Studying the dynamics of visual search behavior

using RT hazard and micro-level speed-accuracy tradeoff functions:

A role for recurrent object recognition and cognitive control processes

Sven Panis[1]

Rani Moran[2,3]

Maximilian P. Wolkersdorfer[1]

Thomas Schmidt[1]

12 Figures & 4 Tables

1 Experimental Psychology Unit, Faculty of Social Sciences, Technische Universität Kaiserslautern, Gottlieb-Daimler-Straße 47, 67663 Kaiserslautern, Germany.

2 Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, 10-12 Russell Square, London WC1B 5EH, United Kingdom.

3 Wellcome Centre for Human Neuroimaging, University College London, London WC1N 3BG, United Kingdom.

Corresponding author: sven.panis@sowi.uni-kl.de

Abstract

Thanks to the work of Anne Treisman and many others, the visual search paradigm has become one of the most popular paradigms in the study of visual attention. However, statistics like mean correct response time (RT) and percent error do usually not suffice to decide between the different search models that have been developed. Recently, to move beyond mean performance measures in visual search, RT histograms have been plotted, theoretical waiting time distributions have been fitted, and whole RT and error distributions have been simulated. Here we promote and illustrate the general application of discrete-time hazard analysis to response times, and of micro-level speed-accuracy tradeoff analysis to timed response accuracies. An exploratory analysis of published benchmark search data from feature, conjunction, and spatial configuration search tasks reveals new features of visual search behavior, such as a relatively flat hazard function in the right tail of the RT distributions for all tasks, a clear effect of set size on the shape of the RT distribution for the feature search task, and individual differences in the presence of a systematic pattern of early errors. Our findings suggest that the temporal dynamics of visual search behavior is resulting from a decision process that is temporally modulated by concurrently active recurrent object recognition, learning, and cognitive control processes, next to attentional selection processes.

Key-words

Visual search, response times, discrete-time hazard analysis, individual differences, speed-accuracy tradeoff, event history analysis

Introduction

The visual search paradigm is one of the most popular paradigms in the study of visual attention as it mimics real search tasks we perform every day (for reviews, see Eckstein, 2011; Humphreys, 2016). In each trial of a standard visual search task a display is presented that contains a spatially arranged set of objects, and participants are asked to press one of two buttons to indicate whether the target (e.g., a red vertical bar) is present or not. The so-called search functions relating the number of items in the display (set size) to the mean correct search response time (RT) are close to linear for both sets of target-present (TP) and target-absent (TA) trials, and their slopes seem to vary on a continuum depending on the difficulty of the search task (Cheal & Lyon, 1992; Liesefeld, Moran, Usher, Müller, & Zehetleitner, 2016). For example, search for a red vertical target among green vertical distractors (feature search; Figure 1A, left) is efficient, and results in search functions with slopes close to 0 ms/item (Figure 1B, left). In contrast, searching for a 2 among 5´s (spatial configuration search; Figure 1A, right) is inefficient, and results in search functions with large positive slopes (Figure 1B, right). Finally, searching for a red vertical target among green vertical and red horizontal distractors (conjunction search; Figure 1A, middle) is of intermediate efficiency, and results in search functions with intermediate slopes (Figure 1B, middle).

To explain visual search behavior, researchers have mainly focused on devising different accounts of the attentional selection process. According to serial search accounts a high-speed attentional spotlight is scanning each object one by one in order to bind its surface features and to recognize it as a target or distractor (Treisman and Gelade, 1980; Wolfe, Cave, & Franzel, 1989). When the target is so salient that it is always scanned first – for example when the target and distractors are very dissimilar in a single surface dimension

such as color – flat search slopes will result. Scanning continues until the target is found or all items are identified as distractors – a serial exhaustive search model (Wolfe, 1994). More recent developments have added grouping processes and feature inhibition processes (Treisman & Sato, 1990; Wolfe, 2007).

According to parallel search accounts all items in the display are attended and identified in parallel. While some parallel search models are based on signal detection theory (Palmer, 1995), others are based on biased competition (Heinke & Backhaus, 2011; Heinke & Humphreys, 2003), similarity, grouping, and recursive rejection (Humphreys & Müller, 1993; Müller, Humphreys, & Donnelly, 1994), neural synchronization (Kazanovich & Borisyuk, 2017), or neurodynamical approaches (Deco & Zihl, 2006; Fix, Rougier, & Alexandre, 2011; Grieben, Tekülve, Zibner, Schneegans, & Schöner, 2018). However, mean correct RT and slopes are not sufficient to distinguish between serial versus parallel processing because both search mechanisms are able to generate efficient and inefficient searches (Townsend, 1990a).

Because mean performance measures such as overall error-rate and mean correct RT can be accounted for by different computational models – the problem of model mimicry – Wolfe, Palmer, and Horowitz (2010) focused on the shape of the RT distributions. They collected very large data sets for three search tasks to set a benchmark: a feature search for a color, a spatial configuration search for a digit 2 among digit 5s, and a color-by-orientation conjunction search (see Figure 1A). Target presence (present or TP/absent or TA) and set size (3, 6, 12, 18) were manipulated for each task to give eight within-subject trial types (TP3, TA3, TP6, …, TA18). About 500 trials were administered to each participant for each trial type. In each trial, the search display remained visible until the observer pressed one of two keys to indicate target present or target absent.

Balota and Yap (2011) distinguish three general approaches for understanding the influences of variables on RT distributions. The first approach is to plot the shape of the RT distribution to determine how a manipulation changes the different regions of the distribution (e.g., histograms, quantile plots, delta plots, hazard plots). For example, Wolfe et al. (2010) plotted histograms by tabulating the RTs in 50 ms-wide bins (see Figure 1C). They found that all distributions were positively skewed, and that variance tracks mean RT (i.e., all distributions broaden as they shift rightward). Furthermore, for the conjunction and spatial configuration tasks, target-absent distributions are generally to the right of target-present distributions, and the variance of the target-absent trials is greater than that of the target-present trials. They concluded that these distributional shapes reject classic, serial self-terminating search models including the Guided Search 2.0 model (Wolfe, 1994) as shown in Figure 1D.

The second approach is to fit a mathematical function to an RT distribution to assess how different parameters of the function are modulated by experimental manipulations (Balota & Yap, 2011). For example, Palmer, Horowitz, Torralba, and Wolfe (2011) fitted four psychologically motivated functions to these benchmark data sets (ex-Gaussian, ex-Wald, Gamma, and Weibull). They found that the three functions with an exponential component were all more successful at modeling the RT distributions than the Weibull. They proposed that these exponential components either reflect residual (non-decision) processes in the generation of response times, or that these residual components are actually encoding something important about the search process itself. However, they were unable to distinguish among these two options.

The third and ultimately the preferred approach discussed by Balota and Yap (2011) is to use a computationally explicit process model that makes specific predictions about the

characteristics of RT distributions. For example, Moran, Zehetleitner, Müller, and Usher (2013) developed the Competitive Guided Search (CGS) model as an extension of the Guided Search 2.0 model (Wolfe, 1994). The main addition was a mechanism to quit searches prematurely in order to explain the large overlap between the empirical distributions in the target-absent conditions (see Figure 1C). Based on several model comparisons using the benchmark data sets from Wolfe et al. (2010), Moran et al. (2013) concluded that the CGS model meets the challenge of accounting for the shape of the RT distributions in the three benchmark search data sets. Furthermore, Moran, Zehetleitner, Liesefeld, Müller, and Usher (2016) found that CGS indeed fits the benchmark data sets better than a flexible, competitive parallel race model.

However, based on another benchmark search data set, Narbutas, Kristan, and Heinke (2017) concluded that the CGS model suffers from a failure to generalize across all display sizes, as did a parallel search model developed by Heinke and Humphreys (2003). Indeed, Cheal and Lyon (1992) already concluded that none of the standard theories of visual search are completely adequate (see also Dutilh et al., 2018).

The structure of this paper is as follows. First we will discuss event history analysis, the standard longitudinal technique to analyze time-to-event data in many scientific disciplines. Event history analysis allows one to study how the effect of an experimental manipulation (here: set size and target presence) on performance changes with the passage of time on one or more time scales. We end the Introduction section by listing our objectives. In the Methods section we explain how we applied the descriptive and inferential statistics from event history analysis to the benchmark data sets of Wolfe et al. (2010). In the Results section we show descriptive plots of the empirical distributions and compare different individuals. Because Balota and Yap (2011) do not discuss the statistical analysis of

RT distributions, we also illustrate how to fit a statistical model to RT distributions and what this reveals about behavioral dynamics. We discuss several new findings in light of existing visual search theories in the Discussion section.

Event history analysis

Event history analysis, a.k.a. survival, hazard, duration, transition, and failure-time analysis, is the standard set of statistical methods for studying the occurrence and timing of events in many scientific disciplines (Allison, 2010; Singer & Willett, 2003). Examples of time-to-event data include RT data, saccade latencies, fixation durations, time-to-force-threshold data, perceptual dominance durations, neural inter-spike durations, etc. To apply event history analysis one must be able to define the event-of-interest (any qualitative change that can be situated in time; here: a button-press response), to define time point zero (here: search display onset), and to measure the passage of time between time zero and event occurrence.

Continuous-time hazard rate function

Luce (1986) mentions that there are several different, but mathematically equivalent, ways to present the information about a continuous random variable T denoting a particular person´s response time in a particular experiment: the cumulative distribution function $F(t) = P(T \leq t)$, its derivative $F´(t)$ known as the probability density function $f(t)$, the survivor function $S(t) = 1 - F(t) = P(T > t)$, and the hazard rate function $\lambda(t) = f(t) / [1 - F(t)] = f(t) / S(t)$. "In principle, we may present the data as estimates of any of these functions and it should not matter which we use. In practice, it matters a great deal, although that fact does not seem to have been as widely recognized by psychologists as it might be" (Luce, 1986, p. 17).

Event history analysis (EHA) has been developed to describe and model *the hazard function of response occurrence*. Hazard quantifies the instantaneous risk that a response will occur at time t, conditional on its nonoccurrence until time t. In other words, it quantifies the likelihood that a response we are still waiting for at time t will occur within the next instant. There are at least five reasons why statisticians and mathematical psychologists recommend focusing on the hazard function in practice, when dealing with a finite sample.

First, "the hazard function itself is one of the most revealing plots because it displays what is going on locally without favoring either short or long times, and it can be strikingly different for *f*'s that seem little different." (Luce, 1986, p. 19). To illustrate this Figure 2 shows the F(t), f(t), S(t), and λ(t) for four theoretical waiting-time distributions. In contrast to λ(t), all F(t) and S(t) distributions look vaguely alike and one cannot describe easily salient features other than the mean and standard deviation. The problem with the density function f(t) is that it conceals what is happening in the right tail of the distribution (Luce, 1986). As discussed by Holden, Van Orden, and Turvey (2009) „Probability density functions can appear nearly identical, both statistically and to the naked eye, and yet are clearly different on the basis of their hazard functions (but not vice versa). Hazard functions are thus more diagnostic than density functions" (p. 331).

Second, because RT distributions may differ from one another in multiple ways, Townsend (1990b) developed a dominance hierarchy of statistical differences between two arbitrary distributions A and B. For example, if F_A(t) > F_B(t) for all t, then both cumulative distribution functions are said to show a complete ordering. Townsend (1990b) showed that a complete ordering on the hazard functions – λ_A(t) > λ_B(t) for all t – implies a complete ordering on both the cumulative distribution and survivor functions – F_A(t) > F_B(t) and S_A(t) < S_B(t) – which in turn implies an ordering on the mean latencies – mean A < mean B. In

contrast, an ordering on two means does *not* imply a complete ordering on the corresponding F(t) and S(t) functions, and a complete ordering on these latter functions does *not* imply a complete ordering on the corresponding hazard functions. This means that stronger conclusions can be drawn from data when comparing the RT hazard functions using event history analysis. For example, when mean A < mean B, the hazard functions might show a complete ordering (i.e., for all t) or only a partial ordering (e.g., only for t < 600 ms).

Third, EHA does not discard right-censored observations (trials for which we do not observe a response during the data collection period so that we only know that the RT must be larger than some value)[1]. Discarding such trials and/or trials with very long RTs will introduce a sampling bias that results in underestimation of the mean. In fact, EHA includes the data from all trials when estimating the descriptive statistics. For inferential statistics, one might exclude some early bins with rarely occurring fast responses (see Methods).

Fourth, hazard modeling allows incorporating *time-varying* explanatory covariates such as heart rate, EEG signal amplitude, gaze location, etc. (Allison, 2010) which is useful for cognitive psychophysiology (Meyer, Osman, Irwin, & Yantis, 1988)[2].

Fifth, hazard is more suited as a measure of the concept of processing capacity, i.e., the amount of work the observer is capable of performing within some unit of time (Wenger & Gibson, 2004). In the context of research on attention, the hazard function can capture the notion of the instantaneous capacity of the observer for completing the task in the next instant, given that the observer has not yet completed the task.

Discrete-time hazard probability function

Unfortunately, estimating the shape of the continuous-time hazard rate function for one observer in one experimental condition is not straightforward because one needs at least

1000 trials for example (Bloxom, 1984; Luce, 1986; Van Zandt, 2000). Furthermore, statistical modeling of continuous time-to-event data requires specialized software to either fit parametric hazard models that are rather restrictive in the shapes they allow (e.g., a Weibull hazard model), or semi-parametric hazard models that completely ignore the shape of the hazard function (e.g., Cox regression). Therefore, we promote the general application of *discrete-time hazard analysis* to RT data, which is straightforward, easy and intuitive, and allows for flexible statistical modeling by logistic regression which is highly familiar to psychologists (Allison, 1982, 2010; Singer & Willett, 1991, 2003; Willett & Singer, 1993, 1995).

To calculate the descriptive statistics – functions of discrete time – one has to set up a life table. A life table summarizes the history of event occurrences for a combination of subject and experimental condition. For illustrative purposes, we present in Table 1 a life table for the 530 trials of one participant in the feature search task for the trial type TP3 (target present and set size 3).

First, the first second after search display onset is divided into 10 contiguous bins of 100 ms (column 1). Then, after counting the number of observed responses in each bin (column 4) the *risk set* must be determined for each bin (column 5). The risk set is equal to the number of trials that have not yet experienced a response at the start of the bin in question. The sample-based hazard estimate in bin t, or $h(t)$ (column 6), is then simply obtained by dividing the number of observed responses in bin t (column 4) by the risk set of bin t (column 5). In discrete time, hazard is defined as the conditional probability of a response occurring in time bin t given it has not yet occurred before, $h(t) = P(T=t|T \geq t)$. The discrete-time hazard function thus tells us the probability that a response we are still waiting for will actually occur in bin t.

Next to the hazard function, EHA also focuses on the survivor function $S(t) = P(T>t) = 1 - F(t)$, because $S(t)$ provides a context for $h(t+1)$ as it shows the proportion of trials not having experienced the response by the end of bin t. For completeness, Table 1 also tabulates the corresponding probability mass function $P(t) = P(T=t)$.[3]

For choice RT data such as the benchmark search data, we want to distinguish different types of events: correct versus incorrect responses. One approach is to assume that each event type has its own hazard function that describes the occurrence and timing of events of that type. One would thus model the $h(t)$ of correct response occurrence separately from the $h(t)$ of error response occurrence. Another approach is to first study the timing of events without distinguishing among event types, and then to study which type of event occurs while restricting the analysis to those cases that experienced an event. A major attraction of this latter approach is that there is no need to assume that the different kinds of events are uninformative for one another (Allison, 2010).[4]

We therefore take the latter, so-called conditional-processes approach (Allison, 2010, pp. 227-229) by extending the $h(t)$ analysis of response occurrence with an analysis of the conditional accuracy function $ca(t) = P(\text{correct}|RT = t)$. The $ca(t)$ is estimated by dividing the number of correct responses in bin t by the total number of observed responses in bin t (Table 1)[5]. Note that $P(t)$ provides a context for $ca(t)$ as $P(t)$ shows on which percentage of trials the $ca(t)$ estimate is based.

Thus, by using $h(t)$ functions of response occurrence in combination with $ca(t)$ functions one can provide an unbiased, time-varying, and probabilistic description of the latency and accuracy of responses based on *all* trials of *any* RT data set. Statistical models for $h(t)$ and $ca(t)$ can each be implemented as generalized linear mixed regression models

predicting event occurrence (1/0) and response accuracy (1/0) in each bin of a selected time range, respectively (Panis & Schmidt, 2016).

