

# ELF: A new measure of response capture

Mathieu Servant<sup>1</sup> · Thibault Gajdos<sup>2</sup>  · Karen Davranche<sup>2</sup>

Published online: 3 November 2017  
© Psychonomic Society, Inc. 2017

**Abstract** Response capture is a widespread and extensively studied phenomenon, in particular in decision tasks involving response conflict. Its intensity is routinely quantified by conditional accuracy function (CAF). We argue that this method might be misleading, and propose an alternative approach, the error location function (ELF). While CAF provides the error rate by bins of reaction time (RT), ELF represents the share of total errors below each quantile of RT. We derive from ELF an index of response capture, the error location index (ELI), which represents the area below the ELF. Using simulations of computational models, we show that ELF and ELI specifically quantify variations in response capture. Finally, we illustrate the usefulness of ELF and ELI through experimental data and show that ELF and CAF can yield to contradictory conclusions.

**Keywords** Computational models · Reaction time analysis · Response time models · Cognitive control and automaticity

## Introduction

One sometimes makes fast decisions before carefully processing the decision at hand, either on the basis of a strong

prior belief, or under the influence of a very salient, though irrelevant, stimulus. Such choices typically generate fast errors, because the decision is not based on the relevant information. Instead, the decision is somehow *captured* by irrelevant cues or beliefs, which yield to refer to such situations as *response captures* (van den Wildenberg et al. 2010).

Experimentally, response capture has been extensively studied in reaction time (RT) paradigms where participants face conflicting stimuli: a relevant one, which determines the answer they should provide, and some irrelevant but salient ones that trigger an automatic response (Stroop, 1935; Eriksen & Eriksen, 1974; Simon, 1990). For instance, in the Simon task, the participants are instructed to respond as fast and as accurately as possible by pressing on a left- or a right-hand key according to a non-spatial attribute of a stimulus presented either on the left or on the right of a fixation point. Performance expressed both in terms of error rate and mean RT is better when the required response corresponds spatially to the irrelevant stimulus location (ipsilateral associations) than when it does not correspond (contralateral associations). However, it is important to note that response captures are not limited to conflict tasks. Fast errors are also observed in simple choice situations where individuals should answer under time pressure (Luce, 1986), and a substantial effort has been made to account for them in computational models (Ratcliff & McKoon, 2008).

Because of its ubiquity, it is important to have a behavioral measure of the intensity of response capture that does not depend on a specific cognitive model. It has been argued that response capture could be measured by the proportion of errors among fast responses (Ridderinkhof, 2002; van den Wildenberg et al. 2010). Accordingly, the conditional accuracy function (CAF), which gives that accuracy rate by bins (bounded by quantiles) of RT (Luce, 1986), is routinely used to measure response capture (Gratton et al. 1988; Ulrich

---

Mathieu Servant, Thibault Gajdos and Karen Davranche contributed equally to this work.

✉ Thibault Gajdos  
thibault.gajdos@univ-amu.fr

<sup>1</sup> Department of Psychological Sciences, Vanderbilt University, Nashville, TN, USA

<sup>2</sup> Aix Marseille University, CNRS, LPC, Marseille, France

et al. 2015; White, Brown, & Ratcliff, 2012).<sup>1</sup> More precisely, one usually use either the slope of the first segment of the CAF (Ridderinkhof, 2002) or its first point, which represents the accuracy rate in the first bin of RT (Wylie et al. 2010, 2012; van Wouwe et al. 2016).

In this article, we propose an alternative measure of response capture. The key idea is to distinguish between the two following statements: (i) “the error rate among the  $p\%$  fastest responses are  $x\%$ ” and (ii) “ $y\%$  of the errors are contained in the  $p\%$  fastest responses”. The first statement is about the accuracy rate below the  $p$ -quantile of RT, and this information is provided by CAF. The second one is about the proportion of errors below the first  $p$ -quantile of RT, and cannot be deduced from CAF. We argue that response capture is better described by the second statement than by the first. This yields us to propose an alternative to the CAF, which we call the error location function (ELF). Importantly, ELF allows defining a very natural index of response capture (error location index, ELI). In the next section, we introduce ELF and ELI in detail. Then we illustrate their significance through numerical simulation of computational decision models. Finally, we illustrate the usefulness of ELF and ELI through experimental data and show that ELF and CAF can yield to contradictory conclusions.

### Error location function (ELF) and error location index (ELI)

We illustrate the logic behind the ELF through a distribution analysis of RT in a conflict task such as the Simon task. However, the following arguments could be easily generalized to other situations. Consider the case of incompatible trials in a Simon task. In such trials, a response capture would be a fast response triggered by the irrelevant stimulus location. On the other hand, slow errors would reflect some other cognitive process, for instance perceptual errors. In this context, a stronger response capture should increase the number of fast errors, without impacting slow errors. Thus, the intensity of response capture is characterized by the proportion of errors contained in the first bin of RT. However, this is not what CAF measures: instead, it provides a measure of the accuracy rate by bin of RT.

