

TURBULENT TIMES: ARGUMENTS FOR ADVANCED
DISTRIBUTIONAL TIME ANALYSES

Vom Fachbereich Sozialwissenschaften
der Rheinland-Pfälzischen Technischen Universität, Campus Kaiserslautern
zur Verleihung des akademischen Grades
Doktor rerum naturalium (Dr. rer.nat.)
genehmigte

D i s s e r t a t i o n

vorgelegt von
Maximilian Philipp Wolkersdorfer

Tag der Disputation: Kaiserslautern, 14.05.2025

Dekan: Prof. Dr. Michael Fröhlich

Vorsitzende/r: Prof. Dr. Marcus Höreth

Gutachter/in: 1. Prof. Dr. Thomas Schmidt
2. Apl. Prof. Dr. Daniela Czernochowski

DE 386

Juli 2025

Acknowledgments

I want to dedicate this work to the people without whom I'd still be preparing my next lecture, my next meeting with the Nachwuchsring, or procrastinating by finding ways to "help" others with what I can't help myself with (not that they asked). Anything but this dissertation.

To my supervisor Thomas Schmidt. Without his continuous support, patience, and our often critical and enlightening discussions – not always related to our scientific work, but his passion for Jazz, movies, or just about the book he is currently reading – my doctoral studies would have been a lot duller.

To Sonja Karnel, the heart and brains of our group. You regularly save us from chaos and disaster.

To Sven Panis. You taught me most of what I know, are part of every line of this dissertation. Without you, none of this would have been possible. Thank you!

To the friends I made on the way. Our nights at the dinner or pool table alone made it worth it. Therefore, thank you to Boran, Chris, Sofia, (newly arrived) Lucia, and Omar, specifically. Not only did the two of us share an apartment (albeit for a short time), but also the grief and struggle that comes with a PhD and life itself. Thank you, for being there and helping me grow, personally and politically.

Last but most importantly, to my family. To my father, who never quit on me, who encouraged me, and was there when I needed him most. You're the kind of father we all hope for, but not all of us are blessed to have. And to Fenia, the love of my life, the one person to believe in me, and the partner I needed, need, and will need, the one who always stood at my side. Thank you from the bottom of my heart.

Publishing notes

The following list contains the published articles under supervision by Prof. Dr. Thomas Schmidt during my work in the Allgemeine Psychologie at the University of Kaiserslautern that are included in this work. In all papers listed here, the author of this thesis (MPW) was an active codeveloper of the presented research, theoretical interpretation, strategical decision making, and methodological application.

Article 1:

Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing response times and other types of time-to-event data using event history analysis: A tool for mental chronometry and cognitive psychophysiology. *I-Perception*, 11(6), 1-24. <https://doi.org/10.1177/2041669520978673>

Reproduced with permission from SAGE under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>).

SP and MPW conceived the idea of a methods article showcasing the advantages of event history analysis. They selected previous results with which the application and power of event history analysis were highlighted. For example, findings from Panis et al. (2020) were selected, in which MPW performed the analysis with SP and jointly drew conclusions from the results. Both authors planned and structured the manuscript. SP wrote the original draft. MPW reviewed and edited the draft extensively before submission and after peer reviews, as did FS and TS.

Article 2:

Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). 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. *Attention, Perception, & Psychophysics*, 82(2), 689–714. <https://doi.org/10.3758/s13414-019-01897-z>

Reproduced with permission from Springer Nature.

SP and MPW conceived the idea of reanalyzing a preexisting benchmark data set. They chose the available visual search data by Wolfe et al. (2010). After performing the analysis and interpreting the results together, they planned and structured the manuscript. SP wrote the original draft. MPW reviewed and edited the draft extensively before submission and after peer reviews, as did RM (an expert in the field of visual search) and TS.

Article 3:

Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential motor activation in a dual-prime paradigm: Insights from conditional accuracy and hazard functions. *Attention, Perception & Psychophysics*, 82(5), 2581–2602. <https://doi.org/10.3758/s13414-020-02010-5>

Reproduced with permission from Springer Nature under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>).

MPW conceived the idea, designed and programmed the experiment, collected the data, ran the analyses and interpreted the results with SP and TS. MPW and SP planned and structured the manuscript. MPW wrote the original draft and created all plots. SP and TS reviewed and edited the draft extensively before submission and after peer reviews.

Article 4:

Schmidt, T., Panis, S., Wolkersdorfer, M. P., & Vorberg, D. (2022). Response inhibition in the Negative Compatibility Effect in the absence of inhibitory stimulus features. *Open Psychology*, 4(1), 219–230.
<https://doi.org/10.1515/psych-2022-0012>

Reproduced with permission from De Gruyter Brill under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>).

TS and DV conceived the idea, based on a previous experiment by DV. TS designed and programmed the experiment, oversaw data collection, and ran mean performance analyses. MPW performed the event history analysis. MPW, SP, and TS interpreted the results together. TS wrote the original draft, in which MPW wrote the sections regarding event history analysis, such as description and interpretation of the findings regarding the conditional accuracy functions. MPW and SP reviewed and edited the draft extensively before submission and after peer reviews.

Additional articles that are not included in this work:

Article 5:

Jubran, O. F., Wolkersdorfer, M. P., Eymann, V., Burkard, N., Czernochowski, D., Herrlich, M., van Leeuwen, C., & Lachmann, T. (2025). Spatiotemporal survival analysis for movement trajectory tracking in virtual reality. *Scientific Reports*, 15(1), 7313. <https://doi.org/10.1038/s41598-025-91471-5>

OFJ, TL, VE, DC, MH, and NB conceived the idea and designed the experiment. OFJ and NB programmed the experiment. OFJ, VE, and NB collected the data. OFJ and MPW developed the method and created the code to analyze the data. Formal analyses were done by OFJ, MPW, CVL, and TL, with MPW being responsible for event history analyses and the analyses of mean performance, as well as the development, application, and interpretation of the new spatiotemporal survival analysis together with OFJ. Visualization was done by OFJ and MPW. A first original draft was written by OFJ and DC. OFJ, MPW, CVL, and TL planned and restructured the manuscript to a more methods-oriented article. This new original draft was written by OFJ, TL, CVL, and MPW, with MPW specifically providing the sections on event history analysis and mean performance measures. VE and DC reviewed the draft before submission and after peer reviews.

Additionally, see author contributions as published:

<https://www.nature.com/articles/s41598-025-91471-5#author-information>

Article 6:

Schmidt, T., Lee, X. Y., & Wolkersdorfer, M. P. (2024). Functional dissociations versus post-hoc selection: Moving beyond the Stockart et al.(2024) compromise. arXiv preprint arXiv:2411.15078.

TS conceived the idea of jointly writing a response to the highly controversial “best practice” recommendations by Stockart et al. (2024). TS wrote the initial manuscript and suggested the conceptual analysis. MPW and XYL extensively discussed the major arguments and the argument style and structure. MPW and XYL reviewed and edited the draft extensively before publication. All three authors continue to be involved in this project, which is going to result in an extensive peer-reviewed publication and foreseeably a short commentary in BEHAVIORAL AND BRAIN SCIENCES.

Table of Contents

Acknowledgments.....	III
Publishing notes	V
Table of Contents	XI
FIGURES FRAME TEXT	XIII
TABLES FRAME TEXT.....	XV
ABBREVIATIONS FRAME TEXT	XVI
Introduction	1
1 A BRIEF HISTORY OF THE ANALYSES OF TIME-TO-EVENT DATA.....	2
2 INTRODUCING EVENT HISTORY ANALYSIS - PANIS, SCHMIDT, ET AL., 2020	6
2.1 <i>Continuous-Time Event History Analysis</i>	8
2.2 <i>Discrete-Time Event History Analysis</i>	13
2.3 <i>Inferential Statistics for discrete-time EHA</i>	20
3 APPLICATION OF DISCRETE-TIME EVENT HISTORY ANALYSIS	22
3.1 <i>Visual Search - Panis, Moran, et al., 2020; Panis, Schmidt, et al., 2020</i>	22
3.2 <i>Response Priming - Wolkersdorfer et al., 2020, & T. Schmidt et al., 2022</i>	27
3.3 <i>Masked (Response) Priming and the Negative Compatibility Effect - T. Schmidt et al., 2022, & Panis, Schmidt, et al., 2020</i>	30
Articles.....	35
ANALYZING RESPONSE TIMES AND OTHER TYPES OF TIME-TO-EVENT DATA USING EVENT HISTORY ANALYSIS: A TOOL FOR MENTAL CHRONOMETRY AND COGNITIVE PSYCHOPHYSIOLOGY	37
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.....	63

— Table of Contents —

TEMPORAL DYNAMICS OF SEQUENTIAL MOTOR ACTIVATION IN A DUAL-PRIME PARADIGM: INSIGHTS FROM CONDITIONAL ACCURACY AND HAZARD FUNCTIONS.....	91
RESPONSE INHIBITION IN THE NEGATIVE COMPATIBILITY EFFECT IN THE ABSENCE OF INHIBITORY STIMULUS FEATURES	115
Discussion	129
4 HIGHLIGHTED FINDINGS	129
4.1 <i>Discrete-time EHA reveals effects that mean performance measures can conceal</i>	<i>129</i>
4.2 <i>Discrete-time EHA reveals the temporal dynamics of response behavior.....</i>	<i>132</i>
4.3 <i>Discrete-time EHA allows to track performance changes on multiple time scales.....</i>	<i>136</i>
5 EVALUATION OF DISCRETE-TIME EHA	137
5.1 <i>Theoretical advantages of discrete-time EHA</i>	<i>137</i>
5.2 <i>Statistical & methodological advantages of EHA.....</i>	<i>140</i>
5.3 <i>Potential disadvantages of discrete-time EHA and how to address them.....</i>	<i>141</i>
5.4 <i>Comparison with other common distributional methods</i>	<i>148</i>
5.5 <i>Recommendations for the application of discrete-time EHA</i>	<i>151</i>
6 OUTLOOK	153
References Frame Text	156
Declaration of authorship.....	172
Curriculum Vitae	173

Figures Frame Text

Figure 1. Four views on four different waiting-time distributions in continuous time. (A) the hazard rate function $\lambda(t)$, (B) the cumulative distribution function $F(t)$, (C) the survivor function $S(t)$, (D) the probability density function $f(t)$. Shown are four theoretical probability distributions (different colors: exponential, Weibull, gamma, log-normal). While the hazard rate 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 (taken from Panis, Schmidt, et al., 2020). 11

Figure 2. Example mean RTs and Accuracy for an idealized priming paradigm. C = consistent, I = inconsistent. 13

Figure 3. (A) The hazard $h(t)$, (B) the cumulative distribution function $F(t)$, (C) the survivor function $S(t)$, (D) the probability mass function $P(t)$, plotted for one subject in both conditions, C = consistent (green), I = inconsistent (red), in an idealized priming paradigm. Time bins are indexed from 1 (0 ms,50 ms] to 10 (450 ms,500 ms]. 17

Figure 4. Conditional accuracy plot for one subject in both conditions, C = consistent (green), I = inconsistent (red), in an idealized priming paradigm. Time bins are indexed from 1 (0 ms,50 ms] to 10 (450 ms,500 ms]..... 19

Figure 5. Taken from Figure 1 in Panis, Moran, et al. (2020). Benchmark visual search data set from Wolfe et al. (2010). (A) Example visual search displays for three search tasks (Fig. 1 in Wolfe et al., 2010). (B) Search functions for the three search tasks (Fig. 2 in Wolfe et al., 2010). Solid lines represent target-present, dashed lines target-absent trials. Lighter lines show individual subject data, darker lines mean data. 23

Figure 6. Stimulus displays and design of Experiment 1 in Wolkersdorfer et al. (2020). After fixating the center of the white lollipop frame, a sequence of two primes and a target is presented, with SOA1-SOA2 combinations of 27/80, 53/53, or 80/27.

..... 29

Figure 7. Experimental design taken from Panis and Schmidt (2016) as cited in Panis, Schmidt, et al. (2020). Three different types of masking: REL, structure masking with response-relevant features; IRREL structure masking with response-irrelevant features; LIN, random noise masking..... 32

Figure 8. Figure 2 from T. Schmidt et al. (2022). a) Time course of a trial. Note that both mask and target remain on screen until the response is completed. b) The four masking conditions. For better legibility, the prime is drawn above the mask and target. In the experiment, it appeared in the central mask cutout so that their contours were adjacent. 33

Permissions

Figure 5 in the Frame Text, and Fig. 1 in Panis, Moran, et al. (2020) are reprinted from *Vision Research*, Volume 50, Issue 14, Jeremy M. Wolfe, Evan M. Palmer, and Todd S. Horowitz, “Reaction time distributions constrain models of visual search”, 1304–1311, © Elsevier Ltd (2009), with permission from Elsevier.

Figure 7 in the Frame Text, and Figure 3A and Figure 5 in Panis, Schmidt, et al. (2020) are reprinted from *Journal of Cognitive Neuroscience*, Volume 28, Issue 11, Sven Panis and Thomas Schmidt, “What Is Shaping RT and Accuracy Distributions? Active and Selective Response Inhibition Causes the Negative Compatibility Effect”, 1651–1671, © Massachusetts Institute of Technology (2016), with permission from MIT Press.

Tables Frame Text

Table 1: Example life table for one subject in the consistent condition of an idealized response priming, with 100 trials administered.14

Table 2: Example life table for one subject in the inconsistent condition of an idealized response priming, with 100 trials administered.15

Abbreviations Frame Text

RT	Reaction or Response Time, 1
EEG	Electroencephalogram, 1
MEG	Magnetoencephalography, 1
fMRI	Functional Magnetic Resonance Imaging, 1
AFM	Additive Factors Method, 2
EHA	Event History Analysis, 2
ANOVA	Analysis of Variance, 4
ms	Milliseconds, 8
ER	Error Rate, 27
SOA	Stimulus-Onset Asynchrony, 27
PCE	Positive Compatibility Effect, 31
NCE	Negative Compatibility Effect, 32
DFT	Dynamic Fields Theory, 139
GEE	Generalized Estimating Equations, 144
GLMM	Generalized Linear Mixed Model, 144
LRP	Lateralized Readiness Potential, 145
ROI	Region of Interest, 145
StSA	Spatiotemporal Survival Analysis, 153

“Wisdom comes from experience. Experience is often a result of lack of wisdom.”

— Terry Pratchett

Introduction

Since its conception, experimental psychology has been on a quest to unravel the cognitive processes involved in human behavior. Of major interest in this endeavor is not only the identification of said processes, but to also reveal their temporal dynamics. In other words, can we identify at which point in time which processes are at play, and how they correspond to human behavior? In order to accomplish such a daring task, researchers for almost 200 years now have developed paradigms and experimental procedures. They have brought forth numerous approaches to analyze the obtained data, made cases in favor of theirs and argued against others. Such is the nature of this work.

Historically, the fundamental measures when researching the temporal dynamics of human behavior have been, and are to this day, reaction or response time (RT) and accuracy, respectively. In its most simple definition, RT refers to the time required to perform an action (Luce, 1986). More specifically, in experimental psychology, it is the time it takes to complete a task (Rouder & Speckman, 2004), most often the time between stimulus presentation and an associated response. Accuracy then evaluates the performance of this action, on a scale from “0 (incorrect)” to “1 (correct)”, depending on the task at hand. Despite the emergence of neuroimaging techniques such as electroencephalography (EEG), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI), RT analyses remain a key tool in the field.

The aim of this work is to give a brief overview of the historic developments associated with these measures (Chapter 1; based in parts on Article 1 & 2). Thus, I

discuss their properties and application, and the methods created to use these measures as a tool to better our understanding of human cognition and behavior. This reaches from Donders' *subtraction method* (Donders, 1969) to Sternberg's *additive factor method* (AFM; Sternberg, 1969, 1984, 2011, 2013). Following this, more recent developments are introduced, focusing on the distributional properties of RT and accuracy, such as *quantile plots* and *Vincentizing* (Ratcliff, 1979; Rouder & Speckman, 2004; Vincent, 1912). Finally, I make my own case for the analyses of the full distribution of response occurrences (Chapter 2; based on Article 1). To do so, I employ discrete-time *event history analysis* (EHA; Allison, 1982, 2010; Panis & Hermens, 2014; Panis & Schmidt, 2016; Singer & Willett, 2003) to various experimental psychology tasks (introduced in Chapter 3; see Articles 1–4). I aim to highlight its advantages, the gains and insights it lends to the field, discuss its shortcomings, and suggest potential developments in the future (Discussion; Articles 1–4).

1 A Brief History of the Analyses of Time-to-Event Data

As mentioned in the previous section, one major goal in experimental psychology and its related fields is to understand the time course of human cognition and behavior. In order to achieve this, an appropriate measure is necessary, as well as the appropriate analyses of the measurements obtained with it.

The foundation for this measurement was laid in the middle of the 19th century by the groundbreaking work of Hermann von Helmholtz (Helmholtz, 1948, 2021; Meyer et al., 1988; Townsend & Ashby, 1983). With his discovery of the conduction velocity of the nervous system in frogs (Helmholtz, 1948), and later the introduction of the *simple reaction time procedure* (Helmholtz, 2021), Helmholtz can be considered

one of the fathers of modern cognitive psychophysiology (Meyer et al., 1988). His work directly influenced the work of Franciscus Cornelius Donders. Donders (1969) developed a technique called the subtraction method. This innovative technique employed three tasks, called Task A, Task B, and Task C. Task A was equivalent to Helmholtz' (2021) simple reaction time procedure, in which a speeded response time to a single stimulus is recorded (Donders, 1969; Meyer et al., 1988; Sternberg, 2011; Townsend & Ashby, 1983). The underlying assumption is that this simple task is accomplished by some process A. Task B refers to a choice reaction time procedure, which now entails multiple stimuli, which have to be discriminated between, resulting in a selection before the response. The idea behind this was that this task would now involve the same process A as Task A before, in addition to the processes of discrimination and selection (both collectively referred to as process B now). If one subtracts the time for Task A from Task B, one would have an estimate for the duration of process B, a combination of discrimination and selection. Finally, Task C can be seen as an attempt to distinguish between the durations of the discrimination and selection processes. By employing a go/no-go reaction time procedure, Donders created a task in which multiple stimuli were deployed but only one response was required (i.e., subjects were tasked to respond to one stimulus, but withhold responses for all others). Therefore, discrimination was still thought to be required for completion of the task, but not selection. Subtracting the RT for Task C from Task B would thus give an estimate of the duration of response selection, while subtracting Task A from Task C would yield an estimate of the duration of discrimination.

As we now know, this method relied upon a plethora of very strong assumptions, which did not all stand the test of time (Meyer et al., 1988). Perhaps the strongest

assumptions were those of strictly successive stages and *pure insertion* (i.e., the idea that by adding or removing a processing stage, other stages are not affected). Nonetheless, the influence and impact that this method had and still has is unquestionable.

Sternberg (1969, 1984, 2011, 2013) expanded on the original work of Donders. Though he still aimed to identify different stages of processing, his AFM does not rely on the assumption of pure insertion. This technique was strongly influenced by the emergence of the serial information processing framework (i.e., assuming that information processing involves a series of successive stages). If a process can be seen as a series of distinct subprocesses (i.e., stages), then the RT related to this process should correspond to the sum of the different stage durations. By applying Analysis of Variance (ANOVA) to mean RTs, the AFM aims to identify these distinct stages and their relations to one another. Let's assume there are two factors, A and B, and they both influence the mean RT in a given task. If we further assume that they do not influence a common stage, then their effects on mean RTs should be additive. If, however, A modulates the effect of B on mean RTs (i.e., they interact with each other), then the two factors must influence at least one common stage. Thus, if one finds at least two additive factors, Sternberg concludes that at least two distinct stages of processing take place. Likewise, if one finds at least two interacting factors, he concludes that they may have a stage of processing in common. By experimentally varying factors, which in turn may lead to different combinations of additive and interactive factors, the goal is to identify the number of distinct processing stages and their purposes (Meyer et al., 1988). Yet, the AFM still holds on to rather strong

assumptions, for instance the assumption that processing stages have no temporal overlap (Meyer et al., 1988).

And while the AFM relies on mean RTs and ANOVAs, this might not always be the best approach. As Whelan (2008) shows, RT data is often not suited for ANOVAs. First, RT distributions are typically not normally distributed, which can result in reduced power when analyzing mean RTs using ANOVAs. In order to increase power, RT data then often requires substantial outlier correction (either through deletion or transformation of data; see Ratcliff, 1993). Second, analyzing the RT distributions instead of merely the means of conditions is often the better and more effective approach (Whelan, 2008).

One such approach utilizes a method called Vincentizing (Ratcliff, 1979; Vincent, 1912; Whelan, 2008). Vincentizing aims to generate an average RT distribution for a sample of subjects. First, the RT distribution is partitioned based on quantiles for each subject. Next, group means for each quantile are calculated by averaging across participants. From these a group RT distribution is created, that is assumed to retain the shape of the “average” participant’s distribution. In other words, RT distributions are normalized across participants. Ratcliff (1979) provides a detailed description and discussion of the method. In order to test different experimental conditions against each other, one applies the method for each condition separately. One can obtain *delta plots* for RT by then calculating the difference in their mean RT for each quantile, and plotting this difference as a function of the average mean RT.

These distributional approaches more readily allow to investigate the time course of cognitive processes. Through them, one can compare effects at different RTs, enabling a more fine-grained view into the temporal dynamics of the task at hand.

Moreover, RT distributions and their properties can discriminate between different models and even falsify models that otherwise predict mean RTs rather well (Ratcliff, 1979).

In light of this, the next chapter aims to present a relatively new approach in psychology, namely EHA. A case will be made for the use of one variant of this more advanced and well-established analysis method. Such a method can maximize the return from the data a researcher might have collected, which, considering the costs and time required to run an experiment, is of high relevance to researchers (Whelan, 2008).

2 Introducing Event History Analysis - Panis, Schmidt, et al., 2020

As shown, the underlying assumptions that directly link differences in mean RT to differences between cognitive operations and processes are rather strong. If longer RTs implicate additional cognitive processes, this heavily relies on the idea that those processes follow the serial information processing framework. However, this hypothesis is not universally accepted. In fact, researchers have criticized the strong claims it makes since its inception (Cisek & Kalaska, 2010; Eriksen & Schultz, 1979; McClelland, 1979; Pieters, 1983; Schöner G., Spencer, J. P., & the DFT Research Group, 2016). Yet, traditional methods derived from it remain dominant in the field, with ANOVAs as the most popular tool when dealing with RT and accuracy data.

It is, however, viable to consider alternative approaches. If one is interested in the temporal dynamics of response activation (e.g., one investigates forced-choice paradigms, priming, etc.), one must not ignore the passage of time when analyzing RTs. van Gelder (1995) postulates that cognition is better described as the operation of

a dynamical system, rather than a series of information processing stages. Under this hypothesis, it is therefore required to continuously track the output of such a dynamical system, if one wants to gain a deeper understanding of human cognition and behavior (Schöner G., Spencer, J. P., & the DFT Research Group, 2016). In light of this, EHA is the most suitable choice (Panis, Schmidt, et al., 2020). EHA is the standard distributional or longitudinal technique for time-to-event data in many scientific fields (Allison, 2010; Panis, Schmidt, et al., 2020; Singer & Willett, 2003). In the research of human behavior and cognition, examples range from developmental psychology (Ha et al., 1997), social psychology (Willett & Singer, 1993), to cognitive psychology (Panis & Wagemans, 2009), and more (see Panis, Schmidt, et al., 2020 for a more extensive overview).

The necessary first step is to define what constitutes an event. This can be any qualitative change in time, ranging from marital status, death, to more experimental variables like saccade onset, the crossing of a point in space, or something as simple as a button-press. Next, one needs to define time point zero, which, in experimental psychology, can be the start of a trial, time of the previous response, onset of fixation, or simply the onset of a stimulus. Finally, one needs to measure the passage of time between this time point and the event of interest (either with discrete or continuous time measures). To illustrate, in case of a standard response priming paradigm, in which subjects have to indicate via button-press the identity of a target stimulus while ignoring any preceding stimuli, the following applies: (1) the event of interest is the button-press, (2) time point zero is the onset of the target stimulus they ought to identify, and (3) RT is the measure between this time point zero and the event.

The following sections will first introduce the background and advantages of a *continuous-time EHA*. Next, discrete-time EHA will be described. Lastly, possible inferential analyses will be presented.

2.1 Continuous-Time Event History Analysis

When dealing with a continuous random variable T (in the present case: a participant's RT in a specific experimental condition, e.g. in a response priming paradigm this is the time between the onset of a target stimulus and the response to it), there is a variety of mathematically equivalent functions to describe it (Luce, 1986): (1) the cumulative distribution function $F(t) = P(T \leq t)$, which yields the cumulative probability of RTs smaller or equal to time point t ; (2) its derivative $F'(t) = f(t)$, the probability density function, which gives the relative likelihood of a given RT; (3) the survivor function $S(t) = 1 - F(t) = P(T > t)$, which provides the cumulative probability of RTs larger than time point t ; (4) and lastly, the hazard rate function $\lambda(t) = f(t) / [1 - F(t)] = f(t) / S(t)$, which calculates the instantaneous risk that a response will occur at time point t , given that it has not occurred yet. Similar to speed, which measures distance covered per unit of time, the hazard rate measures responses given per unit time. To give an example, let's imagine the aforementioned response priming paradigm and RT was measured in milliseconds (ms). Let $\lambda(200) = .1$, in other words, the hazard rate at 200 ms after stimulus presentation equals .1. This translates to an instantaneous rate of response occurrence of .1 events per millisecond after 200 ms of waiting time.

Given these equivalent functions, one might argue that any of those functions would be sufficient to describe the data an experimenter may have collected. However, as Luce (1986, p. 17) put it:

“it matters a great deal”.

EHA has been developed to describe, model, and analyze the hazard function. The following sections will give five reasons as to why the hazard function is the recommended function, if one wants to describe and analyze a finite sample of time-to-event data.

First, generally speaking, the hazard function better captures the concept of processing capacity (Wenger & Gibson, 2004). Processing capacity translates to the amount of effort a subject is able to exert within a specific period of time. While there are measures to assess this capacity (Kok, 2001; Rouder et al., 2011; Verbaten et al., 1997), the hazard function reveals the instantaneous capacity of a subject for the completion of a task within the next time unit, if the task has not been completed yet. This enables the continuous tracking of processing capacity over time.

Second, EHA enables the inclusion of other time-varying variables in modeling. These can be informative covariates like heart rate, EEG signal amplitude, gaze location, etc. (Allison, 2010; Singer & Willett, 2003). Such covariates are of great interest in cognitive psychophysiology, and enable a more holistic analysis of the underlying processes (Meyer et al., 1988).

Third, EHA does not discard right-censored observations. As previously discussed, RT analyses have to deal with outliers, that is, RTs that are either shorter or longer than can be attributed to the process that is being investigated (Ratcliff, 1993; Whelan, 2008). Moreover, RT distributions are known to be positively skewed, which is in the nature of the measurement: if not prevented by the experimental design, RTs can always become slower, but never faster than set by the physical boundaries of the

human body. Therefore, one will often find a quick rise in the distribution of RTs, followed by a long tail. In classical analyses of RTs, such as ANOVAs, the power is diminished by outliers (Ratcliff, 1993). Thus, several methods have been proposed to deal with the issue, most of which include some form of censoring, namely defining cutoffs after which RTs are discarded (in the case of right-censoring) or before which RTs are discarded (in the case of left-censoring). This generally improves power in ANOVAs. However, if the effect of interest actually takes place in the right tail of the distribution, power can be diminished by right-censoring (see Ratcliff, 1993 for a review).

Another common type of right-censoring is called *singly Type I censoring*. Instead of discarding RTs that are deemed “too slow” after data collection, the experiment may introduce a fixed response deadline for all trials, after which responses are no longer collected. “Type I” refers to the fact that the cutoff time is predetermined by the experimenter. “Singly” means that all observations have the same cutoff time (Allison, 2010). Both types of right-censoring have the potential to introduce a sampling bias, which can lead to an underestimation of the true mean RTs. EHA, however, includes the information of trials in which responses were not collected or deemed “too slow”. This is crucial for experiments in which many such trials are to be expected, such as masking paradigms. Similarly, EHA does not require one to discard error trials. Since all forms of random censoring should always be uninformative if one wants to apply any standard statistical method, this is crucial: error responses are often informative. Response channels have been shown to compete with each other (Burle et al., 2004; Eriksen et al., 1985; Praamstra & Seiss, 2005). It is therefore very informative when an

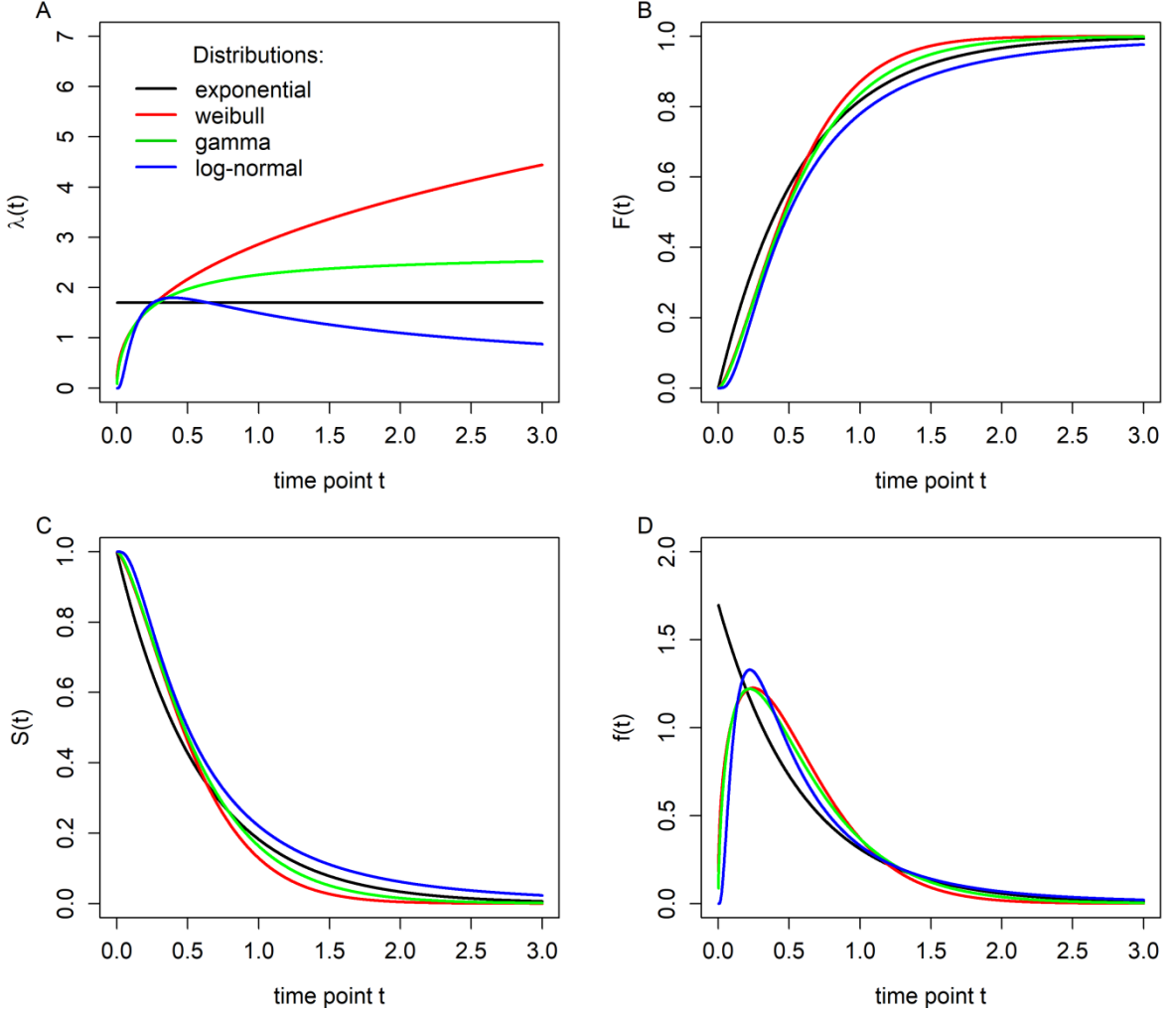


Figure 1. Four views on four different waiting-time distributions in continuous time. (A) the hazard rate function $\lambda(t)$, (B) the cumulative distribution function $F(t)$, (C) the survivor function $S(t)$, (D) the probability density function $f(t)$. Shown are four theoretical probability distributions (different colors: exponential, Weibull, gamma, log-normal). While the hazard rate 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 (taken from Panis, Schmidt, et al., 2020).

error takes place. For this reason, hazard of response occurrence should not be split into correct or incorrect responses, nor should only one of these be analyzed.

Fourth, RT distributions can differ from one another in multiple ways. For that reason, Townsend (1990b) proposed a hierarchy of statistical differences between distributions. Let A and B be two arbitrary distributions. If $F_A(t) > F_B(t)$ for all t, then both cumulative distribution functions are said to show a complete ordering. When there is a complete ordering on the hazard functions – $\lambda_A(t) > \lambda_B(t)$ for all t – there also follows a complete ordering on the cumulative distributions and the survivor

functions, with $F_A(t) > F_B(t)$ and $S_A(t) < S_B(t)$. From this, one can infer also an ordering on the means, with the mean of distribution A being smaller than the mean of distribution B (Townsend, 1990b). However, the reverse is not true. From an ordering on two means does not necessarily follow an ordering on the respective cumulative distributions and survivor functions. Moreover, a complete ordering on these functions does not command a complete ordering on the hazard functions. For example, from mean A < mean B follows $F_A(t) > F_B(t)$, and $S_A(t) < S_B(t)$, whereas hazard functions can show a complete ordering, a partial ordering, they may cross, or even show no ordering at all. Therefore, the hazard functions enable stronger inferences than the other functions on their own.

Finally, the hazard function of response occurrence is one of the most diagnostic functions when describing the distribution of a sample of time-to-event data for additional reasons (Allison, 2010; Luce, 1986; Panis, Schmidt, et al., 2020; Townsend, 1990b). Figure 1 compares $F(t)$, $f(t)$, $S(t)$, and $\lambda(t)$ by plotting four theoretical distributions of RT (exponential, Weibull, gamma, log-normal). While $\lambda(t)$ shows clear differences between all four distributions (Fig. 1, panel A), $F(t)$ and $S(t)$ have considerable overlap (panels B and C). Moreover, $f(t)$ is not able to capture differences in the right tail of the distribution at all (Luce, 1986). Thus, cases exist in which $f(t)$ functions cannot be differentiated, but clear differences in hazard functions are found. Since no such case exists in which one finds differences in probability density functions and not in $\lambda(t)$, Holden et al. (2009) come to the conclusion that hazard functions are the more diagnostic function.