There are also possible disadvantages of discrete-time event history analysis. First, the person-trial-bin oriented data set (see section Methods) can become very large. Second, one needs to explore a few bin sizes. The optimal bin size will depend on the censoring time, the rarity of event occurrence, and the number of repeated measures or trials. Note that the time bins do not have to be of equal size. Third, remember that in hierarchical data like ours, there are two sources of noise: within and between participants. For a distributional analysis it is important to have enough repeated measures per participant and condition (preferably at least 100) to minimize the influence of within-subject noise. Between-subjects variation is a different matter: it can be due to noise, but also due to characteristic differences between individuals (e.g., in speed, capacity, or strategy). Again, high measurement precision in single participants is the only way to deal with this. The analysis of single participants should be regarded as a safeguard against interpreting spurious effects in the pooled data that are actually only generated by a small minority of participants, while at the same time refraining from overinterpreting the individual data patterns. Note that systematic effects will be visible for a majority of participants, while occurrences due to noise will not.

Objectives

The current study is motivated by three goals. First, using a freely available data set we want to illustrate the descriptive and inferential statistics used by discrete-time EHA, and what we can learn from this. As discussed by Whelan (2008) the use of a more advanced analysis method can maximize the return from the obtained data, which is important in view of the time and costs required to run an experiment. Second, using an exploratory approach, we

want to see if the shapes of the diagnostic h(t) and ca(t) functions for the three benchmark data sets will reveal certain, as yet unknown, features of the time-dispersed behavior of searchers. Third, to collect the benchmark data set Wolfe et al. (2010) used a small-N design in which a large number of observations are made on a relatively small number of experimental participants. Smith and Little (2018) argue that, "if psychology is to be a mature quantitative science, then its primary theoretical aim should be to investigate systematic functional relationships as they are manifested at the individual participant level" (p. 2083). Therefore, we will pay attention to individual differences in the time-dispersed search behavior. As discussed below, our results reveal new features of visual search behavior, many – if not all – of which are not considered by current cognitive models of visual search, but can be helpful to inform future models.

Methods

Data sets

We reanalyzed the benchmark data sets provided by Wolfe et al. (2010; http://search.bwh. harvard .edu/new/data_set_files.html). This group collected data in three tasks. In the *feature search* task (9 participants), participants searched for a red vertical rectangle among green vertical rectangles. In the *conjunction search* tasks (10 participants), they searched for a red vertical rectangle among green vertical and red horizontal rectangles. Finally, in the *spatial configuration* task (9 participants), they searched for a numerical digit 2 among 5s (see Figure 1). Four different set sizes (SS; distractors plus target, either 3, 6, 12, or 18) were randomly intermixed. Participants were young adults with corrected or normal acuity; their color vision was ascertained by Ishihara tests. They pressed one key if the target was present (which was the case in 50 % of trials) and another if the target was absent. They were

instructed to respond as quickly and correctly as possible and received feedback after each

trial. Accuracy and RT in ms were recorded. Each participant provided approximately 10

blocks of 400 trials, leading to about 500 trials per participant and search condition. For

further information, see the website.

For the descriptive statistics we downloaded the raw data from the website via the

section named "Fitting Functions". The raw feature, conjunction, and spatial configuration

search data sets contained 35,941, 39,958, and 35,862 rows, respectively. For hazard model

fitting we actually used the downloadable text files (via the section below the one named

"Fitting Functions") because these also contain information about trial and block numbers.

Mean correct RT and percent error

We used the same outlier criteria as Wolfe et al. (2010) in order to calculate the sample

mean RT, mean correct RT, and error percentage for each combination of subject, target

presence, and set size. Specifically, we excluded all trials (N=80) with RT < 200 ms or > 4,000

ms for the feature and conjunction search tasks, and with RT < 200 ms or > 8,000 ms in the

spatial configuration search task.

Event history analysis: descriptive statistics

Starting from the raw data sets without removing outliers, life tables were constructed using

software package R (R Core Team, 2014) for each combination of subject, target presence,

and set size (see Table 1). We used a bin size of 40 ms for the feature and conjunction search

tasks, and a bin size of 80 ms for the spatial configuration search task, to provide high

temporal resolution when visually studying the shape of the empirical distributions (and to

still have an acceptable level of stability in the estimates). We used a censoring time of 1,600,

2,400, and 3,600 ms for the feature, conjunction, and spatial configuration search data sets,

respectively, since most events had occurred by this time in all search conditions. Standard

errors for h(t), P(t), and ca(t) can be estimated using the formula for a proportion p – the

square root of $\{p(1-p) / N\}$ – where N equals the risk set for bin t, the total number of trials,

and the number of observed responses in bin t, respectively. The standard errors for S(t)

were estimated using the recurrent formula on page 350 of Singer and Willett (2003).

Event history analysis: inferential statistics

To test *whether and when* the main and interaction effects including target presence and set

size are significant across participants, we fitted discrete-time hazard models to aggregated

data by implementing generalized linear mixed-effects regression models in R (R Core Team,

2014; function glmmPQL of package MASS) using the complementary log-log (cloglog) link

function (Allison, 2010).[6] An example discrete-time hazard model with three predictors can

be written as follows: cloglog[$h(t)$] $=\ln(-\ln[1-h(t)]) = [\alpha_0 ONE + \alpha_1(TIME-1) + \alpha_2(TIME-1)^2 + \alpha_3(TIME-1)^3] + [\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2(TIME-1)]$. The main predictor variable TIME is the time bin index t

(see Table 1) which is centered on value 1 in this example. The first set of terms within

brackets, the alpha parameters multiplied by their polynomial specifications of (centered)

time, represents the shape of the baseline cloglog-hazard function (i.e., when all predictors

X_i take on a value of zero). The second set of terms (the beta parameters) represents the

vertical shift in the baseline cloglog-hazard for a 1 unit increase in the respective predictor.

For example, the effect of a 1 unit increase in X_1 is to vertically shift the whole baseline

cloglog-hazard function with β_1 cloglog-hazard units. However, if the predictor interacts

linearly with time (see X_2 in the example) then the effect of a 1 unit increase in X_2 is to

vertically shift the predicted cloglog-hazard in bin 1 with β_2 cloglog-hazard units (when TIME-

1 = 0), in bin 2 with $\beta_2 + \beta_3$ cloglog-hazard units (when TIME-1 = 1), etc. To interpret the effects of the predictors, the parameter estimates are anti-logged, resulting in a hazard ratio.

We proceeded as follows. First, for each search task we aggregated the raw data across all subjects, except for the conjunction search task where one observer was ignored (see section Results).

Second, for each task we selected a time range where all subjects provide enough data in each condition, and created between 10 and 20 bins for modeling purposes. For the feature search data we therefore used a censoring time of 800 ms and a bin size of 40 ms. After ignoring the first 240 ms (i.e., the first six 40-ms bins) in which no (or only few) responses occurred, we end up with 14 bins to model. For the conjunction search data we used a censoring time of 1000 ms and a bin size of 50 ms. After ignoring the first 350 ms (i.e., the first seven 50-ms bins) in which no (or only few) responses occurred, we end up with 13 bins to model. For the spatial configuration data we used a censoring time of 2000 ms and a bin size of 80 ms. After ignoring the first 400 ms (i.e., the first five 80-ms bins) in which no (or only few) responses occurred, we end up with 20 bins to model.

Third, trial type TP3 was chosen as the baseline condition in each search task. The main predictor variable TIME was centered on value 10 or bin (360,400], 10 or bin (450,500], and 8 or bin (560,640] for the feature, conjunction, and spatial configuration search task, respectively. For each task, the intercept and the linear effect of TIME were treated as random effects to deal with the correlated data resulting from the repeated measures on the same subjects. Next to dummy-coding the levels of our experimental factors (target presence and set size) we also included TRIAL and BLOCK number as continuous predictors to model across-trial and across-block learning effects in the speed of responses. TRIAL was centered on value 350 (and rescaled by dividing by 10) and BLOCK on value 8 for each task.

Thus, for the feature search task for example, with all effects set to zero, the h(t) model´s intercept refers to the estimated cloglog[h(t)] for bin (360,400] in trial 350 of block 8 when the target is present and set size equals three (see Table 2, column 5, effect nr. 1).

Fourth, to estimate the parameters of the h(t) model, we must create a dataset where each row corresponds to a time bin of a trial of a participant (a person-trial-bin oriented data set). Specifically, each time bin that was at risk for event occurrence in a trial was scored on the dependent variable EVENT (0 = no response occurred; 1 = response occurred), the centered covariates TIME, TRIAL, and BLOCK, the variable SUBJECT, and the dummy-coded dichotomous experimental predictor variables (TA, SS6, SS12, SS18). Thus, for the feature search task for example, all trials with observed RTs > 800 ms were treated as right-censored observations; they provide the information that the response did not occur during the first 800 ms or 20 bins after search display onset (i.e., each of these trials contributes 20 rows, and each row has a value 0 for EVENT).

Just before running glmmPQL, we deleted the rows corresponding to the first 6, 7, and 5 bins for the feature, conjunction and spatial configuration search task, respectively, as mentioned above under step 2. The resulting person-trial-bin oriented data set contained 168,996, 219,762, and 355,261 rows for the feature, conjunction and spatial configuration search task, respectively.

Fifth, for each task, we started with a full multi-level EHA model (46 parameters; with bins at level 1 nested within observers at level 2) encompassing the following effects at level 1: (a) a 7th order polynomial for the shape of the baseline cloglog-hazard function (8 parameters), (b) the effects of Target Absence (TA), SS6, SS12 and SS18 were allowed to interact with time in a quartic fashion (20 parameters), (c) the interaction effects between TA and each of the three set sizes could vary over time in a cubic fashion (12 parameters),

and (d) the linear effects of Trial and Block were allowed to interact with time in a quadratic fashion (6 parameters). For each task we used an automatic backward selection procedure to select a final model. Specifically, during each iteration, the effect with the largest p-value that was not part of any higher-order effect was deleted, and the model refitted. This continued until each of the remaining effects that was not part of any higher-order effect had a $p < .05$ (see highlighted p-values in Tables 2, 3, and 4).

Finally, after model selection, we refitted the final model a number of times with TIME centered each time on another bin, to see explicitly what values the parameter estimates take on according to the final model in these other bins, and whether they represent a significant effect or not (see Tables 2, 3, and 4).

Results

Feature search: Descriptive statistics

In Figure 3 we present the data from one participant in the feature search task. In each experimental condition the hazard function (top row) rises to a peak and then declines before hitting the value 1. We denote time bins by the *endpoint* of the interval they span, so that "bin 280" refers to bin (240,280]. For example, in condition SS3 the estimated $h(280)$ equals 0.42 for TP and 0.014 for TA for this participant. In other words, if the waiting time has increased until 240 ms after display onset without response occurrence, then the (conditional) probability that the (first) response will occur somewhere in bin (240,280] equals 0.42 for TP but only 0.014 for TA; Similarly, if the waiting time has increased until 320 ms after display onset then the estimated $h(360)$ equals .76 for TP and .74 for TA. The effect of Target Presence on $h(t)$ is clearly visible for each set size in the left tail of the distribution (i.e., when hazard is rising), and this effect is decreasing somewhat with increases in set size.

The *ca*(t) functions show that the fastest responses in the target-absent conditions tend to be errors (false alarms) while the slowest responses are error-free. In contrast, in the target-present conditions most emitted responses tend to be correct, except for a small dip in the *ca*(t) functions that reveals a temporarily increased miss rate around the time when the *ca*(t) functions in the target-absent conditions reach 1. This particular participant, however, was the only one in the sample who emitted a response before 800 ms in each trial of each condition.

In Figure 4 we compare the *h*(t) and *ca*(t) estimates between this and three other participants. Comparing individuals reveals that there are two subgroups of observers that show qualitatively different *ca*(t) behavior. Three observers (2, 3, and 7; see Figure 4, top 8 panels) show early false alarms when responses are emitted around 240 ms after search display onset. We define early false alarms as "*ca*(t) ≤ .50 for the earliest emitted responses, for at least two set sizes when the target is absent". At the same time, they show "small dips" in early *ca*(t) for target present, i.e., small temporary increases in the miss rate (early misses) at the time when *ca*(t) for target-absent trials reaches 1. The remaining observers show no systematic errors (see Figure 4, bottom 8 panels). Note that the latter observers emit their fastest responses a bit later compared to those individuals who do show early errors. Interestingly, 8 out of 9 subjects showed a small but systematic effect of set size (i.e., SS3 > SS6 > SS12 > SS18) on *h*(t) for target-present trials in one or more bins before or around the time when hazard reaches its peak (see Figure 4, left *h*(t) panels). Finally, for those subjects who were not as fast as subject number 3 the hazard functions peaked and then declined toward, and stayed hoovering for some time around a non-zero value. Note that as time passes on the standard errors for the *h*(t) estimates automatically increase because the risk set becomes smaller and smaller.

Feature search: Inferential statistics

Table 2 shows the selected hazard model for the feature task (columns 5 to 8). Figure 5

presents the predicted (i.e., model-based) hazard functions (first column), cloglog-hazard

functions (second column), and the corresponding survivor (third column) and probability

mass functions (fourth column), for each set size in target-present (top row) and -absent

trials (bottom row) for trial 350 in block 8.

Because TRIAL and BLOCK are centered, the first six parameter estimates (PE) in

Table 2 model the shape of the cloglog[h(t)] function for TP3 (the chosen baseline condition)

in trial 350 of block 8 using a 5th order polynomial function of TIME (Figure 5, first row,

second column, green line). Because TIME is centered on bin 400, the intercept of our

regression model refers to the predicted cloglog[h(400)] value for TP3 in trial 350 of block 8.

Converting back from cloglogs to hazards, h(400) = .42 (= 1 - exp[-exp(-0.61)]) as shown in

Figure 5 (top left). Parameters 2-6 show a significant linear, quadratic, cubic, quartic, and

quintic effect of TIME on this intercept estimate, such that the predicted response hazard

first quickly increases with increasing waiting time until around 440 ms after display onset,

and then decreases toward a non-zero value: h(280) = 0.04, h(400) = 0.42, h(520) = 0.39, and

h(640) = 0.16. This shows that the hazard of response occurrence changes in a particular

fashion on the across-bin/within-trial time scale.

With respect to the manipulations of interest, we see that in bin 400 and relative to

the reference condition TP3, there is a main effect of removing the target (parameter 7,

column 5, PE = -0.2483, p < .0001). A measure of effect size for a discrete-time cloglog-

hazard model can be obtained by exponentiating the parameter estimates which gives us

hazard ratios (HR; Allison, 2010, p. 242). Thus, compared to the cloglog[h(400)] estimate in

the reference condition, removing the target decreases the estimated cloglog[h(t)] by 0.2483

units, which corresponds to a decrease in response hazard by a factor of 0.78 ($HR(400) =$ exp[-0.2483] = 0.78). There are also main effects in bin 400 of changing the set size to 6 (parameter 12, PE = -0.11, HR = 0.90, $p < .0001$), to 12 (parameter 14, PE = -0.14, HR = 0.87, $p < .0001$), and to 18 (parameter 18, PE = -0.2, HR = 0.83, $p < .0001$). The fact that all these effects are negative indicates that response occurrence slows down.

All these main effects change significantly with TIME (parameters 7 to 20). For example, the effect of Target Absent changes in quartic fashion (parameters 7 to 11) so that it equals -1.70 in bin 280 (HR = 0.18, $p < .0001$), -0.25 in bin 400 (HR = 0.78, $p < .0001$), -0.21 in bin 520 (HR = 0.81, $p < .0001$), and only 0.09 in bin 640 (HR = 1.09, $p = .173$). The effect of SS6 changes in a linear fashion that is marginally significant (parameter 13, $p = .0506$). The effect of SS12 changes in a linear and cubic fashion (parameters 15 to 17). The effect of SS18 changes in a linear fashion (parameters 19 to 20). Increasing the set size leads to a systematic decrease in the estimated h(t) in bins < 500 ms when the target is present (see Figure 5, top left panel). Bins after 500 ms show no significant effects of set size anymore in the target-present conditions. Thus, the h(t) functions show a *partial* ordering with respect to the systematic effects of target presence (i.e., only for t < 600 ms) and set size when the target is present (i.e., only for t < 500 ms). In other words, once the waiting time has increased until 600 ms after display onset, then set-size and target presence have no influence anymore on the hazard of response occurrence.

As expected, there are also interaction effects in bin 400 between Target Absent and each of the 3 set sizes, which change over TIME (parameters 21-28). These positive interaction effects counteract the negative main effects of set size when the target is present. Note that for each set size (SS6, SS12, and SS18) the interaction effect with Target Absent is larger in absolute value than the main effect of each set size (i.e., parameter 21

versus 12, 23 versus 14, and 26 versus 18), both in bin 280 and in bin 400. Thus, increasing

the set size up to 12 leads to a systematic increase in the estimated h(t) in bins < 500 ms

when the target is absent; This can be seen most clearly for bin 280 of the cloglog-hazard

functions in Figure 5 (second row and column). Bins after 500 ms show no significant

interaction effects involving set size.