The distinction between these two approaches is illustrated by the following example. Assume that a subject performs 100 trials of a task, in three different conditions, A, B, and C. Let the RT be divided into five bins. Assume that in condition A, one observes four errors in the first bin, and two errors in each other ones, whereas one observes eight errors in the first bin, and four errors in each other ones in

condition B. Finally, in condition C, the subject only makes two errors in the first bin. Such a situation would yield to the CAF curves reported in the left panel of Fig. 1. The slope of the first segment of the CAF curve is steeper in condition B than in condition A. Thus, on this basis, one could be tempted to conclude that response capture is higher in condition B than in condition A. On the other hand, these slopes are exactly the same in conditions A and C. Since the CAF curve corresponding to condition C is above that of condition A, one might be tempted to conclude that response capture is lower in condition C than in condition A.

However, a closer look leads to question this conclusion. While it is true that the subject makes twice as much errors in condition B than in condition A, the error RTs distribution is exactly the same in both conditions: namely, 1/3 of the errors is located in the first bin, and 1/6 in each other bins. In other words, what we observe when comparing conditions A and B is a homogeneous increase of all errors, fast and slow; but, strictly speaking, a stronger response capture should only increase the number of fast errors, thereby increasing the proportion of fast answers among errors. This is exactly what happens in condition C: there are only two errors, but both are contained in the first RT bin. One could thus suspect that response capture is stronger in condition C than in conditions A and B, while the cognitive mechanisms underlying the difference between conditions A and B are not, or at least not exclusively, related to response capture.

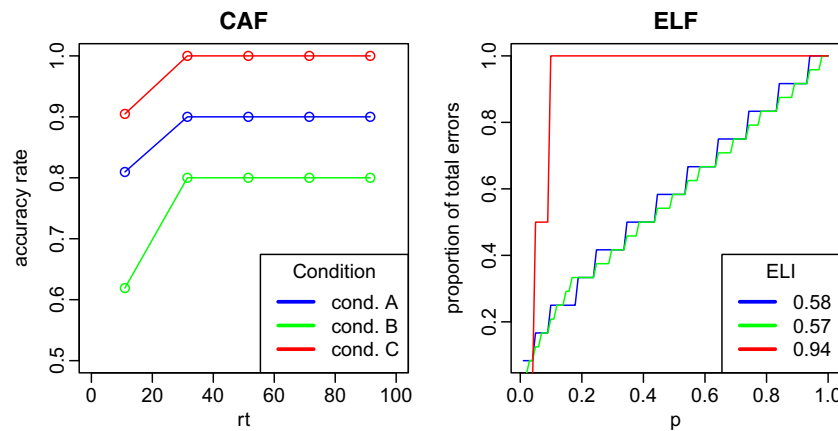
This demonstrates the need to design a tool that specifically measures the intensity of response capture. We suggest to do so by taking literally the idea that response capture can be characterized by the proportion of fast answers among errors. A natural way to implement this idea would be to build relative CAF (RCAF) instead of CAF, by replacing the error rate within each bin by the proportion of all errors that are contained in each bin.<sup>2</sup> However, RCAF still rely on a discretization of the data. The number of bins is somewhat arbitrary, and can influence the outcome. Moreover, it is desirable to get an index of response capture, and it is not obvious how it could be derived from RCAF. This yields us to introduce the error location function (ELF). The ELF represents the proportion of errors located below each quantile  $p$  of the overall RT distribution. Formally, let  $F$  and  $G$  denote the cumulative distribution function of all trials and errors, respectively. Let  $F^{-1}$  be the quantile function of  $F$ , defined as  $F^{-1}(p) = \inf \{x \in \mathbb{R} | F(x) \geq p\}$ . The ELF function is defined by:

$$ELF(F, G)(p) = G(F^{-1}(p)), \quad p \in [0, 1]. \quad (1)$$

Note that the ELF is increasing, and satisfies  $ELF(F, G)(0) = 0$  and  $ELF(F, G)(1) = 1$ . The basic idea of ELF is thus to use the RT distribution of all trials

<sup>1</sup>CAF thus plot the accuracy rate within each bin of RT against the mean RT of each bin.

<sup>2</sup>We thank Andrew Heathcote for this suggestion.



**Fig. 1** CAF curves (*left panel*) corresponding to the three hypothetical conditions A, B, and C. The slope of the first segment is steeper in condition B than in condition A, and is identical in conditions A and C. However, error RTs distributions are exactly the same in conditions

A and B, while all errors are concentrated in the first bin in condition C. The ELF curves (*right panel*) are identical in conditions A and B, but that of condition C is higher. See the text for a detailed description of the *ELI* index reported in right panel

(irrespective of the accuracy) as a reference against the RT distribution of errors. In other words, what ELF measures is the relative RT distribution of errors with respect to overall RT. The choice of this reference is somewhat arbitrary, and was made to stick as closely as possible to the logic behind CAFs. Another natural possibility would be to use the RT distribution of correct trials. This would yield to replace  $F^{-1}$  by the quantile function of correct RTs. All what follows can easily be adapted to this alternative definition. Admittedly, this choice is largely a matter of taste. This being said, both definitions yield to very similar results and interpretations.