2.2 Discrete-Time Event History Analysis

One drawback of the continuous-time hazard rate function is the amount of data one needs to acquire for a good estimate. For example, if one wants to investigate a response priming paradigm, one would need to collect around 1,000 trials per subject per condition (Bloxom, 1984; Luce, 1986; van Zandt, 2000). In experimental psychology, depending on how many conditions one wants or needs to include in an experiment, this can quickly exceed practicality. Fortunately, discrete-time hazard analysis provides a viable alternative that enables the application of logistic regression (Allison, 1982, 2010; Singer & Willett, 1991, 2003; Willett & Singer, 1993, 1995).

When utilizing discrete-time EHA, one has to begin with the creation of a life table per subject per condition. Let's revisit the previous example of a response priming paradigm, in which subjects are tasked with correctly identifying a target stimulus (red or green), preceded by a prime stimulus either linked to the same (consistent) or different response (inconsistent), as fast and accurate as possible.

Figure 2 shows idealized example mean RTs and accuracy. Mean RTs are shorter and responses more accurate in consistent trials than in inconsistent trials,

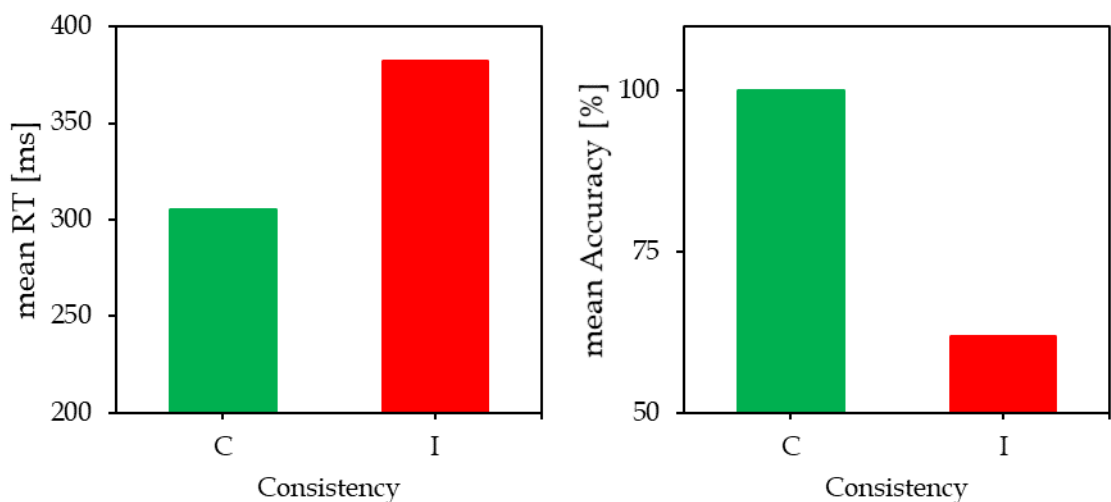


Figure 2. Example mean RTs and Accuracy for an idealized priming paradigm. C = consistent, I = inconsistent.

indicating response conflict elicited by the inconsistent prime preceding the target stimulus, and response facilitation by the consistent prime (more on this in later sections). By merely looking at these mean plots, not much understanding of the underlying processes is gained. However, discrete-time EHA is able to provide more details.

For EHA, the event of interest would be a button-press (indicating either a red or green target). Let there be 100 trials per subject and condition (consistent vs. inconsistent). First, one determines the censoring time. This is typically the response deadline used, or a time point after which you expect no useful responses anymore in any trial of any condition (see the earlier discussion of censoring). To illustrate, let this censoring time be 500 ms. Next, the interval between time point zero and this cutoff is divided up into a sequence of contiguous time bins. Let this be 50 ms bins, resulting in

Table 1: Example life table for one subject in the consistent condition of an idealized response priming, with 100 trials administered.

bin t	responses	$P(t)$	$F(t)$	$S(t)$	risk set	$h(t)$	correct	$ca(t)$
1 = (0,50]	0	0.00	0.00	1.00	100	0.00		
2 = (50,100]	0	0.00	0.00	1.00	100	0.00		
3 = (100,150]	5	0.05	0.05	0.95	100	0.05	5	1.00
4 = (150,200]	10	0.10	0.15	0.85	95	0.11	10	1.00
5 = (200,250]	25	0.25	0.40	0.60	85	0.29	25	1.00
6 = (250,300]	25	0.25	0.65	0.35	60	0.42	25	1.00
7 = (300,350]	15	0.15	0.80	0.20	35	0.43	15	1.00
8 = (350,400]	10	0.10	0.90	0.10	20	0.50	10	1.00
9 = (400,450]	5	0.05	0.95	0.05	10	0.50	5	1.00
10 = (450,500]	5	0.05	1.00	0.00	5	1.00	5	1.00

a total of 10 bins indexed $t = 1$ to 10 (see Tables 1 & 2, column 1). Beginning with the first condition, in this example consistent trials, one counts the number of responses within each bin (see Table 1, column 2). If one divides this count by the total number of trials (100), one obtains the probability of response occurrence for each bin, with the corresponding probability mass function $P(t) = P(T = t)$ (see Table 1, column 3). T is a discrete random variable denoting the rank of the time bin in which the event (response) occurs. From this follows directly the cumulative distribution function $F(t) = P(T \leq t)$, and its complement the survivor function, with $S(t) = P(T > t) = 1 - F(t) = 1 - P(T \leq t)$ (see Table 1, columns 4 and 5). Next, one can determine the *risk set*, the number of trials at the start of bin t without a response, which for $t = 1$ consists of the total number of trials (100, see Table 1, column 6).

Table 2: Example life table for one subject in the inconsistent condition of an idealized response priming, with 100 trials administered.

bin t	responses	$P(t)$	$F(t)$	$S(t)$	risk set	$h(t)$	correct	$ca(t)$
1 = (0,50]	0.00	0.00	0.00	1.00	100	0.00		
2 = (50,100]	0.00	0.00	0.00	1.00	100	0.00		
3 = (100,150]	5.00	0.05	0.05	0.95	100	0.05	0	0.00
4 = (150,200]	10.00	0.10	0.15	0.85	95	0.11	1	0.10
5 = (200,250]	10.00	0.10	0.25	0.75	85	0.12	3	0.30
6 = (250,300]	15.00	0.15	0.40	0.60	75	0.20	8	0.53
7 = (300,350]	20.00	0.20	0.60	0.40	60	0.33	15	0.75
8 = (350,400]	20.00	0.20	0.80	0.20	40	0.50	18	0.90
9 = (400,450]	12.00	0.12	0.92	0.08	20	0.60	12	1.00
10 = (450,500]	5.00	0.05	0.97	0.03	8	0.63	5	1.00

The risk set is then continuously updated: the risk set for bin 2 only consists of *total number of trials – responses in bin 1*, for bin 3 *risk set bin 2 – responses in bin 2*, and so on. If one then divides the number of observed responses in each bin by the corresponding risk set, one obtains $h(t) = P(T = t | T \geq t)$ (see Table 1, column 7). This discrete hazard function of event occurrence gives the conditional probability that the event of interest (here a response) occurs in bin t given that it has not yet occurred before. For example, $h(3) = .05$ in Table 1 corresponds to a conditional probability of 5% that a response is given within time bin 3 for trials that have not been completed after 100 ms have passed. This procedure is then repeated for each subject in each condition. Note, that in the example in Table 2 (same subject, inconsistent trials) not all trials resulted in a response within 500 ms. In three trials the example subject did not give a response before 500 ms of waiting time in this condition. Crucially, these three trials remain in the risk set and are thus accounted for when calculating the discrete hazard of event occurrence (Table 1, row 11, column 7). This highlights how EHA is less affected by right-censoring.

If one is investigating a paradigm that allows to evaluate the performance of an event or response on a scale from “0 (incorrect)” to “1 (correct)”, like the present example, one can expand the $h(t)$ analysis of response occurrence with an analysis of conditional accuracy, i.e. the micro-level speed-accuracy tradeoff function (Allison, 2010; Pachella, 1974; Wickelgren, 1977). The conditional accuracy function is the conditional probability that an observed response is correct, given that it occurs in bin t , with $ca(t) = P(\text{correct} | T = t)$. One can obtain its estimate by counting correct responses in each bin t and dividing the results by the number of collected responses in the respective bin t (Table 1, columns 8 and 9). Combining $h(t)$ functions with $ca(t)$

functions provides a probabilistic description of the latency and accuracy of any sample of (right-censored) event times. Crucially, those descriptions are unbiased, time-varying, and based on all trials.

Figure 3 plots the estimated hazard $h(t)$, the cumulative distribution function $F(t)$, the survivor function $S(t)$, and the probability mass function $P(t)$ for one subject in both conditions of the example. In this idealized data set, one can see now that the hazard function starts rising after 100 ms of waiting time in both conditions (Fig. 3, panel A).

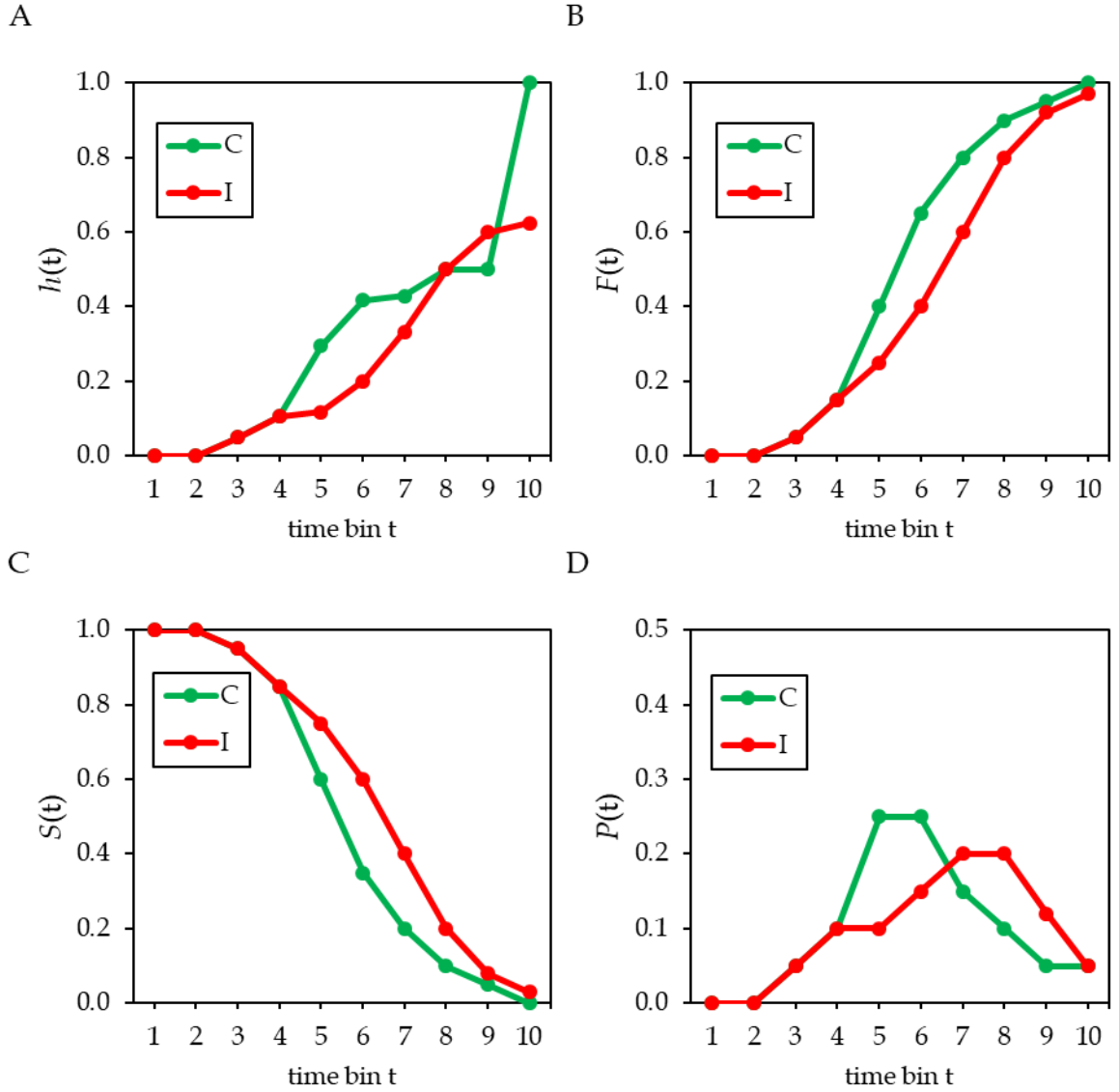


Figure 3. (A) The hazard $h(t)$, (B) the cumulative distribution function $F(t)$, (C) the survivor function $S(t)$, (D) the probability mass function $P(t)$, plotted for one subject in both conditions, C = consistent (green), I = inconsistent (red), in an idealized priming paradigm. Time bins are indexed from 1 (0 ms, 50 ms] to 10 (450 ms, 500 ms].

Thus, this example subject started responding 100 ms after target presentation, irrespective of the condition. Hazards for both conditions overlap until 200 ms have passed. From this point on, an ordering of $h(t)$ is present for the next 200 ms, with $h(t)$ in the consistent condition continuing to accelerate, and slowing down in the inconsistent condition. This ordering is also present in $S(t)$ and $F(t)$, although it persists until the censoring time of 500 ms is reached (Fig. 3, panels B and C). These findings seem to indicate that first responses are void of any response conflict as hazards are synchronized. However, after 200 ms the conflict caused by an inconsistent prime decelerates $h(t)$ in inconsistent trials, indicating the emergence of a response conflict caused by inconsistent prime and target information. After a total of 350 ms of waiting time, hazards synchronize again, indicating the overcoming of the response conflict. Note, that the hazard for the inconsistent condition never reaches 1, while there is a huge jump from .5 in bin 9 to 1 in bin 10. This is due to three responses not given before the censoring time in the inconsistent condition, while the remaining 5 trials are completed in the consistent condition. This highlights that $h(t)$ can potentially, under such circumstances, become less diagnostic in the very right tail of the distribution, as fewer and fewer trials remain without a response. Similarly, $S(t)$ never reaches 0 and $F(t)$ never reaches 1. $P(t)$ shows that most responses in the consistent condition occur between 200 and 300 ms after target presentation, while most responses in the inconsistent condition occur in the following two bins (Fig. 3, panel D). If one were to only look at the latter three functions, one might conclude that the effect of a prime simply results in slowing down response occurrence in the inconsistent condition (i.e., responses occur later than in the consistent condition). Similarly, in the current

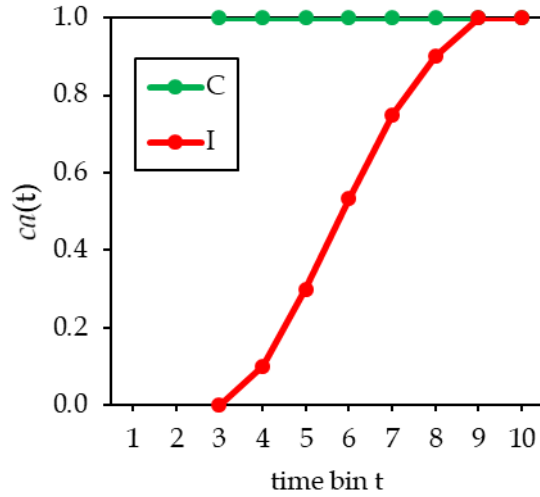


Figure 4. Conditional accuracy plot for one subject in both conditions, C = consistent (green), I = inconsistent (red), in an idealized priming paradigm. Time bins are indexed from 1 (0 ms,50 ms] to 10 (450 ms,500 ms].

example mean RTs are only able to show the existence of a response conflict caused by a preceding prime. However, discrete-time EHA is able to show the time-course of it: (1) responses first emerge 100 ms after target onset, irrespective of condition, (2) response conflict begins 200 ms after target onset, (3) this conflict lasts for 150 ms, and (4) this conflict is resolved 350 ms after target onset.

Adding plots of the conditional accuracy function provides additional insight. Figure 4 shows that the earliest responses in the consistent condition are all correct, while earliest responses in the inconsistent condition are all incorrect. This indicates that these early responses are exclusively towards the identity of the prime stimulus, not the target. As time passes, responses in the consistent condition maintain their accuracy, while responses in the inconsistent condition become more accurate. Remember, hazards decelerated in the inconsistent condition after 200 ms of waiting time. This was interpreted as the onset of response conflict. Incidentally, around this time, responses also begin to become more accurate in the inconsistent condition. For example, after roughly 250 ms, this example subject reaches 50% accuracy in inconsistent trials. This further indicates that target information begins to influence the

response after a waiting time of 200 ms. After 400 ms all responses are correct, showing that at this point target information solely drives the response. In the current example, mean accuracy plots are only able to show the existence of a response conflict. However, discrete-time EHA is able to show its time-course: (1) early responses are to the prime, (2) target information becomes available at around 200 ms, (3) after 400 ms responses are exclusively to the target.

The next section gives a brief overview on how one can apply inferential statistics to confirm these descriptive findings. The focus will be on discrete-time hazard models. Examples on how to perform inferential statistics for continuous time can be found in Allison (2010) and Austin (2017).

2.3 Inferential Statistics for discrete-time EHA

The present work and the associated studies want to study how the discrete-time hazard depends on various predictors, such as consistency between prime and target stimuli (see previous example; Wolkersdorfer et al., 2020), set size in visual search (as in Panis, Moran, et al., 2020; Panis, Schmidt, et al., 2020), mask condition (as in Panis, Schmidt, et al., 2020), etc. For example, one can fit regression models to the data (Singer & Willett, 2003). Considering the previous priming example, let's extend it by introducing a no prime or neutral prime condition. This enables one to compare the previous two conditions (consistent and inconsistent) to a baseline. A discrete-time hazard model could then contain the three predictors time (for example, $TIME = \{1, \dots, 10\}$), consistent prime ($C = \{0,1\}$), and inconsistent prime ($I = \{0,1\}$). The main predictor variable $TIME$ refers to the time bin index t (see Tables 1 and 2). Applying the

complementary log-log (cloglog) link function, an example function might look like the following:

$$\text{cloglog}[h(t)] = \ln(-\ln[1 - h(t)]) = [\alpha_0\text{ONE} + \alpha_1(\text{TIME} - 1) + \alpha_2(\text{TIME} - 1)^2 + \alpha_3(\text{TIME} - 1)^3] + [\beta_1\text{C} + \beta_2\text{I} + \beta_3\text{I}(\text{TIME} - 1)].$$

In this example TIME is centered on value 1. The complementary log-log link is the recommended link when dealing with RT data, where events can occur at any time point within a bin (Singer & Willett, 2003). Let's take a look at the terms of the function. The first bracket represents the shape of the baseline cloglog-hazard function. The alpha parameters are multiplied by their polynomial specifications of time (here linear, quadratic, and cubic), centered on 1. This baseline gives the prediction when all other predictors take on a value of zero, i.e. C = 0 and I = 0. Such is the case when neither a consistent, nor an inconsistent prime is shown, thus the baseline corresponds to the no prime condition modeled across time. The second bracket represents the vertical shift in the baseline cloglog-hazard for a 1 unit increase in the respective predictor. For example, changing from the no prime condition (the baseline) to the consistent condition (C changes from 0 to 1) leads to a shift of the baseline cloglog-hazard function by β_1 cloglog-hazard units. In other words, the cloglog-hazard increases by β_1 in all time bins in the consistent condition compared to the no prime condition. In the current example, I interacts linearly with time, see $\beta_2\text{I} + \beta_3\text{I}(\text{TIME} - 1)$. This predicts that, in bin 1, a change to the inconsistent condition (I changes from 0 to 1) results in a shift of the baseline cloglog-hazard function by β_2 cloglog-hazard units. Note, this is because $\text{TIME} - 1 = 0$ in bin 1. In bin 2 this would result in a shift of the baseline

cloglog-hazard function by $\beta_2 + \beta_3$ cloglog-hazard units (TIME - 1 = 1 in bin 2), and so forth. In other words, the vertical change in cloglog-hazard when changing from the no prime to the inconsistent condition changes linearly with time. While predictors in this example are dichotomous, they can be otherwise discrete or continuous. Furthermore, predictors can be time-invariant (like predictor C), or time-variant (like predictor I). Importantly, due to the application of the cloglog link, anti-logging the parameter estimates results in a hazard ratio, which then allows to interpret the effects of the predictors.

3 Application of discrete-time Event History Analysis

In order to showcase the advantages of discrete-time EHA when one is interested in time-to-event data, the method was applied to various experimental paradigms. The following sections will give a brief overview of these paradigms and respective predictions, in order of inclusion: (1) visual search (investigated in Panis, Moran, et al., 2020, and discussed in Panis, Schmidt, et al., 2020), (2) response priming (employed in Wolkersdorfer et al., 2020, and T. Schmidt et al., 2022), and (3) masked response priming (utilized in T. Schmidt et al., 2022, and discussed in Panis, Schmidt, et al., 2020).

3.1 Visual Search - Panis, Moran, et al., 2020; Panis, Schmidt, et al., 2020

The visual search paradigm aims to enable researchers to investigate search behavior, i.e. how humans find relevant features or objects of interest in an environment full of distractions. It thus often closely resembles and simulates everyday tasks (see Eckstein, 2011, & Humphreys, 2016 for a review of the paradigm). The field generally

differentiates between three most common search tasks, namely (1) feature search, (2) conjunction search, and (3) spatial configuration search, in ascending order of difficulty.

In a typical feature search task, subjects are tasked to decide via button-press whether a target stimulus (e.g., a vertical red bar) is present in a display containing several distractor stimuli (e.g., vertical green bars). Importantly, distractors are identical to a potential target stimulus in all but one feature (here the color), hence the name (see Fig. 5, panel A, left for an example display).

Conjunction search expands the task by including distractors that have some of multiple different features in common with a target. For example, subjects have to search for a vertical red bar in a display containing vertical green and horizontal red

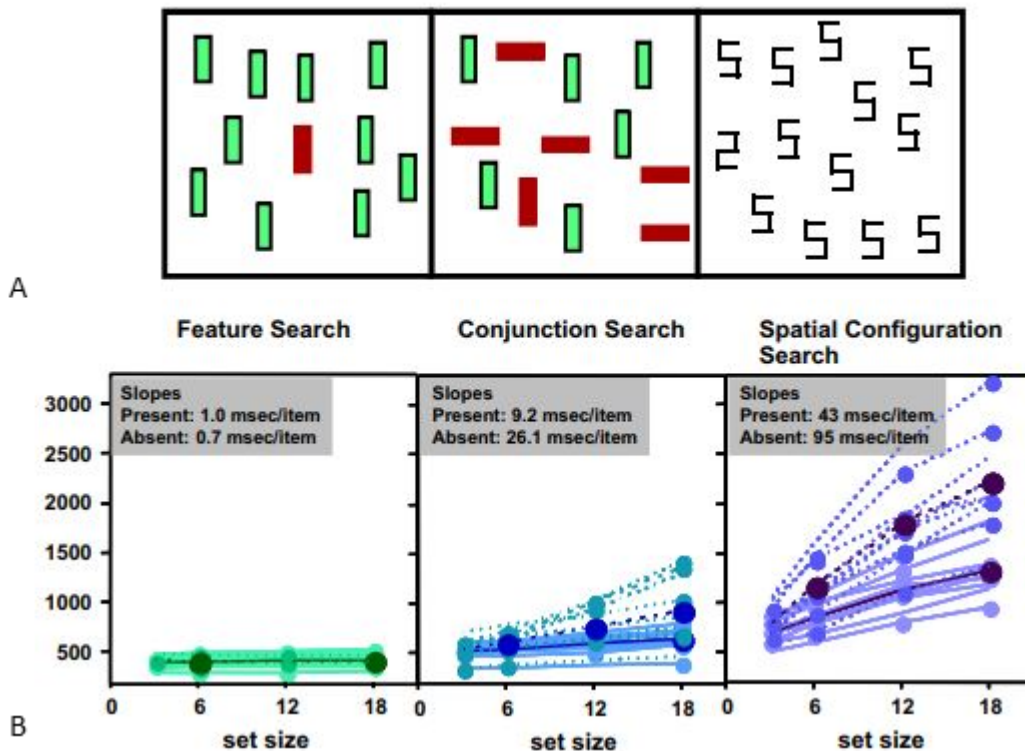


Figure 5. Taken from Figure 1 in Panis, Moran, et al. (2020). Benchmark visual search data set from Wolfe et al. (2010). (A) Example visual search displays for three search tasks (Fig. 1 in Wolfe et al., 2010). (B) Search functions for the three search tasks (Fig. 2 in Wolfe et al., 2010). Solid lines represent target-present, dashed lines target-absent trials. Lighter lines show individual subject data, darker lines mean data.

bars. Only the conjunction of orientation and color makes a stimulus a unique target (see Fig. 5, panel A, center for an example display).

Spatial configuration, on the other hand, includes distractors that share all features but the name-giving spatial configuration with the target. For example, subjects may be tasked with finding a 2 in a display full of 5s (see Fig. 5, panel A, right for an example display).

Commonly, the number of distractors is experimentally varied, as is the presence of a target. Visual search is then typically evaluated with the so-called search function, which makes the mean correct search RT a function of set size (number of items on screen). It has been found that these functions are close to linear in both target-presence conditions, and their slopes vary depending on the difficulty of the search task (Cheal & Lyon, 1992; Liesefeld et al., 2016). Feature search is considered efficient. Typical slopes are close to 0 ms/item, indicating that the inclusion of additional distractors does not slow down the search process. On the other hand, conjunction search shows intermediate efficiency, with intermediate positive slopes. In other words, each additional distractor prolongs the search process. Finally, spatial configuration search is described as inefficient, as each additional distractor severely slows down the search, resulting in large positive slopes in the search functions (see Fig. 5, panel B).

Two accounts regarding the processes of attention selection have been used to explain visual search behavior, namely serial and parallel accounts. Classical visual search theories assume a two-stage process: (1) a parallel stage in which all items or stimuli in the search display are processed simultaneously, and (2) a serial stage in which each item or stimulus is scanned one-by-one (Moran et al., 2016; Treisman &

Gelade, 1980; Wolfe, 2007; Wolfe et al., 1989; Wolfe et al., 2010). For example, *feature integration theory* (Treisman & Gelade, 1980) proposes that a limited set of features, such as color, can be processed in parallel, resulting in the quick identification of the only red vertical bar in a set filled with otherwise green vertical bars. This would explain the apparent independence of set size and RT in feature search. On the other hand, if two or more features are searched for, like in conjunction search, serial attention selection is required. Similarly in the *guided search model* (Wolfe, 1994, 2007; Wolfe et al., 1989), feature search leads to flat search slopes, because the target is so salient in an initial parallel stage that it is always scanned first in a second serial stage. Thus, search RT is independent of set size. In conjunction search, the initial parallel stage identifies stimuli that share relevant features (like color or orientation), and those priority stimuli are scanned in the second stage until the target is found or all items are identified as distractors. This search behavior is referred to as serial exhaustive search (Wolfe, 1994). Because these two-stage accounts rely on a serial component of attention allocation, they are typically referred to as serial models of visual search (Moran et al., 2016).

In contrast, single-stage accounts assume that all items in the display are attended and identified in parallel, omitting a second serial stage, and are henceforth referred to as parallel models of visual search. These accounts are based on a wide range of underlying theories, from signal detection theory to more recent neurodynamical approaches (for an overview, see Panis, Moran, et al., 2020). Much like serial accounts, parallel accounts, through specific modifications, are able to model the differences in the search functions between different search tasks. However, this leads to the following problem: if both accounts are able to generate search functions

for efficient and inefficient searches, how does one decide between the two? It appears that mean RT and the slopes of the search functions are not sufficient tools to distinguish between serial and parallel accounts (Townsend, 1990a). In order to combat this issue, Wolfe et al. (2010) focused on RT distributions and their shapes instead. Balota and Yap (2011) present three approaches when investigating RT distributions: (1) plot the shape of the RT distribution under different variable conditions (i.e., investigate the influence of specific variables on the RT distribution), (2) fit a mathematical function to the RT distribution (i.e., study how experimental variations modulate the parameters of the function), and (3) apply a computationally explicit model to the RT distribution (i.e., test the model predictions).

As discussed in Panis, Moran, et al. (2020), neither of these approaches has yet produced a final model of visual search. In an exploratory study, we applied discrete-time EHA to the hallmark data set of Wolfe et al. (2010). We described how to apply discrete-time EHA to search data and highlighted its advantages. The goal was to reveal differences and similarities between the different searches in $h(t)$ and $ca(t)$ functions, to better the understanding of the processes involved. In particular, would we be able to identify separate stages of search and/or other features of time-dispersed search behavior, on both the overall and the individual subject level? In Panis, Schmidt, et al. (2020), we made further use of the findings obtained in Panis, Moran, et al. (2020). In doing so, we were able to create a guideline on how to apply discrete-time EHA and made a case for its general applicability in the research of human cognition and behavior.

3.2 Response Priming - Wolkersdorfer et al., 2020, & T. Schmidt et al., 2022

Priming is a popular tool in many fields of cognitive psychology. Generally, paradigms associated with the term priming aim to study how a preceding prime stimulus can influence responses to a successive target stimulus. Research has focused on processes at different levels of human cognition, ranging from perceptual (Wiggs & Martin, 1998), to conceptual/semantic (e.g., Schacter & Buckner, 1998), lexical (e.g., Fernández-López et al., 2019), phonological (e.g., Ferrand & Grainger, 1992), and/or motor response levels (e.g., Rosenbaum, 1983).

Wolkersdorfer et al. (2020) focus on a variation of the response priming paradigm (Klotz & Neumann, 1999; Klotz & Wolff, 1995; Vorberg et al., 2003). Together with other conflict paradigms such as the Stroop, Simon, and Eriksen flanker paradigms, the response priming paradigm is widely employed to investigate the structure of cognition using chronometric measures. Briefly introduced in the previous chapters, in a typical response priming experiment, participants are tasked to respond as quickly and as accurately as possible to a target stimulus. The latter is preceded by a so-called prime stimulus, which can either be unmasked or masked (see Chapter 3.3). These two stimuli can either be mapped to the same response in consistent trials or to a different response in inconsistent trials. Typically, consistent trials result in accelerated and more accurate responses. In contrast, inconsistent trials show decelerated and less accurate responses. Calculating the mean RT and mean error rate (ER) differences between these two trial types defines the response priming effect. Vorberg et al. (2003) showed that this priming effect increases linearly with the stimulus-onset asynchrony (SOA) between prime and target stimuli, up to SOAs of around 100 ms.

Several theories have been proposed to explain the response priming effect. *Direct parameter specification* assumes an automatized connection between target stimulus and response that subjects learn during the typical practice blocks (Neumann, 1990). Further, after being prepared to give a motor response, a critical stimulus feature is all that is required to drive the response, even without a conscious percept. It is assumed that a prime stimulus' features can be sufficient to evoke such a fast and automatized response, resulting in the priming effect.

The *action trigger account* further assumes that so-called action triggers (again induced in the practice blocks and the instructions) are maintained in working memory (Kunde et al., 2003). These triggers define release conditions for a motor response. Once a stimulus satisfies these, a response is automatically evoked. Thus, when primes meet the trigger conditions, they can trigger a response, resulting in the priming effect.

Building up on these accounts, T. Schmidt et al. (2006) proposed their *chase theory of response priming*. Their theory is largely based on more recent neurological and behavioral findings, which suggest that sequential visual stimuli evoke sequential feedforward sweeps (Bullier, 2001; Lamme & Roelfsema, 2000; VanRullen & Koch, 2003). This feedforward and sequential activation has since been shown in studies examining neuronal activity in conflict tasks, for example by analyzing lateralized readiness potentials (Eimer & Schlaghecken, 1998; Vath & Schmidt, 2007). Behavioral studies further corroborated these findings. Sequential activation was tracked throughout the time course of pointing movements (F. Schmidt & Schmidt, 2010; T. Schmidt, 2002; T. Schmidt & Schmidt, 2009), and in RT distributions (Panis & Schmidt, 2016). Crucially, these studies showed that initial responses are exclusively triggered

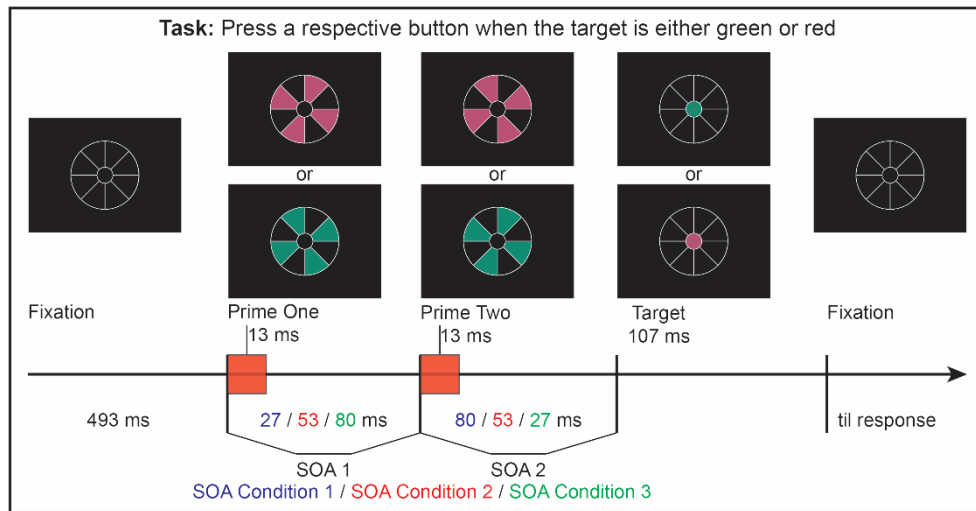


Figure 6. Stimulus displays and design of Experiment 1 in Wolkersdorfer et al. (2020). After fixating the center of the white lollipop frame, a sequence of two primes and a target is presented, with SOA1-SOA2 combinations of 27/80, 53/53, or 80/27.

by prime signals, whereas later responses are influenced by target information. These findings led to the following chase criteria formulated by T. Schmidt et al. (2006): (1) Prime rather than target signals determine the onset and initial direction of the response; (2) target signals influence the response before it is completed; (3) movement kinematics initially depend on prime characteristics only and are independent of all target characteristics. T. Schmidt (2014) gives a precise definition of the criteria and predictions of this rapid-chase theory.

The goal of Wolkersdorfer et al. (2020) was to put these criteria to the test by employing a newly developed variation of the classical response priming paradigm, and by applying discrete-time EHA to the data. Instead of using only one prime stimulus, they extended the design with a second prime stimulus (see Fig. 6). This created a stimulus sequence of a first prime, a second prime, and ultimately a target stimulus. SOAs and consistencies between these stimuli were manipulated in two experiments. If the predictions made by the rapid-chase theory are indeed accurate, the following findings were to be expected: (1) Initial responses would be exclusively

triggered by signals of the first prime; (2) intermediate responses would reflect the competition between first and second prime; (3) late responses would be dominated by target information. This would manifest itself in an ordering of hazard rates, both between consistency and SOA conditions, and time-locked effects of consistency in conditional accuracy functions.

