tag{22}$$

$$p(A = anti, T = t_\delta | t_\delta > 0) = \tag{23}$$

$$p(U_a = t_\delta - \delta_a) \left(p(U_p > t_\delta) + \int_0^{t_\delta} p(U_p = \tau) p(U_s < \tau) d\tau \right).$$

A similar expression holds for the SERIA model. However, in the PROSA model a unit-specific delay is equal to an action-specific delay. By contrast, in the SERIA model both early and late responses can generate pro- and antisaccades. Thus, in the case of the SERIA model, δ_a represents a delay of the late unit that affects both late pro- and antisaccades.

Parametric distributions of the increase rate

The models discussed in the previous sections can be defined independently of the distribution of the rate of each of the units. In order to fit experimental data, we

considered four parametric distributions with positive support for the rates: gamma, inverse gamma, lognormal [19] and the truncated normal distribution (similar to [18]). Table 1 and Fig. 2 summarize these distributions, their parameters, and the corresponding arrival time densities. We considered five different configurations: 1) all units were assigned *inverse gamma* distributed rates, 2) all units were assigned *gamma* distributed rates, 3) the increase rate of the pro and stop unit (or early and the inhibitory unit) were *gamma distributed* but the antisaccade (late) unit's increase rate was *inverse gamma* distributed, 4) all the units were assigned *lognormal* distributed rates or 5) all units were assigned *truncated normal* distributed rates.

Table 1: Parametric density functions of the increase rates.

Name	Parameters	Rate p.d.f.	Arrival time p.d.f.
Gamma	k, θ	$\frac{\theta^{-k}}{\Gamma(k)} e^{-r/\theta} r^{k-1}$	$\frac{\theta^{k}}{\Gamma(k)} e^{-\theta/t} t^{-k-1}$
Inv. gamma	k, θ	$\frac{\theta^{k}}{\Gamma(k)} e^{-\theta/r} r^{-k-1}$	$\frac{\theta^{-k}}{\Gamma(k)} e^{-t/\theta} t^{k-1}$
Log normal	μ, σ^2	$\frac{1}{\sqrt{2\pi}\sigma r} e^{-\frac{1}{2}\left(\frac{\ln r-\mu}{\sigma}\right)^2}$	$\frac{1}{\sqrt{2\pi}\sigma t} e^{-\frac{1}{2}\left(\frac{\ln t+\mu}{\sigma}\right)^2}$
T. normal	μ, σ^2	$\frac{1}{Z} e^{-\frac{1}{2}\left(\frac{r-\mu}{\sigma}\right)^2}$	$\frac{1}{Zt^2} e^{-\frac{1}{2}\left(\frac{t^{-1}-\mu}{\sigma}\right)^2}$

Z is the appropriate normalization constant, i.e., $Z = \int_0^\infty exp\left(-\frac{(r-\mu)^2}{2\sigma^2}\right) dr.$

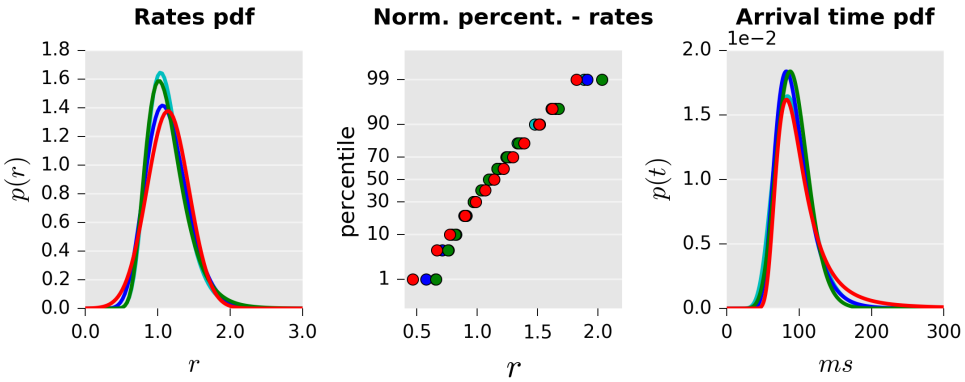

Fig 2. Illustration of probability distributions used to model increase rates.

Left: Distribution of the rates based on different probability density functions: Normal (red), gamma (blue), inverse gamma (green) log-normal (cyan). All distributions were matched to have equal mean and variance. Middle: Probit plots of the same distributions. While the gamma and lognormal distributions are very close to the straight line induced by the normal distribution, the inverse gamma distribution diverges slightly more from linearity. Right: Arrival times distribution (scaled to ms).

All the parametric distributions considered here can be fully characterized by two parameters which we generically refer as k and θ. Hence, the PROSA model is characterized by the parameters for each unit k_p, k_a, k_s, θ_p, θ_a, θ_s. The SERIA model can be characterized by analogous parameters k_e, k_l, k_i, θ_e, θ_l, θ_i and the probabilities of early and late prosaccades π_e and π_l. In addition to the unit parameters, both models included

the non-decision time δ, the antisaccade (or late unit) cost δ_a, and the marginal rate of early outliers η.

Experimental procedures

In this section, we describe the experimental procedures, statistical methods, and inference scheme used to invert the models above. The data is from the placebo condition of a larger pharmacological study that will be reported elsewhere.

Participants

Fifty-two healthy adult males naïve to the antisaccade task were invited to a screening session through the recruitment system of the Laboratory of Social and Neural System Research of the University of Zurich. During screening, and after being debriefed about the experiment, subjects underwent an electrocardiogram, a health survey, a visual acuity test, and a color blindness test. Subjects were excluded if any of the following criteria were met: age below 18 or above 40 years, regular smoking, alcohol consumption the day before the experiment, any possible interaction between current medication and levodopa or benserazide, pulse outside the range 55-100bpm, recreational drug intake in the past 6 months, history of serious mental or neurological illness, or if the medical doctor supervising the experiment deemed the participant not apt. All subjects gave their written informed consent to participate in the study and received monetary compensation.

Procedure

Each subject was invited to two sessions. During both visits, the same experimental protocol was followed. After arrival, placebo or levodopa (Madopar® DR 250, 200mg of levopa + 50 mg benserazide) was orally administered in the form of shape- and color-matched capsules. The present study is restricted to data from the session in which subjects received placebo. Participants and experimenters were not informed about the identity of the substance. Immediately afterwards subjects were introduced to the experimental setup and to the task through a written document. This was followed by a short training block (see below).

The experiment started 70 minutes after substance administration. Subjects participated in three blocks of 192 randomly interleaved pro- and antisaccade trials. The percentages of prosaccade trials in the three blocks were 20%, 50%, or 80%. This yielded three *prosaccade probability* (PP) conditions: PP20, PP50, and PP80. Thus, in the PP20 block, subjects were presented a prosaccade cue in 38 trials, while in all other trials (154) subjects were shown an antisaccade cue. The order of trials was randomized in each

block, but the same order was used in all subjects and sessions. The order of the conditions was counterbalanced across subjects.

Stimulus and apparatus

During the experiment, subjects sat in front of a CRT monitor (Philipps 20B40, distance eye-screen: $\simeq 60cm$, refresh rate: 75Hz). The screen subtended a horizontal visual angle of 38 degrees of visual angle (dva). Eye movements were recorded using a remote infrared camera (Eyelink II, SR-Research, Canada). Participants' head was stabilized with a chin rest. Data were stored at a sampling rate of 500 Hz.

During the task, two red dots (0.25dva), which constituted the saccadic targets, were constantly displayed at an eccentricity of $\pm$12dva. Displaying the saccadic target before the execution of an antisaccade has been reported to affect saccadic velocity and accuracy, but not RTs [20], and arguably decreases the need for sensorimotor transformations [21]. At the beginning of each trial, a gray fixation cross ($0.6\times0.6\ dva$) was displayed at the center of the screen. After a random fixation interval (500 to 1000 ms), the cross disappeared, and the cue instructing either a pro- or an antisaccade trial (see below) was shown centered on either of the red dots. As mentioned above, in each block, subjects were presented with a prosaccade cue in either 20, 50, or 80 percent of the trials. The order of the presentation of the cues was randomized. The cue was a green rectangle ($3.48\times0.8dva$) displayed for 500ms in either horizontal (prosaccade) or vertical orientation (antisaccade). Once the cue was removed and after 1000ms, the next trial started.

Subjects were instructed to saccade in the direction of the cue when a horizontal bar was presented (prosaccade trial) and to saccade in the opposite direction when a vertical bar was displayed (antisaccade trial, see Fig. 3). See [22, 23] for similar task designs.

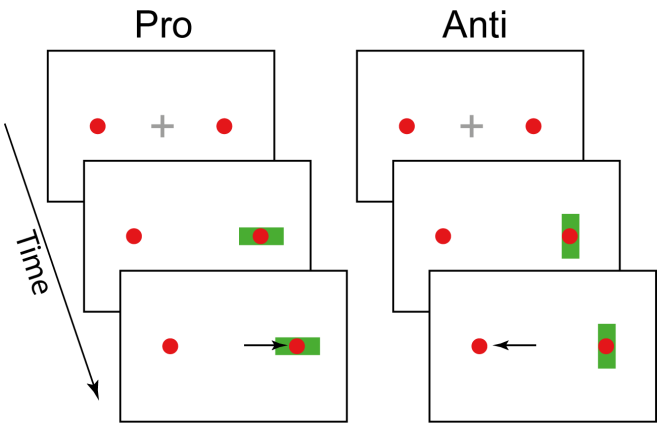

Fig 3. Task design.