One advantage of a discrete-time hazard model is that you can incorporate multiple

time scales. Parameters 29 to 32 show that hazard also varies on the across-trial/within-

block time scale, and on the across-block/within-experiment time scale. First, in bin 400,

each additional series of 10 trials will increase the estimated cloglog[h(t)] value with -0.0023

units (parameter 29, column 5, $p < .0001$), and this effect increases linearly with TIME

(parameter 30, PE = .0009, $p < .0001$). Thus, while the effect of Trial is negative for the left

tail of the distribution (see Table 3, row 29) it is positive for the right tail (e.g., in bin 640).

Second, each additional block will increase the estimated cloglog[h(t)] value with 0.0275

units in bin 400 (parameter 31, column 5, $p < .0001$), and this effect decreases linearly with

TIME (parameter 32, PE = -.0057, $p < .0001$). Figure 6A shows how learning effects operating

on the block-wide and experiment-wide time scales affect the shape of the hazard function

in the baseline condition TP3.

Conjunction search: Descriptive statistics

In Figure 7 we present the data from one participant in the conjunction search task. In each

condition the hazard function rises to a peak, then declines, and finally keeps hoovering

around a non-zero value temporarily. The effect of Target Presence on h(t) is clearly present

in the left-tail of the distributions (i.e., when hazard is rising), and this effect is now clearly

increasing with increases in set size. The ca(t) functions show that responses emitted before

400 ms in the target-absent condition tend to be false alarms. At the same time, they show "small dips" in early ca(t) for target-present trials, i.e., small temporary increases in the miss rate (early misses) at the time when ca(t) for target-absent trials reaches 1.

Inspection of the descriptive functions h(t) and ca(t) showed that individuals can differ in at least three aspects (see Figure 8). First, three observers show a lot of early false alarms (subjects 4, 5, 6), while the remaining observers show few-to-no early false alarms. Those subjects who show early false alarms also show an early, temporary increase in the miss rate (early misses; i.e., a small dip in ca(t) for TP) at the time when ca(t) for target-absent trials reaches 1. Second, three observers (subjects 3, 6, 10) show a very large effect of target presence on h(t) and S(t) for set sizes 12 and 18, while the remaining observers show a smaller effect (compare subjects 10, 2, 4, and 5 in Figure 8). Third, those subjects that show few-to-no early errors tend to emit the earliest responses a bit later compared to those who do show early false alarms. Note that subject 4 was very fast overall. Regardless of these individual differences, for each participant target-present responses were faster on average than target-absent responses and this difference increased with set size. Also, for many subjects misses started to emerge in the later bins for the larger set sizes, while there were few-to-no late false alarms. Finally, the effect of set size on h(t) was visible only in the left tail of the distribution, and not in the flat right tail. Note that the location of the sample means is not systematically related to any feature of the shape of the RT distributions.

Conjunction search: Inferential statistics

Table 3 (columns 3 to 6) shows the selected hazard model for the conjunction task based on the aggregated data of 9 subjects (subject number 4 was ignored because of a lot of missing data in the later bins which led to model fitting failures). Figure 9 presents the predicted

(i.e., model-based) hazard functions (top row), cloglog-hazard functions (second row), and the corresponding survivor (third row) and probability mass functions (bottom row), for each set size in target-present and -absent trials for trial 350 in block 8. The first seven parameter estimates in Table 3 model the shape of the cloglog[h(t)] function for TP3 (the reference condition) in trial 350 of block 8 (Figure 9, top left, green line) using a 6th order polynomial function of TIME. Parameters 2-7 show a significant linear, quadratic, cubic, quartic, quantic, and sextic effect of TIME on this intercept estimate, such that the predicted response hazard first increases with increasing waiting time, and then decreases towards a non-zero asymptote (Figure 9).

With respect to the manipulations of interest, we see that in bin 500 and relative to the reference condition TP3, there is a main effect of removing the target (parameter 8, PE = -0.44, HR = 0.64, $p < .0001$), and main effects of changing the set size to 6 (parameter 13, PE = -0.25, HR = 0.78, $p < .0001$), to 12 (parameter 17, PE = -0.74, HR = 0.48, $p < .0001$), and to 18 (parameter 21, PE = -1.13, HR = 0.32, $p < .0001$). All these main effects change significantly with TIME (parameters 8 to 23). For example, the effect of Target Absent changes in quartic fashion (parameters 9 to 12) so that it equals -0.44 in bin 500 (HR = 0.64, $p < .0001$), 0.11 in bin 650 (HR = 1.12, $p < .001$), 0.18 in bin 800 (HR = 1.20, $p < .005$), and only 0.01 in bin 950 (HR = 1.01, $p = .92$). The effect of SS6 changes in a cubic fashion (parameter 14 to 16), the effect of SS12 changes in a cubic fashion (parameters 18 to 20), and the effect of SS18 changes in a quadratic fashion (parameters 22 to 23).

There are also interaction effects between Target Absent and each of the three set sizes, which change over TIME (parameters 24-33). In contrast to the feature search data, these interaction effects are now negative, and their absolute size in each bin increases with increasing set size. Furthermore, both these interaction effects and the three main effects of

set size are (a) negative in bin 500, (b) decrease in absolute size over time, (c) are larger for larger set sizes, and (d) remain significant for a longer time after display onset for larger set sizes. In other words, the h(t) functions show a *partial* ordering with respect to the systematic effects of set size and target presence (i.e., in general only for t < 1000 ms).

Finally, hazard also varies on the across-trial/within-block time scale, and on the across-block/within-experiment time scale. First, each additional series of 10 trials will increase the estimated cloglog[h(t)] value with 0.0021 units (parameter 34, column 3, p < .0001) in each bin. Second, each additional block will increase the estimated cloglog[h(t)] value with 0.048 units in bin 500 (parameter 35, column 3, p < .0001), and this effect decreases linearly with TIME (parameter 36, PE = -.0021, p < .0001). Figure 6B shows how the effect of Trial affects the shape of the hazard function in the baseline condition within Blocks 1 and 8 when changing from trial 10 to trial 350.

Spatial configuration search: Descriptive statistics

In Figure 10 we present the data from one participant in the spatial configuration search task. Instead of only peaked hazard functions, we now see also monotonically increasing hazard functions for the larger set sizes. The effect of Target Presence on h(t) is clearly present in the left tail of the distributions and the difference between the target-present and -absent hazard functions is lasting longer for larger set sizes. The ca(t) functions show that many responses emitted before 1000 ms in the target-absent condition tend to be false alarms. At the same time, they show "small dips" in early ca(t) for target-present trials. Furthermore, for the larger set sizes 12 and 18 the miss rate starts to increase over time around 1500 ms after display onset.

Comparing individuals (Figure 11) shows that four individuals show many early false alarms coupled with early misses (subjects 1, 7, 8, and 9). The remaining subjects show few-to-no false alarms and their fastest responses appear somewhat later compared to the other subjects. Also, subjects 3 and 7 showed very slow behavior when the target is absent for set sizes 12 and 18 (i.e., hazard functions that start to rise late and at a very low rate). Regardless of these individual differences, for each subject target-present responses were on average faster than target-absent responses and this difference increased with set size. Finally, all subjects show late misses for the larger set sizes, appearing around 1500 ms after display onset.

Spatial configuration search: Inferential statistics

Table 4 (columns 3 to 6) shows the selected hazard model for the spatial configuration task. Figure 12 presents the predicted (i.e., model-based) hazard functions (top row), cloglog-hazard functions (second row), and the corresponding survivor (third row) and probability mass functions (bottom row), for each set size in target-present and -absent trials for trial 350 in block 8. In the baseline condition (TP3) the predicted response hazard first increases with increasing waiting time, and then decreases to a non-zero value (Figure 12).

With respect to the manipulations of interest, we see that in bin 640 and relative to the reference condition TP3, there is a main effect of removing the target (parameter 8, PE = -0.55, HR = 0.58), and main effects of changing the set size to 6 (parameter 13, PE = -0.76, HR = 0.47), to 12 (parameter 18, PE = -1.52, HR = 0.22), and to 18 (parameter 23, PE = -2.05, HR = 0.13) and interaction effects between Target Absent and set size 6 (parameter 28, PE = -0.83, HR = 0.43), set size 12 (parameter 31, PE = -2.06, HR = 0.13), and set size 18 (parameter 34, PE = -1.99, HR = 0.14), with all $p < .0001$.

Furthermore, all these effects interact with TIME in a significantly linear, quadratic, cubic and/or quartic fashion (see Table 4). As a result, the effect of target presence and the systematic effect of set size on h(t) in the target-present condition (i.e., SS3 > SS6 > SS12 > SS18) are gone around 1500 ms after search display onset (parameter rows 8, 13, 18, and 23 in Table 4). In contrast to the conjunction search task, the interaction effects between Target Absent and each set size (parameter rows 28, 31, and 34) do not quickly decrease in absolute size over time (before 1600 ms). In sum, the h(t) functions show a *partial* ordering with respect to the systematic effects of set size and target presence.

Finally, as shown in Figure 6C, each additional series of 10 trials increases the estimated cloglog[h(t)] value in bin 640 with only 0.00001 units (parameter 38, column 3, p = .98) but this effect increases linearly with TIME (parameter 39, PE = .00028, p < .01), so that each additional series of 10 trials increases the estimated cloglog[h(t)] in bin 1600 with 0.00342 units (parameter 38, column 11, p < .001). Second, each additional block will increase the estimated cloglog[h(t)] value with 0.054 units in bin 640 (parameter 40, column 3, p < .0001), and this effect decreases linearly with TIME (parameter 41, PE = -.00145, p < .0001).

Discussion

To study the temporal dynamics of visual search behavior we applied descriptive and inferential discrete-time event history analyses to published benchmark RT data from three search tasks. To study whether correct or error responses occur we also plotted the ca(t) or micro-level speed-accuracy tradeoff functions, next to the discrete-time h(t) or hazard functions of response occurrence.

Based on the results we draw four conclusions. First, event history analysis is a useful statistical technique to analyze RT data as it can detect differences that remain hidden when comparing mean RTs, such as the systematic but temporary effect of set size on h(t) in the feature search task. It is now clear that many – if not all – experimental manipulations lead to effects that change over time, whether in the context of masked response priming (Panis & Schmidt, 2016), simultaneous masking (Panis & Hermens, 2014), or object recognition (Panis, Torfs, Gillebert, Wagemans, & Humphreys, 2017; Panis & Wagemans, 2009; Torfs, Panis, & Wagemans, 2010). While many assume that RTs reflect the cumulative duration of all time-consuming cognitive operations involved in a task (e.g., Liesefeld, 2018; Song & Nakayama, 2009) our results show that fast, medium, and slow RTs can actually index different sets of cognitive operations. Due to the advantages of this method (illustrated in the current work) we recommend that it is used more often in future empirical and simulated RT studies[7]. Second, there are clear individual differences in the presence of a systematic pattern of early false alarms and early misses. Third, the hazard modeling results suggest differences between the underlying processes in the three search tasks, and provide strong constraints for future cognitive modeling efforts. Fourth, there is only a partial ordering of the hazard functions with respect to the effects of set size and target presence, and the hazard functions are relatively flat for the right tail of the RT distributions in all three search tasks.

No pop-out in h(t) for the feature search task

Why is there a systematic but temporary effect of set size (i.e., SS3 > SS6 > SS12 > SS18) on early h(t) for feature search when the target is present (Figure 5) although there is no effect of set size on mean correct RT (Figure 1)? At least three factors related to object recognition

that were not controlled by Wolfe et al. (2010) might be at play. First, the eccentricity of the target varies from trial to trial, and it is known that peripheral targets take a longer time to be recognized than foveal ones. Second, differences in set size are confounded with differences in density. This means that the receptive field of a single high-level visual neuron might only contain 1 or 2 objects for set size 3, but much more objects for set size 18. As color sensitivity is lower in the periphery, it is likely that visual crowding of the eccentric target occurred with large set sizes in many target-present trials. Third, because the search display was presented until response, more eye-movements could have been made with larger set sizes. If this is the case then the distance between the target location and the eye gaze location will have varied across the within-trial time (i.e., gaze-to-target distance is a time-varying covariate).

A small trend for the reversed effect of set size (i.e., SS3 < SS6 < SS12 = SS18) on early h(t) for feature search was found when the target is absent. This finding is consistent with the proposal that distractor-distractor feature similarity, next to target-distractor feature similarity, plays a role in visual search (Duncan & Humphreys, 1989). Because homogeneous distractors tend to group perceptually based on their high feature similarity they can be rejected together and this can explain why target-absent mean RTs sometimes decrease with increasing set size (Cheal & Lyon, 1992; Duncan & Humphreys, 1989; Humphreys & Müller, 1993).

Attentional capture and cognitive control processes in visual search

We noted that a subset of the observers in each task – those who tended to respond very early on some trials – showed early false alarms coupled with early misses. More specifically, we can distinguish at least three states in the ca(t) behavior of these fast-onset responders,

as can be seen clearly in the lower panels of Figures 3, 7, and 10. First, the very fast responses show false alarms (ca(t) ≤ .50) when the target is absent coupled with perfect performance (ca(t) = 1) when the target is present. In other words, these very fast responses display a strong yes-bias, independent from target presence. Second, after this initial ca(t) state the slower – but still relatively fast – responses show perfect accuracy when the target is absent, and a small but temporary increase in the miss-rate when the target is present. Third, after this second ca(t) state responses with intermediate latencies show high accuracy for both target-present and target-absent trials. In the conjunction and spatial configuration search tasks the slower responses in a fourth ca(t) state display a developing "no"-response bias especially for the larger set sizes. In other words, when the search task is difficult, the slower responses show virtually no false alarms and a gradual increase over time of the miss rate.

The results of Kiss, Grubert, and Eimer (2012) provide a likely explanation for the initial yes-bias in the first ca(t) state (see also Lee, Leonard, Luck, & Geng, 2018). They concluded that the attentional selection of targets that are defined by a combination of features – here: "red" and "vertical" in the feature and conjunction search tasks – is a two-stage process: Attention is initially captured by all target-matching features, but is then rapidly withdrawn from distractor objects that share some but not all features with the current target. This means that at the end of the feedforward sweep of the initial neural responses along the ventral and dorsal pathways right after display onset, all elements in the search display will have captured attention to some extent, each signaling the presence of target feature(s) such as red and vertical in the feature and conjunction search tasks, or combinations of left and right curvature in the spatial configuration task. This explains the presence of the early "yes"-response bias in the first ca(t) state of the fast-onset responders.

But why are these early false alarms followed by early temporary misses in the second ca(t) state? If we assume that online error-monitoring processes can detect the task-interfering "yes"-response bias in the earliest response tendencies, then *reactive* cognitive control processes can kick in (Braver, 2012)[8]. Panis and Schmidt (2016) used EHA to show that RT and accuracy distributions are shaped by active and selective response inhibition of premature response tendencies. Thus, it seems that for those participants that display early overt false alarms, this premature "yes"-response tendency is actively (i.e., top-down) and selectively inhibited – resulting in a temporary disinhibition of the competing "no"-response which would lead to an overt no-response if a momentary threshold is crossed –, which explains the observed small, early and temporary increase in the miss rate in target-present trials, and the concurrent almost complete absence of false alarms in the target-absent trials during the second ca(t) state. Crucially, the early difference in h(t) between target-present and -absent conditions might then be caused partially by a response competition process because both responses will be activated in target-absent trials, and not completely by the fact that target absence is confirmed slower on average than target presence as assumed in serial exhaustive search models.

In other words, at any point in time the hazard of response occurrence and conditional accuracy are not only determined by information from the search process but also by cognitive control processes (see Panis & Schmidt, 2016). As time passes on without response occurrence then the chance that target presence is correctly confirmed or rejected increases and this search information is additionally influencing the ongoing decision process (Cisek & Kalaska, 2010). Responses during the third ca(t) state are therefore dominated by information from the search process (i.e., selective response inhibition signals are overridden by response activation signals from the search outcome) and they thus show high

accuracy in both target-present and target-absent trials. Finally, as time passes on response-free and target presence is not yet confirmed or rejected, then search is aborted and a no-bias is developing for the slower responses during a fourth ca(t) state in the conjunction and spatial configuration tasks.

Those observers who show no early errors probably have better *proactive* control in terms of global (or aselective) response inhibition (Panis & Schmidt, 2016)[8]. In other words, these observers are proactively and globally inhibiting both the correct and incorrect response channels until reliable information about the search outcome is available. This hypothesis is consistent with the observation that the earliest responses of these observers are emitted somewhat later in time compared to the earliest responses of the observers who show early errors.

Serial versus parallel selection

While there is a general consensus that the current color feature task relies on parallel selection, and that the current spatial configuration task relies on serial selection, this is not the case for the color-orientation conjunction task. According to feature integration theory (Treisman & Gelade, 1980) attentional selection is serial because of the need to bind both surface features for recognition. However, there are many studies that suggest that certain feature conjunctions can actually be detected in parallel (Eckstein, 1998; McElree & Carrasco, 1999; Mordkoff, Yantis, & Egeth, 1990; Pashler, 1987; Sung, 2008). Although our hazard modeling results provide no answer to this issue, they do show task differences. First, the effect of trial number on hazard was similar for the conjunction and spatial configuration task, and different for the feature search task. Second, the interactions involving set size and time became more complex with task difficulty. These observations argue against the

proposal that differences between search tasks might be due to purely quantitative differences in target discriminability (Haslam, Porter, & Rothschild, 2001; Liesefeld et al., 2016; Wolfe, 1998).