3.3 Masked (Response) Priming and the Negative Compatibility Effect - T.

Schmidt et al., 2022, & Panis, Schmidt, et al., 2020

Visual masking is a procedure in which the visibility of a target stimulus is reduced or eliminated by a mask stimulus (Breitmeyer & Ögmen, 2007). Although masking is possible even when the mask precedes the target (forward masking), most research focuses on the effects of so-called backwards masking (mask presented after the target). Although there are other types of masking (e.g., masking by light, paracontrast masking, etc.), we distinguish between the following two types of masking (Breitmeyer & Ganz, 1976): (1) pattern masking, in which the contours of the mask overlap with the target contours (for examples, see Fig. 7), (2) metacontrast masking, in which the contours of the mask do not overlap with the target (for examples, see Fig. 8). Pattern masking can be further distinguished into noise masking (i.e., random pattern masking without a structural relation between mask and target), and structure masking (i.e., pattern masking in which mask and target share a structural relation). Through manipulation of the SOA between target and mask, two common types of masking functions can be identified, namely *Type-A* and *Type-B* masking (Breitmeyer & Ganz, 1976; Kahneman, 1968; Kolers, 1962). In *Type-A* masking, the masking effect is strongest at short SOAs (i.e., identification of the masked target stimulus is close to

chance level; the target is rendered “invisible”), and diminishes as SOA increases. In Type-B masking, target visibility is intact at short SOAs, the masking effect is strongest at SOAs between 50 – 100 ms, and diminishes again at longer SOAs, resulting in a U-shaped function of target visibility as a function of SOA.

Visual masked priming aims then to investigate the influence of a task-irrelevant and “unconscious” stimulus (i.e., a masked prime stimulus) on the response to a succeeding relevant target stimulus (Ansorge et al., 2014; Marcel, 1983). In the specific case of masked response priming, the masked primes are mapped to the same responses as the target stimuli (Klotz & Neumann, 1999; Vorberg et al., 2003). Remember, a trial is considered consistent when prime and target are mapped to the same response, and inconsistent when not. As discussed in Chapter 3.2, the so-called priming effect increases with the SOA between prime and target, up to an SOA of 100 ms. If one constructs the target stimulus in a way that allows it to simultaneously serve as the mask, and one considers the aforementioned masking functions, this allows for a dissociation between the visibility of a prime stimulus and its effect on the response to a target stimulus (Biafora & Schmidt, 2020; T. Schmidt & Vorberg, 2006; Vorberg et al., 2003). For example, if the priming effect shows a strictly monotonic increase with SOA, but Type-B masking of the prime is found, there exists a dissociation between the prime’s visibility and its influence on response activation. The priming effect, in which inconsistent trials produce longer RTs and more errors than consistent trials, and which increases up to an SOA of 100 ms, is also referred to as the *positive compatibility effect* (PCE). In contrast, when SOAs longer than 100 ms are employed, and especially when the prime is masked, a reversal of this effect has been found. Under such experimental variations, inconsistent trials can produce shorter RTs

and fewer errors than consistent trials, a phenomenon called the *negative compatibility effect* (NCE; Eimer & Schlaghecken, 1998; Lingnau & Vorberg, 2005). Eimer and Schlaghecken (1998) employed such a masked response priming paradigm and analyzed lateralized readiness potentials to trace neuronal preparation of motor activity, a left or right button press response. They found an initial response activation in line with the prime's identity, followed by a temporary activation of the opposite response (antiprime) – considered to be responsible for the emergence of the NCE – and, ultimately, an activation of the response related to the target's identity. Panis and Schmidt (2016) found the same time course of response activation when they analyzed the hazard functions of response occurrence.

The theory of *self-inhibition* proposes that a prime automatically triggers its own inhibition, either when a mask removes “perceptual evidence” for such a prime, or when the SOA between the prime and target is long enough (Eimer & Schlaghecken, 1998, 2003). Inhibition of the primed response in turn results in the transient activation of the opposite response: the disinhibition of the antiprime response. Ultimately, the NCE emerges.

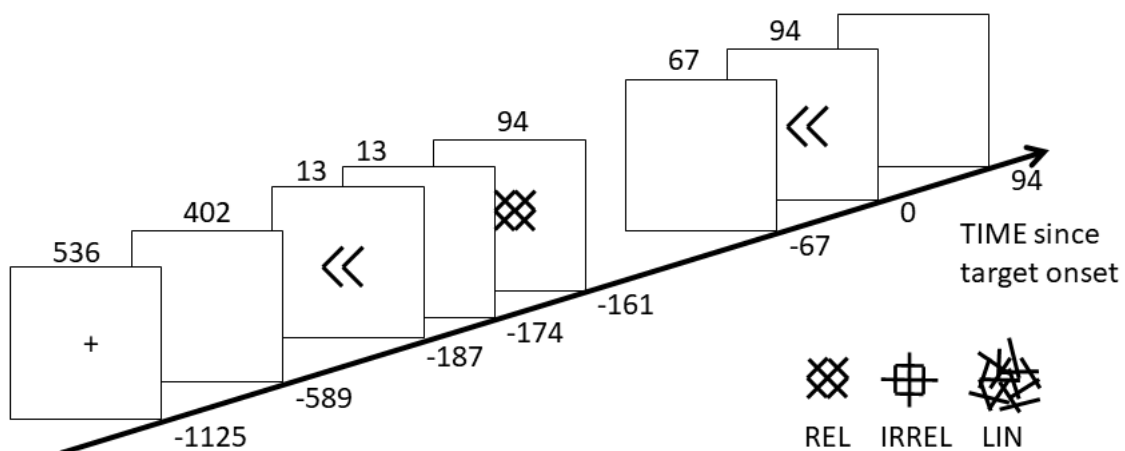


Figure 7. Experimental design taken from Panis and Schmidt (2016) as cited in Panis, Schmidt, et al. (2020). Three different types of masking: REL, structure masking with response-relevant features; IRREL structure masking with response-irrelevant features; LIN, random noise masking.

In contrast, the theory of *object-updating* argues that the NCE reported in Eimer and Schlaghecken (1998) is the result of the stimulus features they used (Lleras & Enns, 2004). The mask was a superposition of the two stimuli that could serve as prime and/or target (<< and >>). Thus, when this mask followed, for example, a “<<” prime, it effectively was merely the addition of “>>” (i.e., arrows pointing in the antiprime direction, compare condition REL in Fig. 7). Therefore, Lleras and Enns (2004) conclude that the original finding of the NCE was simply the result of the positive priming of the antiprime response.

A third account is that of *mask-triggered inhibition* (Jaśkowski & Przekoracka-Krawczyk, 2005). Instead of the prime triggering its own inhibition, a sufficiently

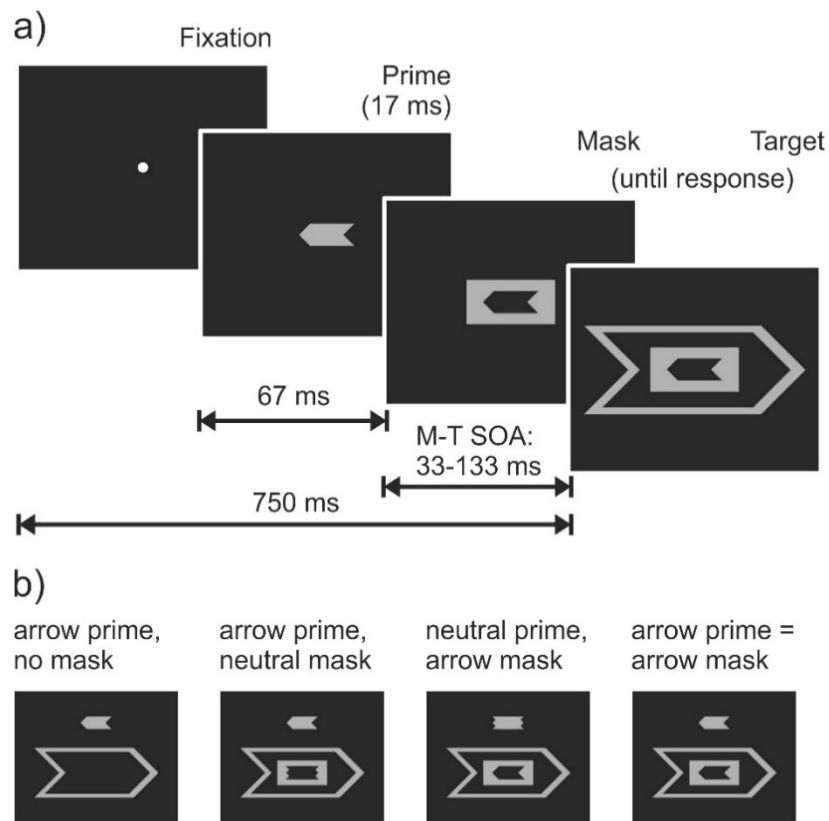


Figure 8. Figure 2 from T. Schmidt et al. (2022). a) Time course of a trial. Note that both mask and target remain on screen until the response is completed. b) The four masking conditions. For better legibility, the prime is drawn above the mask and target. In the experiment, it appeared in the central mask cutout so that their contours were adjacent.

strong mask actively inhibits the prime response (Jaśkowski, 2007, 2008, 2009; Jaśkowski et al., 2008; Jaśkowski & Przekoracka-Krawczyk, 2005). A mask would not need to actually be “good” at masking, its signal would just need to be strong enough to trigger this “emergency break”.

Panis and Schmidt (2016) showed that inhibition, and thus the NCE, is in fact time-locked to the mask and not the prime, supporting the claims of the mask-triggered inhibition over the self-inhibition account. Moreover, they found late inhibition of the primed response strong enough to reverse the strong response priming effect. In Panis, Schmidt, et al. (2020), we used their data set to show the application of and guide through the process of discrete-time EHA.

The goal of T. Schmidt et al. (2022) was to further investigate the NCE by utilizing four different masking conditions (see Fig. 8). Contrasting these conditions allowed testing of the following hypotheses: (1) A neutral mask is sufficient to generate the NCE, (2) the NCE occurs also in the absence of visual features that could elicit an antiprime response, and (3) the NCE occurs even if the mask does not remove perceptual evidence for the prime. By analyzing conditional accuracy functions, we were further able to track the time-course of selective mask-triggered inhibition. Specifically, we revealed that the NCE is even larger than the response priming effect and is directed against the primed response, and showed that response-specific inhibition does not require any positive priming features.

Articles

**Analyzing response times and other types of time-to-event data using
event history analysis: A tool for mental chronometry and cognitive
psychophysiology¹**

¹ *Reproduced with permission from SAGE under the terms of the Creative Commons Attribution 4.0 License
(<https://creativecommons.org/licenses/by/4.0/>).*

Analyzing Response Times and Other Types of Time-to-Event Data Using Event History Analysis: A Tool for Mental Chronometry and Cognitive Psychophysiology

i-Perception

2020, Vol. 11(6), 1–24

© The Author(s) 2020

DOI: 10.1177/2041669520978673

journals.sagepub.com/home/ipe

**Sven Panis** 

Experimental Psychology Unit, Faculty of Social Sciences, Technische Universität Kaiserslautern, Kaiserslautern, Germany

Filipp Schmidt

Abteilung Allgemeine Psychologie, Fachbereich 06, Psychologie und Sportwissenschaft, Justus-Liebig-Universität Gießen, Giessen, Germany

**Maximilian P. Wolkersdorfer and
Thomas Schmidt**

Experimental Psychology Unit, Faculty of Social Sciences, Technische Universität Kaiserslautern, Kaiserslautern, Germany

Abstract

In this Methods article, we discuss and illustrate a unifying, principled way to analyze response time data from psychological experiments—and all other types of time-to-event data. We advocate the general application of discrete-time event history analysis (EHA) which is a well-established, intuitive longitudinal approach to statistically describe and model the shape of time-to-event distributions. After discussing the theoretical background behind the so-called hazard function of event occurrence in both continuous and discrete time units, we illustrate how to calculate and interpret the descriptive statistics provided by discrete-time EHA using two example data sets (masked priming, visual search). In case of discrimination data, the hazard analysis of response occurrence can be extended with a microlevel speed-accuracy trade-off

Corresponding author:

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

Email: sven.panis@sowi.uni-kl.de



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

analysis. We then discuss different approaches for obtaining inferential statistics. We consider the advantages and disadvantages of a principled use of discrete-time EHA for time-to-event data compared to (a) comparing means with analysis of variance, (b) other distributional methods available in the literature such as delta plots and continuous-time EHA methods, and (c) only fitting parametric distributions or computational models to empirical data. We conclude that statistically controlling for the passage of time during data analysis is *equally* important as experimental control during the design of an experiment, to understand human behavior in our experimental paradigms.

Keywords

response times, event history analysis, hazard function, conditional accuracy function, speed-accuracy trade-off, survival analysis, transition analysis

Date received: 31 December 2019; accepted: 16 November 2020

Since the publication of the subtraction method (Donders, 1969) and the additive factors method (Sternberg, 1969), analysis of variance (ANOVA) has become the standard data analysis method in psychology and cognitive (neuro)science for the analysis of response times (RTs). Following these approaches, many researchers interpret differences in RTs between experimental conditions on a difference scale that is assumed to directly capture the time requirements of additional cognitive operations. However, differences in mean RT can only be interpreted that way when assuming that the nature of cognitive processing is captured by the serial information processing framework. Even though the serial information processing framework has been criticized repeatedly in the literature (Cisek & Kalaska, 2010; Eriksen & Schultz, 1979; McClelland, 1979; Pieters, 1983; Schöner et al., 2016), ANOVA continues to be the most popular method to analyze RTs to this day.

As discussed by Van Gelder (1995), there is a viable alternative view on the nature of cognitive processing: Cognition is the behavior of a dynamical system. To understand the behavior of a dynamical system, it is crucial to track its output over time (Schöner et al., 2016). We therefore promote and illustrate the use of a well-established longitudinal or distributional technique known as event history analysis (EHA) for analyzing time-to-event data such as RTs. EHA (also known as survival, hazard, duration, transition, and failure-time analysis) is the name of the standard set of statistical methods for studying *the occurrence and timing of events* in many scientific disciplines (Allison, 2010; Singer & Willett, 2003). While EHA is already applied in many areas of the human sciences, including developmental psychology (Ha et al., 1997), clinical psychology (Corning & Malofeeva, 2004; Greenhouse et al., 1989; Willett & Singer, 1993), social psychology (Griffin, 1993; Keiley & Martin, 2005; Núñez-Antón & Orbe, 2005; Steele et al., 1996, 2004; Stoolmiller & Snyder, 2006), organizational psychology (Morita et al., 1989), and even cognitive psychology (Chechile, 2006; Pannasch et al., 2001; Panis & Wagemans, 2009; Torfs et al., 2010; Wenger & Gibson, 2004; Yang & McConkie, 2001), an introduction to EHA that focuses on its relevance for cognitive (neuro)scientists is still warranted as its use is currently still rather rare. As we will see later, the use of a more advanced and well-established analysis method can maximize the return from the collected data, which is important in view of the costs and time required to run an experiment (Whelan, 2008).

To apply EHA, we must be able to define the event of interest (any qualitative change that can be situated in time), to define time point zero, and to measure the passage of time between time zero and event occurrence in discrete or continuous time units. While sociologists are interested in the occurrence and timing of events such as marriage and divorce—note that some people never marry—and biostatisticians in death, experimental psychologists are interested in events such as button presses (RT analysis), saccade onsets (saccade latency analysis), fixation offsets (fixation duration analysis), and so forth. Typically, time point zero is defined as target display onset time in RT and saccade latency studies. However, sometimes the time of the last response can be defined as time zero for the next response, for example, when studying perceptual dominance durations in studies using ambiguous figures. The onset of fixation is time zero for fixation duration analysis.

The structure of this Methods article is as follows. First, we introduce and explain the concept of hazard, in continuous and discrete time units. Next, we illustrate how to calculate the descriptive statistics in discrete time using a life table, and we discuss two example data sets. We then describe different approaches for obtaining inferential statistics. We end with a discussion of the (dis)advantages of discrete-time EHA, compared with other existing distributional methods.

The Continuous-Time Hazard Rate Function of Event Occurrence

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 RT in a particular experimental condition, including (a) the cumulative distribution function $F(t) = P(T \leq t)$, (b) its derivative $F'(t)$ known as the probability density function $f(t)$, (c) the survivor function $S(t) = 1 - F(t) = P(T > t)$, and (d) 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)

EHA has been developed to describe and model the hazard function of event occurrence (for RT data, the event is a button-press response). For continuous RT data, hazard quantifies the instantaneous risk that a response will occur at time point t , given that it has not occurred before 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. Just as speed is defined as a rate—the distance covered per unit time—so too is the continuous-time hazard $\lambda(t)$. For example, if time is measured in seconds and $\lambda(3.0) = 2$, then the instantaneous rate of event occurrence equals two events per second after 3 seconds of waiting time. 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 of time-to-event data.

First, the hazard function of response occurrence is one of the most diagnostic functions when describing the distribution of a sample of time-to-event data (Allison, 2010; Luce, 1986; Townsend, 1990). For example, “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 1 shows the $F(t)$, $f(t)$, $S(t)$, and $\lambda(t)$ for four theoretical waiting-time distributions. In contrast to $\lambda(t)$, all $F(t)$ and $S(t)$ distributions look vaguely alike, and we cannot easily describe salient features other than the mean and standard deviation. Also, the density

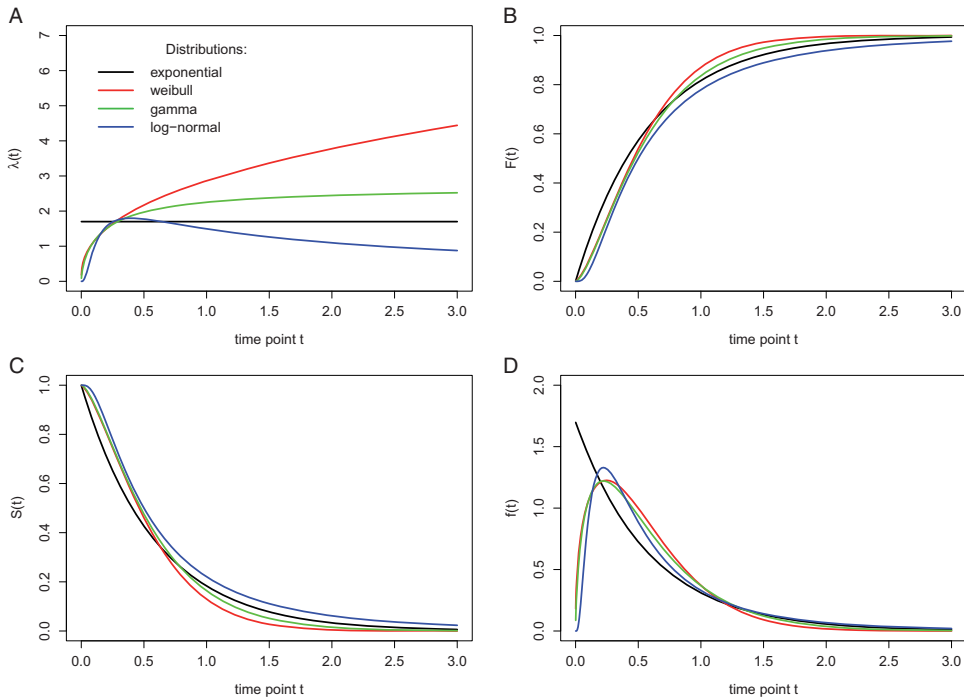


Figure 1. Four views on four different waiting-time distributions in continuous time. The hazard rate function $\lambda(t)$ (A), the cumulative distribution function $F(t)$ (B), the survivor function $S(t)$ (C), and the probability density function $f(t)$ (D) are shown for each of four theoretical probability distributions (different colors: exponential, Weibull, gamma, log-normal). While the hazard rate 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.

function $f(t)$ conceals what is happening in the right tail of the distribution (Luce, 1986). As discussed by Holden et al. (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 (1990) 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 (1990) showed that a complete ordering on the hazard functions— $\lambda_A(t) > \lambda_B(t)$ 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 EHA. For example, when mean A < mean B, the hazard functions might show a complete ordering (i.e., for all t), a partial ordering (e.g., only for $t > 300$ ms, or only for $t < 500$ ms), or they may cross each other one or more times.

Third, EHA does not discard right-censored observations when estimating hazard functions, that is, 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. This is important because although a few right-censored observations are inevitable in most RT tasks, a lot of right-censored observations are expected in experiments on masking, the attentional blink, and so forth, for example.

There are other types of censoring. Left censoring occurs when all that is known about an observation on a variable T is that it is *less* than some value. Interval censoring combines right and left censoring so that all you know about T is that $a < T < b$, for some values of a and b (Allison, 2010). Random censoring occurs when observations are terminated for reasons that are not under the control of the experimenter.

Importantly, all standard statistical methods for time-to-event data require that random censoring be noninformative: For example, a trial that is censored at time c should be representative of all those trials with the same values of the explanatory variables that survive to c (Allison, 2010). For example, the occurrence of 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 et al., 2004; Eriksen et al., 1985; Praamstra & Seiss, 2005). We therefore never recommend to describe or model the hazard of correct response occurrence independently from the hazard of error response occurrence but to extend the hazard of response occurrence with conditional accuracy functions (see later).

The most common type of right-censoring is “singly Type I censoring” that applies when the experiment uses a fixed response deadline for all trials. “Type I” means that the censoring time is fixed and is under the control of the experimenter, and “singly” refers to the fact that all observations have the same censoring time (Allison, 2010). Discarding such trials—or trials with very long RTs in case the experimenter waits for a response on each trial—may introduce a sampling bias that results in underestimation of the mean. In contrast, EHA can include the data information from all trials when estimating the descriptive statistics.

Fourth, hazard modeling allows incorporating *time-varying* explanatory covariates such as heart rate, electroencephalogram (EEG) signal amplitude, gaze location, and so forth (Allison, 2010) which is useful for cognitive psychophysiology (Meyer et al., 1988). For more information, see Singer and Willett (2003, pp. 426–442) and Allison (2010, pp. 243–246).

Finally, hazard is more suited as a measure of the concept of processing capacity, that is, the amount of work the observer is capable of performing within some unit of time (Wenger & Gibson, 2004). 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.

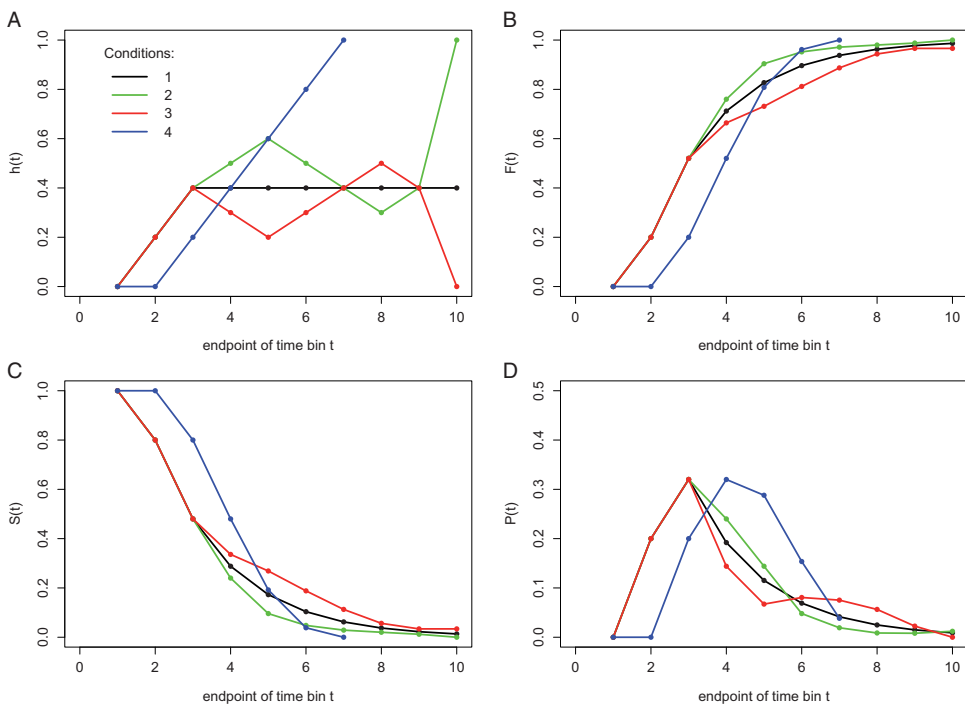
The Discrete-Time Hazard Probability Function of Event Occurrence

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 1,000 trials for example (Bloxom, 1984; Luce, 1986; Van Zandt, 2000). However, by shifting to discrete time, we can trade-off some temporal resolution for increased applicability of EHA in RT studies that typically collect less than 1,000 trials per condition per participant. In this Methods article, we therefore focus on the 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).

In Figure 2A, four hypothetical discrete-time population hazard functions are plotted with time divided in 10 discrete bins (1–10). Each function was constructed by selecting a series of 10 real numbers from the interval $[0,1]$ with replacement, with the only restriction that once “1.0” is selected then the following numbers are set to “missing data”—the reader can construct her or his own example functions. Each hazard function completely describes the shape of a distribution of discrete waiting times. For example, the four theoretical functions in Figure 2A could reflect the true RT distributions of a single participant in four experimental conditions (studied with a small- N design; in which a large number of observations are made on a relatively small number of experimental participants, Smith & Little, 2018); in this example, time might have been measured in discrete time bins of 50 ms each, with a censoring time of 500 ms. Or they might reflect the true distributions of the time it takes to earn a first doctoral degree measured in years for four groups of 100 participants with certain personality characteristics (large- N design), with a censoring time of 10 years. In each case, $h(t)$ gives the conditional probability that the event of interest occurs in bin t given that it has not yet occurred before, or $h(t) = P(T = t | T \geq t)$, where T is a discrete random variable denoting the rank of the time bin in which the event occurs. The discrete-time hazard function of event occurrence thus tells us the probability that the event we are still waiting for (at the start of bin t) will actually occur in bin t .

Figure 2C displays the corresponding discrete-time survivor functions, or $S(t) = P(T > t) = [1 - h(t)] * [1 - h(t-1)] * \dots * [1 - h(1)]$, which gives for each bin the probability



that the response does not occur before the end of bin t . The survivor function is the complement of the cumulative distribution function (Figure 2B), or $S(t) = 1 - F(t) = 1 - P(T \leq t)$. Figure 2D shows the corresponding probability mass functions, or $P(t) = P(T = t) = h(t) * S(t-1)$.

We constructed the hazard functions in Figure 2A in such a way that they show some symmetry. For example, Condition 1 (black line) might represent a neutral priming condition and Conditions 2 and 3 a congruent and incongruent priming condition, respectively. Let us assume for simplicity that each bin is 1 second wide and that the censoring time equals 10 seconds so that we have the following sequence of bins: (0,1], (1,2], ..., (9,10]. For example, the discrete-time hazard for bin 2 in the neutral condition equals .20 (for bins 1–3, the hazard functions for the first three conditions lie on top of each other). In other words, given that time has passed until 1 second after target onset without response occurrence, then there is a conditional probability of .2 that the response occurs sometime during the next second, that is, in the second bin or time segment (1,2]. In short, $h(2) = .2$. When the waiting time has increased to 2 seconds, $h(3) = .4$, and so forth.

If we now compare Conditions 2 and 3 (green and red lines), we see a large positive priming effect in hazard for time segment (3,6] followed by a smaller negative (i.e., inverted) priming effect for time segment (7,8]. Note that while the hazard functions for Conditions 2 and 3 cross two times, the $S(t)$ and $F(t)$ functions do not cross, because they cumulate the (complement of the) current and previous hazard values. This implies that also quantile plots and delta plots—and other types of visualization based on plotting and comparing quantiles from $F(t)$ —would not be able to reveal the crossing that is visible in hazard.

Similarly, note that the symmetry present in the hazard functions for the first three conditions is also absent in the $P(t)$ functions. As a matter of fact, if we would only study $P(t)$, we might conclude incorrectly that the late negative priming effect lasts longer than the early positive priming effect. However, the $P(t)$ values do not give any information on the time course of event occurrence because they denote the probability that the event occurs in bin t given that it can occur in *any* (previous, current, or future) bin. In other words, they simply tell you how many percent of all trials will experience the event in bin t . Note that the $P(t)$ values in Figure 2D do not sum to 1 for Conditions 1 and 3, which is why these are called *subprobability* mass functions (Chechile, 2003); also, the corresponding $h(t)$ and $F(t)$ functions do not reach 1, and the $S(t)$ functions do not reach zero.

Obtaining Descriptive Statistics for Discrete Time Units: The Life Table

To calculate the descriptive statistics—functions of discrete time—for a finite time-to-event data set, one has to set up a *life table*. In the context of a small- N design, the life table summarizes the history of event occurrences for a combination of participant and experimental condition. To set up a life table, you need to determine the censoring time and divide it up into a sequence of contiguous time bins. The fixed censoring time point is typically the response deadline used, or a time point after which you expect no useful responses anymore in any trial of any condition. In this section, we shortly discuss real data from two published experiments using a small- N design, one on masked priming, and one on visual search.

Masked Priming

Panis and Schmidt (2016) asked participants to perform speeded keypress responses to the direction of a 94 ms double arrow target (left/right), within 600 ms (Figure 3A). The central

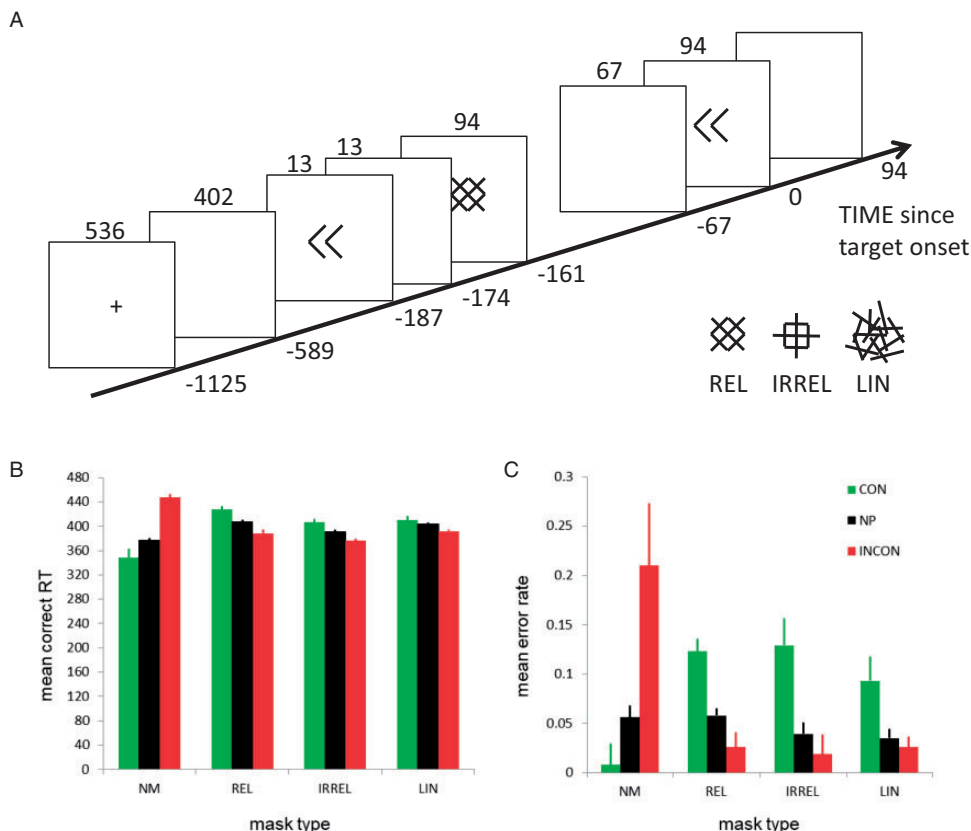


Figure 3. Masked priming example. (A) Trial and mask designs used in Experiment 1 of Panis and Schmidt (2016). A trial with a congruent prime and a relevant mask is shown. Insets show three mask types. Time on the x axis is measured in milliseconds relative to target onset. (B) Mean correct RT. (C) Mean error rate. Error bars represent ± 1 SEM corrected for between-subject variation. NP = no prime; CON = congruent prime; INCON = incongruent prime; NM = no mask; REL = relevant mask; IRREL = irrelevant mask; LIN = random lines mask.

target could be preceded by a central 13 ms double arrow prime that was followed by a 94-ms pattern mask. The factors prime type and mask type were manipulated factorially. The prime could point in the same direction as the target (CONgruent), in the opposite direction (INCONgruent), or no prime was presented (NP). The mask stimulus could be response-relevant (REL), response-irrelevant (IRREL), a set of random lines (LIN), or no mask was presented (NM). Consistent with the literature, the mean correct RT (Figure 3B) and mean error rates (Figure 3C) show a positive priming or congruency effect (PCE) of about 100 ms and 20 percentage points when no mask was presented, but the reversed effect in the presence of relevant or irrelevant masks: a negative congruency effect (NCE) of about -40 ms and -10 percentage points.

Table 1 presents the life table for the data of a single participant in condition NP-NM (no prime, no mask). The first 600 ms after target onset are divided into 15 bins of 40 ms indexed by $t = 1$ to 15. After counting the number of responses in each bin, one can then directly estimate the discrete-time hazard probability function of response occurrence: $h(t) = P(T = t \mid T \geq t)$, where $T \geq t$ denotes the event that the response does not occur before the start of bin

Table 1. Example Life Table for Discrete-Time Statistics.

bin	t	rc	E	RS	h(t)	se[h(t)]	S(t)	se[S(t)]	P(t)	se[P(t)]	# correct	# error	ca(t)	se[ca(t)]
[0,40]	1	0	0	220	0	0	1	0	0	0	0	0	NA	NA
(40,80]	2	0	0	220	0	0	1	0	0	0	0	0	NA	NA
(80,120]	3	0	0	220	0	0	1	0	0	0	0	0	NA	NA
(120,160]	4	0	0	220	0	0	1	0	0	0	0	0	NA	NA
(160,200]	5	0	0	220	0	0	1	0	0	0	0	0	NA	NA
(200,240]	6	0	0	220	0	0	1	0	0	0	0	0	NA	NA
(240,280]	7	0	7	220	0.032	0.012	0.968	0.012	0.032	0.012	2	5	0.29	0.171
(280,320]	8	0	13	213	0.061	0.016	0.909	0.019	0.059	0.016	10	3	0.77	0.117
(320,360]	9	0	26	200	0.130	0.024	0.791	0.027	0.118	0.022	24	2	0.92	0.052
(360,400]	10	0	40	174	0.230	0.032	0.609	0.030	0.182	0.026	40	0	1	0
(400,440]	11	0	48	134	0.358	0.041	0.391	0.028	0.218	0.028	47	1	0.98	0.021
(440,480]	12	0	37	86	0.430	0.053	0.223	0.020	0.168	0.025	37	0	1	0
(480,520]	13	0	32	49	0.653	0.068	0.077	0.010	0.145	0.024	32	0	1	0
(520,560]	14	0	9	17	0.529	0.121	0.036	0.005	0.041	0.013	9	0	1	0
(560,600]	15	4	4	8	0.500	0.177	0.018	0.003	0.018	0.009	4	0	1	0

Note. This life table is based on the 220 trials of Participant 6 in the target-only condition (NP-NM) of Experiment 1 of Panis and Schmidt (2016). For each time bin (column 1) with rank τ (column 2), the number of observed responses (E) are counted, and the risk set (RS) is determined, before estimating (a) the discrete-time hazard function $h(t) = P(T = t | T \geq t)$ as E/RS , (b) the survivor function $S(t) = P(T > t) = [1 - h(1)] * \dots * [1 - h(t)]$, and (c) the probability mass function $P(t) = P(T = t) = h(t) * S(t-1)$. The standard errors for $h(t)$, $P(t)$, and $ca(t)$ are estimated using the familiar formula for a proportion p —the square root of $\{p(1-p)/N\}$ —where N equals $RS(t)$ for $h(t)$, $RS(1)$ for $P(t)$, and $E(t)$ for $ca(t)$. The standard errors for $S(t)$ are estimated using the recurrent formula on page 350 of Singer and Willett (2003). Note that $P(t)$ also equals the number of events in bin t divided by the risk set of the first bin: $P(t) = E(t)/RS(1)$. Four trials are right-censored (rc) at 600 ms after target onset (column 3), that is, no response occurred for these trials during the 600 ms data collection period so that we only know that $RT > 600$ ms. At time point zero, $S(0) = 1$, $P(0) = 0$, and $h(0)$ is undefined (NA). It is important to realize the difference between the probability mass and the hazard function. The (sub)probability mass function calculates the response count in a given time bin relative to the number of all trials, whereas the hazard function calculates the response count relative to the number of all trials that are still response-free up to the start of that time bin.

t. For each bin t , the sample-based estimate of $h(t)$ is obtained by dividing the number of observed responses in bin t by the *risk set* of bin t , which is the number of trials that are still response-free at the start of bin t . Note that the four right-censored observations—trials without response occurrence for which we only know that RT must be larger than 600 ms—do contribute to the risk set of each bin (ignoring such trials creates a sampling bias). Also note how the standard error of $h(t)$ tends to increase as the waiting time increases, because the risk set is becoming rather small for later time bins.