The graph of ELF (the ELF curve) has some interesting properties. First observe that it can equivalently be plotted in the  $(p, G(F^{-1}(p)))$  space (Fig. 2, left panel), or as a parametric curve in the  $(F(t), G(t))$  space (Fig. 2, right panel). In the first case, one reads on the vertical axis the proportion of errors that are faster than the  $p\%$  faster responses (correct and incorrect). In the second case, each point of the curve provides the proportion of overall (on the horizontal axis) and incorrect (on the vertical axis) responses faster than a given RT. This representation has some obvious similarities with the classical ROC curve.

The ELF curve coincides with the first diagonal if and only if  $G = F$ , i.e., error RTs are exactly distributed as RT of all trials. In this case, the corresponding CAF curve would be horizontal.<sup>3</sup> On the other hand, an ELF curve above the first diagonal indicates that most errors are concentrated among fast responses. Conversely, an ELF curve below the first diagonal indicates that most errors are

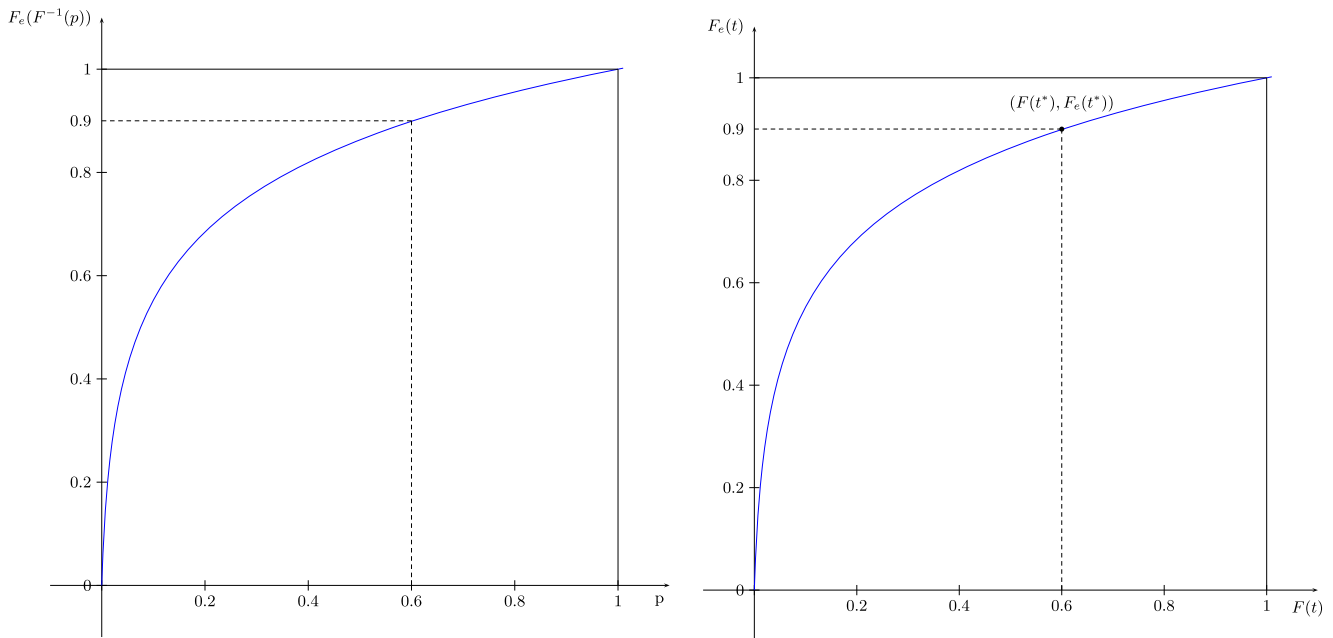
concentrated among slow responses. Thus, one would typically expect the ELF curve to be above the diagonal when subjects exhibit some degree of response capture, as in the case for incompatible trials in conflict tasks. More generally, a “higher” ELF curve indicates a higher concentration of errors among fast responses. The right panel of Fig. 1 represents the ELF curves for the three conditions A, B, and C. As expected, these curves are essentially identical for conditions A and B, reflecting that the distribution of errors is exactly the same in these two conditions. By contrast, the ELF curve corresponding to condition C is higher, which reflects the fact that errors are exclusively concentrated among fast RT, thereby indicating a strong response capture.