Perhaps the question whether search is parallel or serial is ill-posed. It is possible that search actually involves parallel selection early in time as reflected in fast responses < ~500 ms, and serial selection later in time as reflected in slower responses > ~500 ms. Indeed, Li, Kadohisa, Kusunoki, Duncan, Bundesen, and Ditlevsen (2018) found that neurons show parallel processing early after search display onset (related to the initial feedforward sweep of neural activity after display onset), whereas they show serial processing later on (related to attentional effects in recurrent feedback connections where all processing capacities are focused on the attended object; see also Gabroi and Lisman, 2003). It is also possible that sequences of discrete attentional shifts emerge automatically from a parallel neural dynamic architecture that operates in continuous time (Grieben et al., 2018).

Effects of set size due to recurrent object recognition and cognitive control processes

According to Reverse Hierarchy Theory (Hochstein & Ahissar, 2002) feature search "pop-out" is attributed to high-level areas where large receptive fields underlie spread attention detecting categorical differences. Search for conjunctions or fine discriminations depends on reentry to low-level specific receptive fields using focused attention. Similarly, Nakayama and Martini (2011) proposed that visual search relies on object recognition processes, with high level processing occurring very rapidly and often unconsciously. They consider object recognition as a problem of linear classification where high-level areas have to disentangle the representations of different object classes by extracting diagnostic feature dimensions. They propose that search tasks vary on a continuum depending on the computational trade-

off between detail of description (number of feature dimensions) and scope (number of objects). Feature search can be performed in a single glance for many objects (a large attentional window) as only one feature dimension is relevant. Configuration search takes time because many feature dimensions have to be extracted for each display element separately and the small attentional window thus moves serially from element to element. For example, to solve the current spatial configuration task spatial attention has to be focused on each stimulus to extract object-centered spatial reference frame information to distinguish a digit 2 from a digit 5. For conjunction search a few feature dimensions are relevant and therefore an intermediate-sized attentional window is used. For example, to solve the current conjunction task a time-consuming attention-based coupling between two neuronal populations might be necessary (one sensitive to color-position and the other to orientation-position) while only one population is necessary for the feature task (color-position; Grieben et al., 2018). Next to theories based on a strategically modifiable attentional window (Humphreys & Müller, 1993; Theeuwes, 1994; Treisman & Souther, 1985) others have proposed that the size of the attentional window is determined by inherent limitations of the system (Engel, 1977; Geisler & Chou, 1995; Hulleman & Olivers, 2017).

Palmer (1995) distinguished between four causes of a set size effect: (a) preselection factors such as target eccentricity and display density, (b) selection factors such as whether only one object or a group of objects can be selected, (c) postselection factors, and (d) decision processes (see also Liesefeld and Müller, 2019). We can add a fifth cause: increases in set size might result in stronger automatic response activation of the yes-response and a stronger selective response inhibition response due to reactive cognitive control. Similarly, if target recognition during search performance depends on reentry to lower-level populations

then set size will affect performance due to the link between the complexity of the feature

that distinguishes the target from distractors and the receptive field size of the neurons

coding for that feature (VanRullen, Reddy, & Koch, 2004). Future studies can use event

history analysis to study when and how these different factors affect the shape of the hazard

function of response occurrence. Temporally distinguishing the contributions from these

factors to h(t) can be done by adding relevant predictors like target eccentricity, density,

gaze-to-target distance, target-distractor similarity, feature complexity, working memory

capacity, etc. to a hazard model. In other words, by adding the necessary predictors to a

hazard model one can control for variation due to variables irrelevant to the research

question.

Search is aborted rather early

The systematic effect of set size (i.e., SS3 > SS6 > SS12 > SS18) on response hazards lasted

longer for more difficult search tasks. However, the systematic effects of both target

presence and set size on hazard are rather limited in time. That is, we observed a partial

ordering on the hazard functions (i.e., set size and target presence affected only the left tail

of the distributions). For example, for the feature, conjunction, and spatial configuration

search tasks the systematic effects of set size and target presence are gone around 500 ms, 1

s, and 2s after search display onset, respectively.

Thereafter, the system transitions to a state with flat hazard functions without

systematic effects of set size and target presence (see Figures 4, 8, and 11). Horizontally

shaped hazard functions point to exponentially distributed RTs (see Figure 2). Because this is

observed for every search task including feature search, it suggests that the constant hazards

in the right tail of the RT distributions are not related to the visual search process per se, but

to a decision-making process in general (Palmer et al., 2011). Hazard functions that show a peak and a flat right tail have been observed before (Holden et al., 2009). Based on the findings of Shenoy, Sahani, and Churchland (2013) we assume that these flat right tails reflect RT outliers during decision making. Shenoy et al. (2013) described neuronal motor activity from a dynamical systems perspective by studying single-trial neural trajectories in a state-space. They found that the neural state wanders before falling back on track in RT outlier trials so that the monkey hesitated for an abnormally long time before movement onset. Interestingly, Thompson, Hanes, Bichot, and Schall (1996) found that much of the RT variance in search tasks is due to postperceptual motor processing, perhaps to provide the adaptive advantage of allowing for subsequent visual processing and cognitive factors to alter the response choice (e.g., explicitly comparing the presumed target with a few surrounding distractors to confirm target presence) before an irrevocable commitment is made.

Recommendations for experimental design of RT and other time-to-event data studies

Two general recommendations can be made from the viewpoint of event history analysis when designing RT studies. First, always use the same fixed response deadline in each trial, for example 500 ms for single-button detection, and 800 ms for an easy two-button discrimination task. Because hazard analysis deals with right-censored observations, there is no need to wait for very slow responses that are considered meaningless and would be trimmed anyway. As a consequence, event history analysis also allows analyzing RT data in masking paradigms, attentional blink paradigm, etc., that is, in paradigms for which RT is typically not measured, let alone analyzed and reported (because typically no differences in mean RT are found for example). Also, using rather short and fixed response deadlines will

lead to individual distributions that overlap in time, which is important for h(t) and ca(t)

modeling (Panis & Schmidt, 2016). Furthermore, if you wait for a response in each trial and

let the overt response end the trial, then you allow subjects to have control over the trial

(and experiment) duration, which should be avoided unless this is part of the research

question. Second, try to design as many trials as possible per condition because then you can

use small bins and still obtain stable h(t) and ca(t) estimates (i.e., use a small-N design; Smith

& Little, 2018). Also, designing 100 trials per condition, for example, will not result in a large

increase in experiment duration since the response deadline and thus trial duration can be

kept short (see Panis & Schmidt, 2016).

Conclusions

RT and accuracy distributions are a rich source of information on the time course of

cognitive processing. The changing effects of our experimental manipulations with increases

in waiting time become strikingly clear when looking at response hazards and micro-level

speed-accuracy tradeoff functions. An event history analysis of time-to-event data can

strongly constrain the choice between cognitive models of the same phenomenon. We

suggest that future inclusion of recurrent object recognition, learning, and cognitive control

processes in computational models of visual search will improve the ability of such models to

account for RT distributions and to explain the differences in the time-dispersed behavior of

individual searchers.

Acknowledgements

Open practices statement

This article presents a reanalysis of publicly available data. The R code for the descriptive and inferential event-history analyses is available from the first author upon request.

References

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology, 13*, 61-98.

Allison, P. D. (2010). *Survival analysis using SAS: A practical guide, Second Edition*. SAS Institute Inc., Cary, NC, USA.

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science, 20* (3), 160-166.

Bloxom, B. (1984). Estimating response time hazard functions: An exposition and extension. *Journal of Mathematical Psychology, 28*, 401-420.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences, 16* (2), 106-113.

Burle, B., Vidal, F., Tandonnet, C., & Hasbroucq, T. (2004). Physiological evidence for response inhibition in choice reaction time tasks. *Brain and Cognition, 56*, 153-164.

Cheal, M., & Lyon, D. R. (1992). Attention in visual search: Multiple search classes. *Perception & Psychophysics, 52* (2), 113-138.

Chechile, R. A. (2003). Mathematical tools for hazard function analysis. *Journal of Mathematical Psychology, 47*, 478-494.

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience, 33*, 269-298.

Deco, G., & Zihl, J. (2006). The neurodynamics of visual search. *Visual Cognition, 14*, 1006-1024.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96* (3), 433-458.

Dutilh, G., et al. (2018). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. *Psychonomic Bulletin & Review*, published online: https://doi.org/10.3758/s13423-017-1417-2

Eckstein, M. P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science, 9* (2), 111-118.

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision, 11* (5):14, 1-36.

Engel, F. L. (1977) Visual conspicuity, visual search and fixation tendencies of the eye. *Vision Research, 17,* 95–108. doi: 10.1016/0042-6989(77)90207-3.

Eriksen, C. W., Coles, M. G. H., Morris, L. R., & O´hara, W. P. (1985). An electromyographic examination of response competition. *Bulletin of the Psychonomic Society, 23* (3), 165-168.

Fix, J., Rougier, N., & Alexandre, F. (2011). A dynamic neural field approach to the covert and overt deployment of spatial attention. *Cognitive Computation, 3*, 279-293.

Gabroi, D., & Lisman, J. (2003). Recognition by top-down and bottom-up processing in cortex: The control of selective attention. *Journal of Neurophysiology, 90*, 798-810.

Geisler, W. S. & Chou, K. L. (1995) Separation of low-level and high-level factors in complex tasks: Visual search. *Psychological Review, 102,* 356–78. doi: 10.1037/0033-295X.102.2.356.

Grieben, R., Tekülve, J., Zibner, S. K. U., Schneegans, S., & Schöner, G. (2018). Sequences of discrete attentional shifts emerge from a neural dynamic architecture for conjunctive visual search that operates in continuous time. In T. T. Rogers, Rau, M., Zhu, X., & Kalish, C. W. (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 429-434). Downloaded from

http://mindmodeling.org/cogsci2018/papers/0099/index.html

Haslam, N., Porter, M., & Rothschild, L. (2001). Visual search: Efficiency continuum or distinct

processes? *Psychonomic Bulletin & Review, 8*, 742-746. doi:10.3758/BF03196212

Heinke, D., & Backhaus, A. (2011). Modelling visual search with the selective attention for

identification model (VS-SAIM): A novel explanation for visual search asymmetries.

Cognitive Computation, 3, 185-205.

Heinke, D., & Humphreys, G. W. (2003). Attention, spatial representation and visual neglect:

Simulating emergent attention and spatial memory in the Selective Attention for

Identification model (SAIM). *Psychological Review, 110* (1), 29-87.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in

the visual system. *Neuron, 36*, 791-804.

Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times reveals

cognitive dynamics. *Psychological Review, 116* (2), 318-342.

Hulleman, J., & Olivers, C. N. L. (2017). The impending demise of the item in visual search.

The Behavioral and Brain Sciences, 40, e132. doi:10.1017/S0140525X15002794

Humphreys, G. W., & Müller, H. J. (1993). Search via Recursive Rejection (SERR): A

connectionist model of visual search. *Cognitive Psychology, 25*, 43-110.

Humphreys, G. W. (2016). Feature confirmation in object perception: Feature integration

theory 26 years on from the Treisman Bartlett lecture. *The Quarterly Journal of

Experimental Psychology, 69* (10), 1910-1940.

Kazanovich, Y., & Borisyuk, R. (2017). Reaction times in visual search can be explained by a

simple model of neural synchronization. *Neural Networks, 87*, 1-7.

Kiss, M., Grubert, A., & Eimer, M. (2012). Top-down task sets for combined features:

Behavioral and electrophysiological evidence for two stages in attentional object

selection. *Attention, Perception, & Psychophysics, 75* (2), 216-228.

Lee, J., Leonard, C. J., Luck, S. J., & Geng, J. J. (2018). Dynamics of feature-based attentional selection during color-shape conjunction search. *Journal of Cognitive Neuroscience, 30* (12), 1773-1787.

Li, K., Kadohisa, M., Kusunoki, M., Duncan, J., Bundesen, C., & Ditlevsen, S. (2018). Distinguishing between parallel and serial processing in visual attention from neurobiological data. bioRxiv preprint first posted online Aug. 2, 2018; doi: http://dx.doi.org/10.1101/383596.

Liesefeld, H. R. (2018). Estimating the timing of cognitive operations with MEG/EEG latency measures: A primer, a brief tutorial, and an implementation of various methods. *Frontiers in Neuroscience, 12*, Article 765.

Liesefeld, H. R., Moran, R., Usher, M., Müller, H. J., & Zehetleitner, M. (2016). Search efficiency as a function of target saliency: The transition from inefficient to efficient search and beyond. *Journal of Experimental Psychology: Human Perception and Performance, 42* (6), 821-836.

Liesefeld, H.R., & Müller, H.J. (2019). A theoretical attempt to revive the serial/parallel-search dichotomy. *Attention, Perception, & Psychophysics*. Advance online publication. doi: 10.3758/s13414-019-01819-z

Luce, R. D. (1986). *Response times. Their role in inferring elementary mental organization*. New York: Oxford University Press Inc.

McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. *JEP:HPP, 25* (6), 1517-1539.

Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology, 26*, 3-67.

Moran, R., Zehetleitner, M., Müller, H. J., & Usher, M. (2013). Competitive guided search: Meeting the challenge of benchmark RT distributions. *Journal of Vision, 13*(8):24, 1-31.

Moran, R., Zehetleitner, M., Liesefeld, H. R., Müller, H. J., & Usher, M. (2016). Serial vs. parallel models of attention in visual search: accounting for benchmark RT-distributions. *Psychonomic Bulletin & Review, 23*, 1300-1315.

Mordkoff, J. T., Yantis, S., & Egeth, H. E. (1990). Detecting conjunctions of color and form in parallel. *Perception & Psychophysics, 48* (2), 157-568.

Müller, H. J., Humphreys, G. W., & Donnelly, N. (1994). Search via Recursive Rejection (SERR): Visual search for single and dual form-conjunction targets. *Journal of Experimental Psychology: Human Perception and Performance, 20* (2), 235-258.

Nakayama, K., & Martini, P. (2011). Situating visual search. *Vision Research, 51*, 1526-1537.

Narbutas, V., Lin, Y.-S., Kristan, M., & Heinke, D. (2017). Serial versus parallel search: A model comparison approach based on reaction time distributions. *Visual Cognition, 25* (1-3), 306-325.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In: B. Kantowitz (Ed.), *Human information processing*, 41-82. Potomac, MD: Erlbaum.

Palmer, J. (1995). Attention in visual search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science, 4* (4), 118-123.

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance, 37* (1), 58-71.

Panis, S., & Hermens, F. (2014). Time course of spatial contextual interference: Event history analyses of simultaneous masking by nonoverlapping patterns. *Journal of Experimental*

Psychology: Human Perception & Performance, *40 (1)*, 129-144. Doi:10.1037/a0032949

Panis, S., & Schmidt, T. (2016). What is shaping RT and accuracy distributions? Active and selective response inhibition causes the negative compatibility effect. *Journal of Cognitive Neuroscience, 28* (11), 1651-1671.

Panis, S., Torfs, K., Gillebert, C. R., Wagemans, J., & Humphreys, G. W. (2017). Neuropsychological evidence for the temporal dynamics of category-specific naming. *Visual Cognition, 25* (1-3), 79-99. http://dx.doi.org/10.1080/13506285.2017.1330790

Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization and object identification of fragmented object outlines. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 661-687.

Pashler, H. (1987). Detecting conjunctions of color and form: Reassessing the serial search hypothesis. *Perception & Psychophysics, 41* (3), 191-201.

Praamstra, P., & Seiss, E. (2005). The neurophysiology of response competition: Motor cortex activation and inhibition following subliminal response priming. *Journal of Cognitive Neuroscience, 17* (3), 483-493.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Shenoy, K. V., Sahani, M., & Churchland, M. M. (2013). Cortical control of arm movements: A dynamical systems perspective. *Annual Review of Neuroscience, 36*, 337-359.

Singer, J. D., & Willett, J. B. (1991). Modelling the days of our lives: Using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. *Psychological Bulletin, 110 (2)*, 268-290.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modelling change and event occurrence*. New York: Oxford University Press Inc.

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-*N* design. *Psychonomic Bulletin and Review, 25*, 2083-2101.

Song, J.-H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences, 13* (8), 360-366.

Sung K. (2008). Serial and parallel attentive visual searches: Evidence from cumulative distribution functions of response times. *Journal of Experimental Psychology: Human Perception and Performance, 34* (6), 1372-1388.