After a variable fixation period of 500-1000ms (top) the cue (green rectangle) appeared on the screen for 500 ms. The orientation of the cue (horizontal or vertical) indicated the required response (prosaccade or antisaccade).

Prior to the main experiment, participants were trained on the task in a block of 50 prosaccade trials, immediately followed by 50 antisaccade trials. During the training, subjects were automatically informed after each trial whether their response had been correct or not (see below), or whether they had failed to produce a saccade within 500ms after cue presentation (CP). Please note that no feedback was given during the main experimental blocks.

Data preparation

Data were parsed and preprocessed using the Python programming language (2.7). Saccades were detected using the algorithm provided by the eyetracker manufacturer (SR Research), which uses a velocity and acceleration threshold of $22dva/s$ and $3800dva/s^2$ [24]. We only considered saccades with a magnitude larger than $2dva$. RT was defined as the time between CP and the first saccade larger than $2dva$. A prosaccade trial was considered correct if the end position of the saccade was ipsilateral to the cue and, conversely, an antisaccade trial was considered correct if the end position of the saccade was contralateral to the cue.

Trials were excluded from further analysis if a) data were missing, b) a blink occurred between CP and the main saccade, c) the trial was aborted by the experimenter, d) subjects failed to fixate in the interval between fixation detection and CP, e) if a saccade was detected only later than 800ms after CP, f) if the RT was below 50ms, and in the case of an antisaccade if it was below 110ms. Corrective antisaccades were defined as saccades

that a) followed a prosaccade error, b) occurred no later than $900ms$ after CP, and c) had less than $3dva$ horizontal error from the red circle contralateral to the cue.

Besides the fitted non-decision time δ we assumed a fixed non-decision time of $50ms$ for all participants [15]. This was implemented by subtracting $50ms$ of all saccades before being entered into the model. In order to avoid numerical instabilities, RT were rescaled from millisecond to tenths of a second during all numerical analysis. All results are presented in ms.

Classical statistics

Frequentist analyses of RT and ER were performed using a mixed effects generalized linear model with independent variables *subject* (SUBJECT), *prosaccade probability* (PP) with levels PP20, PP50 and PP80, and when pro- and antisaccade trials were analyzed together, *trial type* (TT). The factor SUBJECT was always entered as a random effect, whereas PP and TT were treated as categorical fixed effects. In the case of ER, we used the probit function as link function.

Analyses were conducted with the function *fitglme.m* in MATLAB 9.0. The significance threshold α was set to 0.05.

Modeling

We aimed to answer three questions with the models analyzed here. First, we investigated which of the models proposed here (i.e. PROSA or SERIA) explained the experimental data better, and whether all important qualitative features of the data were captured by the models. We did not have a strong hypothesis regarding the parametric distribution of the data. Hence, comparisons of parametric distributions were only of secondary interest in our analysis. Second, we investigated whether reduced models that kept certain parameters fixed across trial types were sufficient to model the data. Third, we investigated how the probability of a trial type in a block affected the parameters of the model.

Model space

Initially, we defined ten different models as shown in Table 2. Each model was fitted independently for each subject and condition. Since our experimental design included mixed blocks, we allowed for different parameters in pro- and antisaccade trials, i.e., different increase-rate distributions depending on the TT. Under this hypothesis, the

PROSA model had 12 free parameters (6 for each trial type), whereas the SERIA model required 4 further parameters (π_e and π_l in pro- and antisaccade trials). Regarding the non-decision time δ, antisaccade cost δ_a, and rate of outliers η, we assumed equal parameters in both TT. Consequently, the full PROSA model had 15 free parameters whereas the full SERIA model had 19 free parameters.

Table 2. Model families with the respective increase-rate distributions.

	PROSA		
Model	**Prosaccade/ Stop units**	**Antisaccade unit**	**# Param. full/const.**
m_1/m_1^c	Inv. gamma	Inv. gamma	15/13
m_2/m_2^c	Gamma	Gamma	15/13
m_3/m_3^c	Gamma	Inv. gamma	15/13
m_4/m_4^c	Lognorm.	Lognorm.	15/13
m_5/m_5^c	T. norm.	T. norm.	15/13
	SERIA		
	Early/Stop units	**Late unit**	
m_6/m_6^c	Inv. gamma	Inv. gamma	19/13
m_7/m_7^c	Gamma	Gamma	19/13
m_8/m_8^c	Gamma	Inv. gamma	19/13
m_9/m_9^c	Lognorm.	Lognorm.	19/13
m_{10}/m_{10}^c	T. norm.	T. norm.	19/13

Models with parameters constrained to be equal across trial types are referred through the superscript ^{c}.

In addition to the full models, we evaluated restricted versions of each model by constraining parameters to be shared across TT. In the case of the SERIA model, we hypothesized that the parameters of all units were equal irrespective of TT, i.e., that the rate of the units was not affected by the cue presented in a trial. However, we assumed that the probability that an early or late response was a prosaccade was different in pro and antisaccade trials. Therefore, instead of 12 unit parameters (6 per TT), the restricted SERIA model had only 6 parameters for the units' rates. The parameters π_e and π_l were allowed to differ in pro and antisaccade trials. In the case of the PROSA model, similar to [15], it is possible to assume that the parameters of the prosaccade unit remain constant across TT, and that parameters of the stop and antisaccade unit depend on TT, yielding 10 unit parameters.

Prior distributions for model parameters

To complete the definition of our generative models, a prior distribution of the parameters was specified. This distribution reflects beliefs that are independent of the data and provides a form of regularization when inverting a model. In order to avoid any undesired bias regarding the parametric distributions considered here, we reparametrize all but the truncated normal distribution in terms of their mean and variance. We then assumed that the log of the mean and variance of the rate of the units were equally normal distributed (see Table 3). Therefore, the parametric distributions had the same prior in terms of their first two central moments. In the case of the truncated normal distribution, instead of an analytical transformation between its first two moments and its natural parameters μ and σ^2, we defined the prior distribution as a density of μ and $\ln \sigma^2$. To warrant that the μ was positive with high likelihood (96%) we assumed that $\mu \sim N(0.55, 0.09)$. The variance term was distributed as displayed in Table 3. As a further constraint, we restricted the parameter space to enforce that the first two moments of the distributions of rates and RTs existed.

Table 3. Prior probability density functions.

Parameter	Probability density function	Expected value	Variance
μ_r	$\mathcal{N}(\ln \mu_r\,; -1.08, 0.97)$	0.55	0.5
σ_r^2	$\mathcal{N}(\ln \sigma_r^2\,; -2.64, 0.69)$	0.1	0.01
δ	$\mathcal{N}(\ln \delta\,; -1.58, 1.79)$	0.5	1.25
δ_a	$\mathcal{N}(\ln \delta_a\,; -0.87, 1.17)$	0.75	1.25
π_e	$U_{[0,1]}$	0.5	0.08
π_l	$U_{[0,1]}$	0.5	0.08
η	$Beta(\eta; 1,6)$	1/7	0.01

For the non-decision time δ and the antisaccade cost δ_a, the prior was a lognormal distribution equal across all models. Note that the scale of the parameters δ and δ_a in Table 3 is tenths of a second. The distribution of the fraction of early outliers η was assumed to be a Beta distribution with parameters 1 and 6 or equivalently

$$p(\eta) \propto (1-\eta)^5. \tag{24}$$

Finally, we assumed that the parameters π_e, π_i were uniformly distributed in the interval [0,1]. Table 3 displays the parameters used for the prior distributions.

Bayesian inference

Inference on the model parameters was performed using the Metropolis-Hasting algorithm [25]. To increase the efficiency of this sampling scheme, we iteratively modified the proposal distribution during an initial 'burn-in' phase as proposed by [26]. Moreover, we extended this method by drawing from a set of chains at different temperatures and swapping samples across chains. This method, called population MCMC or parallel tempering, increases the statistical efficiency of the Metropolis-Hasting algorithm [27] and has been used in similar contexts before [28]. We simulated 16 chains with a 5-th order temperature schedule [29], drawing a total of 2×10^4 samples per chain, from which the first half was discarded as part of the burn-in phase.

Models were scored using their log marginal likelihood or log model evidence (LME). This is defined as the log probability of the data given a model after marginalizing out all its parameters. When comparing different models, the LME corresponds to the log posterior probability of a model under a uniform prior on model identity. Thus, for a single subject with data *y*, the posterior probability of model *k*, given models 1 to *n* is

$$p(m_k|y) = \frac{p(y|m_k)p(m_k)}{\sum_{i=1}^{n} p(y|m_i)\, p(m_i)} = \frac{p(y|m_k)}{\sum_{i=1}^{n} p(y|m_i)}. \tag{25}$$

Importantly, this method takes into account not only the accuracy of the model but also its complexity, such that overparameterized models are penalized [30]. Widely used approximations to the LME include the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC); these are easy to compute but have a limited concept of complexity (for discussion, see [31]). Here, we computed the LME through sampling using thermodynamic integration [27, 29]. This method provides robust estimates and can be easily computed using samples obtained through population MCMC.