Because we are dealing with two-button discrimination data, the $h(t)$ analysis of response occurrence is extended with an analysis of conditional accuracy, that is, the microlevel speed-accuracy trade-off function (Allison, 2010; Pachella, 1974; Wickelgren, 1977). The conditional accuracy function, or $ca(t) = P(\text{correct} | T = t)$, is the conditional probability that an observed response is correct given that it occurs in bin t and is estimated by dividing the number of correct responses in bin t by the number of observed responses in bin t (Table 1). By using $h(t)$ functions in combination with $ca(t)$ functions, one can provide an unbiased, time-varying, and probabilistic description of the latency and accuracy of *any* sample of (right-censored) event times.

Sample-based estimates of $h(t)$, $S(t)$, $P(t)$, and $ca(t)$ are shown for one participant in Figure 4, for two mask conditions (none and relevant) and three prime types (No Prime, CONgruent, INCONgruent). We refer to each bin using its endpoint, for example, the

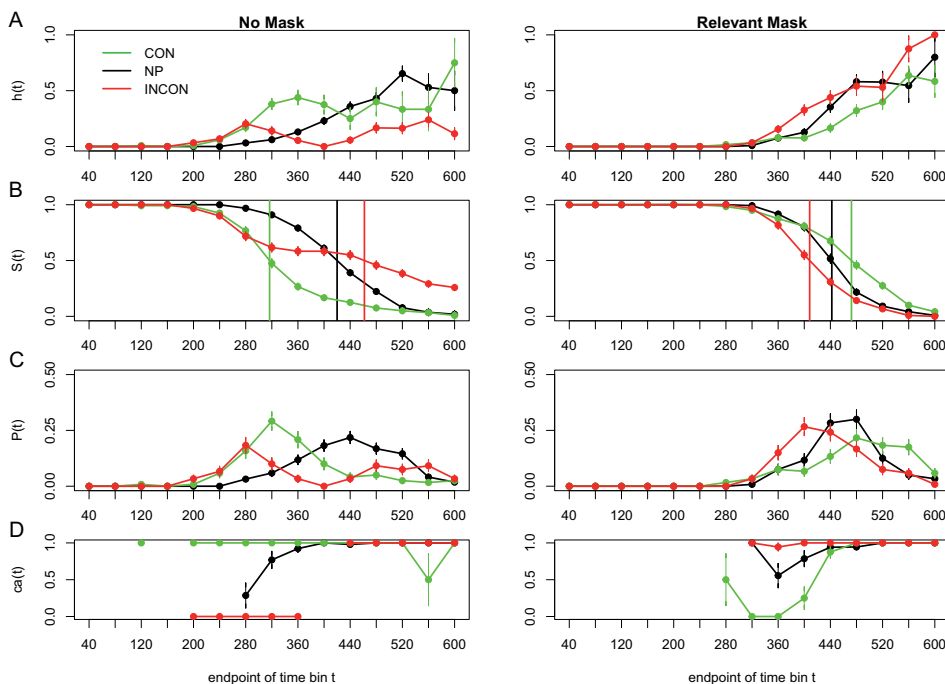


Figure 4. Sample-based estimates for Participant 6 in Experiment I of Panis and Schmidt (2016). For each combination of mask type (no mask and relevant mask) and prime type (congruent, no prime, incongruent), the estimated discrete-time hazard function $h(t)$ is plotted (A), together with the survivor function $S(t)$ (B), the (sub)probability mass function $P(t)$ (C), and the conditional accuracy function $ca(t)$ (D). Time axes are relative to target onset. Error bars represent ± 1 standard error of the respective proportion. CON = congruent prime; NP = no prime; INCON = incongruent prime.

hazard estimate for bin (240,280] is $h(280)$. Figure 4 offers a fascinating view into the microgenesis of primed responses. In the no mask conditions (left panels), response onset is much earlier when primes are present, and the upswing in response hazards is at first identical for consistent and inconsistent primes. If such an early response is emitted, it is always correct for congruent primes and always incorrect for incongruent primes, as shown by the $ca(t)$ functions. This clearly indicates that these initial responses are triggered exclusively by the prime without any contribution from the target (the crucial prediction of the rapid-chase theory of response priming; Schmidt et al., 2006, 2011, 2015).

Once the waiting time has reached 280 ms after target onset without response occurrence, response hazards continue to increase temporarily for congruent primes but start to decline for incongruent primes and eventually even reach zero: in bin (360,400] after target onset, no responses are emitted when the prime is incongruent. In our view, this temporary decline in hazard reflects—at least initially—response competition from the target, which is becoming overtly available in bin 280 and activates the opposite (correct) response as the prime. In other words, this is the phase where the target starts taking over response control from the prime. After bin 400, $h(t)$ starts to increase again in the incongruent condition, and if such a late response is emitted, it is always correct. Thus, the response conflict has been resolved in favor of the target, and these late responses are controlled entirely by the target's identity.

But something else is going on in the relevant mask condition (right panels). The first overt responses only appear around 320 ms after target onset. Overall, response hazards increase faster in incongruent than in congruent trials (with the no-prime condition in between), demonstrating a reversed priming or NCE in response occurrence. Moreover, the earliest emitted responses are typically correct in incongruent trials and typically incorrect in congruent trials: a complete reversal of the pattern in the no mask condition. When the target information becomes available, it now delays responses in the congruent condition around 360 ms after target onset. Following this temporary dip, $h(t)$ sharply increases, and all responses emitted after 480 ms are correct.

The hazard functions for congruent and incongruent trials thus show a partial ordering (i.e., only for $t > 280$ ms in the no mask condition, and for $t > 320$ ms in the relevant mask condition). In other words, the hazard functions reveal the onset time, duration, and shape of the behavioral effect. The differences in means also typically underestimate the duration of the effect in terms of hazard. For example, the within-trial duration of the PCE when the mask is absent is at least 200 ms (5 bins) and that of the NCE when the mask is relevant is at least 160 ms (4 bins). Also, plotting hazard and conditional accuracy functions can reveal important interindividual differences and the time-locking of effects to stimuli or other events. For example, Panis and Schmidt (2016) compared the dynamics of the priming effect in the $ca(t)$ functions for the six different participants and found a high similarity (Figure 5A): Every participant showed a temporary PCE in the no mask condition and a temporary NCE in the various masking conditions. Figure 5B shows the result of a second experiment where the prime-mask and mask-target stimulus-onset-asynchronies (SOAs) were varied independently. The plot shows that three distinct states can be identified when the prime-mask SOA is long (conditions “long–short” and “long–long”): a PCE state time-locked to prime-onset, an NCE state time-locked to mask onset, and an “all correct” state time-locked to target onset. Note that the same three states have been observed in the Lateralized Readiness Potential by Jaśkowski et al. (2008) and Eimer and Schlaghecken (1998). Crucially, the NCE appears ~ 360 ms after mask onset in every condition, an estimate very similar to the 350 ms estimate obtained by looking at pointing movement trajectories (Schmidt et al., 2015).

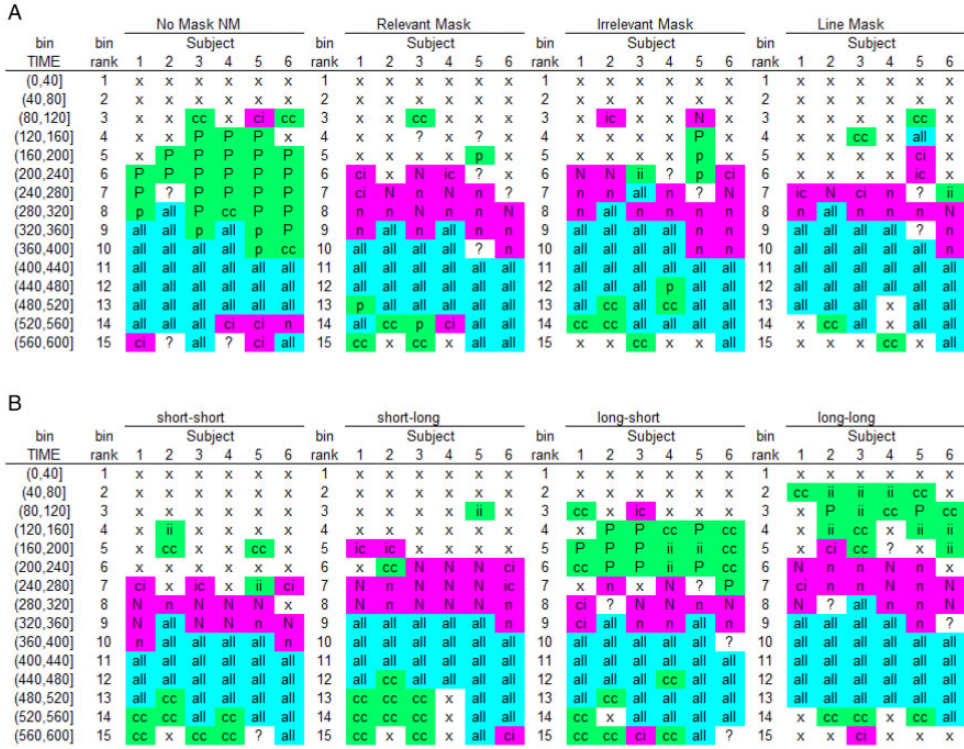


Figure 5. Sample-based $ca(t)$ -state transition plots. For each participant, bin, and mask type (A, Experiment 1) or SOA combination (B, Experiment 2), we first coded the type of difference in observed performance in $ca(t)$ between congruent (CON) and incongruent (INCON) prime conditions and then applied a color code (green = evidence for PCE; pink = evidence for NCE; cyan = no evidence for either). Specifically, for bins where responses are observed for both CON and INCON: “P”: $ca(t) = 1$ for CON and $ca(t) = 0$ for INCON; “p”: CON minus INCON $\geq .2$; “N”: $ca(t) = 0$ for CON and $ca(t) = 1$ for INCON; “n”: CON minus INCON $\leq -.2$; “all”: $ca(t) > .8$ for both CON and INCON. For bins where responses exclusively occur in either CON or INCON: “cc”: $ca(t) = 1$ for CON and no responses for INCON; “ii”: no responses for CON and $ca(t) = 0$ for INCON; “ic”: $ca(t) = 0$ for CON and no responses for INCON; “ci”: no responses for CON and $ca(t) = 1$ for INCON. Remaining bins: “x”: no responses observed in CON and INCON; “?”: other cases. The reader can compare the codes for Participant 6 in Figure 5A (relevant and no mask) with Figure 4D.

NM=no mask.

Panis and Schmidt (2016) concluded that the NCE is neither caused by automatic self-inhibition of the primed response due to backward masking nor by updating response-relevant features of the mask, but by active, selective mask-triggered inhibition. The mask thus acts as a *stop-signal* within the current task context that initiates selective inhibition of the premature prime-triggered response, which temporarily disinhibits the opposite response (thrust reversal; Schmidt et al., 2015). Importantly, these distributional results are compatible with a computational model of the basal ganglia, a subcortical collection of nuclei that are involved in response gating and (selective and global) response inhibition (Wiecki & Frank, 2013).

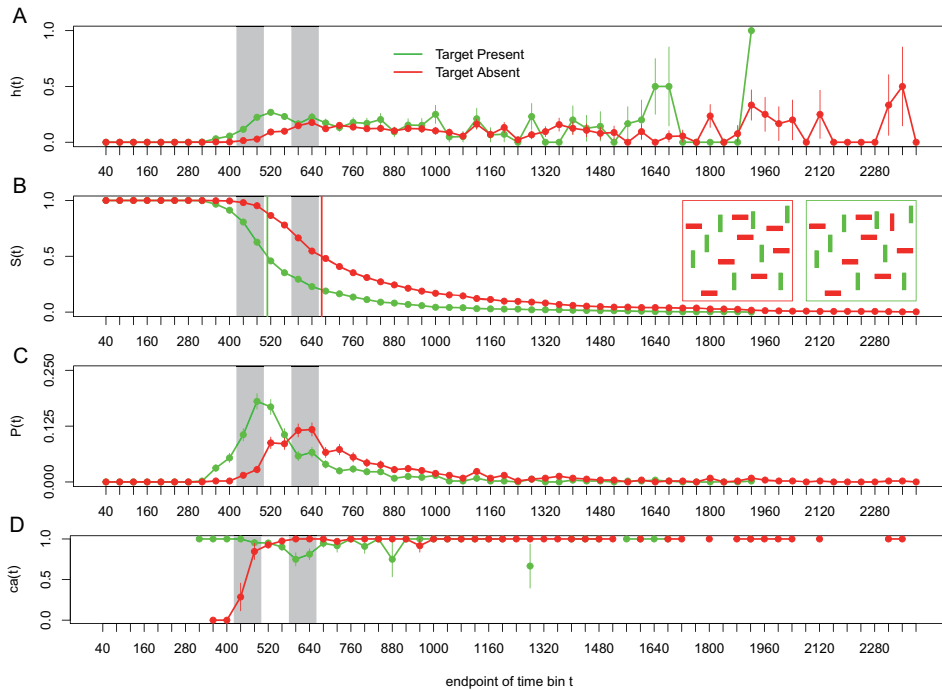


Figure 6. Visual search example. The data for one representative participant in each of the target-present and target-absent conditions of the color-orientation conjunction search task with set size 18 of Wolfe et al. (2010) are plotted as (A) hazard function $h(t)$, (B) survivor function $S(t)$, (C) (sub)probability mass function $P(t)$, and (D) conditional accuracy function $ca(t)$. Both insets in Figure 6B show example displays. The target is a vertical red object. The passage of time is measured discretely using bins of 40 ms starting at search display onset. The vertical lines in Figure 6B show the estimated median response times for the target-present and target-absent conditions. The gray surface areas are used for interpretation (see main text). Error bars represent ± 1 standard error of the respective proportion.

Visual Search

Panis et al. (2020) reanalyzed the benchmark visual search data sets collected by Wolfe et al. (2010). For example, in the color-orientation conjunction search task, 10 participants searched a single display for a red vertical rectangle among green vertical and red horizontal rectangles. Four different set sizes (target plus distractors; 3, 6, 12, or 18) were randomly intermixed. Participants pressed one key if the target was present (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. Figure 6 shows the data for one representative participant in the color-orientation conjunction search task with a set size of 18 objects, using bins of 40 ms and a censoring time of 2,400 ms.

First, there is only a partial ordering of the hazard functions with respect to the effect of target presence (only for $t < 600$ ms), and the hazard functions are relatively flat for the right tail of the RT distributions. Second, false alarms occur mostly early in time, while misses occur mostly for medium-latency responses. The miss rate peaks around 600 ms after search

display onset. As far as we know, none of these features of visual search behavior are predicted by current cognitive models of visual search (Panis et al., 2020).

One tentative interpretation of these data is based on the idea that behavior at any point in time is determined not only by the outcome of the ongoing search process but also by response biases and reactive cognitive control processes (Panis et al., 2020). For example, we can distinguish five phases in the time-dispersed behavior of this observer (the gray surface areas in Figure 6 mark phases two and four). First, if the waiting time has increased until 360 ms after search display onset, then $h(400)$ is higher for target-present than target-absent trials, and all emitted responses are correct for target-present, but incorrect for target-absent trials. The earliest responses thus show a strong yes-bias, regardless of target presence. Kiss et al. (2012) concluded that the attentional selection of targets that are defined by a combination of features—here: “red” and “vertical”—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 suggests that at the end of the initial feedforward sweep of processing right after display onset, all elements in the search display will have captured attention to some extent, each signaling the presence of target features such as red and/or vertical in the conjunction search task. This explains the presence of the early yes-response bias. We also assume that the target is indeed found on a few of the target-present trials (e.g., those where the target is very salient due to spatial grouping processes), which explains the higher hazard for target-present trials. If no early response occurs, however, time passes on, and the search continues.

Second, in the time range 400–480 ms, hazard further increases for both conditions, while $ca(t)$ quickly increases above chance level for target-absent trials and starts to slightly decrease for target-present trials. Thus, while the search process might finish on a subset of trials in this time range, Panis et al. (2020) suggested that online error-monitoring processes can detect the task-interfering yes-response bias in the earliest response tendencies and that reactive cognitive control processes such as active and selective response suppression kick in (Panis & Schmidt, 2016). The active and selective suppression of the premature yes-response tendency can result in a temporary disinhibition of the competing no-response, which would lead to an overt no-response if a momentary threshold is crossed on some trials. Thus, in those trials where the search process has not yet finished, this suppression can lead to overt misses in target-present trials, and it can explain the sharp increase in $ca(t)$ for target-absent trials, which is presumably too early to reflect pure correct rejection decisions.

Third, in the time range 480–560 ms, performance is optimal in the sense that (a) hazard is at its highest level so far, and (b) conditional accuracy is high for both target presence conditions. Around this point in time after display onset, behavior is thus determined mostly by the outcome of the search process. However, for a subset of trials, no overt decision is made and time passes on.

Fourth, in the time range 560–640 ms, the difference in hazard disappears, and a no-bias develops as the miss rate reaches a maximum, and there are no false alarms. In other words, if the waiting time has increased until 560 ms, then $h_{TP}(600)$ equals $h_{TA}(600)$, and $ca_{TP}(600) = .8$ while $ca_{TA}(600) = 1$. Thus, for the more difficult search trials, the suppression effects accumulate—causing hazard to decrease and the miss rate to peak in the target-present condition, while more and more correct rejection decisions occur when the target is absent.

Finally, after 640 ms, hazard functions are flat and most emitted responses are correct. In other words, the system quits the search and finally transitions to a state with flat hazard functions without a systematic effect of target presence. Horizontally shaped hazard functions point to exponentially distributed RTs. Based on the findings of Shenoy et al. (2013),

we assume that these flat right tails reflect RT outliers during decision making. Shenoy et al. (2013) described neuronal motor activity in macaque monkeys 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 et al. (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 before an irrevocable commitment is made. For example, one might keep inspecting a few more items even though the no-response is already selected in the target-absent condition. Similarly, one might explicitly compare the presumed target with a few surrounding distractors to confirm target presence, even though the yes-response is already selected in the target-present condition.

Both these and other discrete-time EHA studies of simultaneous masking (Panis & Hermens, 2014), object recognition (Panis et al., 2017; Panis & Wagemans, 2009; Torfs et al., 2010), spatial cueing (Panis, 2020; Panis & Schmidt, 2020), and priming (Wolkersdorfer et al., 2020) teach us three things: (a) Mean performance measures conceal crucial information about behavioral dynamics such as premature response activation, time-locking, response suppression, and how performance changes as time passes by within *and* across trials, (b) RT and accuracy data reflect different aspects of the time-dispersed decision process (Mulder & van Maanen, 2013), and (c) sometimes one can identify subsets of participants that display qualitatively different behavior (Miller & Schwarz, 2018; Panis, 2020; Panis et al., 2020).

Note that when you measure time in continuous units, the survivor function $S(t)$ can be estimated nonparametrically using the Kaplan–Meier method (Kaplan & Meier, 1958). Estimates of the hazard *rate* function can be obtained based on Kaplan–Meier but are typically smoothed to some extent because they tend to be very choppy when not based on sufficient data (Allison, 2010).

Obtaining Inferential Statistics

There are several approaches for obtaining inferential statistics (Allison, 2010; Austin, 2017). When you simply want to compare survival functions between two groups in continuous time (large- N design), the log-rank and the Wilcoxon tests are available (the latter puts more weight on earlier points in time).

When you want to study how hazard depends on various predictors, you can fit regression models to the data (Singer & Willett, 2003). An example *discrete-time* hazard model with three predictors (TIME, X_1 , X_2) and the complementary log-log (cloglog) link function can be written as follows:

$$\begin{aligned} \text{cloglog}[h(t)] = \ln(-\ln[1 - h(t)]) = & [\alpha_0 \text{ONE} + \alpha_1(\text{TIME} - 1) \\ & + \alpha_2(\text{TIME} - 1)^2 + \alpha_3(\text{TIME} - 1)^3] + [\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2(\text{TIME} - 1)]. \end{aligned}$$

The main predictor variable TIME is the time bin index t (see Table 1) that is centered on value 1 in this example. The complementary log-log link is preferred over the logit link when events can occur in principle at any time point within a bin, which is the case for RT data (Singer & Willett, 2003). 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. Predictors can be discrete, continuous, and time-varying or time-invariant. For example, the effect of a 1 unit increase in X_1 is to vertically shift the whole baseline cloglog-hazard function by 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 by β_2 cloglog-hazard units (when $\text{TIME}-1 = 0$), in bin 2 by $\beta_2 + \beta_3$ cloglog-hazard units (when $\text{TIME}-1 = 1$), and so forth. To interpret the effects of the predictors, the parameter estimates are exponentiated, resulting in a hazard ratio (due to the use of the cloglog link).

In the case of a large- N design without repeated measurements, the parameters of a discrete-time hazard model can be estimated using standard logistic regression software (after expanding the typical person-trial-oriented data set into a person-trial-bin-oriented data set; Allison, 2010). When there is clustering in the data, as in the case of a small- N design with repeated measurements, the parameters of a discrete-time hazard model can be estimated using population-averaged methods (e.g., Generalized Estimating Equations), Bayesian methods, or generalized linear mixed models (Allison, 2010). Examples of the latter can be found in Panis (2020), Panis et al. (2020), Panis and Schmidt (2016, 2020), and Wolkersdorfer et al. (2020). Finding the best random effects structure to generalize beyond the sample is an active area of research (Barr et al., 2013; Cunnings, 2012; Matuschek et al., 2017; Zuor & Ieno, 2016). Note that in case of a small- N design, EHA allows one to test if and how individual performance changes on multiple time scales (e.g., within-trial, across-trial, across-block).

When you treat time continuously, you can fit parametric models (e.g., a lognormal hazard model, an exponential hazard model, and so forth; Figure 1), semiparametric models such as the Cox regression model that ignores the shape of the hazard function and only tests the beta parameters, or piecewise exponential models (Allison, 2010). A piecewise exponential model is useful when (a) event times are measured precisely, (b) you want to estimate the shape of the hazard function, and (c) you do not want to impose a parametric model: Time is divided into intervals, and the hazard rate is assumed to be constant within each interval (i.e., exponentially distributed RTs within each interval).

The use of rather complex regression models to analyze hazard and conditional accuracy functions, and the employment of stepwise techniques to find the best model, harbor the danger of over- or underfitting the data, especially when the model is tested with the same data to which it was fitted. P values from such models have to be treated with the appropriate caution. Therefore, a third approach to obtain inferential statistics is to define different parameters of the descriptive functions (e.g., onset thresholds, divergence and convergence points, inflection points, and so forth) and to use robust techniques such as bootstrapping and jackknifing to compare and test their distributions (Ulrich & Miller, 2001; Wilcox, 2011).

We can shortly illustrate a very simple and immensely useful jackknifing procedure suggested by Ulrich and Miller (2001). Consider the data in Figure 4A (left panel), where we found that the hazard function for incongruent trials experiences a temporary drop in performance (Panis & Schmidt, 2016). If we know from previous experiments that such effects can take place in a certain time window, we can use that window as a region of interest (ROI). The jackknifing procedure now consists of extracting subsamples of the data, each of which contains the average curve for the incongruent trials within the ROI *except for one participant*. Each subsample excludes a different participant so that we have as many subsamples as participants (N). The advantage is that each subsample contains a relatively smooth curve that is based on $N - 1$ participants. It is therefore much easier to extract

parameters of interest from each subsample curve than trying the same for the noisy data of single participants. For example, we can easily find the bottom of the dip in hazard in incongruent trials and extract its time (or amplitude, or both) for each subsample. Those N values can now be put into a table and used for standard ANOVA. Of course, the mean of the subsample curves will be identical to the mean of the individual participants' curves, but the variance will be too small because each participant is included $N - 1$ times. Therefore, all F values have to be corrected by dividing them by a factor of $(N - 1)^2$, and the p values have to be recalculated accordingly (for proofs, see Ulrich & Miller, 2001).

Discussion

The Theoretical and Statistical Advantages of EHA

Many experimental psychologists are still reluctant to embrace EHA and to stop using ANOVA when dealing with time-to-event data. In part, this is due to historical reasons. The computer metaphor of cognition—serial information processing via consecutive stages—was developed by Donders (1969) and became very popular from 1960 onward (Sternberg, 1969, 1984, 2013). During the past decades, however, various distributional methods have been advertised to move beyond the mean (Balota & Yap, 2011; Ridderinkhof, 2002; van Maanen et al., 2019; VanRullen, 2011).

Nevertheless, while many still assume that RTs reflect the cumulative duration of all time-consuming cognitive operations involved in a task (e.g., Liesefeld, 2018; Song & Nakayama, 2009), the results from various discrete-time event history and conditional accuracy analyses show that fast, medium, and slow RTs can actually index different sets of cognitive operations (Figures 4 and 6; cf. van Zoest et al., 2010). Statistically controlling for the passage of time on multiple time scales during data analysis is therefore *equally* important as experimental control during the design of an experiment, to understand human behavior in our experimental paradigms (Panis, 2020; Panis & Schmidt, 2016, 2020).

The distributional data in Figures 4 to 6 are consistent with a dynamic systems approach to cognition according to which cognition involves sequential transitions between stable sensory, motor, and central states (Schöner et al., 2016). To understand the behavioral output of the brain, we must therefore measure quantities— $h(t)$ and $ca(t)$ —that track the motor states over time to study how long they last, how they are replaced by new states, and whether and when different manipulations affect them, to try to infer the spatial-temporal interplay between different cognitive component processes. Averaging these processes over time to look at mean RTs only sometimes preserves the crucial information in the time course of motor activity. More often than not, mean performance measures paint a picture that distorts, conceals, or even reverses the actual dynamical events. One example is the analysis in Figure 5B, which reveals a sequence of positive priming followed by a negative compatibility effect (Panis & Schmidt, 2016). An analysis in terms of mean error rate would necessarily miss at least one of these phases because the effect in mean error rate can only be positive or negative, but not both. It may even miss both phases if integration over time leads to an average that is too small to be significant.

Statistical reasons in favor of EHA include the ability to deal with right-censored observations and time-varying covariates and the fact that hazard provides exactly the kind of information we want to extract from RT data: the instantaneous likelihood of event occurrence given no previous events. We thus recommend to always first plot the $h(t)$ and $ca(t)$ functions of each individual (small- N design) or group of experimental units (large- N design) before making any further data-analytic or computational modeling decision. This practice

would also inform the field about the various shapes the hazard function can take on in different contexts—a big unknown—and this will help in choosing which (combination of) parametric functions we might want to fit to the data, and in knowing how complex our computational models have to be to capture the behavioral dynamics observed empirically (Holden et al., 2009; Townsend & Ashby, 1983; Wickens, 1982).

Issues about bin size optimality play a secondary role at this moment in time in our view, because by working in discrete time—or using interval-censored data—we can make an informed trade-off between the availability of temporal information (smaller bins increase temporal resolution) with the feasibility to perform expensive data collection efforts (small bins can only be used with a large number of repeated measurements in case of a small- N design). In other words, the number and sizes of the time bins used for the analyses can be optimally adapted to each situation, depending on the duration of the data collection period, the rarity of event occurrence, the shape of the empirically observed hazard function, and whether one is using a large- or small- N design (Smith & Little, 2018).

As a standard method, EHA offers a unifying and principled approach to the analysis of time-to-event data that can be flexibly combined with other tools used by cognitive (neuro) scientists. For example, by transforming a sample of time-to-event data into time-series data— $h(t)$ and $ca(t)$ functions—one puts the analysis of behavior on the same footing with respect to time as physiological data such as EEG. Incorporating time-varying covariates (e.g., occipital EEG power in the alpha band) in hazard models of behavioral (or neural) event occurrence extends the set of current approaches to perform cognitive psychophysiology (Meyer et al., 1988). Also, combining EHA with transcranial magnetic stimulation (TMS) allows to read out the time-dispersed effect of a timed TMS pulse in the $h(t)$ and $ca(t)$ functions to answer the question: “When is area x necessary for task y ?”

Finally, as explained by Kelso et al. (2013), it is crucial to first have a precise description of the macroscopic behavior of a system (here: $h(t)$ and $ca(t)$ functions) in order to know what to derive on the microscopic level. For example, fitting parametric functions or computational models to data without studying the shape of the $h(t)$ and $ca(t)$ functions can miss important features in the data (Panis et al., 2020; Panis & Schmidt, 2020). Due to the advantages of EHA, we recommend that it is used more often in future empirical and simulated RT studies. R code to calculate the descriptive statistics and the inferential statistics used by discrete-time EHA for a factorial within-subject design can be downloaded here: https://www.researchgate.net/publication/304069212_What_Is_Shaping_RT_and_Accuracy_Distributions_Active_and_Selective_Response_Inhibition_Causes_the_Negative_Compatibility_Effect.

Discrete-Time EHA Versus Other Distributional Methods

Continuous-Time EHA. Discrete-time methods treat time-to-event data as interval-censored data while continuous-time methods use the exact event times. While learning the discrete-time methods first will ease the learning of the more complex continuous-time methods, they also have a lower temporal resolution. Thus, although 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 semiparametric hazard models that completely ignore the shape of the hazard function, their use might be warranted in particular circumstances. Allison (2010) provides a useful list of considerations when choosing between discrete- and continuous-time methods to perform an EHA. An overview of R functions for a continuous-time EHA can be found here: <https://cran.r-project.org/web/views/Survival.html>.

Quantile Plots and Classic Delta Plots. A quantile plot visualizes a set of quantiles (e.g., the nine deciles) as a function of quantile order. A classic delta plot for RT compares two conditions by subtracting corresponding quantiles and plots each of these (e.g., nine) differences (y axis) as a function of the average of both quantiles in question (x axis). This way we can easily examine in which range of RTs the effect in $F(t)$ is large or small, positive or negative. However, if participants vary strongly in the identity of the time bin in which their fastest emitted responses occur, then quantiles will be very variable among participants, and averaging them will result in a blurring of effects that might otherwise be time-locked to the onset of a stimulus, for example—and effect sizes can also be attenuated. Therefore, we recommend simply plotting the difference in hazards or conditional accuracies for each bin (as in Panis, 2020, Panis & Schmidt, 2020).

Procedures such as Vincentizing (construction of average RT distributions from the average of their quantiles) that are assumed to normalize the RT distributions across participants (Ratcliff, 1979) have not been evaluated positively (Rouder & Speckman, 2004). Instead, we believe that if, for example, the range of RTs and the time course in hazard of an effect are different across participants, then this is theoretically interesting and requires a substantial explanation. Even if it is possible to somehow average those distributions, that does not mean that the underlying processes should be lumped together. Note that individual differences (e.g., in working memory capacity, the time required to stop a response, and so forth) can be taken into account by adding relevant predictors to the participant level of a multi-level hazard model, thus allowing for participant effects and cross-level interactions.

Possible Disadvantages of Discrete-Time EHA. There are also possible disadvantages of discrete-time EHA.

First, the person-trial-bin-oriented data set can become very large.

Second, one needs to explore a few bin sizes to find the optimal size for a particular data set. The optimal bin size will depend on the censoring time, the overall rarity of event occurrence, and the number of repeated measures or trials (small- N design) or the number of participants (large- N design). Note that the time bins do not have to be all of equal size (Panis, 2020).

Third, in hierarchical data from a small- N design, 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-participant noise. Between-participant 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 and the incorporation of covariates at the participant level in a multilevel model is the only way to deal with this. *Power contours* can be used to estimate how many repeated measures are required to reach 80% power for a given sample size N , and vice versa (Baker et al., 2020; see their paper for a useful online tool).

In general, analyzing single participants should be regarded as a safeguard against interpreting spurious effects in the pooled data that are actually only generated by a 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.

Recommendations for Experimental Design of RT and Other Time-to-Event Data Studies

Two general recommendations can be made from the viewpoint of EHA 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. 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 participants to have control over the trial (and experiment) duration, which can be avoided (or systematically controlled).

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 as the response deadline and thus trial duration can be kept short (see Panis & Schmidt, 2016). Note that many more trials are needed if you want to characterize the detailed shape of the right tail of a RT hazard distribution, especially in continuous time.

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 microlevel speed-accuracy trade-off functions. Indeed, working with hazard and conditional accuracy functions, you will discover a whole new layer of the data, and presumably the one where the processes live that actually interest you. An EHA of time-to-event data can strongly constrain the choice between cognitive models of the same psychological phenomenon. Due to the theoretical and statistical advantages of EHA, the fundamental simplicity of the method, and the availability of free software, there is currently no reason anymore to not start using EHA for time-to-event data.

Acknowledgements

We would like to thank Gillian Porter, Tim Meese, Peter Thompson, Frans Verstraten, and Johan Wagemans for inviting us to write a Methods article on event history analysis. We also thank Niko Troje and two anonymous reviewers for their useful comments on previous versions.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projektnummer PA 2947/1-1 (to S. P.).

Open practices statement

The data and R code for the event history analyses are available from the first author upon request.