Theeuwes, J. (1994). Endogenous and exogenous control of visual selection. *Perception, 23*, 429–440. doi:10.1068/p230429

Thompson, K. G., Hanes, D. P., Bichot, N. P., & Schall, J. D. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *Journal of Neurophysiology, 76* (6), 4040-4055.

Torfs, K., Panis, S., & Wagemans, J. (2010). Identification of fragmented object outlines: A dynamic interplay between different component processes. *Visual Cognition, 18* (8), 1133-1164.

Townsend, J. T. (1990a). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1* (1), 46-54.

Townsend, J. T. (1990b). Truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin, 108* (3), 551-567.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Treisman, A., & Sato, S. (1990). Conjunction Search Revisited. *Journal of Experimental Psychology: Human Perception and Performance, 16* (3), 459-478.

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General, 114*, 285–310. doi:10.1037/0096-3445.114.3.285

VanRullen, R., Reddy, L., & Koch, C. (2004). Visual search and dual tasks reveal two distinct attentional resources. *Journal of Cognitive Neuroscience, 16* (1), 4-14.

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review, 7* (3), 424-465.

Wenger, M. J., & Gibson, B. S. (2004). Using hazard functions to assess changes in processing capacity in an attentional cuing paradigm. *JEP:HPP, 30* (4), 708-719.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58*, 475-482.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*, 67-85.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1* (2), 202-238.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9*, 33–39. doi:10.1111/1467-9280.00006

Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In: W. D. Grey (Ed.), *Integrated Models of Cognitive Systems,* 99-119. New York, Oxford University Press, Inc.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *JEP:HPP, 15* (3), 419-433.

Wolfe, J. M., Palmer, E. M., & Horowitz, T. S. (2010). Reaction time distributions constrain models of visual search. *Vision Research, 50*, 1304-1311.

Willett, J. B., & Singer, J. D. (1993). Investigating onset, cessation, relapse, and recovery: Why you should, and how you can, use discrete-time survival analysis to examine event occurrence. *Journal of Consulting and Clinical Psychology, 61 (6)*, 952-965.

Willett, J. B., & Singer, J. D. (1995). It's déjà vu all over again: Using multiple-spell discrete-time survival analysis. *Journal of Educational and Behavioral Statistics, 20*, 41-67.

Footnotes

1 Left censoring occurs when all you know about an observation on a variable T is that it is *less* than some value. Interval censoring combines right and left censoring: An observation on a variable T is interval censored if all you know about T is that a<T<b, for some values of a and b (Allison, 2010). The most common type of right-censoring is "singly Type I censoring" which applies when the experiment uses a fixed response deadline for all trials. Type I means that the censoring time is fixed and under the control of the experimenter, and singly refers to the fact that all observations have the same censoring time (Allison, 2010).

2 As explained in the section Methods, fitting discrete-time hazard models requires an expansion of the standard person-trial oriented data set into a person-trial-bin oriented data set where each bin (of each trial of each participant) that is at risk of event occurrence contributes a single row. Predictors like set size do not vary over time within a trial and therefore will have the same value for each of the rows (i.e., bins) that belong to the same trial. Time-varying predictors like heart rate and EEG time-series data will have different values for each of the rows (e.g., the average of all physiological measurements obtained during a RT bin, possibly lagged to take into account a transmission delay). For more information see Singer and Willett (2003), pp. 426-442, and Allison (2010), pp. 243-246.

3 $S(t) = [1-h(t)]*[1-h(t-1)]*...*[1-h(1)]$, and $P(t) = h(t) * S(t-1)$. At time point zero, $S(0) = 1$, $P(0) = 0$, and $h(0)$ is undefined.

4 Random censoring occurs when observations are terminated for reasons that are not under the control of the experimenter. Standard methods require that random censoring be noninformative: for example, a trial that is censored at time rc should be representative of all those trials with the same values of the explanatory variables that survive to rc (Allison, 2010). For example, an equipment error during a trial will introduce random censoring that

is uninformative. However, when estimating the hazard of correct response occurrence, error responses introduce random censoring (and vice versa) that is very likely informative, because response channels are known to compete with one another (Burle, Vidal, Tandonnet, & Hasbroucq, 2004; Eriksen, Coles, Morris, & O'Hara, 1985; Praamstra & Seiss, 2005). In other words, a trial with an error response at time rc is not representative of all trials with the same values of the explanatory variables that survive to rc as the probability of correct response occurrence will be lower for the former than the latter around time rc (and the probability of correct response occurrence can be larger right after an error response due to error detection and error correction). Informative censoring can lead to severe biases (Allison, 2010).

5 The *ca*(t) function is identical to the micro-level speed-accuracy trade-off (SAT) function based on partitioning RT in bins, as discussed by Wickelgren (1977) and Pachella (1974) in the psychological literature.

6 The complementary log-log link is preferred over the logit link for a discrete-time hazard model when the events can in principle occur at any time during each time bin (Allison, 2010), which is the case for RT data: cloglog$[h(t)]$ = ln{-ln[1-$h(t)$]}. Inverse of the link: $h(t)$ = 1 - exp{-exp{cloglog[$h(t)$]}}.

7 R code to calculate the descriptive statistics and the inferential statistics used by event history analysis can be downloaded here (see Supplementary resources):

https://www.researchgate.net/publication/304069212_What_Is_Shaping_RT_and_Accuracy_Distributions_Active_and_Selective_Response_Inhibition_Causes_the_Negative_Compatibility_Effect

8 Braver (2012) discusses a dual-mechanisms framework for cognitive control. On the one hand, proactive control is about the anticipation and prevention of interference before it

occurs. This happens on a long (across-trial) time scale as goal-relevant information is actively maintained in a sustained manner in the prefrontal cortex in order to bias attention, perception and action. On the other hand, reactive control is about the detection and resolution of interference after stimulus onset. This occurs on a shorter (within-trial) time scale in a transient manner, for example when response and/or stimulus conflict is detected early in time in the anterior cingulate cortex.

Tables

Table 1. A life table for the 530 trials of subject 1 for condition TP3 in the feature task. The censoring time and bin size equal 1000 ms and 100 ms, respectively.

bin ID	bin rank t	#cens	#events	Risk set	$h(t)$	$S(t)$	$P(t)$	#corr	$ca(t)$
(0,100]	1	0	0	530	0	1	0	NA	NA
(100,200]	2	0	0	530	0	1	0	NA	NA
(200,300]	3	0	3	530	0.006	0.994	0.006	3	1
(300,400]	4	0	147	527	0.279	0.717	0.277	137	0.93
(400,500]	5	0	255	380	0.671	0.236	0.481	248	0.97
(500,600]	6	0	84	125	0.672	0.077	0.158	84	1
(600,700]	7	0	18	41	0.439	0.043	0.034	17	0.94
(700,800]	8	0	9	23	0.391	0.026	0.017	8	0.89
(800,900]	9	0	4	14	0.286	0.019	0.008	3	0.75
(900,1000]	10	8	2	10	0.200	0.015	0.004	2	1

Note. #cens = number of observations right-censored at the end of bin t. #events = number of observed responses in bin t. #corr = number of observed correct responses in bin t. NA = undefined. Note that $h(t)$ = $P(t)$ only until the first bin with observed responses. Because 8 trials were right-censored at 1000 ms, $h(t)$ does not reach 1, $S(t)$ does not reach 0, and $P(t)$ does not sum to 1 (Chechile, 2003).

Table 2. Parameter estimates and test statistics for the selected hazard model in the feature search task. The selected model was refitted three times with TIME centered on bin 280, bin 520, and bin 640, respectively.

Nr.	effect	(240,280] PE	p	(360,400] PE	std. err.	t	p	(480,520] PE	p	(600,640] PE	p
1	(Intercept)	-3.3239	0.0000	-0.6147	0.2147	-2.862	0.0042	-0.7098	0.0000	-1.7293	0.0000
2	TIME			0.2405	0.0734	3.276	0.0011				
3	TIME2			-0.1333	0.0033	39.880	0.0000				
4	TIME3			0.0204	0.0010	20.338	0.0000				
5	TIME4			-0.0024	0.0002	10.776	0.0000				
6	TIME5			0.0001	0.0000	8.902	0.0000				
7	TA	-1.6956	0.0000	-0.2483	0.0240	10.323	0.0000	-0.2060	0.0000	0.0895	0.1734
8	TIME:TA			0.0934	0.0138	6.764	0.0000				
9	TIME2:TA			-0.0676	0.0037	18.164	0.0000				
10	TIME3:TA			0.0171	0.0012	13.854	0.0000				
11	TIME4:TA			-0.0011	0.0001	11.254	0.0000				
12	SS6	-0.1641	0.0000	-0.1095	0.0215	-5.083	0.0000	-0.0549	0.0996	-0.0002	0.9959
13	TIME:SS6			0.0182	0.0093	1.954	0.0506				
14	SS12	-0.2735	0.0000	-0.1373	0.0249	-5.514	0.0000	0.0266	0.4717	0.0583	0.3911
15	TIME:SS12			0.0589	0.0118	4.971	0.0000				
16	TIME2:SS12			0.0015	0.0037	0.413	0.6794				
17	TIME3:SS12			-0.0009	0.0004	-2.383	0.0171				
18	SS18	-0.4380	0.0000	-0.1950	0.0230	-8.450	0.0000	-0.0153	0.6557	0.1012	0.0935
19	TIME:SS18			0.0704	0.0118	5.965	0.0000				
20	TIME2:SS18			-0.0035	0.0021	-1.657	0.0975				
21	TA:SS6	0.2926	0.0000	0.1703	0.0306	5.548	0.0000	0.0479	0.2763	-0.0744	0.3334
22	TIME:TA:SS6			-0.0407	0.0128	-3.167	0.0015				
23	TA:SS12	0.5966	0.0000	0.2069	0.0326	6.340	0.0000	0.0142	0.7658	0.0186	0.8260
24	TIME:TA:SS12			-0.0970	0.0176	-5.506	0.0000				
25	TIME2:TA:SS12			0.0109	0.0031	3.444	0.0006				
26	TA:SS18	0.7403	0.0000	0.2631	0.0324	8.112	0.0000	-0.0171	0.7168	-0.1005	0.2102
27	TIME:TA:SS18			-0.1262	0.0174	-7.218	0.0000				
28	TIME2:TA:SS18			0.0109	0.0030	3.621	0.0003				
29	Trial	-0.0050	0.0000	-0.0023	0.0004	-4.902	0.0000	0.0004	0.4731	0.0032	0.0054
30	TIME:Trial			0.0009	0.0001	4.713	0.0000				
31	Block	0.0449	0.0000	0.0275	0.0018	14.693	0.0000	0.0102	0.0001	-0.0071	0.1300
32	TIME:Block			-0.0057	0.0007	-7.322	0.0000				

Note. TIME2 = TIME*TIME; TIME3 = TIME*TIME*TIME; etc. Highlighted *p*-values indicate effects that had to be significant to stay in the model. PE = parameter estimate.

Table 3. Parameter estimates and test statistics for the selected hazard model in the conjunction search task. The selected model was refitted three times with TIME centered on bin 650, bin 800, and bin 950, respectively. Same conventions as in Table 2.

		(450,500]				(600,650]		(750,800]		(900,950]	
Nr.	effect	PE	std. err.	t	p	PE	p	PE	p	PE	p
1	(Intercept)	-0.83067	0.1529	-5.429	0.0000	-0.79880	0.0000	-1.26129	0.0000	-1.73059	.0000
2	TIME	0.31784	0.0335	9.460	0.0000						
3	TIME2	-0.16165	0.0062	25.758	0.0000						
4	TIME3	0.02151	0.0028	7.619	0.0000						
5	TIME4	-0.00003	0.0010	-0.034	0.9724						
6	TIME5	-0.00022	0.0001	-1.635	0.1020						
7	TIME6	0.00001	0.0000	2.262	0.0236						
8	TA	-0.44065	0.0277	15.898	0.0000	0.11170	0.0008	0.18047	0.0032	0.01110	.9168
9	TIME:TA	0.38436	0.0188	20.403	0.0000						
10	TIME2:TA	-0.09695	0.0097	-9.899	0.0000						
11	TIME3:TA	0.01177	0.0020	5.795	0.0000						
12	TIME4:TA	-0.00057	0.0001	-4.468	0.0000						
13	SS6	-0.25043	0.0264	-9.467	0.0000	0.02804	0.4185	0.26464	0.0000	0.08835	.5171
14	TIME:SS6	0.05858	0.0161	3.621	0.0003						
15	TIME2:SS6	0.01828	0.0067	2.697	0.0070						
16	TIME3:SS6	-0.00229	0.0006	-3.329	0.0009						
17	SS12	-0.73870	0.0272	27.115	0.0000	-0.25413	0.0000	0.13561	0.0177	0.27681	.0122
18	TIME:SS12	0.16024	0.0183	8.730	0.0000						
19	TIME2:SS12	0.00327	0.0067	0.484	0.6279						
20	TIME3:SS12	-0.00094	0.0006	-1.535	0.1246						
21	SS18	-1.12963	0.0280	40.245	0.0000	-0.55897	0.0000	-0.11059	0.0200	0.21551	.0262
22	TIME:SS18	0.21059	0.0159	13.206	0.0000						
23	TIME2:SS18	-0.00679	0.0021	-3.108	0.0019						
24	TA:SS6	-0.41021	0.0406	10.087	0.0000	-0.31111	0.0000	-0.23745	0.0038	0.22465	.1574
25	TIME:TA:SS6	0.08325	0.0312	2.663	0.0077						
26	TIME2:TA:SS6	-0.02440	0.0110	-2.209	0.0271						
27	TIME3:TA:SS6	0.00255	0.0009	2.608	0.0091						
28	TA:SS12	-1.03899	0.0530	19.572	0.0000	-0.71853	0.0000	-0.61508	0.0000	-0.26466	.0397
29	TIME:TA:SS12	0.19453	0.0428	4.536	0.0000						
30	TIME2:TA:SS12	-0.03783	0.0125	-3.021	0.0025						
31	TIME3:TA:SS12	0.00286	0.0009	2.937	0.0033						
32	TA:SS18	-1.17587	0.0555	21.154	0.0000	-0.99663	0.0000	-0.81739	0.0000	-0.63816	.0000
33	TIME:TA:SS18	0.05974	0.0153	3.892	0.0001						
34	Trial	0.00210	0.0005	3.537	0.0004	0.00210	0.0004	0.00210	0.0004	0.00210	.0004
35	Block	0.04790	0.0019	25.204	0.0000	0.04160	0.0000	0.03529	0.0000	0.02899	.0000
36	TIME:Block	-0.00210	0.0005	-4.007	0.0001						

Table 4. Parameter estimates and test statistics for the selected hazard model in the spatial configuration search task. The selected model was refitted three times with TIME centered on bin 960, bin 1280, and bin 1600, respectively. Same conventions as in Table 2.