Besides comparing the evidence of each model, we also performed a hierarchical or random effects analysis described in [31, 32]. This method can be understood as a form of soft clustering in which each subject is assigned to a model using the LME as assignment criterion. Here, we report the expected probability of the model r_i, which represents the percentage of subjects that are assigned to the cluster representing model

i. This hierarchical approach is robust to population heterogeneity and outliers, and complements reporting the group-level LME. Finally, we compared families of models [33] based on the evidence of each model for each subject summed across conditions.

Implementation

All likelihood functions were implemented in the *C* programming language using the GSL numerical package (v. 1.13). Integrals without an analytical form or well-known approximations were computed through numerical integration using the Gauss-Kronrod-Patterson algorithm [34] implemented in the function *gsl_integration_qng*. The sampling routine was implemented in MATLAB (v. 8.1) and is available as a module of the open source software package TAPAS (www.translationalneuromodeling.org/tapas).

Results

Behavior

Forty-seven subjects (age: 23.8 ± 2.9) completed all blocks and were included in further analyses. A total of 27072 trials were recorded, from which 569 trials (2%) were excluded (see Table 4).

Table 4. Summary of trials.

	Valid	Blink	Missing	Aborted	FE	Late S.	Early S.	Total
Total	26503	188	60	42	249	0	30	27072
Mean	563.9	4.0	1.3	0.9	5.3	0.0	0.6	576
Std.	9.9	5.1	2.5	1.5	5.0	0.0	1.3	-
Min.	536	0	0	0	0	0	0	-
Max.	576	22	15	6	19	0	8	-

FE: Fixation errors. Late saccades are saccades elicited after 800ms. Early saccades are prosaccades elicited before 50ms after CP or antisaccades elicited before 110ms after CP.

Both ER, and RT showed a strong dependence on PP (Fig 4 and Table 5). The mean RT of correct prosaccade and antisaccade trials were analyzed independently with two ANOVA tests with factors SUBJECT and PP. We found that, in both prosaccade ($F_{2,138} = 46.9, p < 10^{-5}$) and antisaccade trials ($F_{2,138} = 37.3, p < 10^{-5}$) the effect of PP was significant. With increasing PP, prosaccade RT diminished, whereas the RT of correct antisaccades increased. Similarly, there was significant effect of PP on ER (ANOVA with factors SUBJECT and PP) in both prosaccade ($F_{2,138} = 376.1, p < 10^{-5}$) as well as in antisaccade ($F_{2,138} = 347.0, p < 10^{-5}$) trials.

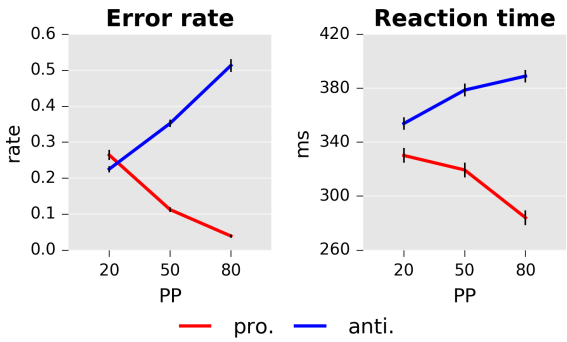

Fig 4. Error rate and reaction times as a function of prosaccade trial probability (PP).

Left panel: Mean error rates for prosaccade and antisaccade. Right panel: Mean RT in ms. Error bars indicate standard errors of the mean.

Table 5. Summary of mean RTs and ERs.

	Reaction times (ms)		
	PP 20	PP 50	PP80
Pro.	330(72)	319(67)	284(59)
Pro. error	326(68)	329(46)	336(57)
Anti.	354(60)	378(57)	389(61)
Anti. error	234(50)	231(47)	225(31)
	Error rates %		
Pro.	26(15)	11(8)	4(4)
Anti.	23(17)	35(21)	51(20)

Standard deviations are shown in brackets.

Modeling

Model comparison results

In a first step, we considered the models outlined in Table 2. The LME over all participants (fixed effects analysis) and the posterior probability of all models and all subjects are presented in Fig 5. Independently of the particular parametric distribution of the units, the SERIA model showed higher evidence compared to the PROSA model. A random effects family-wise model comparison [33] resulted in an expected frequency of $r = 98\%$ for the SERIA model family ($r = 2\%$ for PROSA). In addition, constraining the parameters

to be equal across trial types increased the model evidence irrespective of the parametric distribution assigned to the units (Fig 5). Here, the family-wise model comparison showed that models with constrained parameters had an expected frequency of $r = 93\%$. Over all 20 models, m_8^c showed the highest LME with a difference of $\Delta LME > 78.2$ compared to all other models. Following [35], a difference in LME larger than 3 corresponds to strong evidence, roughly equivalent to a p-value of 0.05.

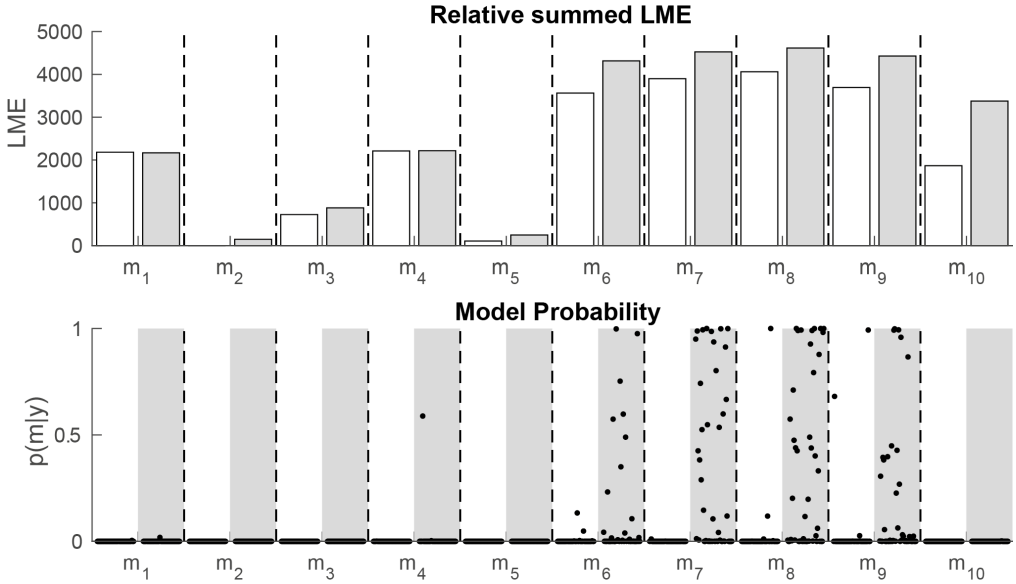

Fig 5. Summary of model comparison.

Top: Summed LME across all subjects for all 20 models. White bars show models with all parameters free, grey bars models with restricted parameters. LMEs are normalized by subtracting the lowest LME (m_2). Model m_8^c clearly exceeds all other models (ΔLME>78.3). Bottom: Illustration of model probability for all individual subjects. The posterior model probabilities for all subjects are shown as black dots for all models individually. In white shading are models with all parameters free, grey bars models with restricted parameters. Note that nearly all subjects show high model probabilities for SERIA models with restricted parameters.

In addition to the initially hypothesized models, we performed an additional unplanned, post hoc analysis on a refinement of the constrained SERIA family of models, in which we fixed the probability of an early antisaccade to a small number ($\pi_e = 1 - 0.005 \approx 1 - e^{-5}$). Hence, this family of models had 11 free parameters. The relative LME is displayed in Fig 6. We found that the most restrictive model was favored ($r = 88\%$) when compared to the original ($r = 5\%$) and constrained models ($r = 7\%$). When restricted to the models evaluated post hoc, there was very strong evidence in favor of m_8^c with

fixed π_e as compared to other models (ΔLME>133). If not otherwise stated, in the following we restrict the analysis to this model, which we denote as $\widetilde{m}_8^c$.

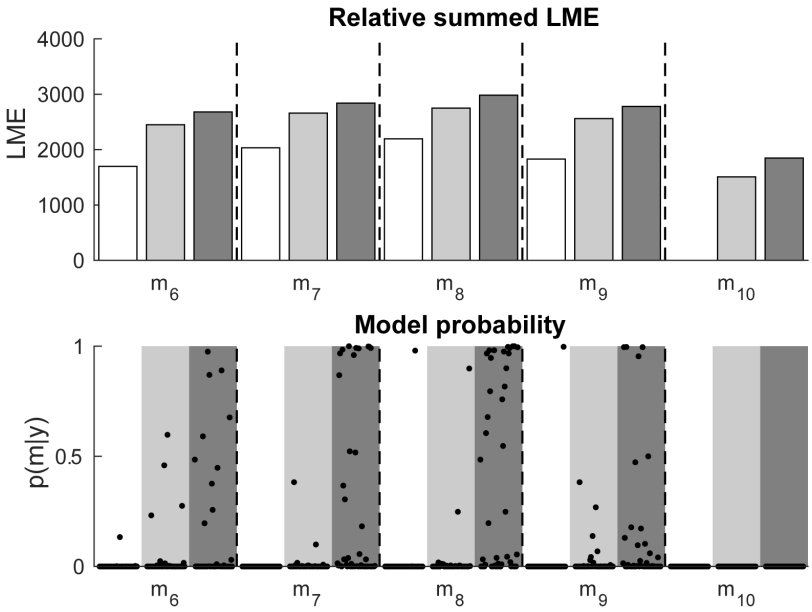

Figure 6: Post hoc model comparison between SERIA models.