This suggests taking the area below the ELF curve as an index of the strength of response capture. This yields us to define the error location index (ELI) as:

$$ELI(F, G) = \int_0^1 G(F^{-1}(p))dp. \tag{2}$$

The ELI is easy to interpret, by observing that it can be rewritten as  $ELI(F, G) = \int_0^{+\infty} G(t)f(t)dt$ , where  $f$  is the density function of correct RTs. Thus, ELI can be interpreted as the expectation that a uniformly drawn incorrect response is faster than a uniformly drawn (correct or incorrect) trial. Thus, if  $ELI = 1$ , all errors are concentrated among the fastest trials, which corresponds to a very strong response capture. On the other hand, if  $ELI = 0$ , all errors are concentrated among the slowest trials, which is the converse of response capture. Finally, when errors RT are evenly distributed among all RT, there is no response capture and  $ELI = \frac{1}{2}$ . Observe that the converse is not true:  $ELI = \frac{1}{2}$  could correspond to an ELF curve that crosses the first diagonal. It would be the case, for instance, if errors are

<sup>3</sup>Note that, because of the discretization involved in the computation of CAF, the converse is not always true: CAF are invariant to changes of the distribution of error RTs within bins.



**Fig. 2** Example of an ELF curve. *Left panel:* the curve is represented in the  $(p, G(F^{-1}(p)))$  space. We have, for instance:  $ELF(F, G)(0.6) = 0.9$ . This means that 90% of the errors are concentrated among the 60% faster answers. *Right panel:* the curve is

represented in the  $(F(t), G(t))$  space. Each point of the curve gives the proportion of all responses (*horizontal axis*) and errors (*vertical axis*) faster than a given RT

exclusively concentrated among both fastest responses (due to response capture) and slowest ones (due to some other cognitive mechanism influencing error RTs). In such a case, interpreting ELF and ELI as measuring response capture might be misleading. This comes as the cost of the non-parametric nature of ELF and ELI: because they avoid any specific assumption about the mechanisms generating RTs, they cannot disentangle response capture from other factors influencing the distribution of error RTs. While, *ceteris paribus*, a stronger response capture necessarily generates a higher ELI, the converse is not true. As any other non-parametric measures, ELF and ELI should thus be used with caution, and related to other measures within a specific theoretical framework.

One convenient feature of the CAF function is that it allows for simple group summaries, by simple averaging across subjects. Fortunately, ELF has a similar property. Assume there are  $n$  subjects, and let  $F_i$  and  $G_i$  denote the cumulative distribution function of all trials and errors of subject  $i$ , respectively. By definition, subject  $i$ 's ELF curve is equal to  $ELF_i(p) = G_i(F_i^{-1}(p))$ . Let's define the ELF group function as

$$ELF_{group}(p) = \frac{1}{n} \sum_{i=1}^n ELF_i(p).$$

It is easy to plot  $ELF_{group}$ , as it is simply the mean of the individual  $ELF$  curves. Now let  $ELI_{group}$  be the area under the group ELF curve. We then have:

$$\begin{aligned} ELI_{group} &= \int_0^1 ELF_{group}(p) dp = \int_0^1 \frac{1}{n} \sum_{i=1}^n ELF_i(p) dp \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^1 ELF_i(p) dp = \frac{1}{n} \sum_{i=1}^n ELI_i. \end{aligned}$$

Thus, the group ELI, which is the area under the group ELF curve, is simply the mean of individual ELIs. We use this method in Section “[Application: color saturation manipulation in an Eriksen task](#)”.

R codes implementing ELF and ELI are provided in the [Appendix](#). It should be noted for empirical applications that, because ELF and ELI strongly rely on cumulative distribution functions of correct responses and errors, one needs to ensure that there are enough data for these empirical cumulative distributions to be meaningful. One way to cope with this issue consists of computing confidence intervals for ELI using a standard bootstrap procedure. One should refrain from using ELI if these confidence intervals are too large. We apply this method in Section “[Application: color saturation manipulation in an Eriksen task](#)” when illustrating ELF and ELI with actual data.

## Relation to computational models

We now illustrate the use of ELF and ELI through the simulation of computational models of decision making. The rationale behind this exercise is the following. Computational models are generally designed to reflect to some extent some cognitive processes. Their parameters can thus be interpreted as measuring some characteristics of cognitive processes. One can therefore use these models as benchmarks, and observe for instance how specifically ELF and ELI react to changes of parameters that are thought to influence response capture.

We start with the popular drift diffusion model (DDM, Ratcliff 1978). This model assumes that evidence in favor of a given option in a two-alternative forced choice is accumulated, and that a choice is made when accumulated evidence reaches a given threshold. Formally, this accumulation process is modeled by a stochastic diffusion process. The drift rate of this process ( $\mu_c$ ) depends on the quality of sensory information, subject's attention, and discrimination ability. The threshold ( $b$  for a correct choice, and  $-b$  for an incorrect one, conditional on a given alternative) depends on the speed–accuracy trade-off made by the subject. Finally, the starting point of the accumulation process (set here to zero) depends on the prior bias in favor of a choice alternative. On top of this process of evidence accumulation, the DDM includes a random non-decision time. We do not include this component into account for the sake of clarity. We simulate response data using the DDM model with three sets of parameters based on Matzke & Wagenmakers's (2009) survey of parameter values estimated in empirical studies (see Table 1). The model is simulated in seconds, with a diffusion coefficient equal to 1. The first set of parameters is our baseline condition. The second set corresponds to a “low threshold” condition, in which the upper threshold  $b$  is reduced from 0.5 to 0.4. An important property of the basic DDM model is that the RT distribution of errors and correct trials are identical. Therefore, this model cannot account for fast or slow errors. Laming (1968) showed that one can predict errors to be faster than correct responses by introducing inter-trial variability in the starting point of