ORCID iD

Sven Panis  <https://orcid.org/0000-0002-6321-583X>

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.
- Austin, P. C. (2017). A tutorial on multilevel survival analysis: Methods, models and applications. *International Statistical Review*, 85(2), 185–203. <https://doi.org/10.1111/insr.12214>
- Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2020, July 16). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*. Advance online publication. <http://dx.doi.org/10.1037/met0000337>
- 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.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bloxom, B. (1984). Estimating response time hazard functions: An exposition and extension. *Journal of Mathematical Psychology*, 28, 401–420.
- 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.
- Chechile, R. A. (2003). Mathematical tools for hazard function analysis. *Journal of Mathematical Psychology*, 47, 478–494.
- Chechile, R. A. (2006). Memory hazard functions: A vehicle for theory development and test. *Psychological Review*, 113(1), 31–56.
- Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, 33, 269–298.
- Corning, A. F., & Malofeeva, E. V. (2004). The application of survival analysis to the study of psychotherapy termination. *Journal of Counseling Psychology*, 51(3), 354–367. <https://doi.org/10.1037/0022-0167.51.3.354>
- Cummings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), 369–382.
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, 30, 412–431. [https://doi.org/10.1016/0001-6918\(69\)90065-1](https://doi.org/10.1016/0001-6918(69)90065-1)
- Eimer, M., & Schlaghecken, F. (1998). Effects of masked stimuli on motor activation: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 24(6), 1737–1747.
- 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.
- Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics*, 25, 249–263.
- Greenhouse, J. B., Stangl, D., & Bromberg, J. (1989). An introduction to survival analysis: Statistical methods for analysis of clinical trial data. *Journal of Consulting and Clinical Psychology*, 57(4), 536–544.
- Griffin, W. A. (1993). Event history analysis of marital and family interaction: A practical introduction. *Journal of Family Psychology*, 6(3), 211–229.
- Ha, J. C., Kimpo, C. L., & Sackett, G. P. (1997). Multiple-spell, discrete-time survival analysis of developmental data: Object concept in pigtailed macaques. *Developmental Psychology*, 33(6), 1054–1059.
- Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.
- Jaśkowski, P., Białuńska, A., Tomanek, M., & Verleger, R. (2008). Mask- and distractor-triggered inhibitory processes in the timing of motor responses: An EEG study. *Psychophysiology*, 45, 70–85.

- Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457–481.
- Keiley, M. K., & Martin, N. C. (2005). Survival analysis in family research. *Journal of Family Psychology*, 19(1), 142–156. <https://doi.org/10.1037/0893-3200.19.1.142>
- Kelso, J. A. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior and brain coordination. *Neural Networks*, 37, 120–131.
- 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.
- 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.
- Luce, R. D. (1986). *Response times. Their role in inferring elementary mental organization*. Oxford University Press Inc.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315.
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287–330.
- Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, 26, 3–67.
- Miller, J., & Schwarz, W. (2018). Implications of individual differences in on-average null effects. *Journal of Experimental Psychology: General*, 147(3), 377–397. <https://doi.org/10.1037/xge0000367>
- Morita, J. G., Lee, T. W., & Mowday, R. T. (1989). Introducing survival analysis to organizational researchers: A selected application to turnover research. *Journal of Applied Psychology*, 74, 280–292.
- Mulder, M. J., & van Maanen, L. (2013). Are accuracy and reaction time affected via different processes? *PLoS One*, 8(11), e80222. <https://doi.org/10.1371/journal.pone.0080222>
- Núñez-Antón, V., & Orbe, J. (2005). Statistical time to event analysis in the social sciences: Modeling hazard rate and duration in finance. *Methodology*, 1, 104–118.
- Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing* (pp. 41–82). Erlbaum.
- Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct routes kept in check by inhibitory control processes. *Open Psychology*, 2, 238–279. <https://doi.org/10.1515/psych-2020-0107>
- 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. <https://doi.org/10.1037/a0032949>
- Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). 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. *Attention, Perception, & Psychophysics*, 82, 689–714. <https://doi.org/10.3758/s13414-019-01897-z>
- 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., & Schmidt, T. (2020, May 25). *What is causing “inhibition of return” in spatial cueing tasks? Temporally disentangling multiple cue-triggered effects on multiple time scales using response history and conditional accuracy analyses*. <https://doi.org/10.31234/osf.io/udpvs>
- 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.

- Pannasch, S., Dornhoefer, S., Unema, P. J., & Velichkovsky, B. M. (2001). The omnipresent prolongation of visual fixations: Saccades are inhibited by changes in situation and in subject's activity. *Vision Research*, 41(25–26), 3345–3351.
- Pieters, J. P. M. (1983). Sternberg's additive factor method and underlying psychological processes: Some theoretical considerations. *Psychological Bulletin*, 93, 411–426.
- 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.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 86, 446–461.
- Ridderinkhof, K. R. (2002). Activation and suppression in conflict tasks: Empirical clarification through distributional analyses. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action. Attention & performance*, Vol. XIX (pp. 494–519). Oxford University Press.
- Rouder, J. N., & Speckman, P. L. (2004). An evaluation of the Vincentizing method of forming group-level response time distributions. *Psychonomic Bulletin & Review*, 11, 419–427.
- Schmidt, T., Haberkamp, A., Veltkamp, G. M., Weber, A., Seydell-Greenwald, A., & Schmidt, F. (2011). Visual processing in rapid-chase systems: Image processing, attention, and awareness. *Frontiers in Psychology*, 2(169), 1–16.
- Schmidt, T., Hauch, V., & Schmidt, F. (2015). Mask-triggered thrust reversal in the negative compatibility effect. *Attention, Perception, & Psychophysics*, 77, 2377–2398.
- Schmidt, T., Niehaus, S., & Nagel, A. (2006). Primes and targets in rapid chases: Tracing sequential waves of motor activation. *Behavioral Neuroscience*, 120(5), 1005–1016.
- Schöner, G., Spencer, J. P., & the DFT Research Group. (2016). *Dynamic thinking. A primer on dynamic field theory*. Oxford University Press
- 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*. 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.
- Steele, F., Diamond, I., & Wang, D. (1996). The determinants of the duration of contraceptive use in China: A multilevel multinomial discrete hazards modelling approach. *Demography*, 33, 12–33.
- Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics. *Statistical Modelling*, 4, 145–159.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276–315.
- Sternberg, S. (1984). Stage models of mental processing and the additive-factor method. *The Behavioral and Brain Sciences*, 7(1), 82–84.
- Sternberg, S. (2013). The meaning of additive reaction-time effects: Some misconceptions. *Frontiers in Psychology*, 4, Article 744. <https://doi.org/10.3389/fpsyg.2013.00744>
- Stoolmiller, M., & Snyder, J. (2006). Modeling heterogeneity in social interaction processes using multilevel survival analysis. *Psychological Methods*, 11(2), 164–177. <https://doi.org/10.1037/1082-989X.11.2.164>
- 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. (1990). Truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108(3), 551–567.
- Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge University Press.
- Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, 38, 816–827.
- Van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, 92(7), 345–381.
- van Maanen, L., Katsimpokis, D., & van Campen, A. D. (2019). Fast and slow errors: Logistic regression to identify patterns in accuracy-response time relationships. *Behavior Research Methods*, 51, 2378–2389.
- VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, 2(362), 1–6. <https://doi.org/10.3389/fpsyg.2011.00365>
- Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, 7(3), 424–465.
- van Zoest, W., Hunt, A. R., & Kingstone, A. (2010). Visual representations in cognition: It's about time. *Current Directions in Psychological Science*, 19(2), 116–120.
- Wenger, M. J., & Gibson, B. S. (2004). Using hazard functions to assess changes in processing capacity in an attentional cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 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.
- Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. Freeman.
- Wiecki, T. V., & Frank, M. J. (2013). A computational model of inhibitory control in frontal cortex and basal ganglia. *Psychological Review*, 120, 329–355.
- Wilcox, R. (2011). *Introduction to robust estimation & hypothesis testing* (3rd ed.). Elsevier.
- 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.
- Wolfe, J. M., Palmer, E. M., & Horowitz, T. S. (2010). Reaction time distributions constrain models of visual search. *Vision Research*, 50, 1304–1311.
- Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential motor activation in a dual-prime paradigm: Insights from conditional accuracy and hazard functions. *Attention, Perception, & Psychophysics*, 82, 2581–2602. <https://doi.org/10.3758/s13414-020-02010-5>
- Yang, S.-N., & McConkie, G. W. (2001). Eye movements during reading: A theory of saccade initiation times. *Vision Research*, 41(25–26), 3567–3585.
- Zuur, A. F., & Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*, 7, 636–645.

How to cite this article

Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing response times and other types of time-to-event data using event history analysis: A tool for mental chronometry and cognitive psychophysiology. *i-Perception*, 11(6), 1–24. <https://doi.org/10.1177/2041669520978673>

**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 processes²**

² *Reproduced with permission from Springer Nature.*



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 Panis¹ · Rani Moran^{2,3} · Maximilian P. Wolkersdorfer¹ · Thomas Schmidt¹

Published online: 15 January 2020
© The Psychonomic Society, Inc. 2020

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 not usually 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 results from a decision process that is temporally modulated by concurrently active recurrent object recognition, learning, and cognitive control processes, next to attentional selection processes.

Keywords 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; Fig. 1a, left) is efficient, and results in search functions with slopes close to 0 ms/item (Fig. 1b, left). In contrast, searching for a 2 among 5's (spatial configuration search; Fig. 1a, right) is inefficient, and results in search functions with large positive slopes (Fig. 1b, right). Finally, searching for a red vertical target among green vertical and red horizontal distractors (conjunction search; Fig. 1a, middle) is of intermediate efficiency, and results in search functions with intermediate slopes (Fig. 1b, middle).

✉ Sven Panis
sven.panis@sowi.uni-kl.de

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, UK

³ Wellcome Centre for Human Neuroimaging, University College London, London WC1N 3BG, UK

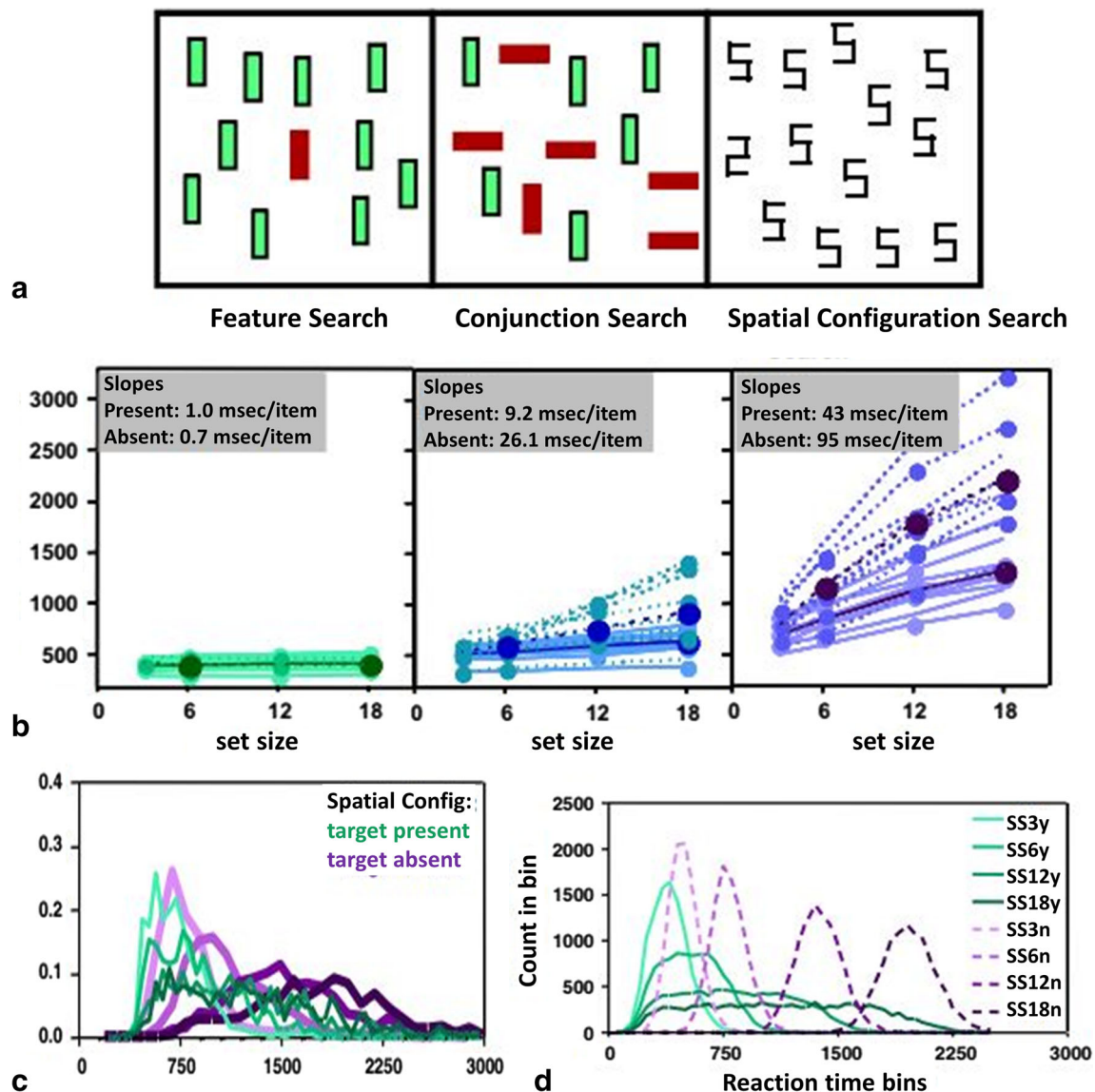


Fig. 1 Benchmark visual search data set from Wolfe, Palmer, and Horowitz (2010). **a** Example visual search displays for three search tasks (copy of Fig. 1 in Wolfe et al., 2010). Participants search for the red vertical bar in the feature (left) and conjunction (middle) search tasks, and for a digital 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).

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 Fig. 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 Fig. 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 Fig. 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, Lin, 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, Fig. 2 shows the $F(t)$, $f(t)$, $S(t)$, and $\lambda(t)$ for four theoretical waiting-time distributions. In contrast to $\lambda(t)$, 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 – $\lambda_A(t) > \lambda_B(t)$ 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.

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.

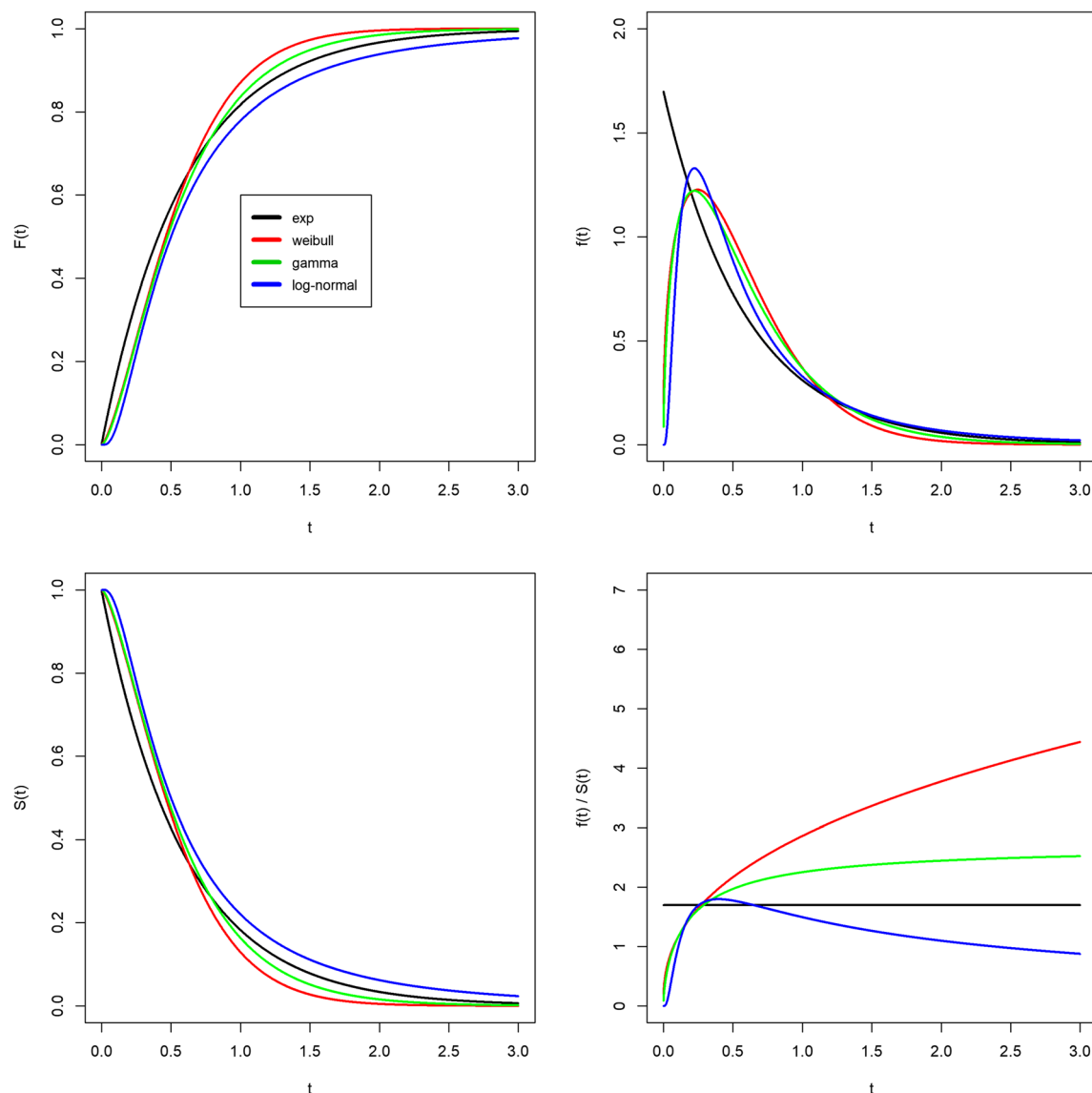


Fig. 2 Four views on waiting-time distributions. The cumulative distribution function (*top left*), the density function (*top right*), the survivor function (*bottom left*) and the hazard rate function (*bottom right*) are shown for each of four theoretical probability distributions (exponential,

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

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 ten 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

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 tradeoff (SAT) function based on partitioning RT in bins, as discussed by Wickelgren (1977) and Pachella (1974) in the psychological literature.

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

#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)

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 (nine participants), participants searched for a red vertical rectangle among green vertical rectangles. In the *conjunction search* tasks (ten participants), they searched for a red vertical rectangle among green vertical and red horizontal rectangles. Finally, in the *spatial configuration* task (nine participants), they searched for a numerical digit 2 among 5s (see Fig. 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 ten 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 > 8000 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 1600, 2400, and 3600 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: $\text{cloglog}[h(t)] = \ln(-\ln[1-h(t)]) = [\alpha_0 \text{ONE} + \alpha_1(\text{TIME} - 1) + \alpha_2(\text{TIME} - 1)^2 + \alpha_3(\text{TIME} - 1)^3] + [\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2(\text{TIME} - 1)]$. 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 $\text{TIME}-1 = 0$), in bin 2 with $\beta_2 + \beta_3$ `cloglog`-hazard units (when $\text{TIME}-1 = 1$), 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

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: $\text{cloglog}[h(t)] = \ln\{-\ln[1-h(t)]\}$. Inverse of the link: $h(t) = 1 - \exp\{-\exp\{\text{cloglog}[h(t)]\}\}$.

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

		(240,280]		(360,400]				(480,520]		(600,640]	
Nr.	effect	PE	<i>p</i>	PE	std. err.	<i>t</i>	<i>p</i>	PE	<i>p</i>	PE	<i>p</i>
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.

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 (eight 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 (six 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).

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.

Nr.	effect	(450,500]				(600,650]		(750,800]		(900,950]	
		PE	std. err.	<i>t</i>	<i>p</i>	PE	<i>p</i>	PE	<i>p</i>	PE	<i>p</i>
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						

Results

Feature search: Descriptive statistics

In Fig. 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

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]				(880,960]		(1200,1280]		(1520,1600]	
		PE	std. err.	<i>t</i>	<i>p</i>	PE	<i>p</i>	PE	<i>p</i>	PE	<i>p</i>
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						

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

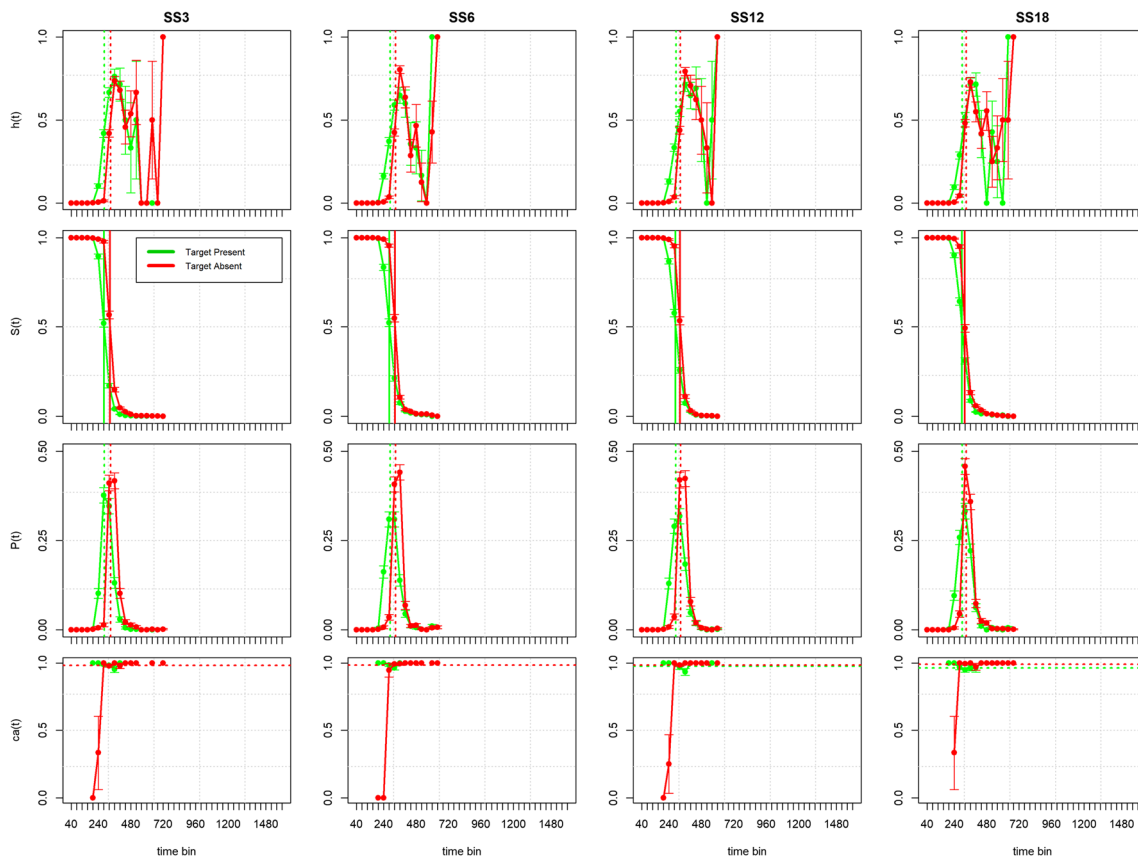


Fig. 3 Descriptive statistics for subject number 3 in the feature search task. *Top to bottom:* $h(t)$, $S(t)$, $P(t)$, and $ca(t)$ for each set size (columns) and target presence (green = target present, red = Target Absent). *Vertical*

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

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 Fig. 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 Fig. 4, top eight panels) show early false alarms when responses are emitted around 240 ms after search display onset. We define early false alarms as “ $ca(t) \leq .50$ 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 Fig. 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, eight out of nine 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 Fig. 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 hovering 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

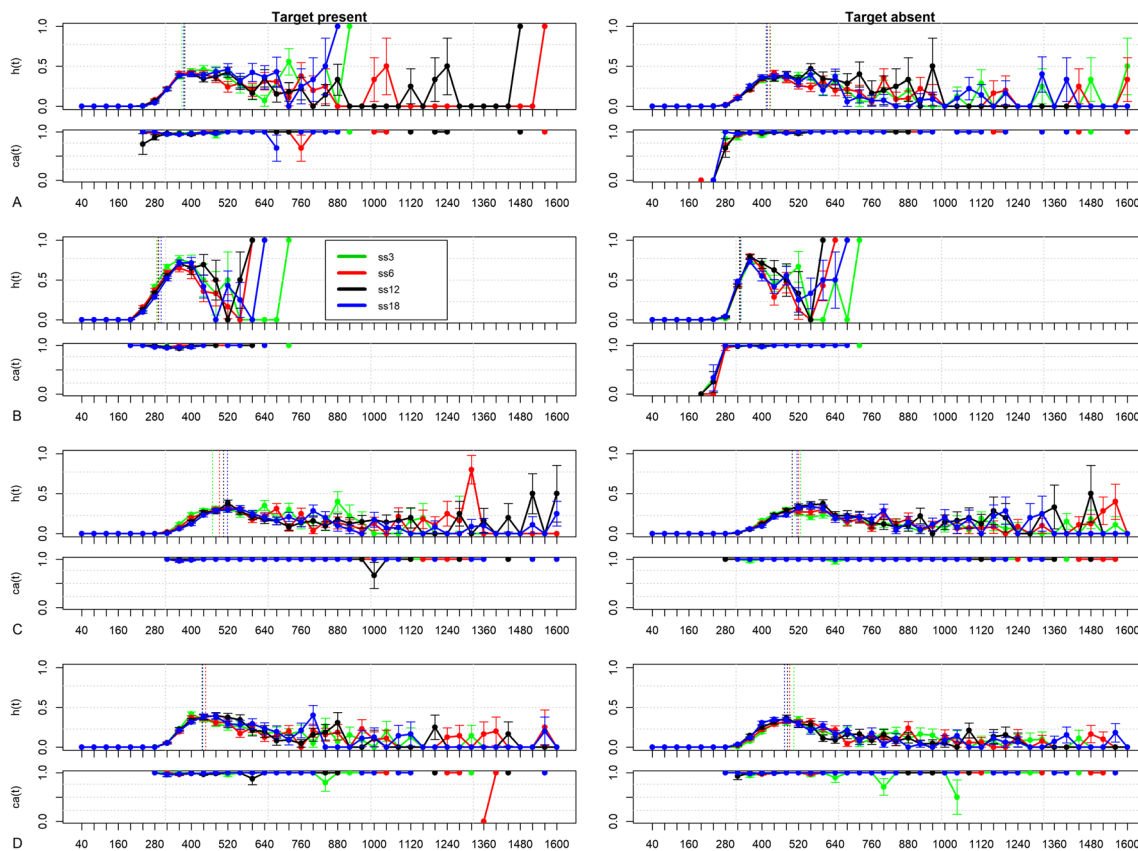


Fig. 4 Inter-individual differences in feature search. Estimates of $h(t)$ and $ca(t)$ for four participants in the feature search task, for target-present (left column) and target-absent (right column) trials and each set size (green =

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

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 Fig. 5, top left panel). Bins after 500 ms show no

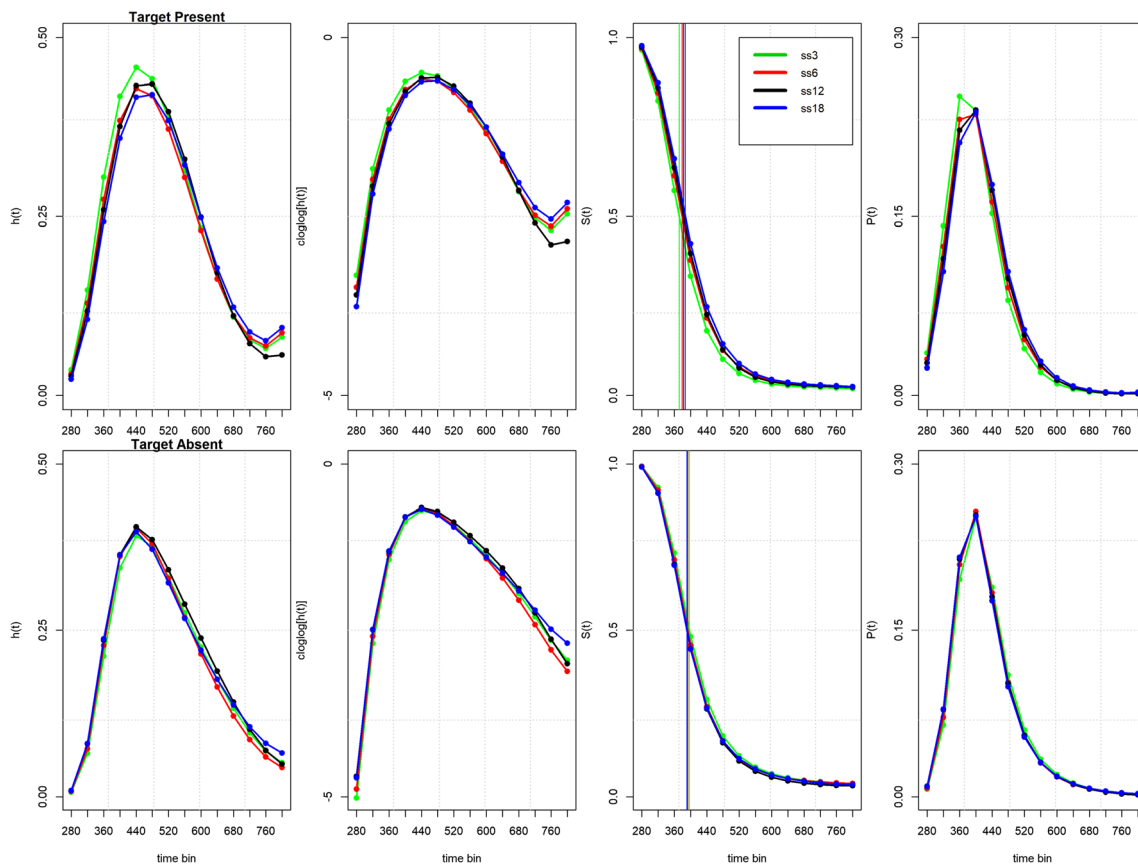


Fig. 5 Hazard model predictions for feature search. Predicted $h(t)$ functions for trial 350 in block 8 (*first column*) and the corresponding cloglog-hazard functions (*second column*), $S(t)$ (*third column*) and $P(t)$

(*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 $S(t)_{.50}$

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 three 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 anymore.

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 2, 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 Fig. 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 hovering

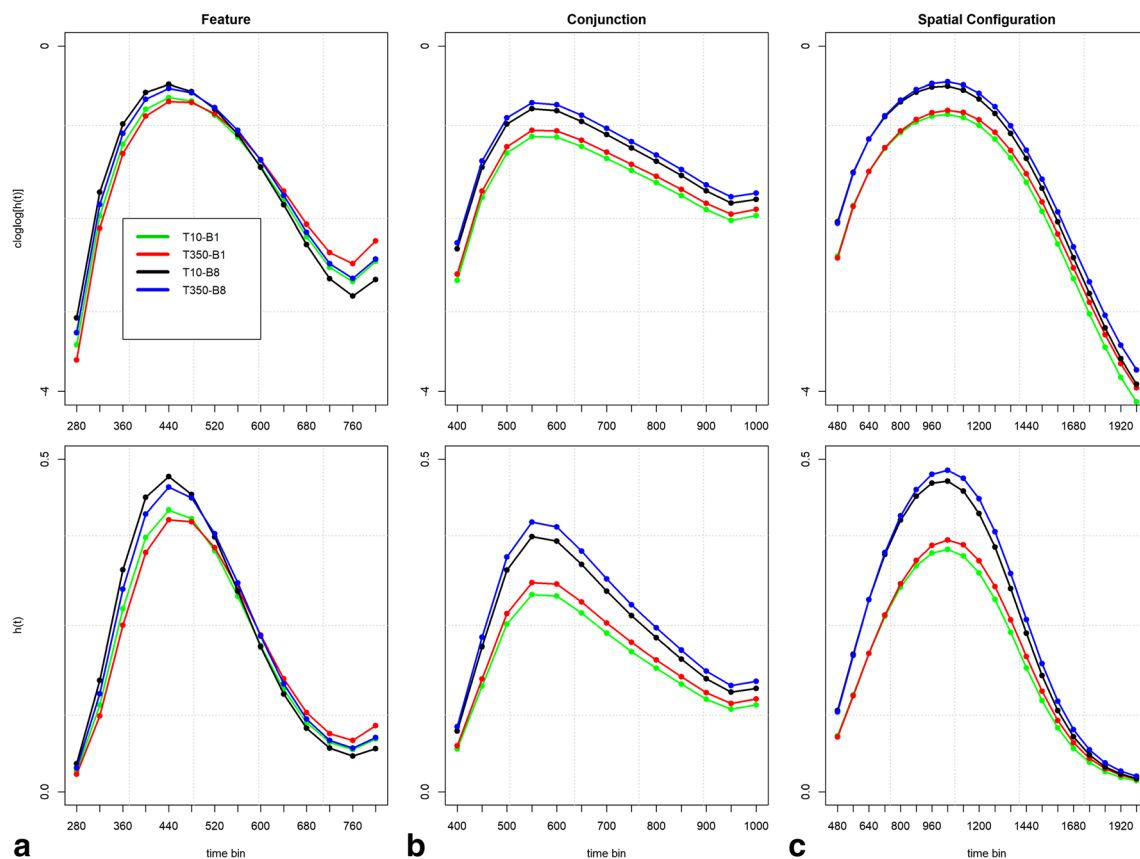


Fig. 6 Effect of practice on event occurrence. The model-based effects of Trial (T) and Block (B) are shown for the baseline condition (TP3) for the feature (A), conjunction (B) and spatial configuration (C) search tasks

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 Fig. 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 nine 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

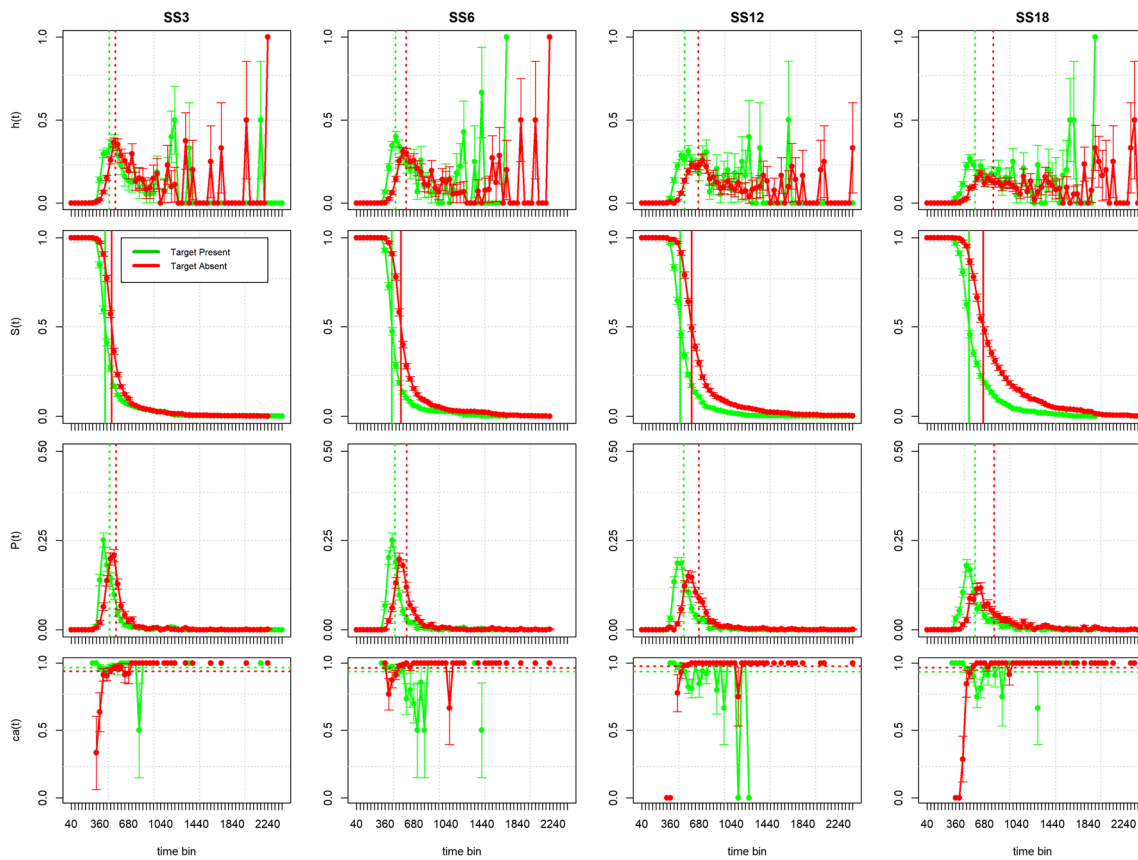


Fig. 7 Descriptive statistics for subject number 5 in the conjunction search task. Same conventions as in Figure 3

seven parameter estimates in Table 3 model the shape of the $\text{cloglog}[h(t)]$ function for TP3 (the reference condition) in trial 350 of block 8 (Fig. 9, top left, green line) using a 6th-order polynomial function of TIME. Parameters 2–7 show a significant linear, quadratic, cubic, quartic, quintic, 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 (Fig. 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 9–12) 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 18–20), and the effect of SS18 changes in a quadratic fashion (parameters 22–23).