Top: LME over all participants for the 5 SERIA models. White bars show models with all parameters free, light grey bars models with restricted parameters, and dark gray bars models with fixed prosaccade probability for early responses. LMEs are normalized by subtracting the lowest LME (m_{10}). Model $\widetilde{m}_8^c$ with fixed prosaccade probability for early responses clearly exceeds all other models (ΔLME>133). Bottom: Illustration of the subject wise model probability. The posterior model probability for all subjects are shown as black dots for all models individually. White shaded areas contain models with all parameters free, light grey areas models with restricted parameters and dark grey areas models with fixed early prosaccade probability. Note that nearly all subject show high model probabilities for SERIA models with restricted parameters.

Fits of four subjects using the maximum *a posteriori* (MAP) parameter estimates of the best PROSA model m_4 and the highest scoring model $\widetilde{m}_8^c$ (SERIA) are depicted in Fig 7 and Fig 8, respectively. Although model m_4 was the best model in the PROSA family, it clearly did not explain the apparent bimodality of the prosaccade RT distributions. Instead, RTs were explained through wider distributions. We further examined the model fits in Fig 9 by plotting the weighted fits collapsed across subjects. The histogram of RTs clearly shows a large number of late prosaccades whose distribution is similar to the distribution of the antisaccade RTs. Model $\widetilde{m}_8^c$ captures well the shape of these distributions

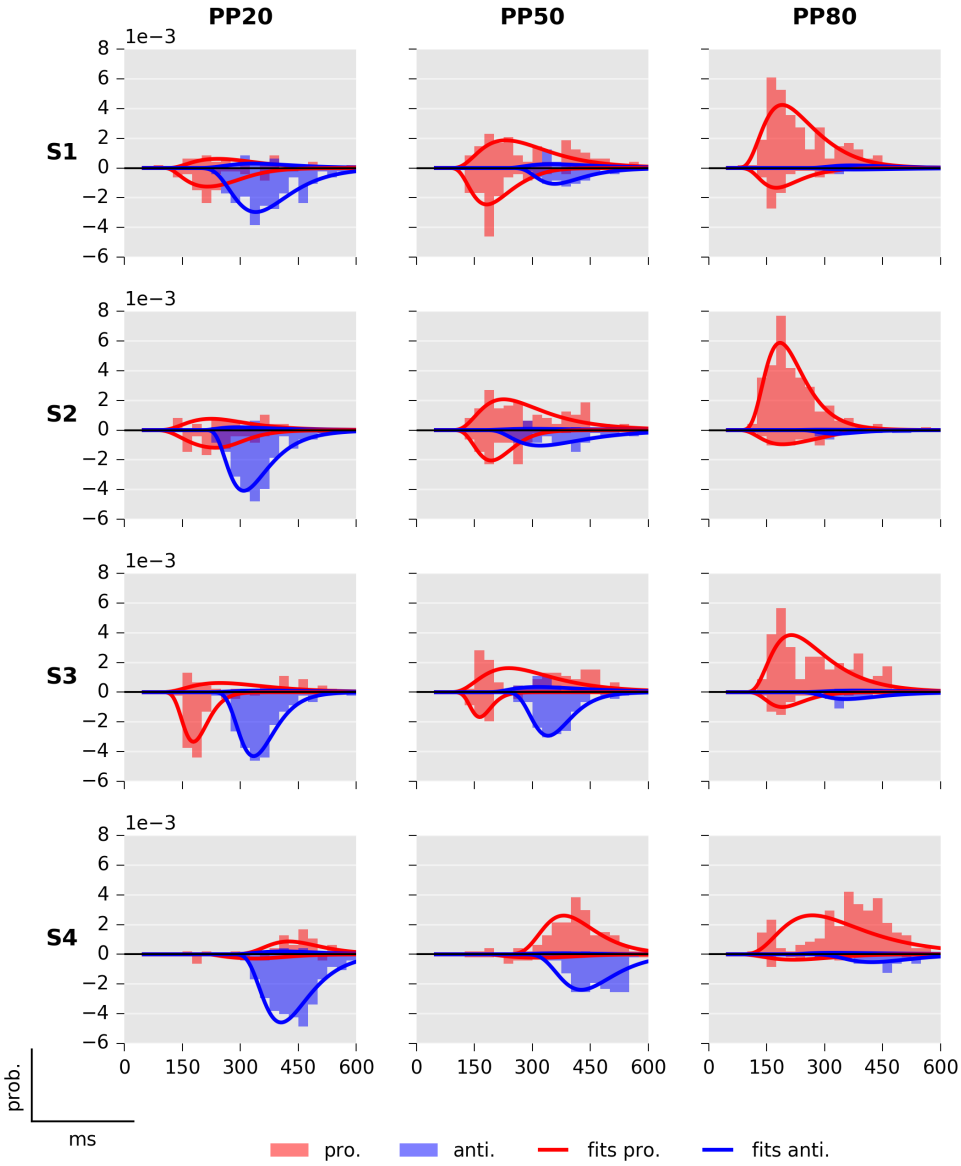

Fig 7. Fits of best PROSA model m_4.

Columns display the normalized histogram of the RTs of pro- (red) and antisaccades (blue) in each of the conditions. Rows correspond to individual subjects (named S1 to S4 for display purpose). Prosaccade trials are displayed on the upper plane, whereas antisaccade trials are displayed in the bottom plane. Thus, blue bars in the upper plane and red bars in the bottom plane indicate errors. The RT distributions based on the MAP estimates are displayed in red (prosaccades) and blue (antisaccades) lines.

453

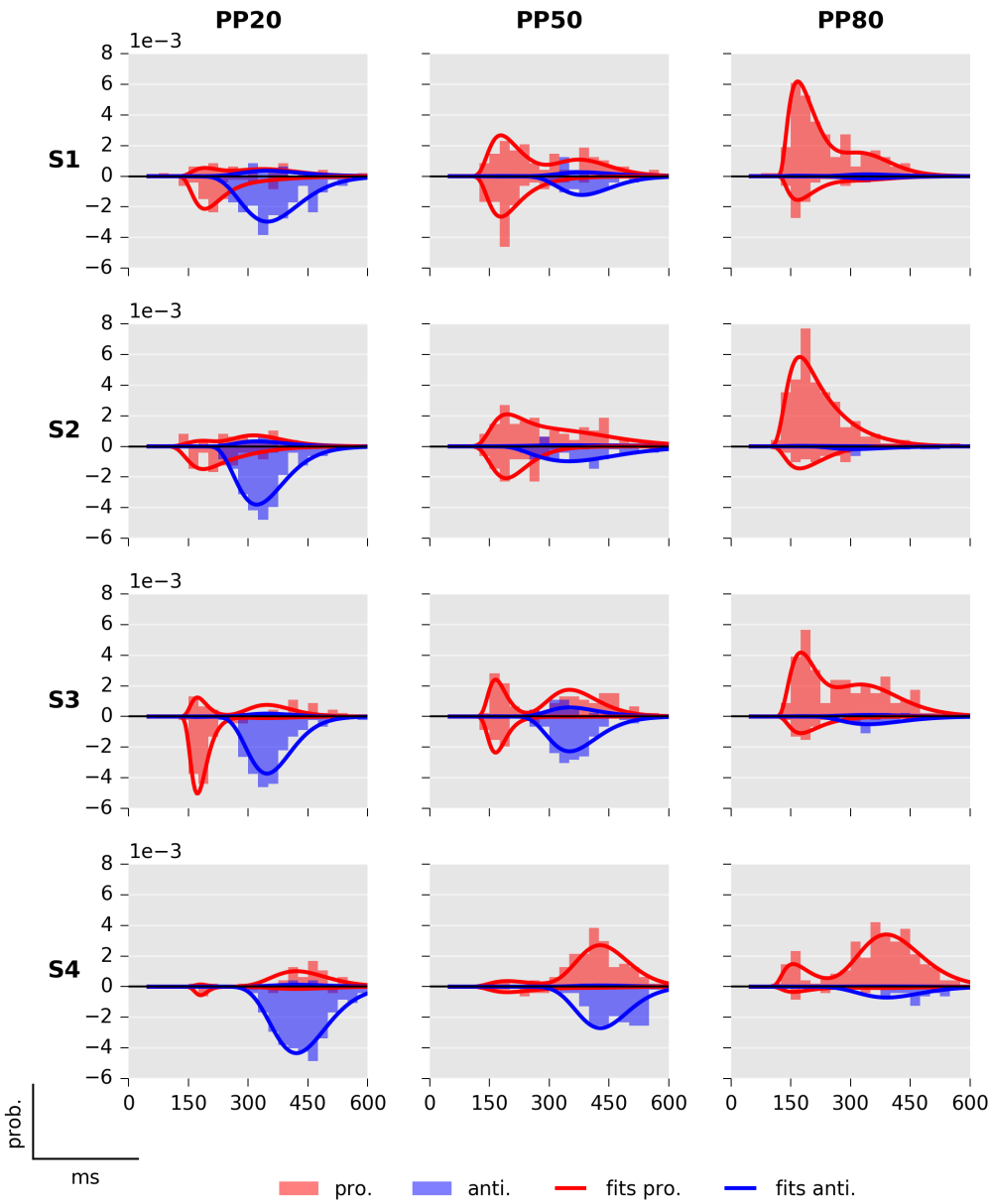

Fig 8. Fits of best SERIA model ($\widetilde{\mathrm{m}}_8^c$).