Nr.	effect	(560,640] PE	std. err.	t	p	(880,960] PE	p	(1200,1280] PE	p	(1520,1600] PE	p
1	(Intercept)	-1.07600	0.2361	-4.556	0.0000	-0.43265	0.0302	-0.69074	0.0001	-1.79917	0.0000
2	TIME	0.32380	0.0158	20.425	0.0000						
3	TIME2	-0.06204	0.0046	13.255	0.0000						
4	TIME3	0.00842	0.0015	5.598	0.0000						
5	TIME4	-0.00095	0.0002	-4.049	0.0001						
6	TIME5	0.00004	0.0000	3.100	0.0019						
7	TIME6	-0.00000	0.0000	-2.191	0.0284						
8	TA	-0.54943	0.0261	21.022	0.0000	-0.32379	0.0000	-0.11251	0.0553	0.57951	0.0000
9	TIME:TA	0.15184	0.0144	10.480	0.0000						
10	TIME2:TA	-0.03992	0.0048	-8.264	0.0000						
11	TIME3:TA	0.00456	0.0005	7.811	0.0000						
12	TIME4:TA	-0.00013	0.0000	-6.801	0.0000						
13	SS6	-0.75653	0.0281	26.903	0.0000	-0.75585	0.0000	-0.20832	0.0004	0.76806	0.0000
14	TIME:SS6	-0.04975	0.0141	-3.510	0.0004						
15	TIME2:SS6	0.00782	0.0058	1.336	0.1812						
16	TIME3:SS6	0.00145	0.0008	1.785	0.0742						
17	TIME4:SS6	-0.00007	0.0000	-2.332	0.0197						
18	SS12	-1.51813	0.0326	46.565	0.0000	-1.68899	0.0000	-1.12858	0.0000	0.19046	0.1336
19	TIME:SS12	-0.07066	0.0175	-4.015	0.0001						
20	TIME2:SS12	-0.00604	0.0068	-0.877	0.3801						
21	TIME3:SS12	0.00389	0.0008	4.377	0.0000						
22	TIME4:SS12	-0.00015	0.0000	-4.823	0.0000						
23	SS18	-2.05118	0.0377	54.352	0.0000	-2.15339	0.0000	-1.52742	0.0000	-0.19681	0.1253
24	TIME:SS18	-0.06195	0.0211	-2.933	0.0034						
25	TIME2:SS18	-0.00244	0.0077	-0.315	0.7525						
26	TIME3:SS18	0.00347	0.0009	3.528	0.0004						
27	TIME4:SS18	-0.00014	0.0000	-4.020	0.0001						
28	TA:SS6	-0.83385	0.0458	18.193	0.0000	-0.50727	0.0000	-0.47555	0.0000	-0.73869	0.0000
29	TIME:TA:SS6	0.11850	0.0209	5.655	0.0000						
30	TIME2:TA:SS6	-0.00921	0.0020	-4.465	0.0000						
31	TA:SS12	-2.05803	0.0869	23.670	0.0000	-1.29614	0.0000	-0.99413	0.0000	-1.15197	0.0000
32	TIME:TA:SS12	0.24795	0.0294	8.410	0.0000						
33	TIME2:TA:SS12	-0.01437	0.0023	-6.046	0.0000						
34	TA:SS18	-1.98806	0.1155	17.204	0.0000	-1.40328	0.0000	-1.39360	0.0000	-1.67945	0.0000
35	TIME:TA:SS18	0.24138	0.0551	4.379	0.0000						
36	TIME2:TA:SS18	-0.02670	0.0076	-3.488	0.0005						
37	TIME3:TA:SS18	0.00072	0.0003	2.406	0.0161						
38	Trial	0.00001	0.0008	0.020	0.9835	0.00115	0.0496	0.00228	0.0004	0.00342	0.0003
39	TIME:Trial	0.00028	0.0001	2.603	0.0092						

| 40 | Block | 0.05400 | 0.0019 | 27.109 | 0.0000 | 0.04819 | 0.0000 | 0.04238 | 0.0000 | 0.03656 | 0.0000 |
| 41 | TIME:Block | -0.00145 | 0.0002 | -5.342 | 0.0000 | | | | | | |

Figure captions

Figure 1. Benchmark visual search data set from Wolfe, Palmer, and Horowitz (2010). (A) Example visual search displays for three search tasks (copy of Fig. 1 in Wolfe et al., 2010). Participants search for the red vertical bar in the feature (left) and conjunction (middle) search tasks, and for a digital 2 among digital 5´s in the spatial configuration search task (right). (B) Mean correct RT for target-present (solid lines) and target-absent trials (dashed lines). Lighter lines show data for individual observers and darker lines show mean data (copy of Fig. 2 in Wolfe et al., 2010). (C) Empirical RT distributions for one observer in the spatial configuration search task. Set size is coded by lightness from the lightest lines, set size 3, through set sizes 6 and 12 to the darkest, set size 18 (copy of the lower panel in Fig. 4 in Wolfe et al., 2010). (D) Simulated RT distributions from a serial, self-terminating search model for target-present (solid) and target-absent (dashed) trials. Lighter lines represent smaller set sizes (copy of Fig. 7 in Wolfe et al., 2010).

Figure 2. Four views on waiting-time distributions. The cumulative distribution function (top left), the density function (top right), the survivor function (bottom left) and the hazard rate function (bottom right) are shown for each of four theoretical probability distributions (exponential, Weibull, gamma, log-normal). While the hazard function for the exponential is flat, it keeps increasing for the Weibull, it increases to an asymptote for the gamma, and it reaches a peak and then gradually decreases to an asymptote for the log-normal.

Figure 3. Descriptive statistics for subject number 3 in the feature search task. Top to bottom: h(t), S(t), P(t), and ca(t) for each set size (columns) and target presence (green=Target Present, red=Target Absent). Vertical lines in the h(t), S(t), and P(t) plots represent the sample mean RT, the estimated median RT or S(t)$_{.50}$, and the sample mean correct RT, respectively. Horizontal lines in the ca(t) plots represent overall accuracy.

Figure 4. Inter-individual differences in feature search. Estimates of h(t) and ca(t) for four participants in the feature search task, for Target-Present (left column) and Target-Absent (right column) trials and each set size (green=SS3, red=SS6, black=SS12, blue=SS18). Vertical lines in the h(t) plots represent the sample mean RT. (A) Subject 2. (B) Subject 3. (C) Subject 4. (D) Subject 8.

Figure 5. Hazard model predictions for feature search. Predicted h(t) functions for trial 350 in block 8 (first column) and the corresponding cloglog-hazard functions (second column), S(t) (third column) and P(t) (right column) functions, for Target-Present (top row) and Target-Absent (bottom row) trials. Vertical lines in the S(t) plots represent the estimated median RT or S(t)$_{.50}$.

Figure 6. Effect of practice on event occurrence. The model-based effects of Trial (T) and Block (B) are shown for the baseline condition (TP3) for the feature (A), conjunction (B) and spatial configuration (C) search tasks.

Figure 7. Descriptive statistics for subject number 5 in the conjunction search task. Same conventions as in Figure 3.

Figure 8. Inter-individual differences in conjunction search. Estimates of h(t) and ca(t) for four participants in the conjunction search task. (A) Subject 4. (B) Subject 5. (C) Subject 2. (D) Subject 10. Same conventions as in Figure 4.

Figure 9. Hazard model predictions for conjunction search. Predicted h(t) functions for trial 350 in block 8 (top row) and the corresponding cloglog-hazard functions (second row), S(t) (third row) and P(t) (bottom row) functions, for Target-Present (left column) and Target-Absent (right column) trials. Vertical lines in the S(t) plots represent the estimated median RT or S(t)$_{.50}$.

Figure 10. Descriptive statistics for subject number 8 in the spatial configuration search task. Same conventions as in Figure 3.

Figure 11. Inter-individual differences in spatial configuration search. Estimates of h(t) and ca(t) for four participants in the spatial configuration search task. (A) Subject 1. (B) Subject 8. (C) Subject 3. (D) Subject 4. Same conventions as in Figure 4.