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 ten trials will increase the estimated $\text{cloglog}[h(t)]$ value with 0.0021 units (parameter 34, column 3, $p < .0001$) in each bin. Second, each additional block will increase the estimated $\text{cloglog}[h(t)]$ 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.

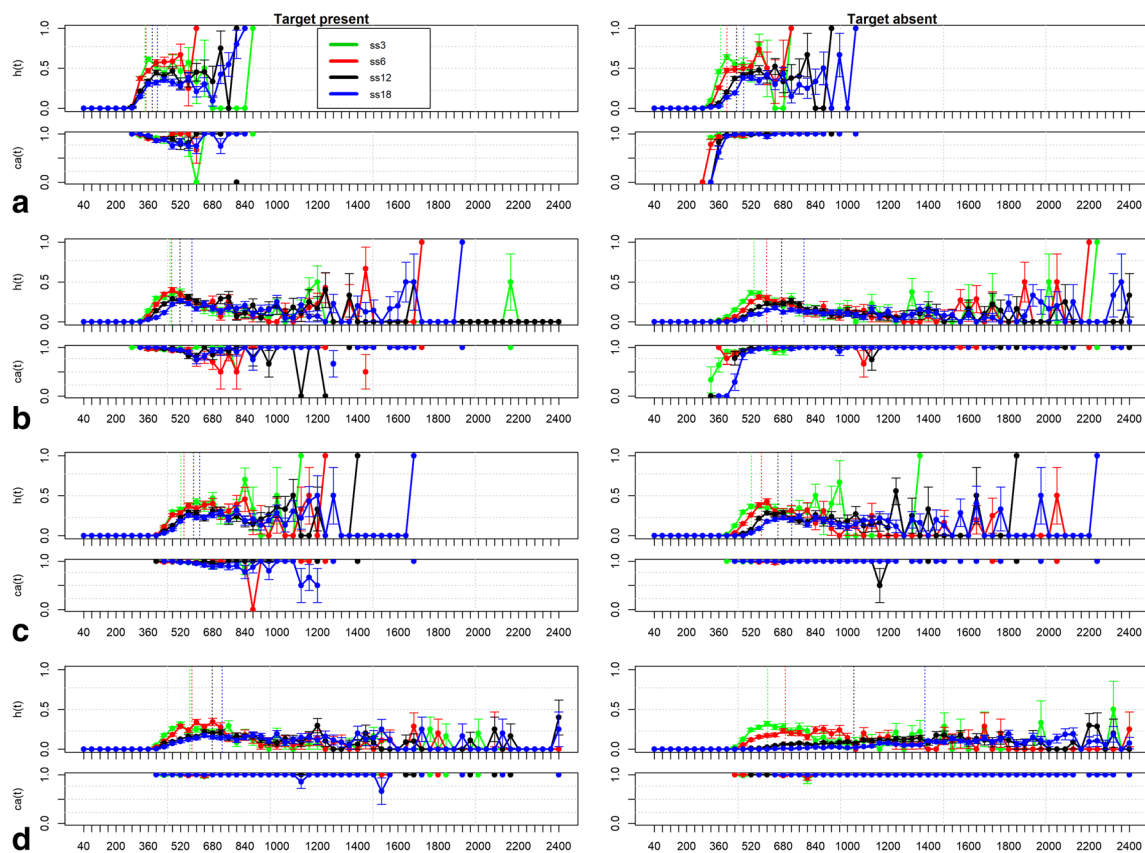


Fig. 8 Inter-individual differences in conjunction search. Estimates of $h(t)$ and $ca(t)$ for four participants in the conjunction search task. **a** Subject 4. **b** Subject 5. **c** Subject 2. **d** Subject 10. Same conventions as in Fig. 4

Spatial configuration search: Descriptive statistics

In Fig. 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 (Fig. 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 3–6) 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 (Fig. 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 =

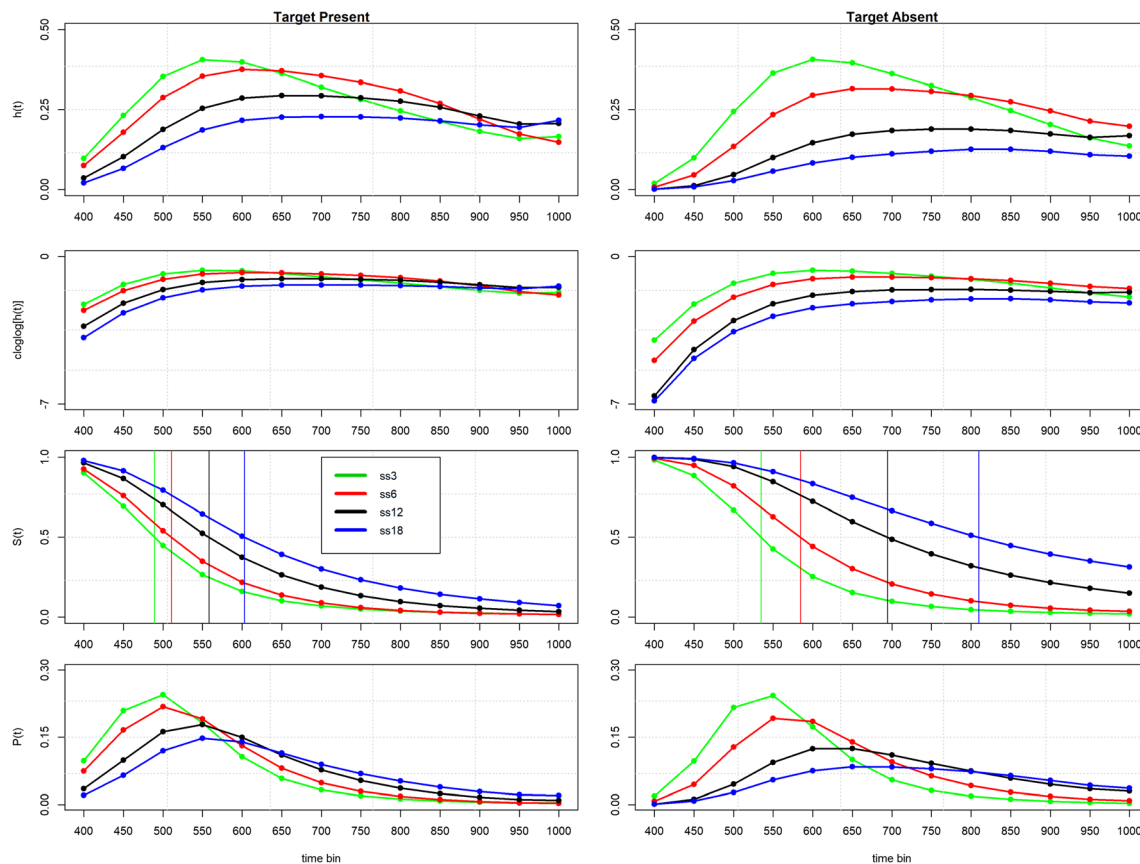


Fig. 9 Hazard model predictions for conjunction search. Predicted $h(t)$ functions for trial 350 in block 8 (top row) and the corresponding cloglog-hazard functions (second row), $S(t)$ (third row) and $P(t)$ (bottom row)

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 Fig. 6c, each additional series of ten 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 ten trials increases the estimated $cloglog[h(t)]$ in bin 1600 with 0.00342 units (parameter 38, column 11, $p < .001$). Second,

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 $S(t)_{.50}$

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

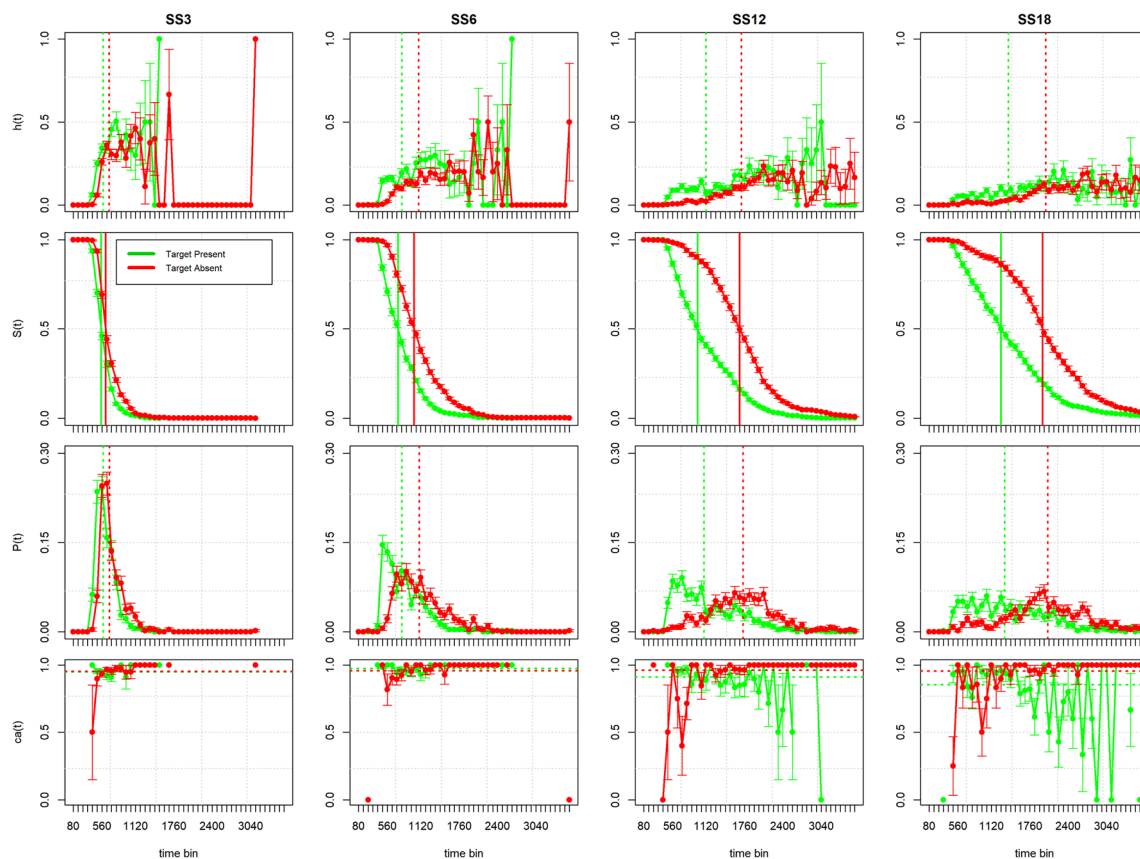


Fig. 10 Descriptive statistics for subject number 8 in the spatial configuration search task. Same conventions as in Fig. 3

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 studies⁷. 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.

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

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 (Fig. 5) although there is no effect of set size on mean correct RT (Fig. 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

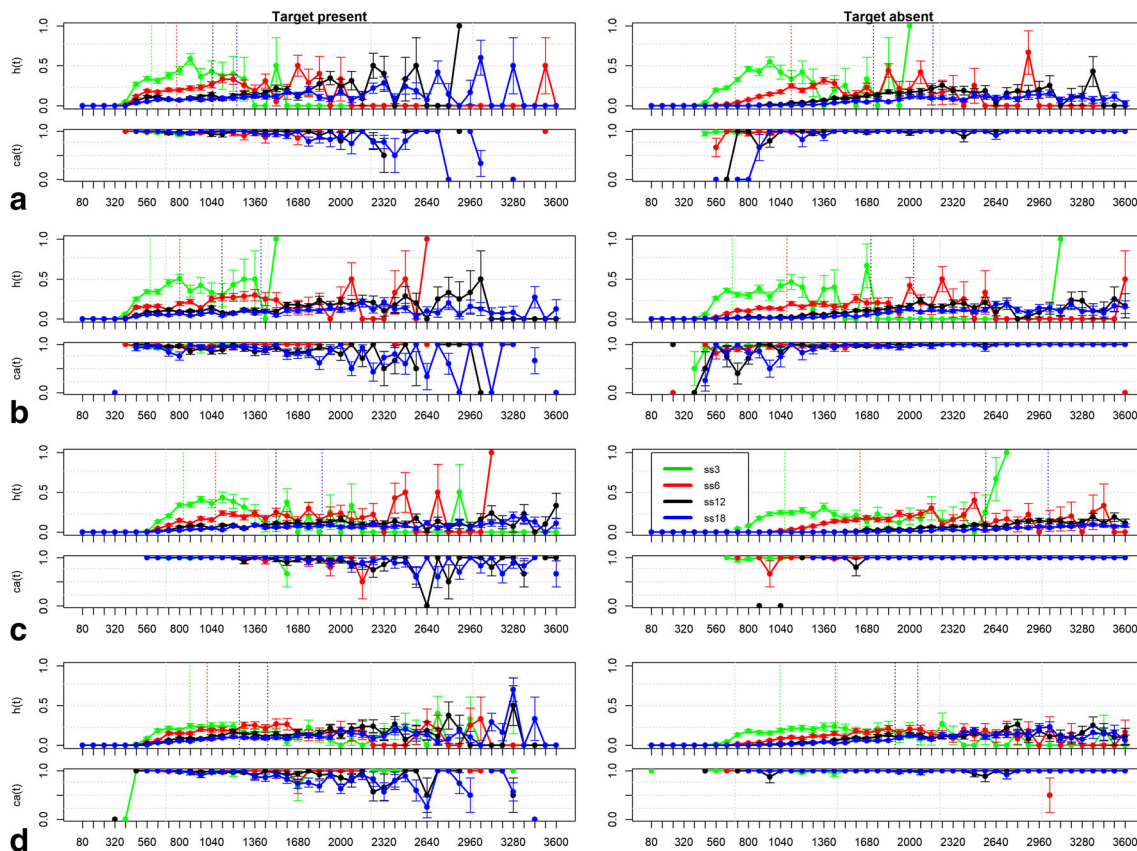


Fig. 11 Inter-individual differences in spatial configuration search. Estimates of $h(t)$ and $ca(t)$ for four participants in the spatial configuration search task. **a** Subject 1. **b** Subject 8. **c** Subject 3. **d** Subject 4. Same conventions as in Fig. 4

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 ($ca(t) \leq .50$) 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

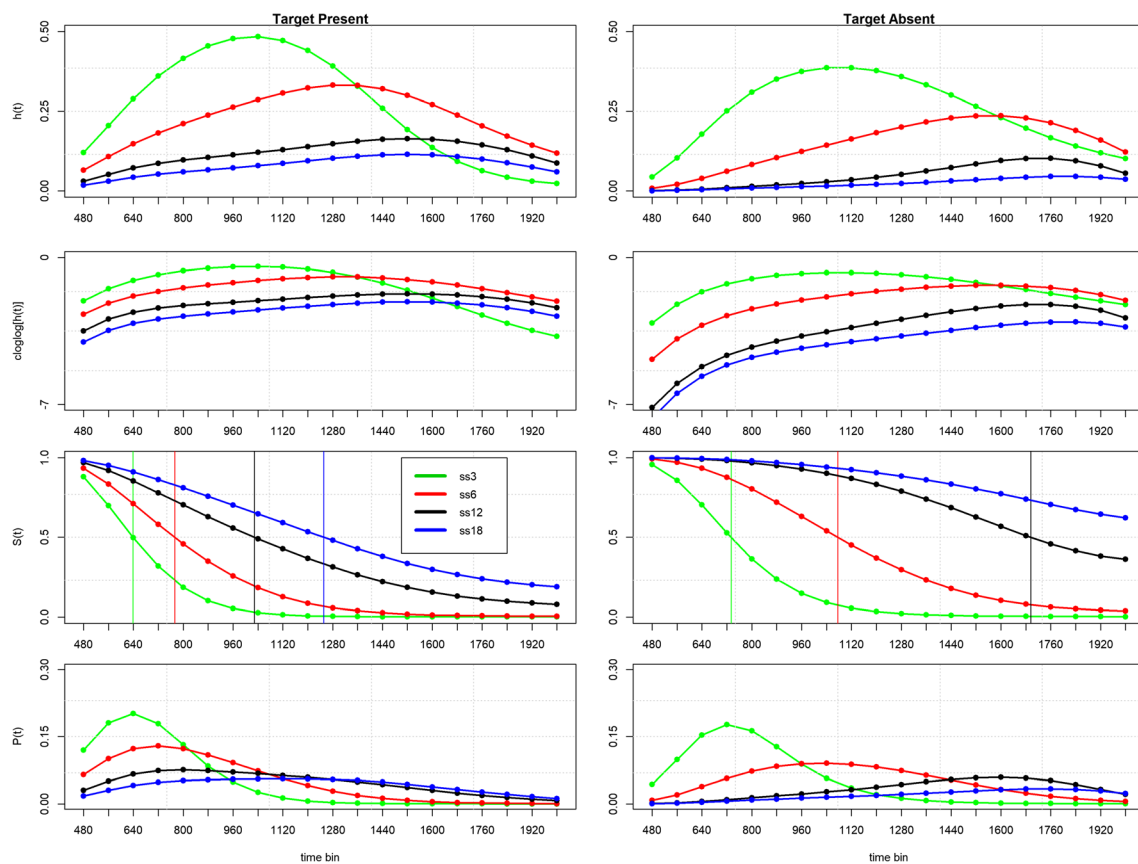


Fig. 12 Hazard model predictions for spatial configuration search. Same conventions as in Fig. 9

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

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.

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 tradeoff 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 2 s 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 Figs. 4, 8, and 11). Horizontally shaped hazard functions point to exponentially distributed RTs (see Fig. 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 We wish to thank Heinrich René Liesefeld, Hermann J. Müller, and an anonymous reviewer for their helpful comments on previous drafts. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer PA 2947/1-1 (to S.P.).

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. [https://doi.org/10.1016/0042-6989\(77\)90207-3](https://doi.org/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. <https://doi.org/10.1037/0033-295X.102.2.356>.
- Griegen, 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. <https://doi.org/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. <https://doi.org/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; 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. <https://doi.org/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), 157–168.
- 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 non-overlapping patterns. *Journal of Experimental Psychology: Human Perception & Performance*, 40 (1), 129–144. <https://doi.org/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. <https://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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Temporal dynamics of sequential motor activation in a dual-prime
paradigm: Insights from conditional accuracy and hazard functions³**

³ *Reproduced with permission from Springer Nature under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>).*



Temporal dynamics of sequential motor activation in a dual-prime paradigm: Insights from conditional accuracy and hazard functions

Maximilian P. Wolkersdorfer¹ · Sven Panis¹ · Thomas Schmidt¹

Published online: 12 March 2020

© The Author(s) 2020

Abstract

In response priming experiments, a participant has to respond as quickly and as accurately as possible to a target stimulus preceded by a prime. The prime and the target can either be mapped to the same response (consistent trial) or to different responses (inconsistent trial). Here, we investigate the effects of two sequential primes (each one either consistent or inconsistent) followed by one target in a response priming experiment. We employ discrete-time hazard functions of response occurrence and conditional accuracy functions to explore the temporal dynamics of sequential motor activation. In two experiments (small-*N* design, 12 participants, 100 trials per cell and subject), we find that (1) the earliest responses are controlled exclusively by the first prime if primes are presented in quick succession, (2) intermediate responses reflect competition between primes, with the second prime increasingly dominating the response as its time of onset is moved forward, and (3) only the slowest responses are clearly controlled by the target. The current study provides evidence that sequential primes meet strict criteria for sequential response activation. Moreover, it suggests that primes can influence responses out of a memory buffer when they are presented so early that participants are forced to delay their responses.

Keywords Feedforward sweep · Response priming · Event history analysis · Reaction-time analysis · Visuomotor

Priming paradigms are very popular in many fields of cognitive psychology to study how exposure to a prime stimulus influences the response to a subsequently presented target stimulus. In general, the representations that mediate priming can be located at perceptual (Wiggs & Martin, 1998), conceptual/semantic (e.g., Schacter & Buckner, 1998), lexical (e.g., Fernández-López, Marcet, & Perea, 2019), phonological (e.g., Ferrand & Grainger, 1992), and/or motor response levels (e.g., Rosenbaum, 1983). In this paper we focus on the so-called response priming paradigm (Klotz & Neumann, 1999; Klotz & Wolff, 1995; Vorberg, Mattler, Heinecke, Schmidt, &

Schwarzbach, 2003). In a typical response priming experiment, a participant has to respond as quickly and as accurately as possible to a target stimulus preceded by a (masked or unmasked) prime stimulus. The prime and the target can either be mapped to the same response (consistent trial) or to different responses (inconsistent trial). While consistent trials typically show accelerated and more accurate responses, inconsistent trials show decelerated and less accurate responses, respectively. The differences between consistent and inconsistent trials in both mean reaction times (RTs) and overall error rates (ERs) define the *response priming effect*. Characteristically, this priming effect increases linearly with stimulus-onset asynchrony (SOA) for SOAs of up to about 100 ms (Vorberg et al., 2003). Response priming effects are believed to be mostly mediated by motor response conflicts (Schmidt, Haberkamp, & Schmidt, 2011; Schmidt, 2002). However, how a rapid sequence of visual stimuli is processed and converted into motor action is still under debate. In order to gain insights into the covert temporal dynamics of our visual system and the online transfer of visual signals into overt behavior, we employ event history analysis, a longitudinal technique to perform a distributional analysis.

✉ Maximilian P. Wolkersdorfer
max.wolkersdorfer@sowi.uni-kl.de

Sven Panis
sven.panis@sowi.uni-kl.de

Thomas Schmidt
thomas.schmidt@sowi.uni-kl.de

¹ Faculty of Social Sciences, Experimental Psychology Unit,
University of Kaiserslautern, Erwin-Schrödinger-Str. Geb. 57,
D-67663 Kaiserslautern, Germany

Multiple-prime paradigm

What if instead of only one prime, a sequence of primes is preceding a target stimulus? A number of previous studies have touched upon this question. Jaśkowski, Skalska, and Verleger (2003) presented five pairs of squares sequentially, with an SOA of 35 ms, so that each stimulus masked the previous one via metacontrast. The last and largest pair was the target, and observers had to decide whether the left or right square contained a gap. The first four pairs could serve as masked primes that contained a gap in the same (consistent) or opposite (inconsistent) square as the target. They found that the priming effect in mean correct RT increases with the number of primes presented in a sequence of successively masked stimuli. Because all of the primes within a single trial were either consistent or inconsistent to the target, this result would be expected from the accumulation of prime information (Miller, 1982). Jaśkowski et al. (2003) concluded that “motor activation evoked by a series of primes does accumulate, facilitating or inhibiting motor responses to the target” (p. 913).

Similarly, Breitmeyer and Hanif (2008) showed that when two successively presented prime stimuli are both consistent to a target in terms of shape (square versus diamond), mean RTs are faster than when only one of the two primes is consistent. Furthermore, they found that the priming effects from the second prime dominate over those of the first prime. That is, if the first prime was consistent and the second inconsistent to the target (condition “CI”), mean RT increased much more than when the first prime was inconsistent and the second consistent (condition “IC”). This contradicts the idea that due to the longer Prime1-target SOA, the first prime should cause a larger priming effect than the succeeding second prime. They argue that the second prime instead updates and overrides the effects of the first prime.

Grainger, Scharnowski, Schmidt, and Herzog (2013) employed two 20-ms Vernier stimuli as primes. In a series of experiments, they found that (1) two primes presented in immediate succession at the same location integrate before activating a motor response, and do not cause sequential activation; (2) two identical primes yield larger priming effects than single primes; (3) one consistent and one inconsistent prime presented simultaneously at different locations cancel each other's effects. More importantly, in the varying-primes condition of their Experiment 3, they presented two lateralized Vernier primes and a central Vernier target, kept the Prime 1–target SOA constant at 200 ms, and varied the interprime interval (and thus also Prime 2–target ISI). For interprime intervals of 30 and 80 ms Prime 2 clearly dominated, but for an interprime interval of 150 ms (and a corresponding Prime 2–target ISI of 30 ms) Prime 1 dominated slightly. The authors propose that all visual stimuli enter a time-selective buffer stage, integrate, and only then initiate a motor response. Instead of activating their associated responses in strict

sequence, their joint impact is determined by their relative dominance in the motor buffer.

However, it has been suggested that—in the context of response-conflict paradigms such as response priming and flanker effects—sequential visual stimuli elicit sequential feedforward sweeps (Bullier, 2001; Lamme & Roelfsema, 2000; VanRullen & Koch, 2003). These fast and bottom-up processes can activate motor responses in a strictly sequential manner (T. Schmidt et al., 2011). Moreover, since both prime and target in a response priming paradigm activate their respective motor responses, response conflict arises if prime and target are inconsistent, thus leading to an increase in RT (Schmidt, 2014). Several studies have demonstrated the existence of this feedforward and sequential activation, in both neuronal activity, such as lateralized readiness potentials (Eimer & Schlaghecken, 1998; Vath & Schmidt, 2007), and overt behavior, such as the time course of pointing movements (Schmidt & Schmidt, 2010; Schmidt, 2002; Schmidt & Schmidt, 2009) and response-time distributions (Panis & Schmidt, 2016). In particular, these studies demonstrated that the first responses are exclusively triggered by prime properties, independent of the target, whereas only later responses are influenced by target properties.

Schmidt, Niehaus, and Nagel (2006) hence proposed a *chase theory of response priming* in which they formulated the *chase criteria* of such a feedforward system: (1) Prime rather than target signals determine the onset and initial direction of the response; (2) target signals influence the response before it is completed; (3) movement kinematics initially depend on prime characteristics only and are independent of all target characteristics (see Schmidt, 2014, for precise definitions of criteria and predictions). Such a simple feedforward-sweep model seems to account very well for response priming effects at short SOAs (up to 100 ms), but would predict unrealistically high error rates for longer SOAs (because in inconsistent trials, the prime would always have enough time to drive the wrong response to completion). Therefore, priming effects at longer SOAs are more plausibly carried by the content of a response buffer that carries information from both primes, but is dominated by the second one (Grainger et al., 2013). This buffer would allow participants to delay their responses, waiting out the target.

Event history analysis

The aims of the current study were to trace sequential priming effects over the time course of a trial to see (a) whether sequential primes actually initiate sequential response activation, (b) whether that sequence conforms to the chase criteria at short SOAs, and (c) how the influence of the first prime changes when the interprime interval is prolonged. In order to investigate the temporal dynamics of response activation, one

must take the passage of time into account when analyzing behavioral output. Here, we make use of a relatively new approach to analyze reaction time data: Event history analysis (EHA; Allison, 1982, 2010; Luce, 1986; Panis & Schmidt, 2016; Singer & Willett, 2003). In EHA, it is assumed that for each time point since target onset in each trial of an experiment, there is a risk for the response to occur. The time after target onset is subdivided into a series of nonoverlapping and contiguous time bins indexed by t , $t \in \{1 \dots n\}$, and for each time bin, the discrete-time *hazard* probability of response occurrence is estimated. The hazard probability $h(t)$ is defined as the conditional probability that a response occurs sometime within bin t given that no response has been emitted in previous bins: $h(t) = P(T = t \mid T \geq t)$ (Allison, 1982, 2010; Luce, 1986; Panis, Torfs, Gillebert, Wagemans, & Humphreys, 2017; Panis & Wagemans, 2009). The survival function $S(t) = P(T > t)$ estimates the probability that no response has been emitted by the time Bin t is completed. In addition, $P(t) = P(T = t)$ gives the unconditional probability that a response (no matter whether correct or incorrect) occurs within Bin t .¹ Since correct and incorrect response occurrences are not independent (Burle, Vidal, Tandonnet, & Hasbroucq, 2004; Praamstra & Seiss, 2005), we calculate the conditional accuracy $ca(t) = P(\text{correct response} \mid T = t)$, the probability that a response emitted in time Bin t is correct. Together, $h(t)$ and $ca(t)$ give an unbiased description of the time course of the latency and accuracy of responses (Panis & Hermens, 2014; Panis & Schmidt, 2016).

Current study

Here, we investigate the effects of two sequential primes followed by one target on response occurrence and accuracy in a response priming experiment. Our goal was to investigate (a) whether sequential primes actually initiate sequential response activation, or integrate in a buffer before a response is emitted, (b) whether that response activation sequence conforms to the rapid-chase criteria at short SOAs, and (c) how the influence of the first prime changes when the SOAs are all prolonged.

We designed a stimulus layout where two primes can be presented in sequence without mutual interference and without masking. Further, we varied the timing of the stimuli by keeping the Prime 1 target (P1-T or SOA1) SOA fixed and moving the onset of Prime 2, resulting in different combinations of Prime 1–Prime 2 (P1–P2) and Prime 2–target (P2–T or SOA2) SOAs. Each prime could either be consistent or inconsistent to the target. In a first experiment we investigated quick successions of primes and target, a second experiment used prolonged stimulus-onset asynchronies. We reasoned from the idea that when the P1–T SOA is short (Experiment 1), participants can rely on feedforward response activation

and give speeded responses without using the response buffer. In contrast, when the P1–T SOA is long (Experiment 2), participants are forced to withhold responses in order to avoid errors triggered by inconsistent primes, and in that situation the response buffer can influence the response.

Experiment 1

Method

We constructed a stimulus arrangement dubbed the ‘lollipop’ that allows us to present a sequence of primes and targets without any spatial overlap or masking (see Fig. 1). The lollipop consisted of a large circle subdivided into eight segments that would contain the primes. A circle in the center of the lollipop contained the target and served as fixation point. Participants were instructed to give speeded responses to the color of the target—red or green—with two successive primes appearing prior to its onset. For the first prime, every other lollipop segment briefly changed color simultaneously (all either red or green). For the second prime, the previously unoccupied segments all briefly turned red or green simultaneously, independent of the color of the first prime.

Participants

Twelve participants (seven female, ages 22–36 years, $M = 28.2$ years) were recruited out of the pool of students of the University of Kaiserslautern. They participated in one 60-minute session for each experiment and were rewarded with course credits. All of them had normal or corrected-to-normal vision (17% with correction). Each participant gave informed consent and was treated in accordance with the ethical standards of the American Psychological Association.

Apparatus and stimuli

Participants sat comfortably on a chair in front of a 17-inch VGA cathode-ray monitor (refresh rate of 75 Hz, resolution of $1,280 \times 1,024$) in a dimly lit room, such that their faces were at a distance of roughly 80 cm from the screen. Responses were collected with a USTC Response Time Box (Li, Liang, Kleiner, & Lu, 2010). Microsoft Windows XP served as the operating system and the experiments were written in MATLAB, using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

Prime and target stimuli appeared inside the lollipop frame, which was present throughout the trial (see Fig. 1). The frame was shown in white (54.3 cd/m^2 , line width 2 pixels) against a black background (0.03 cd/m^2) and consisted of a central circle ($\varnothing 0.8 \text{ cm}$, 0.57°) for the target and a larger circle ($\varnothing 2.4 \text{ cm}$, 1.72°) for the primes. The large circle was subdivided into

eight 45° segments by horizontal, vertical and diagonal lines. The first prime (P1) was presented by filling-in four non-contiguous segments with the same color (either red, 11.0 cd/m², $x = .45$, $y = .30$, or green, 11.0 cd/m², $x = .24$, $y = .40$). The second prime (P2) was then presented in the remaining segments. The two sets of segments were randomly assigned to colors and primes. As a target stimulus (T), the inner small circle of the frame was filled with either red or green color.

Procedure

Experiment 1 lasted 60 minutes. The experiment started with one practice and two experimental blocks with 50 trials each in which no prime [N] was displayed. This had the purpose to accustom the participants to the procedure. After completion of this task, prime conditions were administered to the participants. Each prime could either be consistent (same color) or inconsistent (different color) with the target. There were two single-prime conditions, consistent [C] and inconsistent [I], and four double-prime conditions, consistent–consistent [CC], consistent–inconsistent [CI], inconsistent–consistent [IC], and inconsistent–inconsistent [II]. (Throughout this paper, we always code consistency relative to the target.) Again, participants had to complete one practice block, this time followed by 25 experimental blocks, with 56 trials each. Each block contained eight single-prime trials and 48 double-prime trials. Altogether, this led to participants completing 100 trials each for the no-prime, two one-prime and twelve double-prime (three SOA \times four prime combinations) conditions.²

Each trial began with the onset of the lollipop frame (see Fig. 1). After 493 ms of fixation, P1 was presented in either

red or green for 13 ms, except for the no-prime trials during which all segments remained black (such that the SOA structure was maintained even when one or both primes were absent). After a P1–P2 SOA of 27, 53, or 80 ms, either a red or green P2 was presented for another 13 ms, except for the no-prime and single-prime trials during which all segments stayed black. Finally, after a P2–T SOA of 80, 53, or 27 ms, a red or green target followed. As a result, the SOA between P1 and T was always 107 ms. The target stayed on-screen for 107 ms. Participants were instructed to fixate the target circle at the center of the frame (see Fig. 1) and to respond to the target color as quickly and accurately by pressing one of two response buttons with their left or right index finger, while all other stimuli were irrelevant. After detection of the manual response, a feedback display was shown for 500 ms, followed by a blank screen for 360 ms before the next trial started. Participants received a “too slow” feedback message if their RT was slower than 999 ms. During practice trials they received an additional “wrong” feedback message if their response was incorrect and “correct” if their response was correct. Additionally, after each block participants received feedback on their performance (percentage correct, number of errors, mean reaction time) and could take a short rest if desired. Color-to-button mapping was fixed for each participant and counterbalanced across participants. All stimulus conditions, except for the blocked no-prime condition, occurred randomly and equiprobably over the course of a session.