Fits of four subjects (same as in Fig. 7) using the best scoring model of the SERIA family ($\widetilde{m}_8^c$), in which the parameters were fixed across trial types and the probability of early antisaccades was fixed to a small number. Fits are displayed for the three PP conditions. For more details see Fig 7.

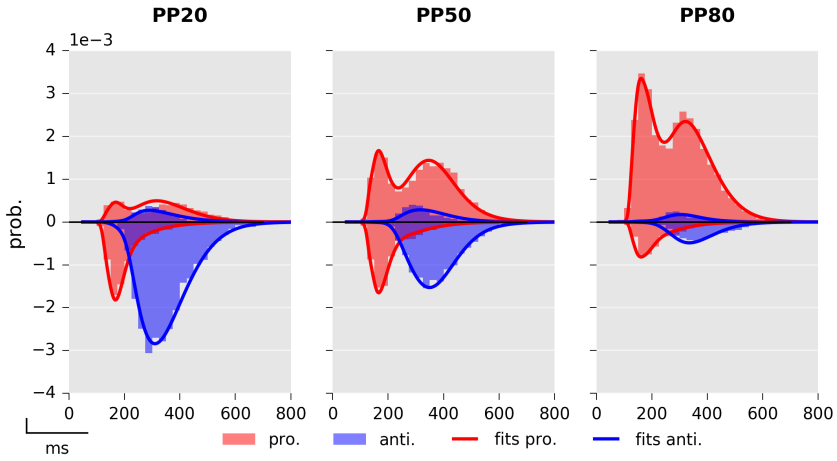

Fig 9. Fits using MAP estimates of the best model ($\widetilde{m}_8^c$).

The fits and RT histograms for each condition are collapsed over subjects. For more details see Fig. 7.

Corrective antisaccades

The RTs of antisaccades that follow an error prosaccade were not directly modeled. However, we hypothesized that corrective antisaccades are delayed late responses. A total of 2989 corrective antisaccades were included in the analysis. The mean (±std) end time of the erroneous prosaccades was 268(±63)ms. The mean RT of corrective antisaccades was 447(±103)ms, and the weighted mean arrival time of the late unit was 361ms. Fig 10 displays the histogram of the end time of all prosaccade errors, the RT of all corrective antisaccades and the time shifted (+86ms) predicted arrival times of the late unit. Since we did not have a strong hypothesis regarding the magnitude of the delay of the corrective antisaccades, we selected the time shift to be the difference between the empirical and predicted mean arrival time of the late unit. Visual inspection strongly suggests that the distribution of corrective antisaccade RTs is well approximated by the distribution of the late responses. Since the difference between corrective antisaccades' RT and the expected arrival time of the late is relatively short (86ms), this suggests that the plan for a corrective antisaccade was started before the initial incorrect prosaccade has finished.

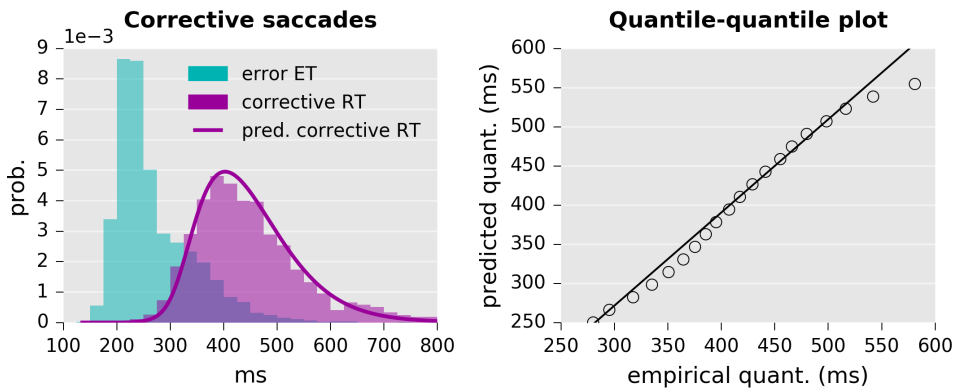

Fig 10. Empirical and predicted RT of corrective antisaccades.

Left: End time of erroneous prosaccades, corrective antisaccades RTs, and time shifted predicted arrival times distribution of the late unit. The time shift was selected to match the empirical and predicted mean RT. Right: Quantile-quantile plot of the empirical distribution of corrective antisaccades, predicted distribution, and a linear fit to the central 95% quantiles. There is a large deviation only at the tail of the distribution.

Effects of prosaccade probability on model parameters

The effect of PP on the parameters of the model was investigated by examining the MAP estimates of the best scoring model $\widetilde{m}_8^c$. Initially, we considered the question of whether the mean arrival time of each of the units changed as a function of PP. This corresponds to

$$E[U_i|k_{MAP}^i, \theta_{MAP}^i] + \delta_{MAP}^i \tag{26}$$

where i is an index over the units and δ_{MAP} is the estimated delay. Note that for model $\widetilde{m}_8^c$ this value can be analytically computed and is equal in pro- and antisaccade trials. Fig 11 left displays the mean arrival times of each of the units. The expected arrival times were submitted to three separate ANOVA tests, which revealed that PP had a significant effect on the late ($F_{2,138} = 13.3, p < 10^{-5}$), the inhibition ($F_{2,138} = 33.3, p < 10^{-5}$), and the early unit ($F_{2,138} = 3.1, p = 0.047$), although the effect on this unit was relatively weak. We then considered the differences across conditions through planned post hoc tests on each condition for each of the units (see Table 7). The arrival times of the early unit did not change significantly between condition PP20 and PP50, but decreased significantly in the PP80 condition as compared to the first two. The arrival times of the late unit increased significantly between the PP50 as compared to all other conditions, but there

was no significant difference in the PP20 and PP80 conditions. Regarding the inhibitory unit, we found that it significantly changed across all conditions.

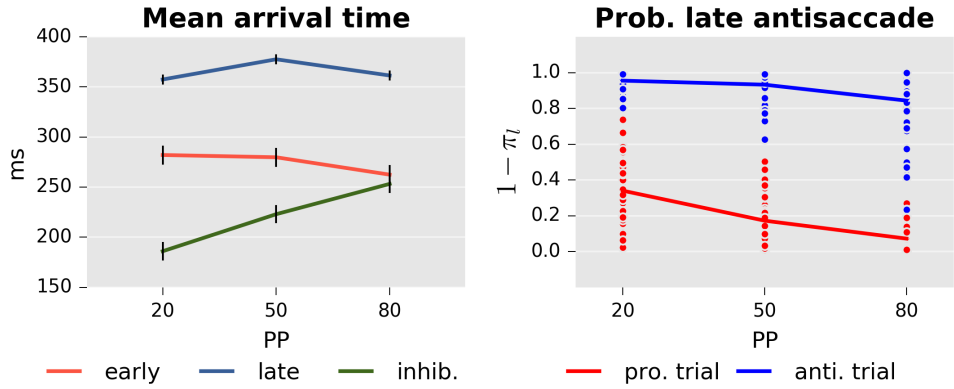

Fig 11. Model parameters.

Left: Mean arrival time and standard error of the late, early, and inhibition units. Right: Probability of a late antisaccade $1-\pi_l$ in prosaccade (red) and antisaccade (blue) trials in each condition.

Table 6. Post hoc comparison of the effect of PP.

	Early unit			Late unit			Inhib. unit		
Contrast	mean	t_{138}	p	mean	t_{138}	p	mean	t_{138}	p
PP80 -PP50	-17.4	-2.0	0.04	-16	-3.9	<0.01	30.2	3.6	<0.01
PP80 -PP20	-19.6	-2.2	0.02	4.0	0.96	0.33	67.2	8.1	<0.01
PP50 -PP20	-0.2	-0.2	0.79	20.1	4.75	<0.01	36.9	4.4	<0.01

Effect of PP on the mean arrival time for each of the units in ms.

The arrival times of the late unit showed a peak at PP50 condition suggesting an effect of the uncertainty associated with it. The uncertainty about trial type is highest in the PP50 condition, but equal in the two other conditions. To test this, we performed an unplanned analysis in which we entered PP as a linearly increasing regressor and included the Shannon entropy or uncertainty associated with each block as a further factor. The Shannon entropy is defined as

$$H = -PP \times \ln PP - (1-PP) \times \ln(1-PP). \tag{27}$$

Since in our initial analysis PP was entered as a categorical variable, this corresponds to a nested model with one fewer degree of freedom. Results are summarized in Table 8. While there was a significant effect of PP (but not UNCERTAINTY) on the early and inhibitory unit, there was a significant effect of UNCERTAINTY (but not PP) on the late unit.