Figure 12. Hazard model predictions for spatial configuration search. Same conventions as in Figure 9.
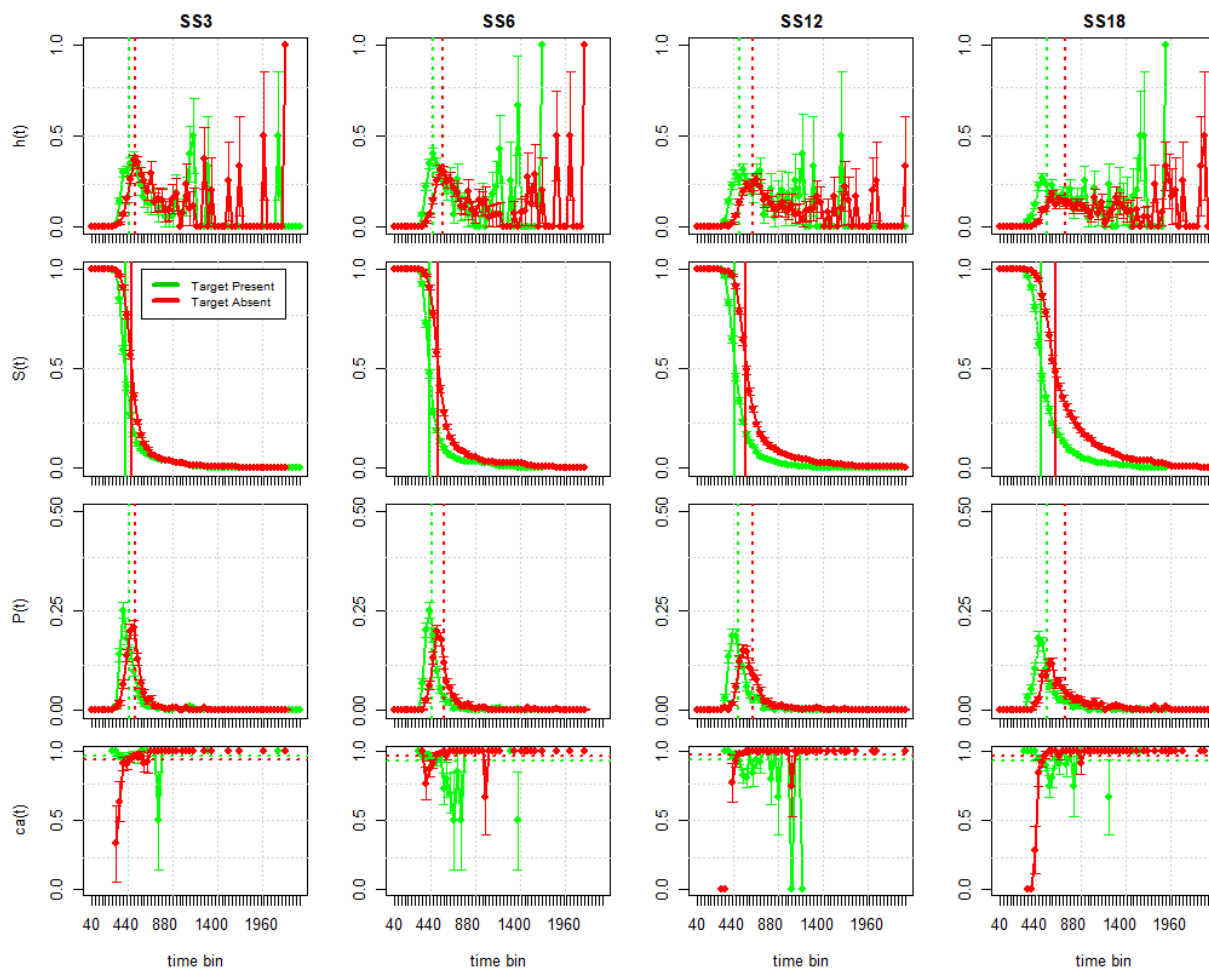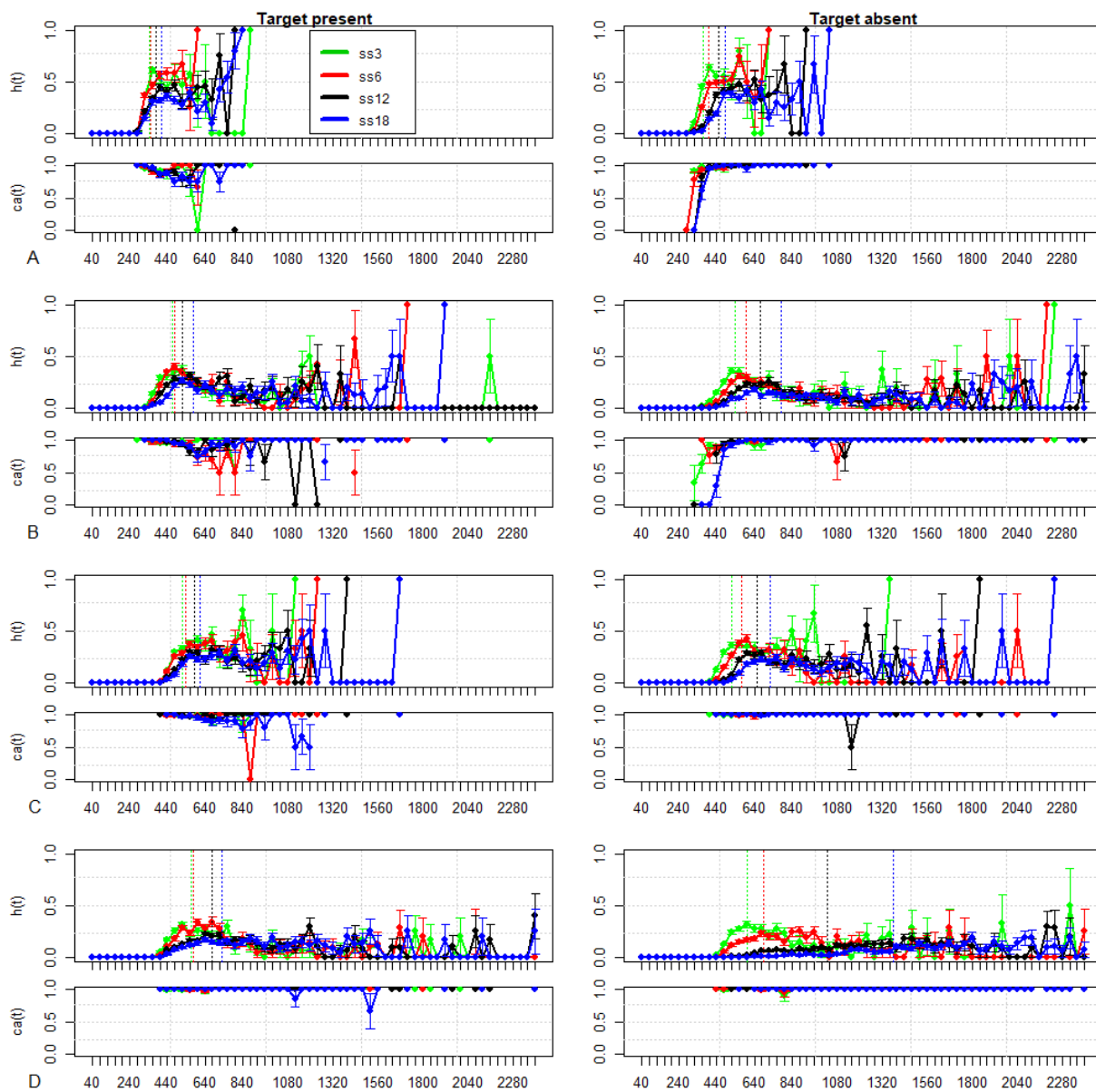
Figures
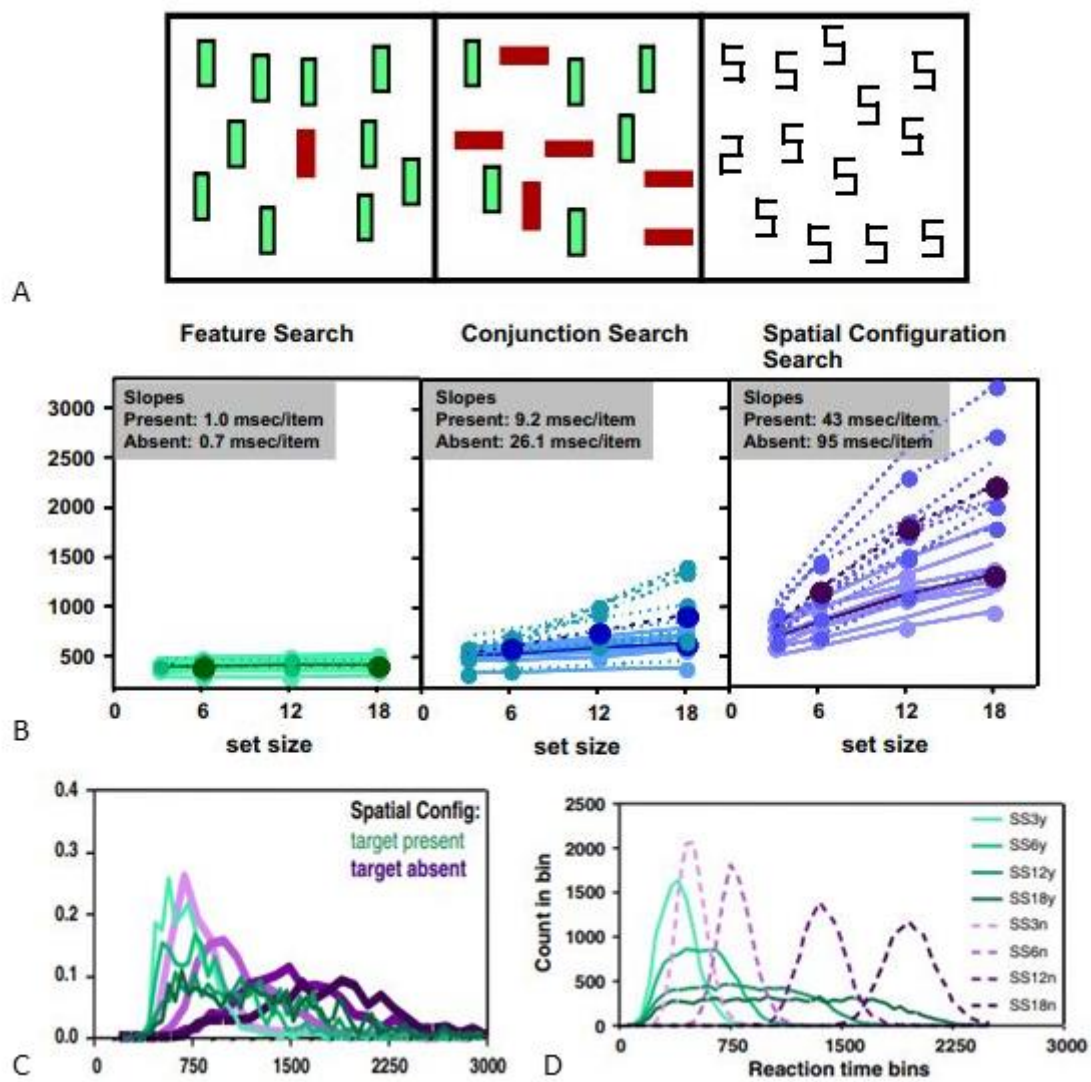
Figure 1
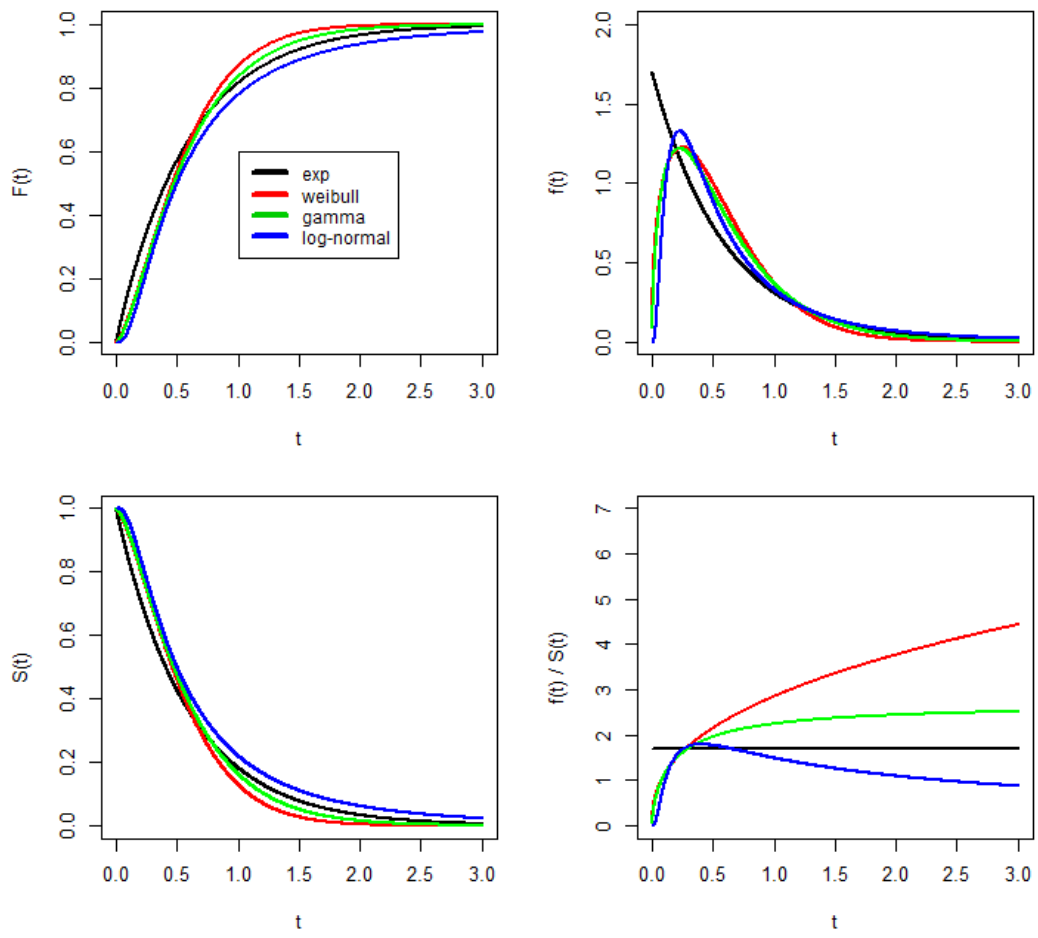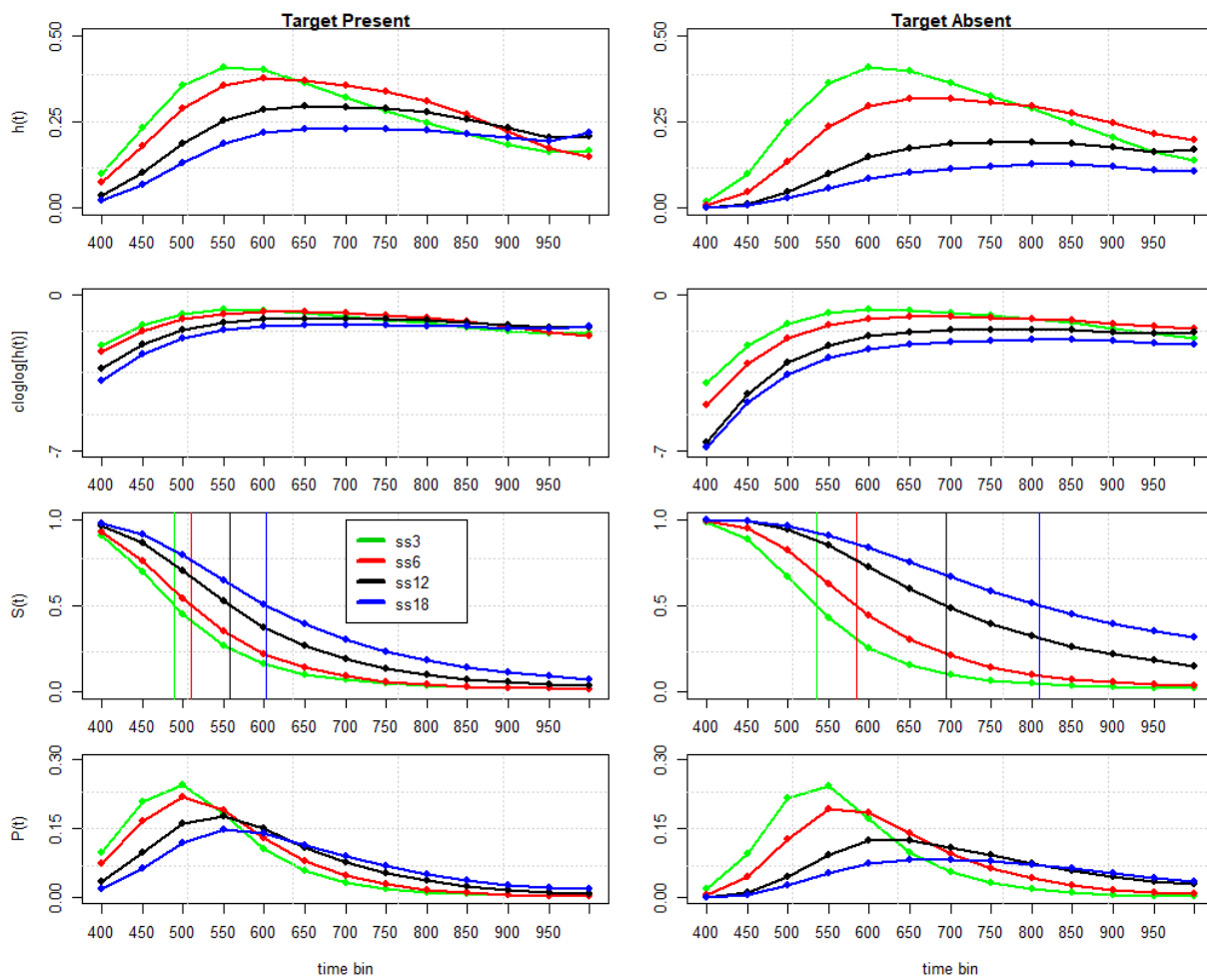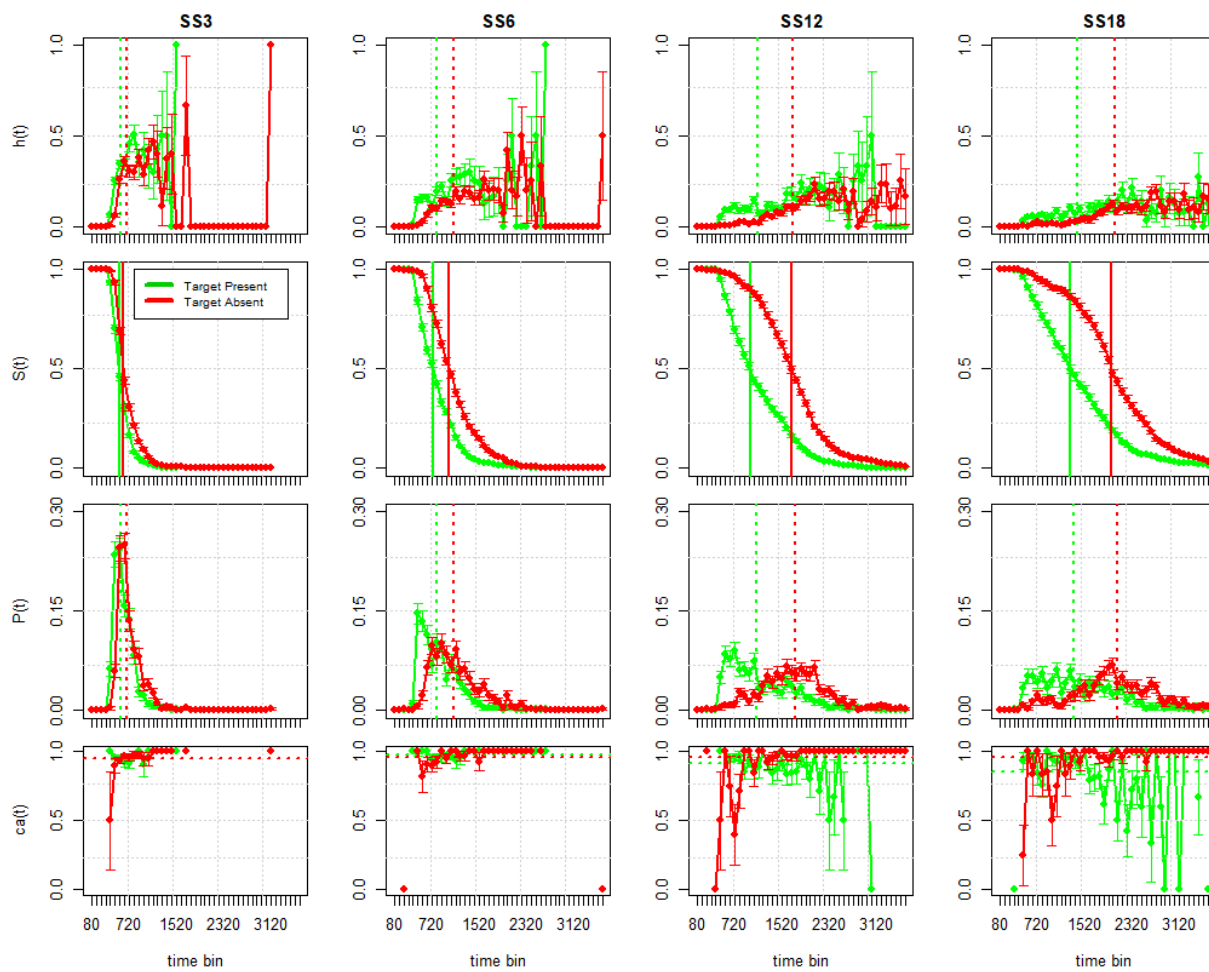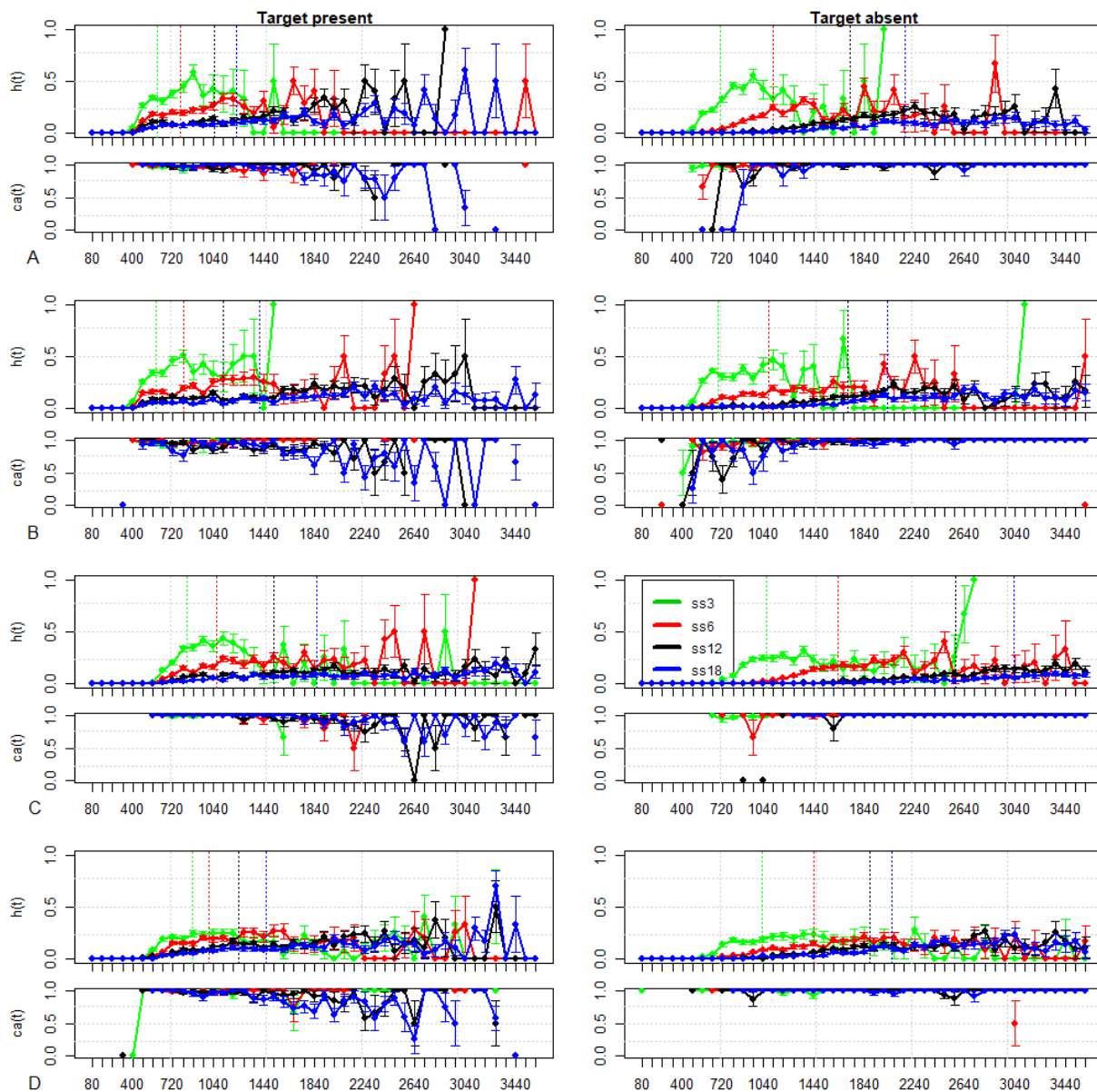