the evidence accumulation process. We thus consider a third condition (“high variability”), adding inter-trial variability in the starting point of the evidence accumulation process (uniformly distributed with mean 0 and range 1; see Ratcliff and Rouder (1998)). The CAF, ELF curves, and ELI indices corresponding to these three conditions are shown in Fig. 3.

We can observe that in both the high variability and low threshold conditions, the CAF curves are different from the baseline. Focusing on the first point of the CAF curves, one might be tempted to conclude that response capture is stronger in both the high variability and low threshold conditions than in the baseline condition. However, we know that RT distributions of errors and correct responses are actually the same in the low threshold condition. The ELF curves, on the other hand, tell a totally different story: they are identical in the baseline and low threshold conditions, and higher in the high variability condition. These observations are confirmed by the values of the ELI, reported in the right panel of Fig. 3. This is of course consistent with the fact that the variability of the starting point of the DDM (and not the threshold of the decision rule) generates faster errors than correct responses. Incidentally, this analysis suggests that the ELI might be a behavioral counterpart of the variability of the starting point of the DDM.

While very versatile, it has been demonstrated that the standard DDM cannot account for the pattern of data observed in conflict tasks, namely faster errors than correct responses in incompatible trials only (e.g., Hübner, Steinhäuser, and Lehle, 2010; White, Ratcliff, and Starns, 2011). Recent efforts have been made to fill this gap. The diffusion model for conflict tasks (DMC) proposed by Ulrich et al. (2015) is particularly interesting, insofar as it is a natural extension of the DDM, with parameters that can naturally be related to response capture and control. In a nutshell, the DMC is an extension of the DDM, where the rate of evidence accumulation is allowed to vary with time. More precisely, the decision process is assumed to be the sum of two diffusion processes: a controlled process, with a constant drift ( $\mu_c$ ), and an automatic process, with a drift that first increases, and then progressively vanishes. Formally, dynamics of the variable drift rate are described by a rescaled Gamma density function, with a peak amplitude  $A$ , a shape parameter  $\alpha$  and a characteristic time  $\tau$ .<sup>4</sup> The peak amplitude ( $A$ ) is related to the strength of the automatic activation that triggers response capture. The shape and characteristic time are related to the efficiency of the

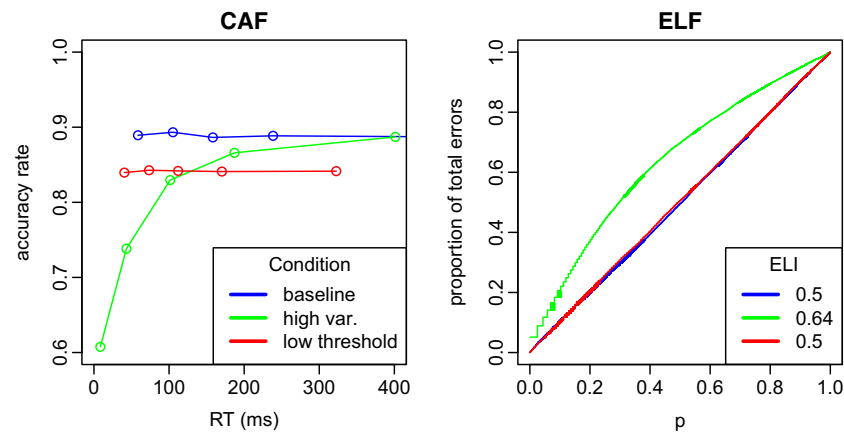
**Table 1** Parameters used in the simulations of the DDM model, with decision threshold  $b$ , drift  $\mu_c$ , and variability of the starting point  $s_z$ . The number of trials in each simulation is  $n_{\text{trials}}$

|                  | $b$ | $\mu_c$ | $s_z$ | $n_{\text{trials}}$ |
|------------------|-----|---------|-------|---------------------|
| Baseline         | .5  | 2       | 0     | 80000               |
| Low threshold    | 0.4 | 2       | 0     | 80000               |
| High variability | 0.5 | 2       | 1     | 80000               |

<sup>4</sup>The expected mean of the variable drift rate at time  $t$  is thus given by:

$$Ae^{-\frac{t}{\tau}} \left[ \frac{te}{(\alpha-1)\tau} \right]^{\alpha-1}.$$





**Fig. 3** Simulated CAF and ELF curves for the DDM model

suppression of the automatic activation. Suppression efficiency can be quantified by the time between the peak latency of the automatic activation  $t_{max} = \tau(\alpha - 1)$  and its 90th percentile ( $t_{90}$ ). We simulate data from the DMC with four sets of parameters (see Table 2) chosen in a plausible range (Ulrich et al. 2015; White, Servant, & Logan, 2017). The model is simulated in seconds, with a diffusion coefficient equal to 1.