Table 7. Effect of PP and UNCERTAINTY on the units' arrival time.

			Early unit		Late unit		Inhib. unit	
Contrast	DF 1	DF 2	*F*	*p*	*F*	*p*	*F*	*p*
PP	1	138	5.2	0.02	0.9	0.33	66.6	<0.01
UNCERTAINTY	1	138	1.0	0.31	25.8	<0.01	0.2	0.63

Finally, we examined how the probability of a late antisaccade $1 - \pi_l$ (Fig 11, right) depended on PP and TT. The estimated parameters for both pro- and antisaccade trials were analyzed with a model with factors SUBJECT, TT, PP and the interaction between TT and PP. An ANOVA test demonstrated that both PP ($F_{2,276} = 33.6, p < 10^{-4}$) and TT ($F_{1,276} = 658.7, p < 10^{-5}$) had a significant effect, but there was no evidence for an interaction between the two factors ($F_{2,276} = 0.8, p = 0.44$), suggesting that PP affected the probability of a late antisaccade similarly in pro- and antisaccade trials.

Discussion

In this study, we provided a formal treatment of error rates (ER) and reaction times (RT) in the antisaccade task using a probabilistic model. We applied the model to the data from an experiment with 3 mixed blocks with different probabilities of pro- and antisaccades trials. Model comparison showed that a novel model that allows for late pro- and antisaccades, explains our experimental findings better than a model in which all late responses are assumed to be antisaccades. The parameter estimates of the hidden units of the model showed that changes in the inhibition unit and changes in the probability of late prosaccades (π_l) explained most of the overt changes in behavior caused by our experimental manipulation, i.e., differences in PP. Moreover, we found that while inhibition was highly sensitive to the PP in a block, late responses were sensitive to the uncertainty associated with that block.

Influence of trial probability on reaction times and error rates

Our results show that both RT and ER depend on PP. While this was a highly significant factor in our study, there are mixed findings in previous reports. ER in antisaccade trials was found to be correlated with trial type probability in several studies [23,36,37]. However, this effect might depend on the exact implementation of the task [37,38]. Changes in prosaccade ER similar to our study have been reported by [23] and [39]. Studies in which the type of saccade was signaled at fixation prior to the presentation of the peripheral cue do not always show this effect [37]. The results on RTs are less consistent in the literature. Our findings of increased anti- and decreased prosaccade RTs with higher prosaccade trial probability are in line with the overall trend in [23] and with studies in which the cue was presented centrally [23,37]. Often, there is an additional increase in RT in the PP50 condition [23,37,1], which was visible in our data as a slight increase in RT in the PP50 condition on top of the linear effect of PP. Overall, RTs in our study were relatively slow compared to studies with the task cue separated from the spatial cue [36, 37, 39]. However, a study with a similar design and added visual search reported even slower RTs in both pro- and antisaccades [23].

Interpretation of model comparison results

The formal comparison of generative models can offer insight into the mechanisms underlying eye movement behavior [11] and might be relevant in translational

541 neuromodeling applications, such as computational psychiatry [41-45]. Here, we have
542 presented what is to our knowledge the first formal statistical comparison of models of
543 the antisaccade task. For this, we formalized the model introduced in [15] and proceeded
544 to develop a novel model that relaxes the one-to-one association of early and late
545 responses with pro- and antisaccades, respectively. All models and estimation techniques
546 presented here are openly available under the GPLv3.0 license as part of the open source
547 TAPAS package (www.translationalneuromodeling.org/tapas).

548 Bayesian model comparison yielded three conclusions at the family level. First, the SERIA
549 models were clearly favored when compared to the PROSA models. Second, models in
550 which race-to-threshold parameters were constrained to be equal across trial types had
551 a higher LME than models in which all parameters were free. Hence, the effect of the cue
552 in a single trial was limited to the probability of making a late prosaccade, and did not
553 directly affect the race-to-threshold process. Third, early responses were nearly always
554 prosaccades. Crucially, these three conclusions hold in a family comparison across all
555 parametric distribution of the increase rate of the units.

556 One less obvious but important consequence of our modeling findings is that the decision
557 to make a late pro- or antisaccade was not ruled by the same race process that governed
558 RTs. This follows from the main postulate of the SERIA model, namely, the conditional
559 independence of actions and RTs given response type (early or late). Thus, two
560 independent and qualitatively different decision processes lead to an antisaccade: the
561 race-to-threshold process between early and late responses, and the independent
562 decision process that generates different late responses (pro- vs antisaccades). A similar
563 separation of eye movement processes into a 'where' and 'when' component has been
564 proposed by [46], although mainly in conceptual terms.

565 *Parametric distribution of reaction times*

566 The parametric distribution of oculomotor RTs has been discussed in great detail in the
567 literature (e.g., [47,48]). Here, we did not aim at determining the most suitable
568 distribution, but rather opted for a practical approach by evaluating different models
569 with a reduced number of parametric distributions and based our conclusions on the
570 model with the highest LME. Nevertheless, one can consider the relationship of the
571 models presented here with other families of parametric distributions. In particular, the
572 linear relationship

$$\frac{s_i}{r_i} = t \tag{28}$$

seems to be inconsistent with the observation that RT are likely to be explained by stochastic accumulation processes (see for example [49, 50]). However, it can be shown that if RTs follow a generalized inverse normal distribution (GIN) of the form

$$GIN(t; \lambda, \kappa, \psi) = \frac{(\psi/\kappa)^{\lambda}}{2K_{\lambda}(\sqrt{\kappa\psi})} t^{\lambda-1} \exp\left(-\frac{1}{2}(\kappa t^{-1} + \psi t)\right) \tag{29}$$

where $\lambda \leq 0$, and K_{λ} is a modified Bessel function of the second kind, there exists a continuous diffusion process whose first hit distribution (FHD) follows the GIN [51]. A particular case of this distribution is the Wald distribution for which $\lambda = -\frac{1}{2}, \kappa = 0$. It is the FHD of the Brownian diffusion process with drift

$$X_t = -\sqrt{\sigma}\psi t + \sigma W_t \tag{30}$$

where W_t denotes a Wiener process, $x_0 > 0$, and the absorbing boundary a is zero. More relevant here, when $\psi = 0$ the distribution reduces to an inverse gamma distribution, the FHD of the process

$$X_t = \sqrt{\sigma}(2\lambda - 1)\, t^{-1} + \sigma W_t \tag{31}$$

with $x_0 > 0$ and boundary $a = 0$ (for a detailed mathematical treatment see [51]). Thus, if the rates of a ballistic, linear process are assumed to be gamma distributed, the RTs follow a distribution that is formally equivalent to a first hit model with stochastic updates and fixed rates. While the model presented here is a ballistic accumulation model, this equivalence suggests that it is *compatible* with a diffusion process with infinitesimal mean change proportional to t^{-1}.

Other antisaccade models

In broad terms, three families of antisaccade models can be distinguished. The first set of models is based on a race-to-threshold mechanism with independent saccadic and stop units. These models build on the seminal work by [13] on the stop-signal paradigm. According to these authors, a 'go' signal triggers a stochastic 'race' process that generates a response once it reaches threshold. Critically, a stop signal triggers a second process

595 that inhibits the first 'go' response if it is the first to reach threshold. Importantly, the pace
596 of both units is mutually independent. This model was further extended by [15], who
597 included a third unit such that an antisaccade is generated when a reflexive prosaccade
598 is inhibited by an endogenously triggered stop process. Note that the original 'horse-race'
599 model has also been modified [14] to account for different competing response actions,
600 similarly as in the antisaccade task. The SERIA model proposed here belongs to this
601 family.

602 A second type of model relies on lateral or mutual inhibition of competing pro- and
603 antisaccade units. In this direction, Cutsuridis and colleagues [52] proposed that lateral
604 inhibition is implement by inhibitory connections in the intermediate layers of the
605 superior colliculus. Thus, saccades are the result of accumulation processes, but these are
606 not independent of each other. Crucially, no veto-like stop signal is required. Although no
607 formal model fitting has been proposed for this model, qualitative agreement with data
608 suggests that it might capture behavioral patterns relevant in translational applications
609 [53, 54]. Since no probabilistic version of this model is available, it is currently not
610 possible to decide on the grounds of model comparison whether mutually dependent or
611 independent race processes best explain behavioral findings.

612 Finally, several models that incorporate detailed physiological mechanisms have been
613 proposed [17, 55-57]. These models cannot easily be assigned to one of the above
614 categories, as they often employ both an inhibitory mechanism that stops or withholds
615 the reactive responses as well as competition between actions. In addition, while more
616 realistic models possess a more fine-grained representation of the underlying
617 neurobiology, they rely on a large number of parameters and it is difficult to fit them to
618 behavioral data (for discussion, see [11]).