Figure 2
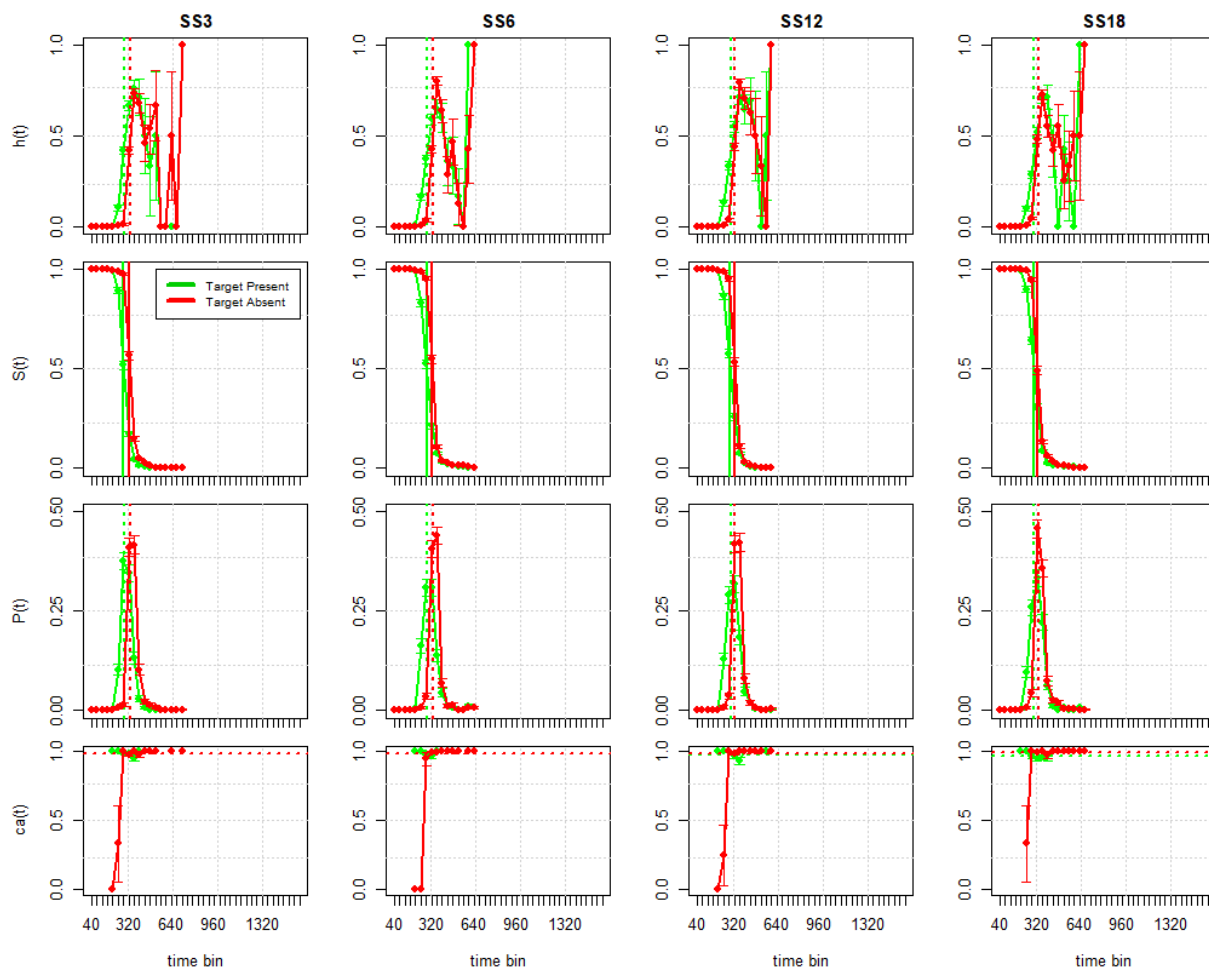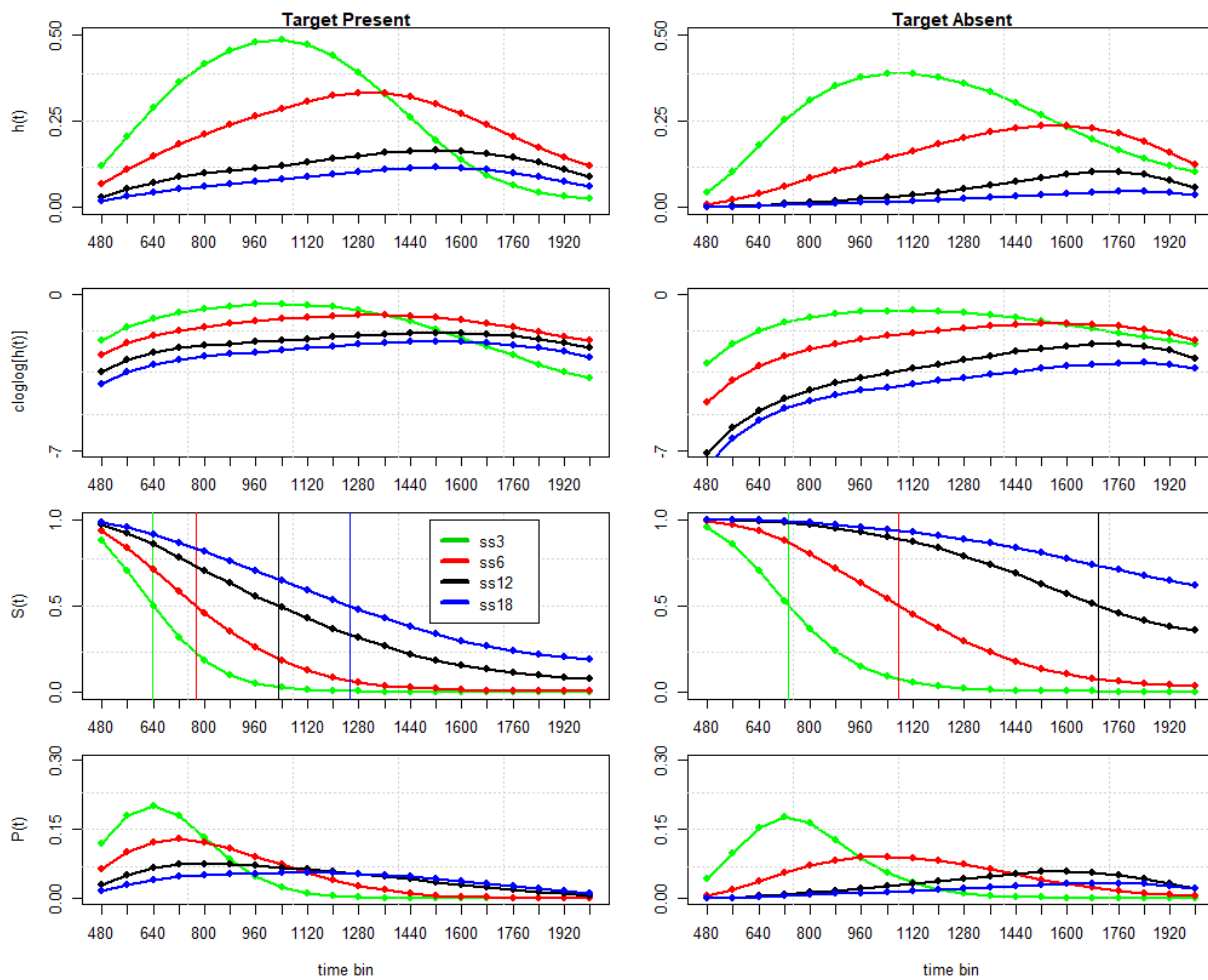
Figure 3
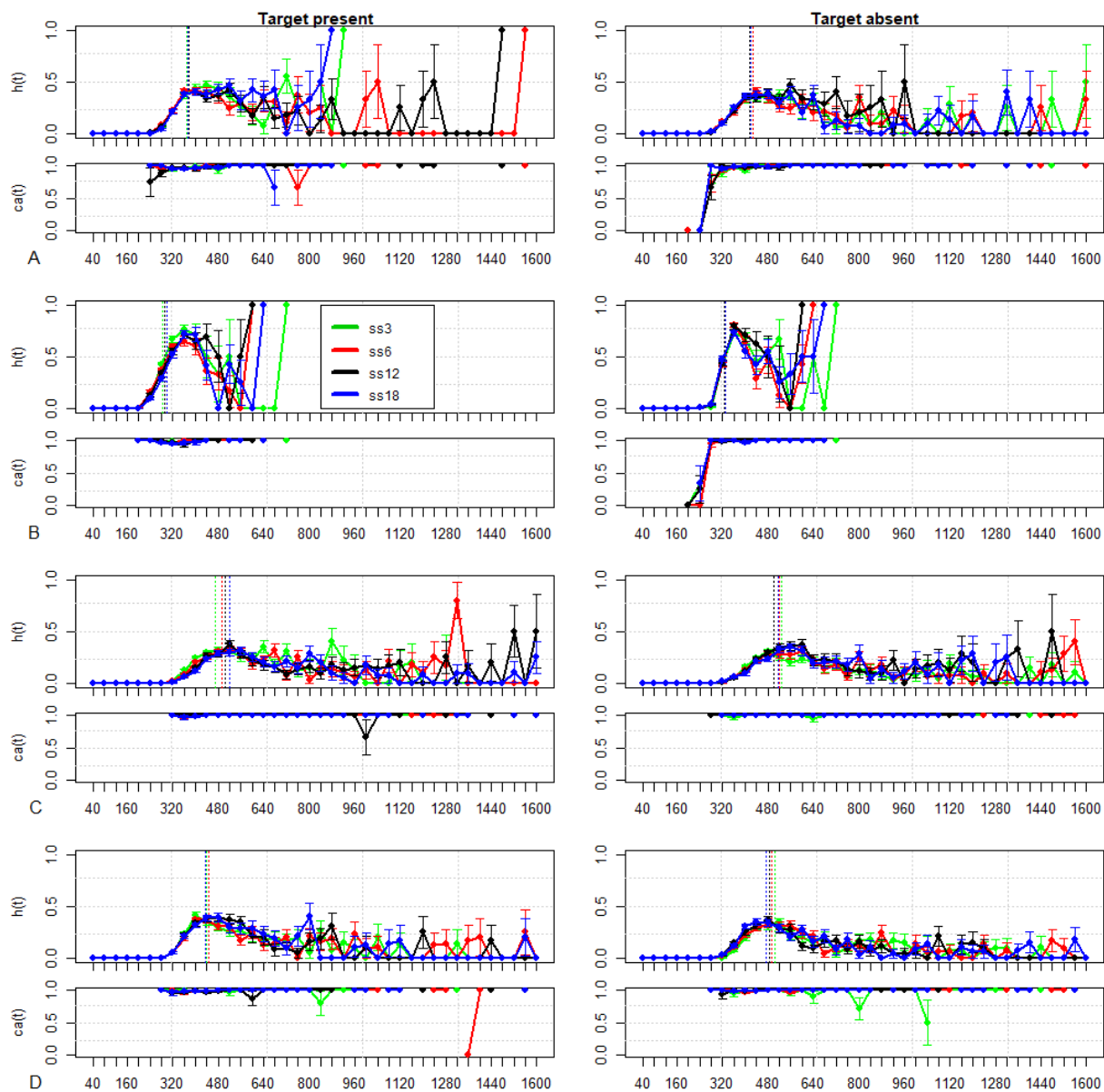
Figure 4
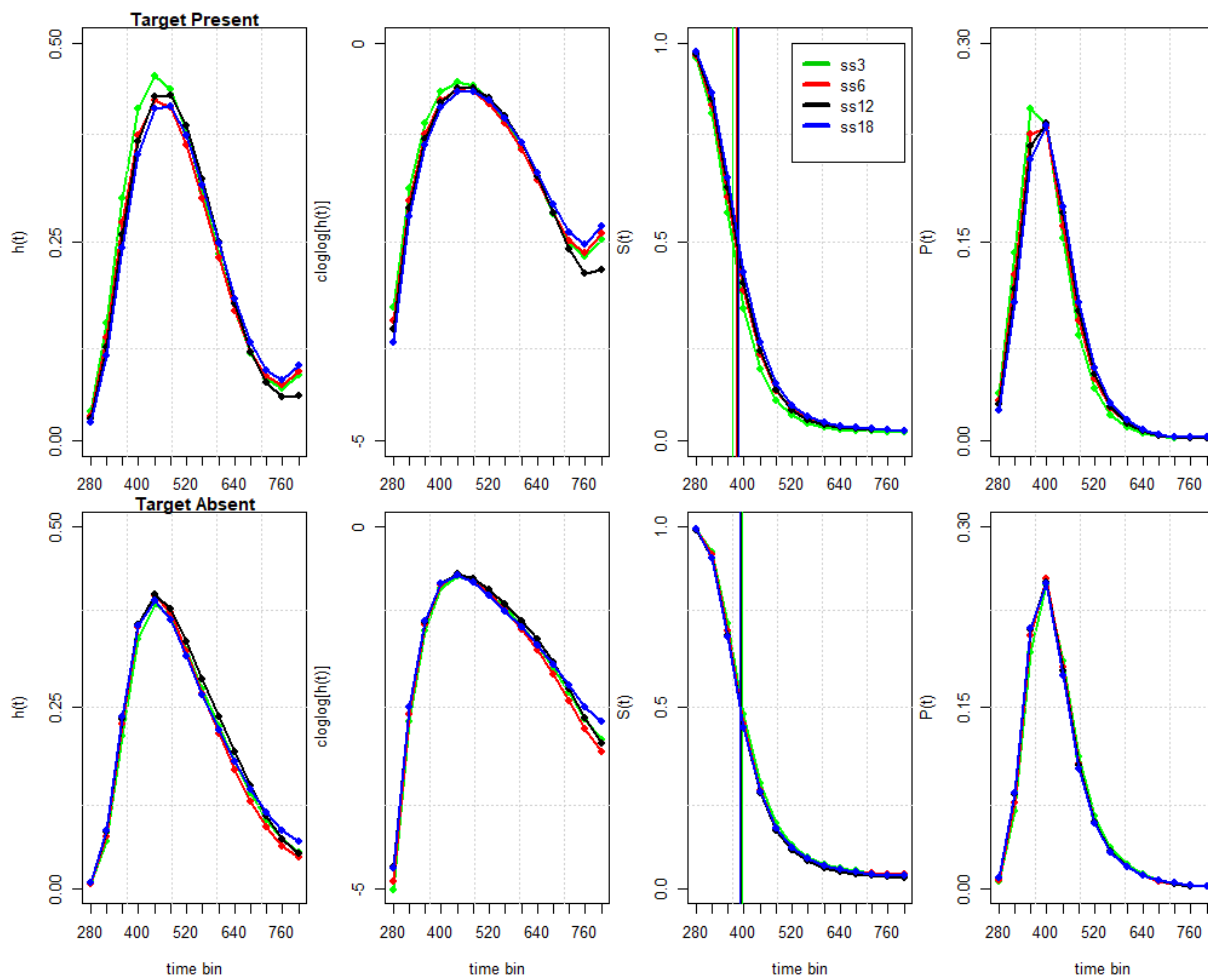
Figure 5
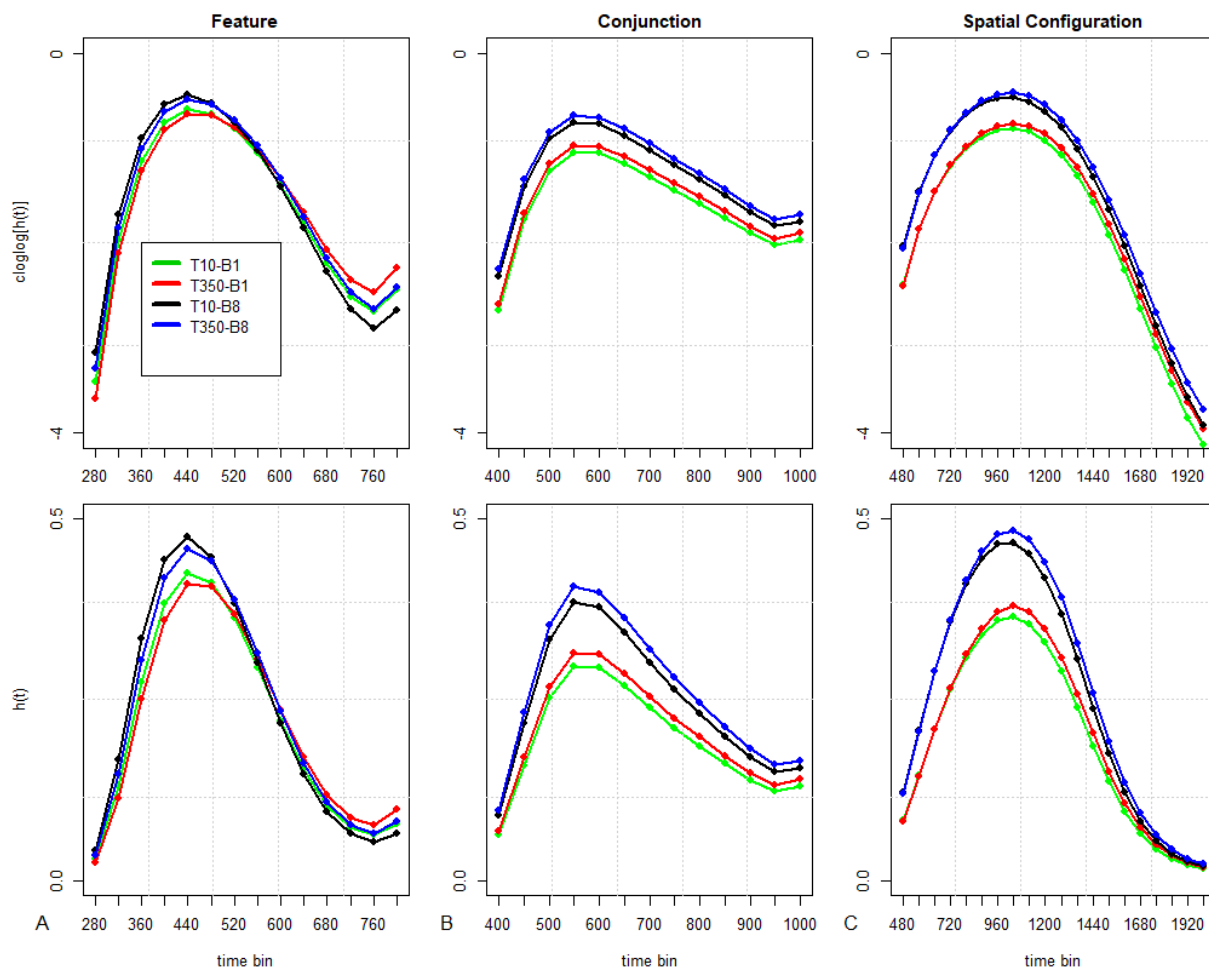
Figure 6

Figure 7

Figure 8

Figure 9

Figure 10

Figure 11

Figure 12