The first set of parameters corresponds to our baseline. In the strong automatic activation condition, the peak of the automatic activation is set at 0.35, instead of 0.2 in the baseline condition. Therefore, the automatic activation has a larger influence on the response. In the low threshold condition, the upper decision bound is reduced from 0.6 (baseline condition) to 0.5. As in the DDM, this corresponds to a change in the speed–accuracy trade-off, with faster but less accurate answers. Finally, in the high control condition, the suppression of the automatic activation is more efficient ( $t_{90} - t_{max} = 0.14$  seconds instead of 0.30 seconds in the baseline condition). The CAF and ELF curves corresponding to these four conditions are shown in Fig. 4.

The slopes of the initial segments of the CAF in the strong automatic and low threshold conditions are very similar. By contrast, the ELF curve is much higher in the strong automatic activation condition than in the low threshold condition. This suggests that the ELF curve actually captures something related to automatic response capture that is not measured by CAF curves. These results are reflected in the ELI, reported in the right panel of Fig. 4.<sup>5</sup>

<sup>5</sup>One might be tempted, in view of the above results, to use ELF to assess model fit. This would not be a good idea, as ELF totally neglects some important information, such as the overall error rate.

### Application: color saturation manipulation in an Eriksen task

Servant, Montagnini, and Burle (2014) studied the influence of the salience of the target in an Eriksen task (experiment 1). We use their data to illustrate the usefulness of ELF and ELI. We show that ELF and CAF yield contradictory results. Incidentally, we also show that the results derived from CAF curves are sensitive to the choice of the number of bins.

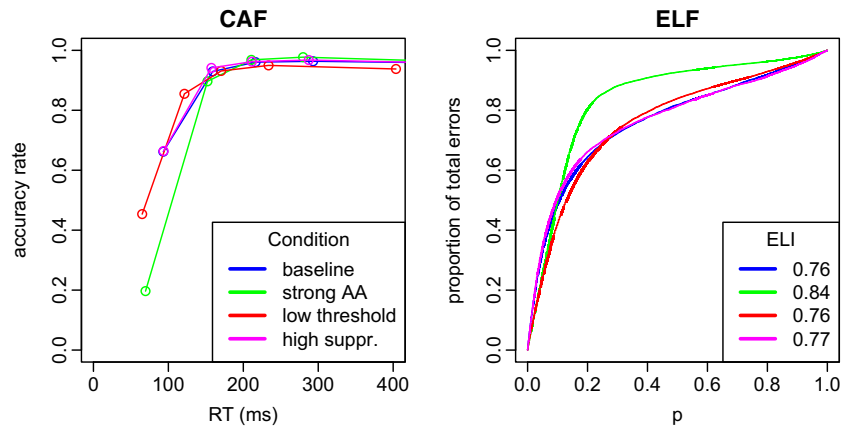
All analyses are performed in R (version 3.2.4 (R Core Team, 2016)). Because CAF slopes and first bins, as well as ELI indices, are typically not normally distributed, their means across conditions are compared with permutation tests.

### Experimental procedure

A brief overview of critical details about participants and experimental procedure used by Servant, Montagnini, and Burle (2014, experiment 1) is recounted below. Twelve students participated in the experiment. Participants had to

**Table 2** Parameters used in the simulations of the DMC model, with  $b$  the decision threshold,  $\mu_c$  the drift of the controlled process,  $A$ ,  $\tau$  and  $\alpha$  the peak amplitude, shape and characteristic time of the automatic process. The number of trials in each simulation is  $n_{trials}$

|                             | $b$  | $\mu_c$ | $A$  | $\alpha$ | $\tau$ | $n_{trials}$ | $t_{90} - t_{max}$ |
|-----------------------------|------|---------|------|----------|--------|--------------|--------------------|
| Baseline                    | 0.6  | 2.3     | 0.2  | 2        | 0.08   | 80000        | 0.30               |
| Strong automatic activation | 0.6  | 2.3     | 0.35 | 2        | 0.08   | 80000        | 0.30               |
| Low threshold               | 0.5  | 2.3     | 0.2  | 2        | 0.08   | 80000        | 0.30               |
| High control                | 0.60 | 2.3     | 0.2  | 2.5      | 0.53   | 80000        | 0.14               |



**Fig. 4** Simulated CAF and ELF curves for the DMC model

discriminate the color of a central target circle while ignoring the color of a pair of irrelevant flanking circles. The color saturation of the target circle was manipulated using six levels (15, 25, 35, 45, 60, and 80%). Participants performed 24 blocks of 96 trials in a single session. Half of the trials were congruent (same color for target and distractors), and the other half were incongruent (different colors for targets and distractors). First-order sequential effects were controlled by a pseudo-randomization procedure, and color saturation levels were balanced across sequences of trials. Participants were instructed to answer as quickly and accurately as possible.