619 Regarding neurobiologically realistic models, the model proposed by [17] is the most
620 similar to the SERIA model. It posits two different mechanisms that interact in the
621 generation of antisaccades: an action selection module and a remapping module that
622 controls the cue-action mapping. As a consequence, this model allows for the generation
623 of late errors that follow a similar RT distribution as correct antisaccades. Consistent with
624 this observation, the SERIA model can quantitatively distinguish between inhibition and
625 decision (cue-action mapping) errors (Fig 12, left panel). A less obvious similarity
626 between the SERIA model and [17] is that different cues do not lead *directly* to different

dynamics in the action module, but only in the so-called 'remapping' module. Similarly, our model comparison results show that different cues (i.e., trial types) do not affect the race process but only the late cue-action mapping expressed in the parameter π_l.

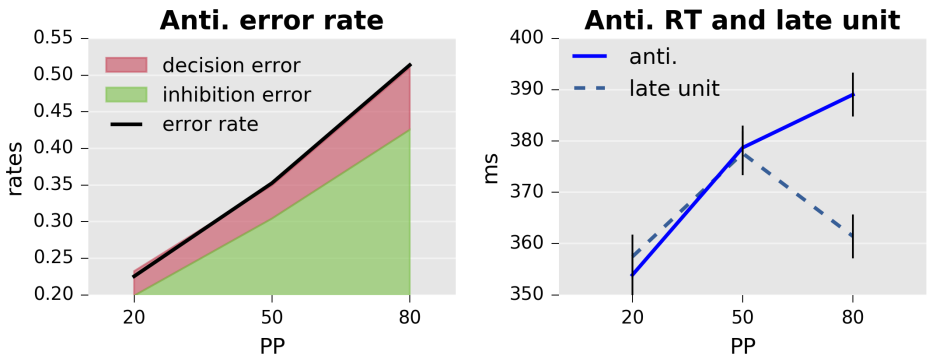

Fig 12. Error sources and late unit arrival time against antisaccade RT.

Left: Error rate (black line) split into the two causes predicted by the model. Inhibition errors are early actions that always trigger prosaccades. Similarly as described by [17], decision errors occur when a late response leads to a prosaccade. Right: Mean late unit arrival time and mean antisaccade RTs. Although mean RT increases with antisaccade probability this is due to slower inhibition, not to slower late responses. On the contrary, as uncertainty decreases, late responses are faster.

Parameter changes across trial types

One of the most salient results presented here is that models in which the parameters of the units were constrained to be equal across trial types had a larger LME than models in which all the parameters were free, suggesting that the race units were not affected by the cue presented on a single trial. However, while visual inspection of the predicted likelihood under the MAP parameters showed that most of the prominent characteristics of the data were explained correctly, some more subtle effects were not captured accurately by the model, for example, the distribution of late prosaccades in prosaccade trials in the PP20 and PP50 conditions. One possible explanation is that restricting the parameters across trial types made the model too rigid to capture this effect. Fig 13 compares the fitted RT distributions for models m_8 and $\widetilde{m}_8^c$. Although removing the constraint on the parameters did improve the fit, the differences are marginal and, thus, did not justify the additional model complexity. One might suspect that the distribution of late prosaccades was influenced by factors not included in the model such as unidirectional switch costs [58] that would be more prominent in the PP20 and PP50 conditions. Nevertheless, the differences in LME strongly suggest that the cue presented

on a given trial had only a marginal effect on the putative race processes that generates early and late responses. In fact, this example illustrates the protection against overfitting provided by the LME, as this is a case in which simpler models were preferred over more complex models despite of slightly less accurate fits.

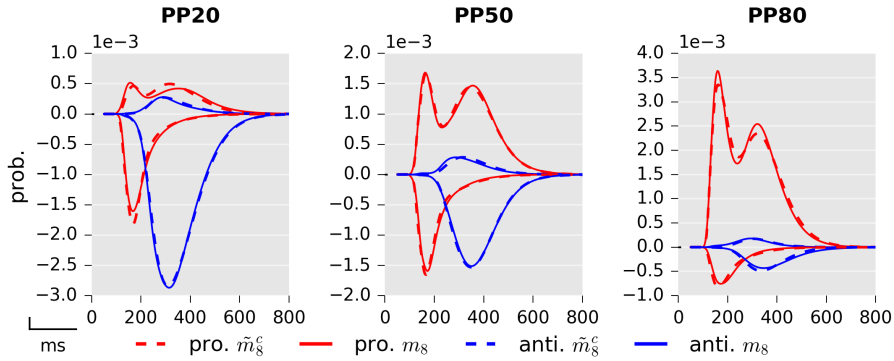

Fig 13: Comparison between constrained and unconstrained models.

Comparison between models m_8 (solid lines; all unit parameters are free) and $\tilde{m}_8^c$ (broken lines; unit parameters are equal across trial types). Only minor differences were observed, mainly in the PP20 condition.

The effect of trial type probability

It is far from obvious why TT probability affects RT and ER in the antisaccade task. One possible explanation is that increased probability leads to higher preparedness for either pro- or antisaccades. Such a theory posits an intrinsic trade-off between preparations for one of the two action types that leads to higher RTs and ERs in low probability trials. Thus, a trade-off theory predicts that the arrival times of early and late responses should be anticorrelated. Although this hypothesis can explain our behavioral findings in terms of summary statistics, our model suggests a more complicated picture.

The main explanation of our results is the effect of TT probability on the inhibitory unit and the probability of a late prosaccade. A higher probability of antisaccade trials leaded to faster inhibition and to a higher number of late prosaccades. This resulted in higher mean RT in prosaccade trials when PP is low. In the case of antisaccades, although the mean arrival times of the late unit increased in the conditions with highest uncertainty (Fig 12 right panel), the increased arrival time of the inhibitory unit on the PP80 condition skewed the antisaccade distribution towards higher RTs.

Regarding possible neural correlates of the effect of uncertainty in the responses of the late unit, a recent study [39] investigated the changes in BOLD signal in a task design

similar to ours, in which subjects performed pro- and antisaccades in mixed blocks with PP of either 25, 50, or 75 percent. When examining the interaction between the TT and PP factors, four clusters showed a significant activation: precuneus/right middle occipital gyrus, medial superior frontal gyrus, fusiform gyrus, and right inferior/middle frontal gyrus. Post hoc analysis revealed that in the prosaccade trials, these areas showed an increased activation with prosaccade probability, while there were no significant changes in antisaccade trials. Visual inspection ([39], Fig. 4) suggests that the pattern of activation change in the antisaccade trials resembled the uncertainty function that characterizes the arrival time of the late responses in our data. Unfortunately, the authors did not test for the effect of uncertainty, and thus, we can only speculate that these areas might be involved in the generation of late responses.

Corrective antisaccades

Although not a primary goal of our model, we considered the question of predicting corrective antisaccades. This problem has received some attention recently [16], as more sophisticated models of the antisaccade task have been developed. A natural hypothesis is that the distribution of these RTs should be similar to the distribution of the late responses. We speculated that these are generated by the very same mechanism that triggers late responses. Our results strongly suggest that this is the case, as suggested by visual examination (see Fig 10). The time delay of the corrective antisaccades indicates that, on average, corrective antisaccades are not the result of the late unit being restarted at the end time of the erroneous prosaccade, as this would lead to much higher RTs. Rather, the planning of a corrective antisaccade might be started much before the end of the execution of an erroneous prosaccade.

Summary

Here we have presented a novel model of the antisaccade task. While the basic structure of the model follows the layout of a previous model [15,16], we have introduced two crucial advancements. First, we postulated that late responses can trigger both pro- and antisaccades, which are selected by an independent decision process. Interestingly, a recent neural network model [17] introduced a comparable solution based on attractor network dynamics that can yield late erroneous prosaccades. Second, the generative nature of our model allows for Bayesian model inversion, which enables the comparison of different models and families of models on formal grounds. To our knowledge this has

not been done for any of the previous models of the antisaccade task. This is of relevance for translational applications that aim at better understanding psychiatric diseases by means of computational modeling.

The application of the model to a large data set yielded several novel results. First, the race process triggered by different cues is almost identical. Moreover, different PP had very different effects on the individual units, which was not obvious from the linear analysis of the mean RT and ER. In particular, late responses are mostly affected by uncertainty but not by PP. Crucially, our modeling allowed us to look at a mechanistic explanation or the effects of PP by examining the individual race units. In future work we aim to disentangle the mechanisms of behavioral differences caused by different drugs and psychiatric illnesses using formal Bayesian inference.

Acknowledgments

We thank Saee Paliwal for comments on the manuscript.

714 **References**

715 [1] Hallett PE. Primary and secondary saccades to goals defined by instructions. Vision
716 Res 1978;18:1279–96.

717 [2] Munoz DP, Everling S. Look away: the anti-saccade task and the voluntary control of
718 eye movement. Nat Rev Neurosci 2004;5:218–28.

719 [3] Hutton SB, Ettinger U. The antisaccade task as a research tool in psychopathology: a
720 critical review. Psychophysiology 2006;43:302–13.

721 [4] Fukushima J, Fukushima K, Chiba T, Tanaka S, Yamashita I, Kato M. Disturbances of
722 voluntary control of saccadic eye movements in schizophrenic patients. Biol Psychiatry
723 1988;23:670–7.

724 [5] Curtis CE, Calkins ME, Grove WM, Feil KJ, Iacono WG. Saccadic disinhibition in
725 patients with acute and remitted schizophrenia and their first-degree biological
726 relatives. Am J Psychiatry 2001;158:100–6.