Analysis of mean error rate and mean correct RT

In a first step, mean reaction times (RT) and error rates (ER) were inspected. We performed two sets of analyses. First, one-

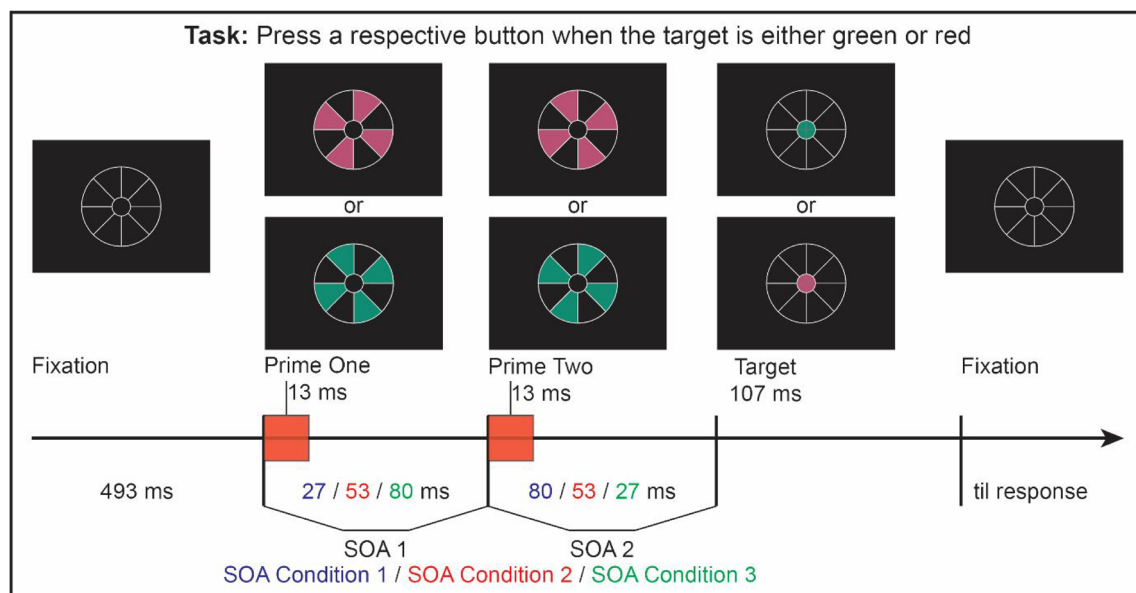


Fig. 1 Stimulus displays and design. After fixating the center of the white lollipop frame, a sequence of two primes and a target is presented, with SOA1–SOA2 combinations of 27/80, 53/53, or 80/27

Table 1 Selected hazard model for Experiment 1

Effect	(175,200]		(250,275]				(300,325]		(375,400]	
	PE	<i>p</i>	PE	SE	<i>t</i>	<i>p</i>	PE	<i>p</i>	PE	<i>p</i>
1 Intercept	−4.650	0.0000***	−2.361	0.260	−9.084	0.0000***	−1.318	0.0000***	−0.620	0.0000***
2 TIME			0.637	0.039	16.504	0.0000***				
3 TIME ²			−0.054	0.004	−14.811	0.0000***				
4 TIME ³			−0.003	0.001	−3.114	0.0018**				
5 TIME ⁴			0.000	0.000	3.366	0.0008***				
6 TRIAL	−0.003	0.5553	0.005	0.003	1.919	0.0549	0.011	0.0000***	0.019	0.0000***
7 TIME:TRIAL			0.003	0.001	3.418	0.0006***				
8 C	0.179	0.1651	0.729	0.055	13.260	0.0000***	0.549	0.0000***	−0.029	0.6932
9 TIME:C			−0.006	0.025	−0.252	0.8007				
10 TIME ² :C			−0.050	0.009	−5.615	0.0000***				
11 TIME ³ :C			0.004	0.002	2.820	0.0048**				
12 I	0.113	0.3951	−0.506	0.070	−7.242	0.0000***	−0.688	0.0000***	−0.308	0.0000***
13 TIME:I			−0.196	0.041	−4.785	0.0000***				
14 TIME ² :I			0.042	0.010	4.291	0.0000***				
15 TIME ³ :I			0.008	0.003	2.915	0.0036**				
16 TIME ⁴ :I			−0.001	0.000	−3.736	0.0002***				
17 N	0.666	0.0000***	0.476	0.061	7.775	0.0000***	0.050	0.3661	−0.355	0.0000***
18 TIME:N			−0.187	0.024	−7.732	0.0000***				
19 TIME ² :N			−0.024	0.007	−3.373	0.0007***				
20 TIME ³ :N			0.006	0.001	4.673	0.0000***				
21 II	0.285	0.0012**	−0.403	0.054	−7.423	0.0000***	−0.623	0.0000***	−0.396	0.0000***
22 TIME:II			−0.193	0.025	−7.688	0.0000***				
23 TIME ² :II			0.035	0.006	5.714	0.0000***				
24 TIME ³ :II			0.005	0.002	3.146	0.0017**				
25 TIME ⁴ :II			−0.001	0.000	−3.747	0.0002***				
26 CC	0.389	0.0000***	0.627	0.049	12.860	0.0000***	0.407	0.0000***	−0.084	0.1254
27 TIME:CC			−0.054	0.016	−3.404	0.0007***				
28 TIME ² :CC			−0.034	0.005	−6.597	0.0000***				
29 TIME ³ :CC			0.003	0.001	3.835	0.0001***				
30 CI	−0.256	0.0000***	−0.256	0.032	−8.003	0.0000***	−0.256	0.0000***	−0.256	0.0000***
31 SOA_53_53	−0.191	0.0000***	−0.191	0.032	−5.985	0.0000***	−0.191	0.0000***	−0.191	0.0000***
32 SOA_80_27	−0.223	0.0027**	−0.327	0.046	−7.053	0.0000***	−0.336	0.0000***	−0.261	0.0000***
33 TIME:SOA_80_27			−0.017	0.012	−1.359	0.1742				
34 TIME ² :SOA_80_27			0.006	0.002	2.585	0.0097**				
35 II:SOA_53_53	0.124	0.0267*	0.124	0.056	2.216	0.0267*	0.124	0.0267*	0.124	0.0267*
36 II:SOA_80_27	0.245	0.0001***	0.245	0.062	3.928	0.0001***	0.245	0.0001***	0.245	0.0001***
37 CC:SOA_53_53	0.167	0.0021**	0.167	0.054	3.078	0.0021**	0.167	0.0021**	0.167	0.0021**
38 CC:SOA_80_27	0.287	0.0000***	0.287	0.061	4.693	0.0000***	0.287	0.0000***	0.287	0.0000***
39 CI:SOA_80_27	0.212	0.0001***	0.212	0.055	3.845	0.0001***	0.212	0.0001***	0.212	0.0001***
SD Intercept	1.122		.892				.679		.431	
SD TIME	.124		.124				.124		.123	
Correlation	−.937		−.899				−.818		−.427	

Note. Parameter estimates (PE) and test statistics. During model selection, TIME was centered on bin 275. The selected model was refitted three times with TIME centered on bin 200, 325, and 400, respectively. *SD* = standard deviation

way repeated-measures ANOVAs, with the factor consistency (consistent, inconsistent, no prime), were performed for

single-prime and no-prime conditions, one for each of the two dependent variables, RT and ER. A total of 3,600 trials

Table 2 Selected $ca(t)$ model for Experiment 1. Parameter estimates (PE) and test statistics

Effect	(150,175]		(200,225]		(250,275]				(375,400]	
	PE	<i>p</i>	PE	<i>p</i>	PE	SE	<i>t</i>	<i>p</i>	PE	<i>p</i>
1 Intercept	−2.379	0.0000***	−0.076	0.7026	1.918	0.157	12.187	0.0000***	3.229	0.0000***
2 TIME					0.797	0.062	12.806	0.0000***		
3 TIME ²					−0.118	0.015	−7.696	0.0000***		
4 TIME ³					−0.006	0.003	−1.993	0.0463*		
5 TIME ⁴					0.002	0.000	4.087	0.0000***		
6 C	5.213	0.0000***	2.906	0.0000***	1.339	0.232	5.764	0.0000***	0.659	0.1446
7 TIME:C					−0.599	0.109	−5.516	0.0000***		
8 TIME ² :C					0.093	0.027	3.369	0.0008***		
9 I	0.407	0.5240	−2.040	0.0000***	−2.405	0.188	−12.783	0.0000***	−0.126	0.6165
10 TIME:I					0.188	0.088	2.139	0.0324*		
11 TIME ² :I					0.148	0.025	5.903	0.0000***		
12 TIME ³ :I					−0.019	0.004	−4.559	0.0000***		
13 N	3.456	0.0000***	1.421	0.0000***	0.112	0.178	0.629	0.5292	0.013	0.9667
14 TIME:N					−0.473	0.072	−6.584	0.0000***		
15 TIME ² :N					0.091	0.018	5.130	0.0000***		
16 II	−0.900	0.1208	−2.376	0.0000***	−2.438	0.146	−16.694	0.0000***	−0.139	0.4832
17 TIME:II					0.227	0.073	3.108	0.0019**		
18 TIME ² :II					0.106	0.021	4.951	0.0000***		
19 TIME ³ :II					−0.012	0.003	−3.655	0.0003***		
20 CC	5.349	0.0000***	3.189	0.0000***	1.624	0.175	9.269	0.0000***	0.309	0.2298
21 TIME:CC					−0.634	0.074	−8.596	0.0000***		
22 TIME ² :CC					0.074	0.016	4.659	0.0000***		
23 CI	4.742	0.0000***	0.487	0.0502	−1.440	0.164	−8.805	0.0000***	−0.471	0.0513
24 TIME:CI					−0.494	0.070	−7.041	0.0000***		
25 TIME ² :CI					0.207	0.024	8.478	0.0000***		
26 TIME ³ :CI					−0.014	0.004	−3.927	0.0001***		
27 SOA_53_53	−0.461	0.0553	−0.391	0.0226*	−0.320	0.115	−2.780	0.0054**	−0.144	0.3858
28 TIME:SOA_53_53					0.035	0.039	0.893	0.3717		
29 SOA_80_27	−1.177	0.0179*	−1.246	0.0000***	−1.137	0.168	−6.750	0.0000***	−0.091	0.7183
30 TIME:SOA_80_27					0.099	0.075	1.322	0.1862		
31 TIME ² :SOA_80_27					0.022	0.013	1.736	0.0826		
32 II:SOA_80_27	−0.367	0.6955	0.626	0.1782	1.074	0.234	4.595	0.0000***	−0.190	0.5370
33 TIME:II:SOA_80_27					0.088	0.133	0.661	0.5085		
34 TIME ² :II:SOA_80_27					−0.068	0.023	−2.928	0.0034**		
35 CC:SOA_80_27	2.505	0.0002***	1.928	0.0001***	1.351	0.322	4.199	0.0000***	−0.092	0.8261
36 TIME:CC:SOA_80_27					−0.289	0.105	−2.750	0.0060**		
37 CI:SOA_53_53	1.327	0.0015**	0.957	0.0013**	0.587	0.195	3.012	0.0026**	−0.339	0.2108
38 TIME:CI:SOA_53_53					−0.185	0.068	−2.733	0.0063**		
39 CI:SOA_80_27	4.495	0.0000***	3.388	0.0000***	2.281	0.256	8.911	0.0000***	−0.486	0.1392
40 TIME:CI:SOA_80_27					−0.553	0.085	−6.521	0.0000***		
SD Intercept	.373		.324		.353				.353	
SD TIME	.081		.081		.081				.081	
Correlation	−.496		−.068		.398				.406	

Note. During model selection TIME was centered on bin 275. The selected model was refitted three times with TIME centered on bin 175, 225, and 400, respectively

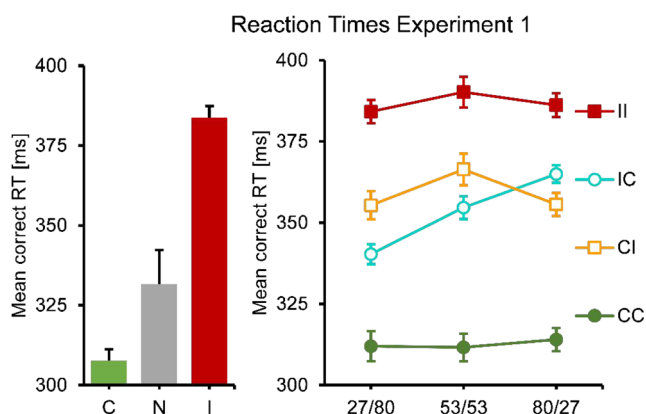


Fig. 2 Mean correct RT results for Experiment 1. No and single-prime conditions: Left panel, error bars resemble the standard error of the mean, consistency conditions on the x-axes. Double-prime conditions: Right panel, error bars resemble the standard error of the mean, separate lines for consistency conditions, SOA conditions on the x-axes

were initially available for analysis. Trials with reaction times faster than 100 ms or slower than 999 ms (0.5%) were excluded from the analysis. Further, error trials (10.92%) were excluded from RT analysis.

Second, two $3 \text{ (SOA)} \times 4 \text{ (consistency)}$ repeated-measures ANOVAs were performed for all double-prime conditions, one each for RT and ER. A total of 14,400 trials were initially available for analysis. Trials with reaction times faster than 100 ms or slower than 999 ms (0.53%) were excluded from the analysis. In addition, error trials (13.53%) were excluded from RT analysis. To follow up significant interaction effects, one-way repeated-measures ANOVAs, with the four-level factor consistency (CC, CI, IC, II) were performed separately for each SOA condition. Greenhouse–Geisser-corrected p values were used. To satisfy ANOVA requirements error rates were arcsine transformed. Additional within-subjects contrasts were calculated to further investigate significant main effects.

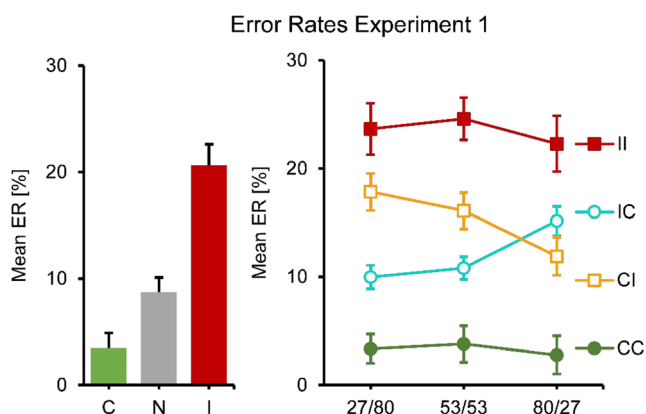


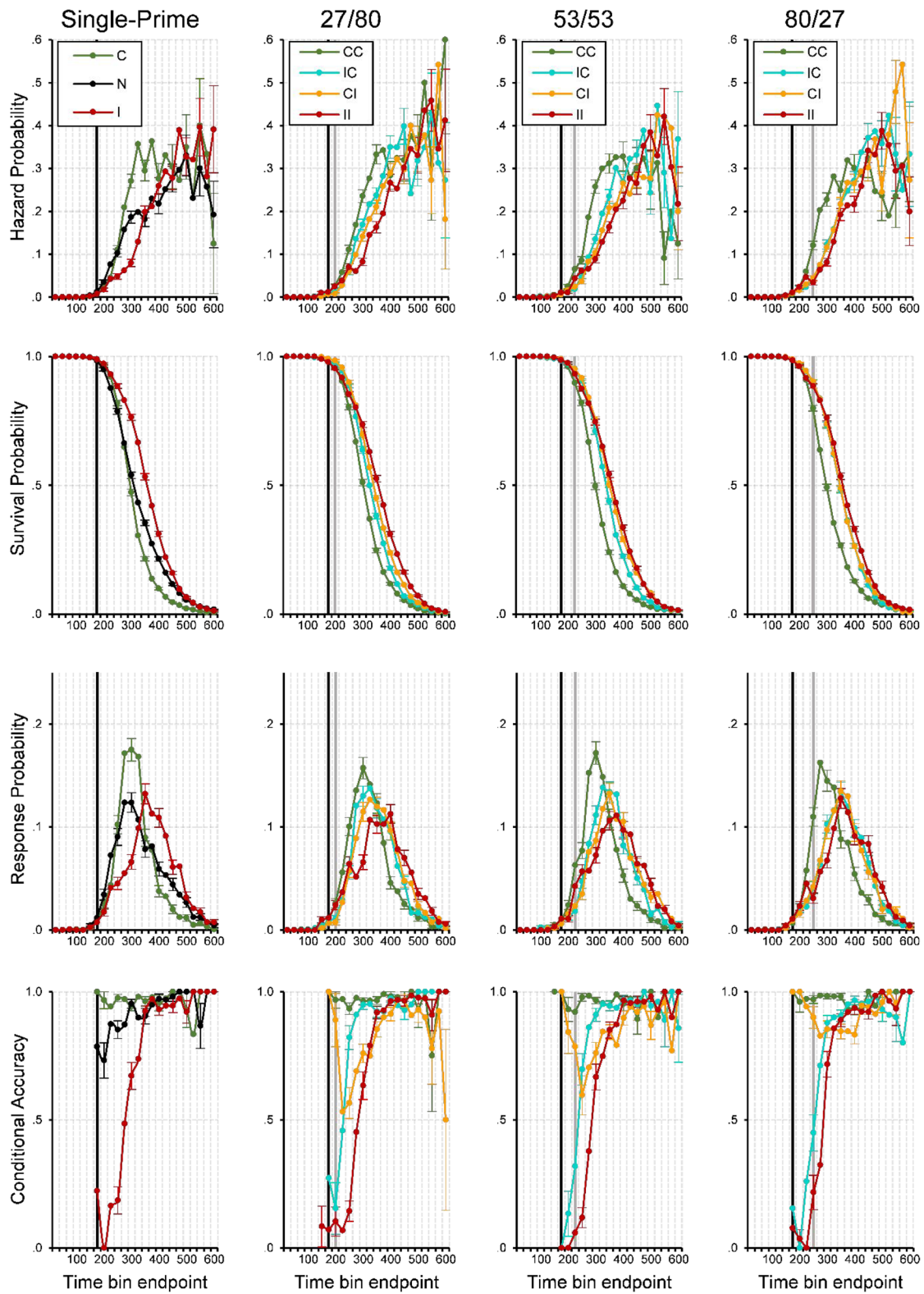
Fig. 3 Mean ER results for Experiment 1. No and single-prime conditions: Left panel, error bars resemble the standard error of the mean, consistency conditions on the x-axes. Double-prime conditions: Right panel, error bars resemble the standard error of the mean, separate lines for consistency conditions, SOA conditions on the x-axes

Event history analysis

Sample-based descriptive estimates of hazard function $h(t)$, survival function $S(t)$, probability function $P(t)$, and conditional-accuracy function $ca(t)$ were calculated for each combination of condition. For the purpose of visually inspecting the descriptive functions data was pooled across participants to reduce noise, after checking that each participant showed similarly timed effects. A censoring time of 600 ms was used because only a limited amount of responses occurred afterwards. To provide a high temporal resolution and still obtain stable estimates a bin size of 25 ms was used. In other words, the first 600 ms after target onset were divided into 24 time bins of 25 ms indexed by $t = 1$ to 24. Trials with RTs longer than 600 ms were treated as right-censored observations. Time bins are denoted by the endpoint of the interval they span, such that Bin 11 = Bin 275 = (250,275].

Next, discrete-time hazard models and conditional accuracy models were estimated by computing linear mixed-effects regression models in *R* (R Core Team, 2014; function `glmmPQL`³ of package MASS; see also Panis & Schmidt, 2016). For the hazard models we used the *complementary log-log (cloglog)* link.⁴ An example discrete-time hazard model with three predictors can be written as follows: $\text{cloglog}[h(t)] = \ln(-\ln[1 - h(t)]) = [\alpha_0 \text{ONE} + \alpha_1(\text{TIME} - 1) + \alpha_2(\text{TIME} - 1)^2 + \alpha_3(\text{TIME} - 1)^3] + [\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2(\text{TIME} - 1)]$. The main predictor variable TIME is the time bin index t 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 $\text{TIME} - 1 = 0$), in Bin 2 with $\beta_2 + \beta_3$ cloglog-hazard units (when $\text{TIME} - 1 = 1$), and so forth. To interpret the effects of the predictors, the parameter estimates are antilogged, resulting in a hazard ratio.

For our data we centered TIME on Bin 275 during model selection. TRIAL number was included as a predictor (centered on Trial 1,000, rescaled by dividing by 100), in order to account for across-trial learning effects in $h(t)$. 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.⁵ The IC-27/80 condition (P1: inconsistent, P2: consistent, SOA1: 27 ms, SOA2: 80 ms) was



chosen as a baseline condition. Because TIME and TRIAL are centered, the intercept of the hazard regression model refers to Bin 275 in Trial 1,000 of the IC-27/80 condition.

To estimate the parameters of an $h(t)$ model, we must create a dataset where each row corresponds to a time bin of a trial of a participant (a subject-trial-bin oriented data set).

Fig. 4 Sample-based estimates of $h(t)$, $S(t)$, $P(t)$, and $ca(t)$ aggregated across all participants in Experiment 1, for the first 24 bins (or 600 ms) after target onset. Bin width equals 25 ms. First column: Black lines represent the no-prime condition, green lines the consistent single-prime condition, and red lines the inconsistent single-prime condition. Second to last column: Each column represents a different SOA condition. Green lines represent consistent-consistent conditions, cyan lines inconsistent-consistent conditions, orange lines consistent-inconsistent conditions, red lines inconsistent-inconsistent conditions. Black vertical lines highlight bins at ~250–275 ms after onset of P1, grey vertical lines after onset of P2. Note that we only plotted a $ca(t)$ estimate if the corresponding hazard for that bin was larger than .005. For better visibility only every second error bar is depicted

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, the variable SUBJECT, and the dummy-coded dichotomous experimental predictor variables (C, I, N, II, CC, CI, SOA_53_53, SOA_80_27). Only the time range between 125 and 450 ms was modeled, because most responses occurred in this range. Trials with RTs longer than 450 ms were treated as right-censored observations, and trials with RT smaller or equal to 125 ms were discarded. The expanded (subject-trial-bin oriented) data set contained 157,656 rows.

For $ca(t)$ modeling, the original dataset was used where each row corresponds to one trial of one participant ($1,500 \times 12 = 18,000$ trials). We used the same model but applied the *logit* link⁶, and included only those trials with an observed response between 125 and 450 ms in the data set. In other words, trials with RT shorter than 125 ms and longer than 450 ms were discarded (11.63 % of the 18,000 trials).

For both models, we started with a full model containing all fixed effects of interest (main and interaction effects of the dichotomous predictors), and their interactions with TIME (linear, quadratic, cubic, and quartic). In a step-by-step backward selection procedure, this full model was reduced to the final, selected model. More precisely, in each iteration, the effect with the largest $p > .05$ that was not part of any higher order effect left the model before the next fit. Finally, after model selection, we refitted the selected 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 (see Tables 1 and 2). This way, it becomes more explicit what the interaction effects including TIME imply, because we are able to study the effect of the various predictor variables at different time points.

Predictions

We expected primes to have sequential effects that are traceable over time in the conditional accuracy functions. Because

P1–T SOAs in Experiment 1 are short, the sequence of response activations should conform to the chase criteria, so that the earliest responses are controlled exclusively by the first prime, while later responses are consecutively controlled by the second prime, and the slowest responses by the target. The earliest responses should therefore be correct whenever P1 is consistent with the target and incorrect whenever it is inconsistent. In contrast, the slowest responses should be controlled mainly by the target and thus all be correct. Intermediate responses should be influenced by the second prime. From previous data, we expected that the second prime would dominate the response at the shortest P1–P2 SOA (i.e., the longest P2–T SOA), and this dominance of the second prime should decrease with increasing P1–P2 SOA because the first prime has progressively more time to activate a response before the second prime occurs, while the second prime has progressively less time before the target occurs (Grainger et al., 2013).

Results

Analysis of mean error rate and mean correct RT

An analysis of the single prime conditions showed that responses were faster and more accurate when primes were consistent rather than inconsistent, with the no-prime condition in between (see Figs. 2 and 3, left panel). One-way repeated-measures ANOVAs showed significant differences in RT, $F(1.45, 15.90) = 33.39$, $p < .001$, as well as error rates, $F(1.84, 20.25) = 35.56$, $p < .001$. In RTs as well as error rates, all means were significantly different from each other, all $p \leq .001$, except for the RT difference between consistent and no-prime conditions ($p = .061$).

In a next step, double-prime conditions were analyzed. RTs and error rates showed a similar overall pattern: Responses were fastest and most accurate for two consistent primes, slowest and least accurate for two inconsistent primes, and in between when primes were mixed (conditions CI and IC). In RTs, a two-way repeated-measures ANOVA showed a significant main effect of consistency (with levels CC, CI, IC, II), $F(2.09, 22.98) = 64.94$, $p < .001$, a significant main effect of SOA $F(1.85, 20.35) = 10.43$, $p = .001$, and a significant interaction, $F(4.00, 44.04) = 7.86$, $p < .001$ (see Fig. 2, right panel). This pattern was broken down into two separate ANOVAs, one for identical (CC, II) and one for different primes (CI, IC). The first one (CC versus II) only showed a significant main effect of consistency, $F(1.00, 11.00) = 113.30$, $p < .001$. This effect was constant across SOA, with no main effect of SOA or an interaction. The second test (CI versus IC) showed that RT increased with SOA, $F(1.78, 19.32) = 27.76$, $p < .001$. There was no main effect of consistency, but a significant interaction, $F(1.46, 16.10) = 12.69$, $p = .001$. IC was faster

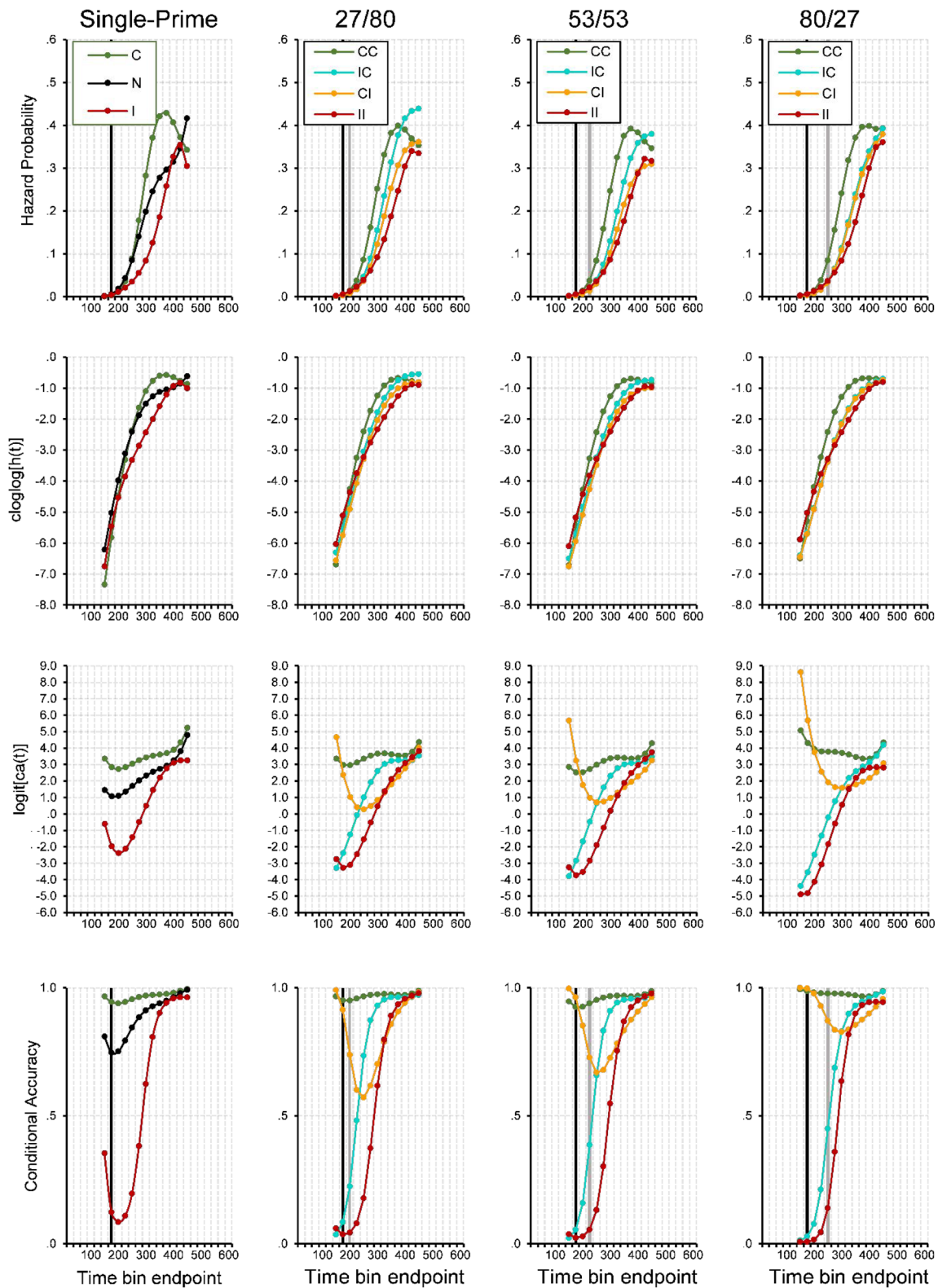


Fig. 5 Model predictions. Predicted hazard (first row), $cloglog[h(t)]$ (second row), $logit[ca(t)]$ (third row), and conditional accuracy functions (fourth row) for trial 1,000 of Experiment 1. Again, black vertical lines highlight bins at ~250–275 ms after onset of P1, grey vertical lines after onset of P2

than CI when the first SOA was short, but slower when it was long.

This pattern was even clearer (and almost perfectly symmetrical) in the error rates. An ANOVA of all double-prime

conditions showed no main effect of SOA, but a significant main effect of consistency, $F(1.73, 19.07) = 37.84$, $p < .001$, and a significant interaction $F(3.38, 37.18) = 5.61$, $p = .002$. The follow-up analysis of CC versus II conditions only showed a main effect of consistency, $F(1.00, 11.00) = 59.40$, $p < .001$, but no SOA or interaction effects. The follow-up analysis of CI versus IC conditions only showed an interaction, $F(1.98, 21.74) = 13.87$, $p < .001$, but no main effects: IC was more accurate than CI when the first SOA was short, but less accurate when it was long.

Event history analysis: Descriptive statistics

In the single-prime conditions (see the first column in Fig. 4), the fastest responses occurred around 150 ms after target onset. Thereafter, we saw a steady increase in response hazards, which was delayed for inconsistent compared with consistent primes. This led to a marked priming effect in $h(t)$ of about 150 ms duration, and also in median RT (i.e., when the survivor function crosses .5) and mean RT. When most responses had occurred and the survival probability was low, response hazard was still at a high constant level. Strikingly, early responses were virtually always correct whenever the prime was consistent, but incorrect whenever it was inconsistent, showing that responses were exclusively determined by the prime, not the target. In inconsistent trials, conditional accuracy then quickly increased from almost zero to almost one, showing how the target took control over the response.⁷

Let us now take a look at the double-prime conditions where the P1–P2 SOA was short and the P2–T SOA was long (27/80; see the second column in Fig. 4), so that the impact of the second prime should be high relative to the first prime. Again, the fastest responses occurred around the same time in all priming conditions, around 150 ms after target onset or about 250 ms after P1 onset. However, initial response hazards in CI and IC conditions were lower than in CC and II conditions around 150–200 ms. This likely reflected early response competition due to conflicting prime information, as both primes activated opposite responses. After about 250 ms without response occurrence, the hazard functions began to differentiate and followed the order observed in mean RTs: CC was fastest, followed by IC, CI, II. This was evident in the hazard, survivor, and probability mass functions. The most diagnostic information, however, was in the conditional accuracy functions. Not surprisingly, the earliest responses were virtually all correct when both primes were consistent and all incorrect when both primes were inconsistent, which again showed that the first prime determined the earliest responses. In the II condition, conditional accuracy then quickly increased as the target took control over the response. This also occurred in the IC condition, demonstrating that the first prime alone controlled the earliest responses; but the following increase in accuracy occurred earlier than in the

II condition, demonstrating that the consistent second prime influenced the response as well. Exactly the reverse process occurred in the CI condition. Here, response accuracy was nearly perfect at first because of the consistent first prime, then decreased as the inconsistent second prime became effective, and then increased again as the target finally took control over the response.

What happened when the SOAs shifted first to 53/53 and then to 80/27? The overall pattern in $ca(t)$ remained the same: CC was always fastest with near-perfect accuracy, II was always slowest with conditional accuracy rising from very low to very high values, IC always showed an earlier increase in conditional accuracy, time-locked to the second prime's appearance. The most important change occurred in the CI condition. As the P1–P2 SOA became larger and the P2–T SOA became correspondingly shorter, the influence of the inconsistent second prime diminished, and the temporary drop in conditional accuracy became smaller. As with the effects observed in IC, this nadir in conditional accuracy was time locked to the second prime's presentation.

Event history analysis: Inferential statistics

To see whether these observed differences are significant we fitted hazard and conditional accuracy models to the aggregated data. Table 1 shows the selected hazard model, and Table 2 the selected $ca(t)$ model. Figure 5 shows predicted (i.e., model-based) hazard $\text{cloglog}[h(t)]$, $\text{logit}[ca(t)]$, and conditional accuracy functions for Trial 1,000 (note that choosing another trial number would not change the priming effects because we did not include interaction effects including TRIAL). The first five parameters in Table 1 model the shape of the $\text{cloglog}[h(t)]$ function in the baseline condition, IC-27/80 in Trial 1,000 (see Fig. 5, row 2, column 2, blue line). The intercept of -2.361 cloglog-hazard units corresponds to an estimated hazard of .09 in Bin 275. This intercept increases over time in a linear, quadratic, cubic and quartic fashion (see the Parameters 2 to 5 in Table 1), so that the intercept changes from -4.65 in Bin 200 to $-.62$ in Bin 400 (see row 1 in Table 1).

Most importantly, compared with condition IC, changing to CC increases the estimated cloglog-hazard in Bin 275 by .627 units (Parameter 26), changing to CI decreases it by .256 units (Parameter 30), and changing to II decreases it by .403 units (Parameter 21; all $p < .0001$). While the main effects of CC and II in Bin 275 change in magnitude over time (parameter estimates in rows 27–29, 22–25), the effect of CI is time invariant. For example, note that in Bin 200 conditions, II and CC have positive parameter estimates that significantly differ from condition IC (see the parameter estimates in rows 21 and 26 in Table 1, column 3). This means that the hazard of response occurrence is lower in Bin 200 in mixed prime conditions.