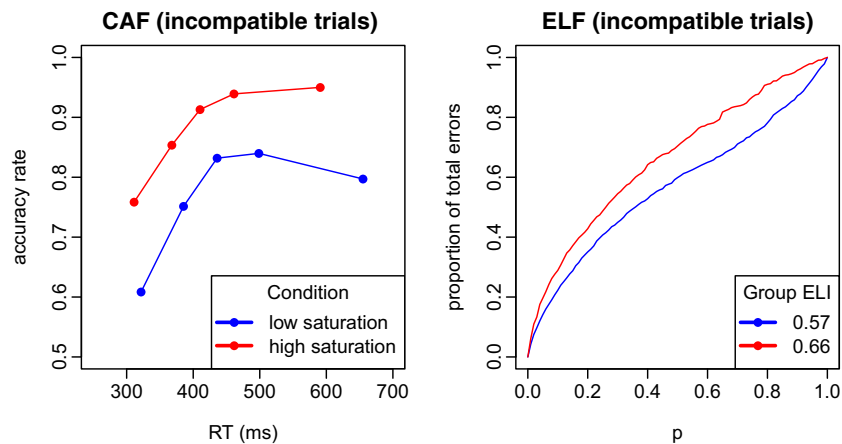
**Analysis of response capture with CAF**

We split incongruent trials into low (15, 25, and 35%) and high (45, 60, and 80%) saturation levels, and we investigated the impact of saturation level (high versus low) on response capture. Responses were faster in the high saturation condition for errors (mean RT = 370 ms, SD = 103 ms; low saturation condition: mean RT = 431 ms, SD = 154 ms;

$p < 0.001$ ) and correct responses (mean RT = 436 ms, SD = 119 ms; low saturation condition: mean RT = 468 ms, SD = 128 ms;  $p < 0.001$ ). In addition, accuracy was higher in the high saturation condition (mean accuracy = 0.88, SD = 0.32; low saturation: mean accuracy = 0.77, SD = 0.32;  $p < 0.001$ ). On average, each participant made 67.6 errors (SD = 38, min = 5, max = 121) in the high saturation condition, and 135 errors in the low saturation condition (SD = 84.2, min = 45, max = 278).

We first computed conventional CAF curve for each condition (high and low saturation levels), using five bins. The left panel of Fig. 5 shows these CAF curves averaged across participants. We found that the initial slope of the CAF was significantly lower for high saturation (mean = 1.62, SD = 1.25) than low saturation levels (mean = 2.14, SD = 1.03;  $p = 0.03$ ). In addition, the accuracy rate in the first bin was significantly larger in the high saturation condition (mean = 0.76, SD = 0.15; low saturation: mean = 0.61, SD = 0.23,  $p < 0.001$ ).

Thus, CAF analyses suggest that response capture is stronger with low saturation levels. It is worth noting that



**Fig. 5** Average CAF and ELF curves for the incongruent trials in the low and high saturation conditions

these results are sensitive to the choice of the number of bins used to compute the CAF curve. If one uses ten bins instead of five, the slopes of the CAF curves are not significantly different (high saturation: mean = 2.14, SD = 2.12; low saturation: mean = 3.38, SD = 2.23;  $p = 0.4$ ), while the accuracy rate of the first bin remains significantly higher in the high saturation condition (high saturation: mean = 0.70, SD = 0.17; low saturation: mean = 0.53, SD = 0.25;  $p = 0.004$ ).

### Analysis of response capture with ELF

We now turn to ELF and ELI. We observed that all participants made at least five errors in each condition. We implemented a bootstrap procedure to evaluate whether this is enough to provide meaningful ELI. We bootstrapped 9999 times participants RTs in each saturation condition, and computed the mean ELI across participants for each bootstrap sample. We then computed the mean and 95% confidence intervals of these indices. Given that these confidence intervals seem reasonably small, we compared ELI indices in the low and high saturation levels conditions. We found that ELI was larger for the high saturation condition (high saturation: mean = 0.66,  $CI = [0.63, 0.70]$ ; low saturation: mean = 0.57,  $CI = [0.55, 0.59]$ ,  $p = 0.03$ ).

Thus, ELI shows that errors are less concentrated among fast responses when saturation decreases, suggesting a lower amount of response capture. This contradicts the conclusion drawn from CAF.

### Conclusions

Conditional accuracy function (CAF) is routinely used as a way to quantify the strength of response capture, particularly in times of response conflicts. We argued that this could be misleading, and probably results from a confusion between two distinct concepts: (i) the accuracy rate within the first RT bin, which can be measured by CAF but is not specifically related to response capture, and (ii) the proportion of errors that are faster than a given quantile of the overall RT distribution, which is a natural measure of the strength of response capture, but cannot be deduced from CAF.