727 [6] Harris MS, Reilly JL, Keshavan MS, Sweeney JA. Longitudinal studies of antisaccades
728 in antipsychotic-naive first-episode schizophrenia. Psychol Med 2006;36:485–94.

729 [7] Reilly JL, Frankovich K, Hill S, Gershon ES, Keefe RS, Keshavan MS, et al. Elevated
730 antisaccade error rate as an intermediate phenotype for psychosis across diagnostic
731 categories. Schizophr Bull 2014;40:1011–21.

732 [8] Radant AD, Millard SP, Braff DL, Calkins ME, Dobie DJ, Freedman R, et al. Robust
733 differences in antisaccade performance exist between COGS schizophrenia cases and
734 controls regardless of recruitment strategies. Schizophr Res 2015;163:47–52.

735 [9] Crawford TJ, Sharma T, Puri BK, Murray RM, Berridge DM, Lewis SW. Saccadic eye
736 movements in families multiply affected with schizophrenia: the Maudsley Family
737 Study. Am J Psychiatry 1998;155:1703–10.

738 [10] Radant AD, Dobie DJ, Calkins ME, Olincy A, Braff DL, Cadenhead KS, et al.
739 Antisaccade performance in schizophrenia patients, their first-degree biological
740 relatives, and community comparison subjects: data from the COGS study.
741 Psychophysiology 2010;47:846–56.

742 [11] Heinzle J, Aponte EA, Stephan KE. Computational models of eye movements and
743 their application to schizophrenia. Current Opinion in Behavioral Sciences 2016;11:21–
744 9. doi:http://dx.doi.org/10.1016/j.cobeha.2016.03.008.

745 [12] Huys QJ, Maia TV, Frank MJ. Computational psychiatry as a bridge from
746 neuroscience to clinical applications. Nat Neurosci 2016;19:404–13.

747 [13] Logan GD, Cowan WB, Davis KA. On the ability to inhibit simple and choice reaction
748 time responses: a model and a method. J Exp Psychol Hum Percept Perform
749 1984;10:276–91.

750 [14] Logan GD, Van Zandt T, Verbruggen F, Wagenmakers EJ. On the ability to inhibit
751 thought and action: general and special theories of an act of control. Psychol Rev
752 2014;121:66–95.

[15] Noorani I, Carpenter RH. Antisaccades as decisions: LATER model predicts latency distributions and error responses. Eur J Neurosci 2013;37:330–8.

[16] Noorani I, Carpenter RH. Re-starting a neural race: anti-saccade correction. Eur J Neurosci 2014;39:159–64.

[17] Lo CC, Wang XJ. Conflict Resolution as Near-Threshold Decision-Making: A Spiking Neural Circuit Model with Two-Stage Competition for Antisaccadic Task. PLoS Comput Biol 2016;12:e1005081.

[18] Carpenter RH, Williams ML. Neural computation of log likelihood in control of saccadic eye movements. Nature 1995;377:59–62.

[19] Gorea A, Rider D, Yang Q. A unified comparison of stimulus-driven, endogenous mandatory and 'free choice' saccades. PLoS ONE 2014;9:e88990.

[20] Edelman JA, Valenzuela N, Barton JJ. Antisaccade velocity, but not latency, results from a lack of saccade visual guidance. Vision Res 2006;46:1411–21.

[21] Zhang M, Barash S. Neuronal switching of sensorimotor transformations for antisaccades. Nature 2000;408:971–5.

[22] Sato TR, Schall JD. Effects of stimulus-response compatibility on neural selection in frontal eye field. Neuron 2003;38:637–48.

[23] Chiau HY, Tseng P, Su JH, Tzeng OJ, Hung DL, Muggleton NG, et al. Trial type probability modulates the cost of antisaccades. J Neurophysiol 2011;106:515–26.

[24] Stampe D. Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. Behavior Research Methods, Instruments, & Computers 1993;25:137–42. doi:10.3758/BF03204486.

[25] Robert C, Casella G. Monte carlo statistical methods. Springer Science & Business Media; 2013.

[26] Shaby E, Well M. Exploring and adaptative Metropolis algorithm. Dep. of Statistical Sciences Duke University; 2011.

[27] Gelman A, Carlin JB, Stern HS , Rubin DB. Bayesian data analysis. Chapman; Hall/CRC; 2003.

[28] Aponte EA, Raman S, Sengupta B, Penny WD, Stephan KE, Heinzle J. mpdcm: A toolbox for massively parallel dynamic causal modeling. J Neurosci Methods 2016;257:7–16.

[29] Calderhead B, Girolami MA. Estimating bayes factors via thermodynamic integration and population MCMC. Computational Statistics & Data Analysis 2009;53:4028–45. doi:10.1016/j.csda.2009.07.025.

[30] MacKay DJC. Information theory, inference, and learning algorithms. Cambridge University Press; 2003.

[31] Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. Bayesian model selection for group studies. Neuroimage 2009;46:1004–17.

[32] Rigoux L, Stephan KE, Friston KJ, Daunizeau J. Bayesian model selection for group studies - revisited. Neuroimage 2014;84:971–85.

[33] Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, et al. Comparing families of dynamic causal models. PLoS Comput Biol 2010;6:e1000709.

[34] Patterson TNL. The optimum addition of points to quadrature formulae. Mathematics of Computation 1968;22:847–56.

[35] Kass RE, Raftery AE. Bayes factors. Journal of the American Statistical Association 1995;90:773–95. doi:10.2307/2291091.

[36] Massen C. Parallel programming of exogenous and endogenous components in the antisaccade task. Q J Exp Psychol A 2004;57:475–98.

[37] Pierce JE, McDowell JE. Effects of preparation time and trial type probability on performance of anti- and pro-saccades. Acta Psychol (Amst) 2016;164:188–94

[38] Pierce JE, McDowell JE. Modulation of cognitive control levels via manipulation of saccade trial-type probability assessed with event-related BOLD fMRI. J Neurophysiol 2016;115:763–72.

[39] Pierce JE, McDowell JE. Effects of preparation time and trial type probability on performance of anti- and pro-saccades. Acta Psychol (Amst) 2016;164:188–94.

[40] Cornwell BR, Mueller SC, Kaplan R, Grillon C, Ernst M. Anxiety, a benefit and detriment to cognition: behavioral and magnetoencephalographic evidence from a mixed-saccade task. Brain Cogn 2012;78:257–67.

[41] Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. Trends Cogn Sci (Regul Ed) 2012;16:72–80.

[42] Wang XJ, Krystal JH. Computational psychiatry. Neuron 2014;84:638–54.

[43] Stephan KE, Mathys C. Computational approaches to psychiatry. Curr Opin Neurobiol 2014;25:85–92.

[44] Paulus MP, Huys QJ, Maia TV. A roadmap for the development of applied computational psychiatry. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 2016.

[45] Huys QJ, Maia TV, Paulus MP. Computational psychiatry: From mechanistic insights to the development of new treatments. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 2016;1:382–5.

[46] Findlay JM, Walker R. A model of saccade generation based on parallel processing and competitive inhibition. Behav Brain Sci 1999;22:661–74.

[47] Feng G. Is there a common control mechanism for anti-saccades and reading eye movements? Evidence from distributional analyses. Vision Res 2012;57:35–50.

[48] Palmer EM, Horowitz TS, Torralba A, Wolfe JM. What are the shapes of response time distributions in visual search? J Exp Psychol Hum Percept Perform 2011;37:58–71.

[49] Gold JI, Shadlen MN. The neural basis of decision making. Annu Rev Neurosci 2007;30:535–74.

[50] Ratcliff R, Smith PL, Brown SD, McKoon G. Diffusion Decision Model: Current Issues and History. Trends Cogn Sci (Regul Ed) 2016;20:260–81.

[51] Barndorff-Nielsen O, BlÆsild P, Halgreen C. First hitting time models for the generalized inverse gaussian distribution. Stochastic Processes and Their Applications 1978;7:49–54.

[52] Cutsuridis V, Smyrnis N, Evdokimidis I, Perantonis S. A neural network model of decision making in an antisaccade task by the superior colliculus. Neural Networks 2007;20:690–704.

[53] Cutsuridis V, Kumari V, Ettinger U. Antisaccade performance in schizophrenia: a neural model of decision making in the superior colliculus. Front Neurosci 2014;8:13.

[54] Cutsuridis V. Neural competition via lateral inhibition between decision processes and not a STOP signal accounts for the antisaccade performance in healthy and schizophrenia subjects. Front Neurosci 2015;9:5.

[55] Brown JW, Bullock D, Grossberg S. How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. Neural Netw 2004;17:471–510.

[56] Heinzle J, Hepp K, Martin KA. A microcircuit model of the frontal eye fields. J Neurosci 2007;27:9341–53.

[57] Wiecki TV, Frank MJ. A computational model of inhibitory control in frontal cortex and basal ganglia. Psychol Rev 2013;120:329–55.

[58] Weiler J, Heath M. Task-switching in oculomotor control: unidirectional switch-cost when alternating between pro- and antisaccades. Neurosci Lett 2012;530:150–4.

Supporting Information

S1_Dataset.zip. Table of Data. Spreadsheet including all reaction times, actions and errors that entered the analysis. More details are included directly in the file.