We thus proposed an alternative way to represent RT distributions, namely the error location function (ELF), which precisely represents the RT distribution of errors compared to the RT distribution of all trials. Importantly, ELF does not rely on a discretization of the data. We further showed that the area under the ELF curve provides a natural index of error location. This index, termed error location index (ELI) to ELF can be easily interpreted as the expectation of the proportion of incorrect responses faster than a randomly (according to the uniform distribution) chosen trial.

It should be noted for empirical applications that because ELF and ELI strongly rely on cumulative distribution functions of correct responses and errors, one needs to make sure that there are enough data for these empirical cumulative distributions to be meaningful.

Using simulation of computational models, we illustrated the scope of ELF and ELI. The results of our simulations show that ELF can appropriately distinguish situations where there is a strong response capture, while a more conventional interpretation of CAF would yield to incorrect conclusions. We applied ELF method to real data from Servant et al. (2014), who studied the influence of the salience of the target in an Eriksen task. We show that ELF and CAF yield to opposite conclusions. We also show that the results derived from CAF curves are sensitive to the choice of the number of bins. We thus hope that ELF and ELI, which are both conceptually easy to understand and computationally easy to implement, will prove to be helpful for scientists interested in the measure of the strength of response capture. As a note of caution, we remind the reader that ELF and ELI are non-parametric measures, and should be interpreted in combination with other measures derived from specific theoretical frameworks.

**Acknowledgements** We thank Andrew Heathcote and two anonymous reviewers for insightful comments.

### Appendix: R codes for ELF and ELI

The code we provide to compute ELF crucially relies on the quantile function. This function allows to choose the method used to compute quantiles among the nine methods presented in Hyndman and Fan (1996). The scripts below could be adapted accordingly.

```
# Compute ELF
elf <- function (rt, accuracy) {
  p=(1:length(rt))/length(rt)
  errors <- rt[accuracy==0]
  correct <- rt[accuracy==1]
  if (length(correct) == length(rt)) { L <- NA }
  else {
    c_error <- ecdf(errors)
    L <- c_error(quantile(rt, probs = p))
    return(L)
  }
}

# Compute ELI
eli <- function (L){
  n <- length(L)
  V <- (1/n)*sum(L)
  return(V)
}
```



## References

- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Attention, Perception, & Psychophysics*, *16*(1), 143–149.
- Gratton, G. et al. (1988). Pre- and poststimulus activation of response channels: a psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 331.
- Hübner, R., Steinhauser, M., & Lehle, C. (2010). A dual-stage two-phase model of selective attention. *Psychological Review*, *117*(3), 759.
- Hyndman, R. J., & Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician*, *50*(4), 361–365.
- Laming, D. R. J. (1968). Information theory of choice-reaction times. Luce, R. D. (1986). Response times: Their role in inferring elementary mental organization. 8. Oxford University Press on Demand.
- Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*(5), 798–817.
- R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356.
- Ridderinkhof, K. R. (2002). Activation and suppression in conflict tasks: Empirical clarification through distributional analyses. In *Attention and performance XIX: Common mechanisms in perception and action* (pp. 494–519).
- Servant, M., Montagnini, A., & Burle, B. (2014). Conflict tasks and the diffusion framework: Insight in model constraints based on psychological laws. *Cognitive Psychology*, *72*, 162–195.
- Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. *Advances in Psychology*, *65*, 31–86.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643.
- Ulrich, R. et al. (2015). Automatic and controlled stimulus processing in conflict tasks: superimposed diffusion processes and delta functions. *Cognitive Psychology*, *78*, 148–174.
- van den Wildenberg, W. P. M. et al. (2010). To head or to heed? Beyond the surface of selective action inhibition: a review. *Frontiers in Human Neuroscience*, *4*, 222.
- van Wouwe, N. C. et al. (2016). Dissociable effects of dopamine on the initial capture and the reactive inhibition of impulsive actions in Parkinson's disease. *Journal of Cognitive Neuroscience*, *28*(5), 710–723.
- White, C. N., Brown, S., & Ratcliff, R. (2012). A test of Bayesian observer models of processing in the Eriksen flanker task. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 489.
- White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*, *63*(4), 210–238.
- White, C. N., Servant, M., & Logan, G. D. (2017). Practical considerations for using conflict-based diffusion models to interpret choice RT data *Psychonomic Bulletin & Review*. in press.
- Wylie, S. A. et al. (2010). The effect of Parkinson's disease on the dynamics of online and proactive cognitive control during action selection. *Journal of Cognitive Neuroscience*, *22*(9), 2058–2073.
- Wylie, S. A. et al. (2012). Dopamine agonists and the suppression of impulsive motor actions in Parkinson disease. *Journal of Cognitive Neuroscience*, *24*(8), 1709–1724.