The effect of changing the SOA combination from 27/80 to 53/53 is to decrease the estimated cloglog-hazard by .191 units in all bins (Parameter 31). The estimated cloglog-hazard decreases even further when SOA combination is changed to 80/27 (Parameters 32–34). In other words, response occurrence slows down with decreasing P2–T SOA for condition IC. However, this effect is much smaller or absent for CC and II due to interactions with 53/53 (Parameters 35 and 37) and 80/27 (Parameters 36 and 38). Furthermore, with SOA combination 80/27 the difference between CI and IC is gone (due to Parameter 39 neutralizing the effect of Parameter 30). Finally, the hazard model also shows a significant effect of TRIAL on the estimated cloglog-hazard in bins after 275 ms after target onset.

The first five parameters in Table 2 model the shape of the $\text{logit}[ca(t)]$ function in the baseline condition, IC-27/80 in Trial 1,000 (see Fig. 5, row 3, column 2, blue line). The intercept of 1.918 corresponds to an estimated $ca(t)$ of .87 in Bin 275. This intercept increases over time in a linear, quadratic, cubic and quartic fashion (Parameters 2–5).

Most importantly, compared with condition IC, changing to CC increases the estimated $\text{logit}-ca(t)$ in Bin 275 by 1.624 units (Parameter 20), changing to CI decreases it by 1.44 units (Parameter 23), and changing to II decreases it by 2.438 units (Parameter 16; all $ps < .0001$). The main effects of CC, CI, and II in Bin 275 change over time (Parameters 16–26), so that relative to IC, the positive effect of CC decreases over time, the negative effect of II first increases and then decreases, and the effect of CI shifts from positive to negative to zero. For example, note that in Bin 175 conditions CC and CI have positive parameter estimates that significantly differ from

condition IC while II is not significantly different (compare rows 20 and 23 with row 16 in Table 2, column 3). This means that the conditional accuracy of these early responses is almost zero for II and IC, and almost one for CI and CC, thus reflecting first prime identity (see Fig. 5, row 4).

Increasing the P1–P2 SOA leads to a decrease in the estimated $\text{logit}-ca(t)$ in each bin (Parameters 27–31). Confirming the change in the temporary drop in conditional accuracy for condition CI in Fig. 4 are the (early and positive) interactions between CI, SOA combination, and TIME (Parameters 37–40).

Summary

As expected, mean RT and mean ER analyses of the single-prime and no-prime conditions revealed that the stimulus-set used was sufficient to produce the common finding in response priming experiments: faster and more accurate responses in consistent trials and slower and less accurate responses in inconsistent trials. Similarly, when two primes were presented, responses were fastest and most accurate for two consistent primes, slowest and least accurate for two inconsistent primes, and in between when primes were mixed. The event history analysis showed that sequential primes in fact initiate sequential response activation: (1) earliest responses were controlled exclusively by the first prime, (2) intermediate responses reflected competition between the primes where the identity of the second prime increasingly dominated the response as P2–T SOA increased, (3) this latter effect was tracking the onset of the second prime, both in magnitude and timing, and (4) only the slowest responses were clearly controlled by the target.

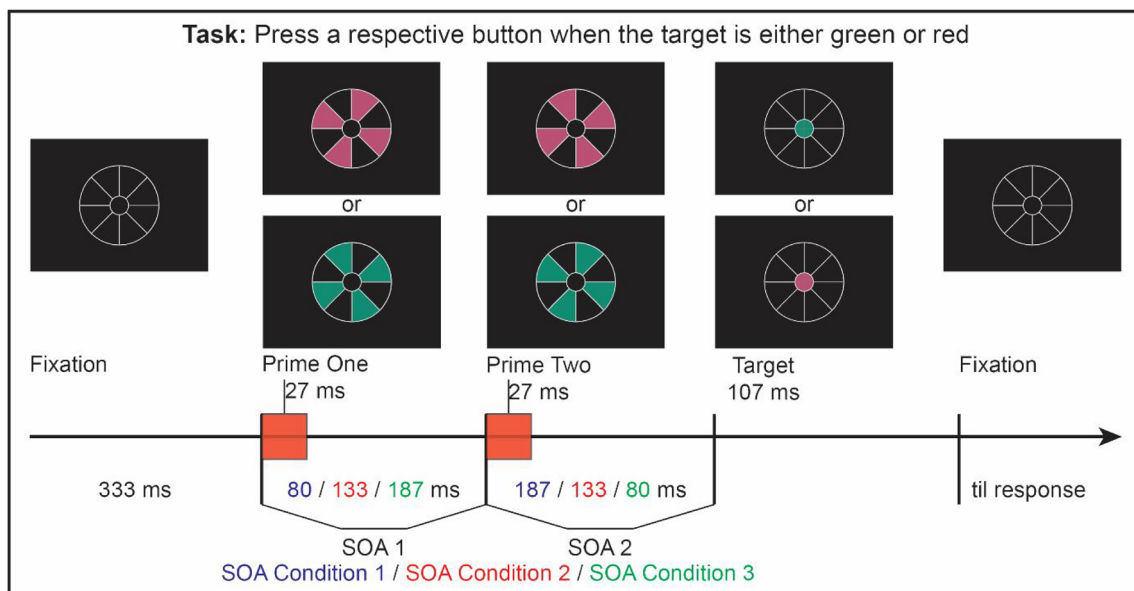


Fig. 6 Stimulus displays and design. After fixating the center of the white lollipop frame a sequence of two primes and a target is presented, with SOA1–SOA2 combinations of 80/187, 133/133, or 187/80

Table 3 Selected hazard model for Experiment 2

		(150,175]		(250,275]				(300,325]		(375,400]	
effect		PE	<i>p</i>	PE	SE	<i>t</i>	<i>p</i>	PE	<i>p</i>	PE	<i>p</i>
1	Intercept	−5.941	0.0000***	−2.162	0.259	−8.355	0.0000***	−1.190	0.0000***	−0.843	0.0000***
2	TIME			0.666	0.037	18.245	0.0000***				
3	TIME ²			−0.089	0.004	−23.066	0.0000***				
4	TIME ³			−0.002	0.001	−2.467	0.0136*				
5	TIME ⁴			0.001	0.000	6.131	0.0000***				
6	TRIAL	0.008	0.0000***	0.008	0.002	4.099	0.0000***	0.008	0.0000***	0.008	0.0000***
7	C	0.128	0.4836	0.315	0.050	6.343	0.0000***	0.268	0.0000***	0.022	0.7489
8	TIME:C			0.000	0.026	−0.001	0.9995				
9	TIME ² :C			−0.012	0.005	−2.337	0.0194*				
10	I	−0.279	0.1733	−0.442	0.056	−7.876	0.0000***	−0.383	0.0000***	−0.118	0.0377*
11	TIME:I			0.006	0.029	0.217	0.8282				
12	TIME ² :I			0.012	0.005	2.409	0.0160*				
13	N	1.282	0.0000***	0.337	0.053	6.368	0.0000***	0.034	0.5100	−0.208	0.0024**
14	TIME:N			−0.180	0.022	−8.171	0.0000***				
15	TIME ² :N			0.014	0.004	3.248	0.0012**				
16	II	1.188	0.0000***	−0.569	0.059	−9.696	0.0000***	−0.790	0.0000***	−0.394	0.0000***
17	TIME:II			−0.236	0.028	−8.398	0.0000***				
18	TIME ² :II			0.063	0.006	10.072	0.0000***				
19	TIME ³ :II			0.001	0.001	0.662	0.5082				
20	TIME ⁴ :II			−0.001	0.000	−2.521	0.0117*				
21	CC	0.754	0.0000***	0.335	0.038	8.733	0.0000***	0.194	0.0000***	0.069	0.1961
22	TIME:CC			−0.082	0.019	−4.397	0.0000***				
23	TIME ² :CC			0.006	0.004	1.522	0.1281				
24	CI	−0.100	0.5469	−0.529	0.047	−11.279	0.0000***	−0.599	0.0000***	−0.358	0.0000***
25	TIME:CI			−0.095	0.027	−3.494	0.0005***				
26	TIME ² :CI			0.026	0.006	4.155	0.0000***				
27	TIME ³ :CI			0.003	0.002	1.769	0.0769				
28	TIME ⁴ :CI			−0.001	0.000	−2.630	0.0085**				
29	SOA_80_187	0.780	0.0000***	0.350	0.049	7.174	0.0000***	0.200	0.0000***	0.056	0.3904
30	TIME:SOA_80_187			−0.086	0.020	−4.252	0.0000***				
31	TIME ² :SOA_80_187			0.005	0.003	1.621	0.1050				
32	SOA_133_133	0.275	0.0000***	0.160	0.032	5.045	0.0000***	0.103	0.0000***	0.018	0.6524
33	TIME:SOA_133_133			−0.029	0.010	−2.884	0.0039**				
34	II:SOA_80_187	−0.933	0.0000***	−0.080	0.078	−1.025	0.3055	0.137	0.0550	0.202	0.0222*
35	TIME:II:SOA_80_187			0.143	0.030	4.809	0.0000***				
36	TIME ² :II:SOA_80_187			−0.017	0.005	−3.336	0.0008***				
37	II:SOA_133_133	−0.381	0.0028**	−0.169	0.068	−2.496	0.0126*	−0.063	0.2253	0.096	0.1665
38	TIME:II:SOA_133_133			0.053	0.018	2.925	0.0034**				
39	CC:SOA_80_187	−0.818	0.0000***	−0.127	0.064	−1.982	0.0474*	−0.007	0.9181	−0.107	0.2798
40	TIME:CC:SOA_80_187			0.098	0.029	3.364	0.0008***				
41	TIME ² :CC:SOA_80_187			−0.019	0.006	−3.154	0.0016**				
42	CI:SOA_80_187	−0.268	0.0534	−0.093	0.069	−1.337	0.1811	−0.005	0.9291	0.127	0.1213
43	TIME:CI:SOA_80_187			0.044	0.021	2.082	0.0374*				
	SD Intercept	1.301		.888				.696		.460	
	SD TIME	.111		.111				.111		.111	
	Correlation	−.954		−.898				−.828		−.528	

Note. Parameter estimates (PE) and test statistics. During model selection TIME was centered on bin 275. The selected model was refitted three times with TIME centered on bin 175, 325, and 400, respectively

Experiment 2

Method

Participants

All participants from the first experiment also took part in the second experiment (see Participants section for Experiment 1). Experiment order was counterbalanced.

Apparatus and stimuli

The same apparatus and stimuli were employed (see Apparatus and Stimuli section in Experiment 1).

Procedure

Again, each trial began with the onset of the lollipop frame (see Fig. 6). This time, after 333 ms of fixation, P1 was

Table 4 Selected $ca(t)$ model for Experiment 2

Effect	(175,200]		(250,275]				(300,325]		(375,400]	
	PE	<i>p</i>	PE	SE	<i>t</i>	<i>p</i>	PE	<i>p</i>	PE	<i>p</i>
1 Intercept	1.861	0.0000***	2.827	0.176	16.059	0.0000***	3.139	0.0000***	3.603	0.0000***
2 TIME			0.188	0.054	3.499	0.0005***				
3 TIME ²			−0.025	0.013	−1.968	0.0491**				
4 TIME ³			0.005	0.003	1.877	0.0605				
5 TIME ⁴			0.000	0.000	−0.784	0.4328				
6 C	0.700	0.0378*	0.223	0.207	1.079	0.2807	−0.095	0.5991	−0.573	0.0359*
7 TIME:C			−0.159	0.061	−2.599	0.0094**				
8 I	−3.641	0.0000***	−1.287	0.193	−6.660	0.0000***	−0.322	0.1310	0.218	0.4510
9 TIME:I			0.603	0.095	6.345	0.0000***				
10 TIME ² :I			−0.060	0.020	−2.998	0.0027**				
11 N	−0.723	0.0000***	−0.723	0.158	−4.584	0.0000***	−0.723	0.0000***	−0.723	0.0000***
12 II	−6.031	0.0000***	−2.997	0.173	−17.332	0.0000***	−1.590	0.0000***	−0.402	0.0576
13 TIME:II			0.827	0.064	12.927	0.0000***				
14 TIME ² :II			−0.062	0.013	−4.679	0.0000***				
15 CC	0.961	0.0011**	0.111	0.151	0.731	0.4645	0.229	0.1297	0.313	0.1589
16 TIME:CC			−0.022	0.073	−0.298	0.7660				
17 TIME ² :CC			0.059	0.019	3.039	0.0024**				
18 TIME ³ :CC			−0.009	0.004	−2.637	0.0084**				
19 CI	−3.660	0.0000***	−3.678	0.188	−19.545	0.0000***	−2.233	0.0000***	−1.129	0.0000***
20 TIME:CI			0.721	0.103	7.017	0.0000***				
21 TIME ² :CI			0.072	0.025	2.823	0.0048**				
22 TIME ³ :CI			−0.044	0.006	−6.856	0.0000***				
23 TIME ⁴ :CI			0.004	0.001	4.837	0.0000***				
24 SOA_80_187	0.972	0.0034**	−0.013	0.170	−0.076	0.9396	−0.386	0.0178*	−0.574	0.0081**
25 TIME:SOA_80_187			−0.240	0.065	−3.706	0.0002***				
26 TIME ² :SOA_80_187			0.028	0.022	1.291	0.1967				
27 TIME ³ :SOA_80_187			0.000	0.004	−0.128	0.8984				
28 SOA_133_133	−0.097	0.7321	0.051	0.162	0.317	0.7513	0.034	0.8416	−0.166	0.4399
29 TIME:SOA_133_133			0.015	0.056	0.260	0.7949				
30 TIME ² :SOA_133_133			−0.012	0.012	−0.978	0.3280				
31 II:SOA_80_187	1.084	0.0000***	1.084	0.207	5.239	0.0000***	1.084	0.0000***	1.084	0.0000***
32 II:SOA_133_133	0.724	0.0008***	0.724	0.215	3.365	0.0008***	0.724	0.0008***	0.724	0.0008***
33 CI:SOA_80_187	−1.783	0.0058**	1.598	0.254	6.289	0.0000***	1.640	0.0000***	0.281	0.3841
34 TIME:CI:SOA_80_187			0.369	0.119	3.089	0.0020**				
35 TIME ² :CI:SOA_80_187			−0.205	0.040	−5.113	0.0000***				
36 TIME ³ :CI:SOA_80_187			0.016	0.006	2.538	0.0112*				
37 CI:SOA_133_133	−0.146	0.7966	0.672	0.230	2.922	0.0035**	0.769	0.0010**	0.242	0.4493
38 TIME:CI:SOA_133_133			0.138	0.111	1.242	0.2142				
39 TIME ² :CI:SOA_133_133			−0.045	0.023	−1.987	0.0470*				
SD Intercept	.231		.390				.496		.655	
SD TIME	.053		.053				.053		.053	
Correlation	.988		.996				.998		.999	

Note. Parameter estimates (PE) and test statistics. During model selection TIME was centered on bin 275. The selected model was refitted three times with TIME centered on bin 200, 325, and 400, respectively

presented in either red or green for 27 ms, except for the no-prime trials, during which all segments remained black. After

a P1–P2 SOA of 80, 133, or 187 ms, either a red or green P2 was presented for 27 ms, except for the no-prime and single-

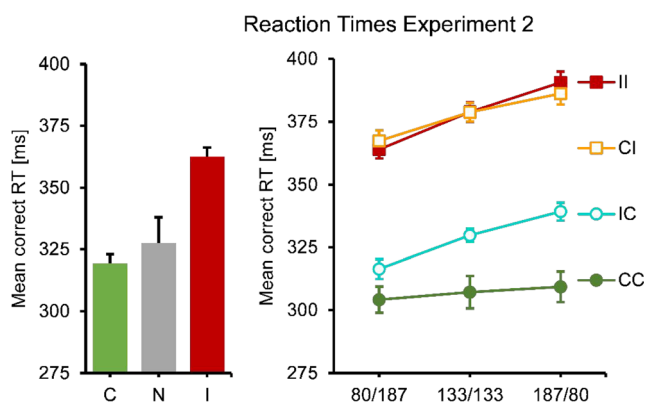


Fig. 7 Mean correct RT results for Experiment 2. No and single-prime conditions: Left panel, error bars resemble the standard error of the mean, consistency conditions on the x-axes. Double-prime conditions: Right panel, error bars resemble the standard error of the mean, separate lines for consistency conditions, SOA conditions on the x-axes

prime trials, during which all segments remained black. After a P2–T SOA of 187, 133, or 80 ms, respectively, a red or green target was presented. In this experiment, P1–P2 and P2–T SOAs always added up to a P1–T SOA of 267 ms. The target stayed on-screen for 107 ms.

Analysis of mean error rate and mean correct RT

In a first step, two sets of analyses were performed. First, one-way repeated-measures ANOVAs, with the factor consistency (consistent, inconsistent, no prime), were performed for single-prime and no-prime conditions, one for each of the two dependent variables, RT and ER. A total of 3,600 trials were initially available for analysis. Trials with reaction times faster than 100 ms or slower than 999 ms (0.58%) were excluded from the analysis. Further, error trials (then 8.16%) were excluded from RT analysis.

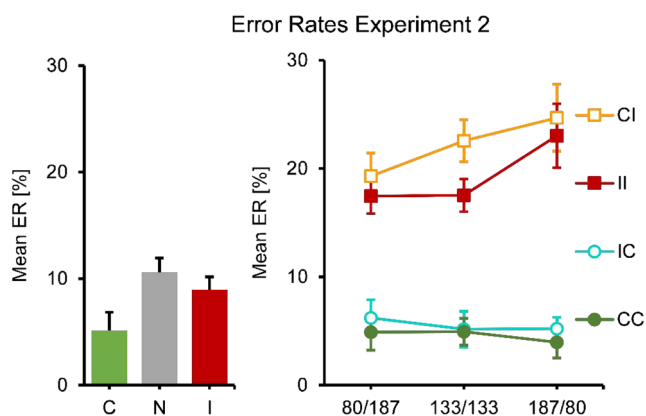


Fig. 8 Mean ER results for Experiment 2. No and single-prime conditions: Left panel, error bars resemble the standard error of the mean, consistency conditions on the x-axes. Double-prime conditions: Right panel, error bars resemble the standard error of the mean, separate lines for consistency conditions, SOA conditions on the x-axes

Second, two $3 \text{ (SOA)} \times 4 \text{ (consistency)}$ repeated-measures ANOVAs were performed for all double-prime conditions, one each for RT and ER. Initially, a total of 14,400 trials were available for analysis. Trials with reaction times faster than 100 ms or slower than 999 ms (0.93%) were excluded from the analysis. Further, error trials (then 12.88%) were excluded from RT analysis. To follow up on significant interaction effects, one-way repeated-measures ANOVAs, with the four-level factor consistency, were performed separately for each SOA condition.

Greenhouse–Geisser-corrected p values were used. To satisfy ANOVA requirements, error rates were arc-sine transformed. Additional within-subjects contrasts were calculated to further investigate significant main effects.

Event history analysis

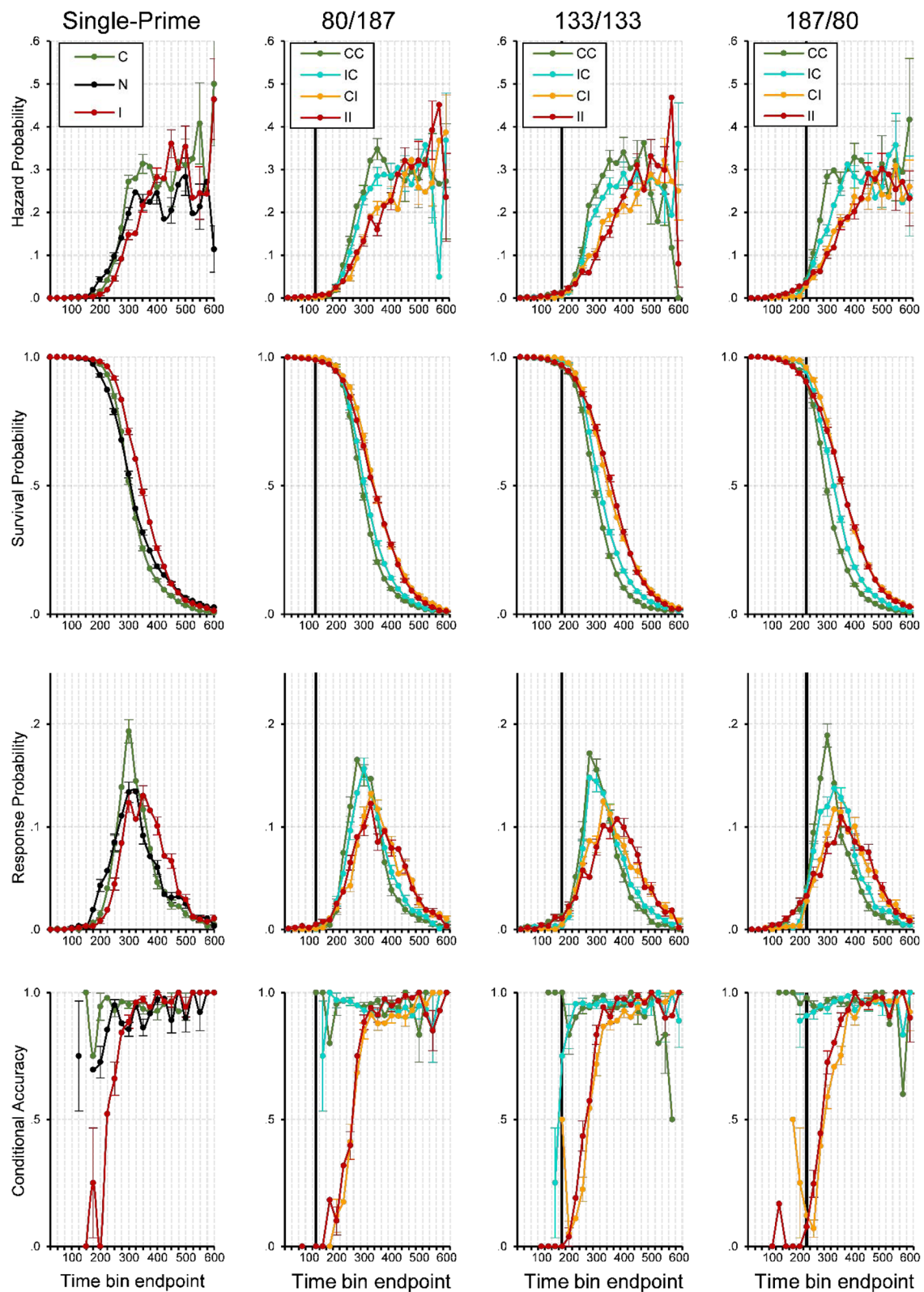
First, descriptive statistics were calculated as in Experiment 1 (see Event History Analysis section in Experiment 1). Next, for hazard modeling purposes, we censored the trials at 450 ms after target onset, and discarded the first five bins, since the most informative events occurred within 125 to 450 ms. The final data set for fitting $h(t)$ models contained 153,286 rows.

Finally, for $ca(t)$ modeling, the original data set was used where each row corresponds to one trial of one participant ($1,500 \times 12 = 18,000$ trials). Trials with a response latency below 125 ms or above 450 ms were deleted (12.36% of the data), in order to avoid problems of linear separability during model fitting. The final data set for the $ca(t)$ model contained 15,775 rows.

The estimation procedures were the same for both models as in Experiment 1, except that now the IC-187/80 condition (P1: inconsistent, P2: consistent, P1–P2 SOA: 187 ms, P2–T SOA: 80 ms) was chosen as a baseline condition. In summary, with all effects set to zero, the $h(t)$ model's intercept refers to the estimated $\text{cloglog}[h(t)]$, and the $ca(t)$ model's intercept to the estimated $\text{logit}[ca(t)]$, for bin 275 in Trial 1,000 of the IC-187/80 condition. Again, we refitted the selected model a number of times, with TIME centered each time on another bin (see Tables 3 and 4).

Predictions

Because P1–T SOAs in Experiment 2 are long, responses are no longer expected to conform to the chase criteria because participants have to wait out the target in order to safeguard against errors provoked by the primes, so that early primes can influence responses only out of the memory buffer that carries information from both primes but is dominated by the second



one (Grainger et al., 2013). Therefore, we expected that early responses would no longer be controlled exclusively by the first prime, but jointly by both primes, with the second prime

becoming more dominant as the P2–T SOA increased. The latest responses should be controlled mainly by the target and thus all be correct.

Fig. 9 Sample-based estimates of $h(t)$, $S(t)$, $P(t)$, and $ca(t)$ aggregated across all participants in Experiment 2, for the first 24 bins (or 600 ms) after target onset. Bin width equals 25 ms. (First column) Black lines represent the no-prime condition, green lines the consistent single-prime condition, and red lines the inconsistent single-prime condition. (Second to last column) Each column represents a different SOA condition. Green lines represent consistent-consistent conditions, cyan lines inconsistent-consistent conditions, orange lines consistent-inconsistent conditions, red lines inconsistent-inconsistent conditions. Black vertical lines highlight bins at ~275–300 ms after onset of P2. Note that we only plotted a $ca(t)$ estimate if the corresponding hazard for that bin was larger than .003. For better visibility only every second error bar is depicted

Results

Analysis of mean error rate and mean correct RT

Analysis of the single prime conditions showed that responses were faster and more accurate when primes were consistent rather than inconsistent. The no-prime condition was intermediate in response times, but higher than the other two in error rate (see Figs. 7 and 8, left panel). One-way repeated-measures ANOVAs showed significant differences in RT, $F(1.34, 14.75) = 12.60$, $p = .002$, as well as error rates, $F(1.82, 20.02) = 3.95$, $p = .039$. In RTs, the differences between all means were significant, all $ps \leq .002$, except the one between consistent and no primes. In ER, only the difference between consistent and no primes was significant, $p = .035$.

In a next step, double-prime conditions were analyzed (see Figs. 7 and 8, right panel). Responses were fastest and most accurate for two consistent primes and slowest and less accurate for two inconsistent primes. The CI condition was virtually identical to the II condition in response times, but slightly higher in error rate. The IC condition was similar to the CC condition in error rates, but slower in terms of response times. In RTs, a two-way repeated-measures ANOVA showed a significant main effect of consistency (with levels CC, CI, IC, II), $F(1.55, 17.00) = 59.88$, $p < .001$, a significant main effect of SOA, $F(1.62, 17.83) = 87.63$, $p < .001$, and a significant interaction, $F(3.35, 36.88) = 3.66$, $p = .018$, that seems to be based on the less steep increase in RT with SOA in the CC condition. We broke down this pattern post hoc into two separate ANOVAs, one for inconsistent and one for consistent second primes. The first one (II versus CI) only showed that response time increased with SOA, $F(1.73, 19.01) = 87.91$, $p < .001$. The second test (CC versus IC) showed that responses were faster for CC than for IC, $F(1.00, 11.00) = 15.44$, $p = .002$, that RT increased with SOA, $F(1.66, 18.30) = 15.66$, $p < .001$, and that the increase was steeper for IC than for CC, $F(1.68, 18.46) = 6.11$, $p = .012$.

The same strategy was used for the error rates. An ANOVA of all dual-prime conditions showed no main effect of SOA, but a significant main effect of consistency, $F(1.83, 20.08) = 45.15$, $p < .001$, and an interaction effect, $F(4.26, 46.82) =$

2.55, $p = .048$. The analysis of CC versus IC conditions gave no significant effects, and neither did the analysis of II versus CI conditions.

Event history analysis: Descriptive statistics

In the single-prime conditions (first column in Fig. 9), the first responses occur after about 200 ms, which is a bit later than in Experiment 1 and in line with the prediction that participants have to safeguard against errors. After that, there is an increase in response hazards that is steeper for consistent than for inconsistent primes, leading to an advantage in mean and median RT. Again, around 400 ms after target onset, this priming effect is gone. As in Experiment 1, early responses are mostly correct when the single prime is consistent, but incorrect when it is inconsistent, showing that early responses are still determined by the prime, not the target.⁸

Let us now look at the double-prime conditions where the P1–P2 SOA is short and the P2–T SOA is long (80/187, second column in Fig. 9), so that the impact of the second prime should be high relative to the first prime. Again, although the very earliest responses occur around the same time in all priming conditions, initial response hazards in CI and IC conditions are lower than in CC and II conditions, reflecting early response competition between both prime-triggered responses (see also the survivor functions). After about 250 ms, both groups begin to differentiate and now follow the order observed in mean RTs: CC, IC, and then CI and II. Again, the most diagnostic information is in the conditional accuracy functions, which show a markedly different pattern than in Experiment 1. The earliest responses are still predominantly correct when both primes are consistent and predominantly incorrect when both primes are inconsistent, showing that the earliest responses are not determined by the target, but by information in the primes. However, conditional accuracy functions for CC and IC are virtually identical, as are those of II and CI. In other words, the earliest systematic responses reflect only the second prime and not the first, probably because the P1–T SOA is too long. However, if observers would respond faster, we believe that the very first responses would reflect the first prime, just as in Experiment 1.

Although these early effects seem to be driven largely by the second prime, comparison with the 133/133 and 187/80 SOA conditions shows that the effects on hazard cannot be attributed to the second prime alone. While the CC, CI, and II condition show highly similar time courses in every condition, this is not true for the IC condition: The longer the first SOA and the shorter the second one, the more delay appears in condition IC compared with CC (the same effect that is evident in average RT). This effect shows that the first prime has an influence on the timing of the response. Moreover, it

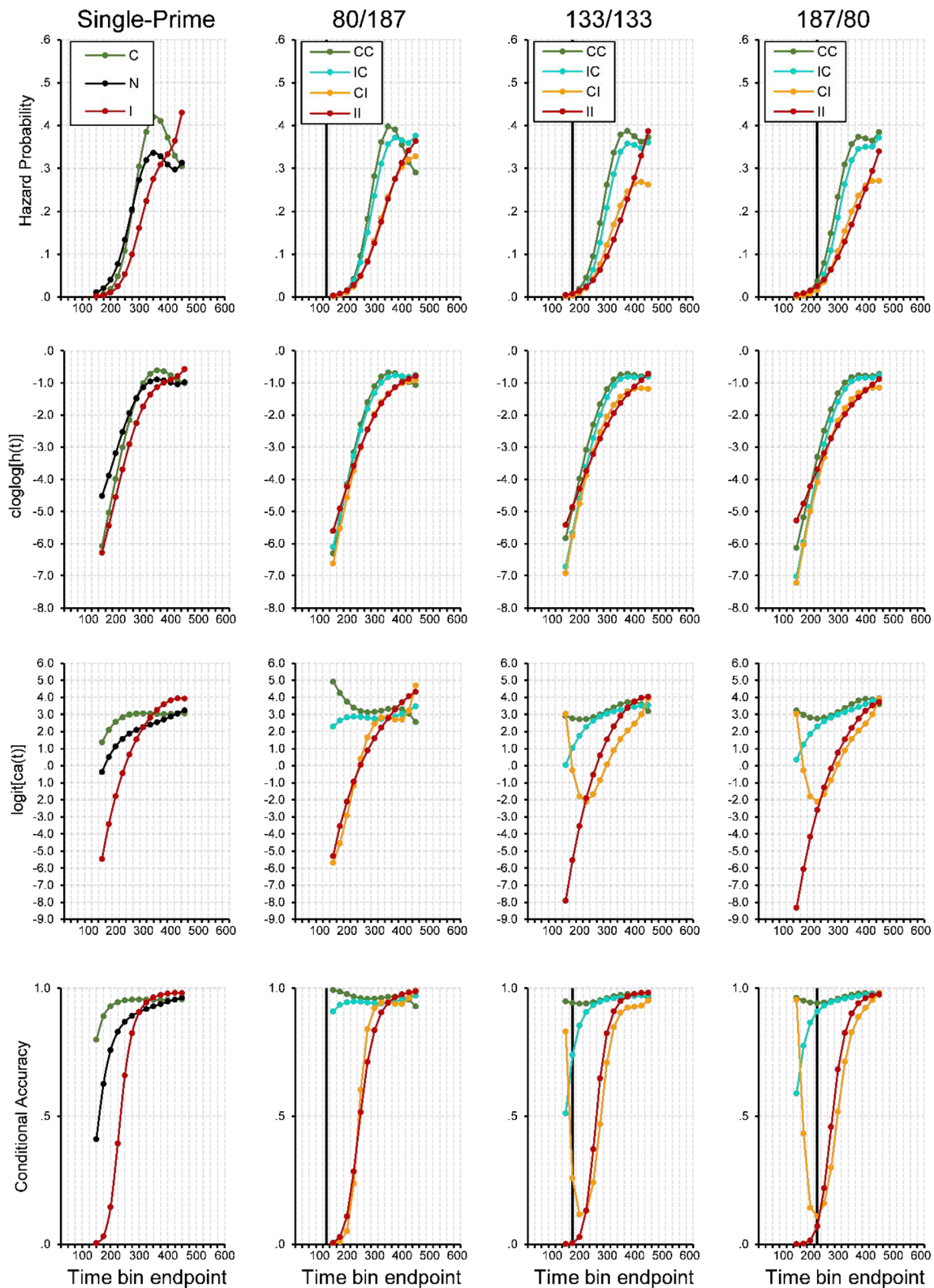


Fig. 10 Model predictions. Predicted hazard (first row), $cloglog[h(t)]$ (second row), $logit[ca(t)]$ (third row), and conditional accuracy functions (fourth row) for trial 1,000 of Experiment 2. Black vertical lines highlight bins at ~275–300 ms after onset of P2

appears that with shorter P2–T SOA, effects of the second prime become visible in the conditional accuracy functions.

Earliest responses are increasingly more incorrect in the IC condition and increasingly more accurate in the CI condition.

Event history analysis: Inferential statistics

Table 3 shows the selected hazard model, and Table 4 the selected $ca(t)$ model. Figure 10 shows predicted (i.e., model-based hazard) $\text{cloglog}[h(t)]$, $\text{logit}[ca(t)]$ and conditional accuracy functions for Trial 1,000. The first five parameters in Table 3 model the shape of the $\text{cloglog}[h(t)]$ function in the baseline condition, IC-187/80 in Trial 1,000 (see Fig. 10, row 2, column 4, blue line). Most importantly, compared with condition IC, changing to CC increases the estimated cloglog -hazard in Bin 275 by .335 units (Parameter 21), changing to CI decreases it by .529 units (Parameter 24), and changing to II decreases it by .569 units (Parameter 16; all $ps < .0001$). While the main effect of CC in Bin 275 decreases over time, the main effects of II and CI increase over time initially (Parameters 17–20, 25–28). Similar to Bin 200 in Experiment 1, in Bin 175 the estimated cloglog -hazard is higher for CC and II than the mixed conditions IC and CI.

The effect of changing the SOA combination from 187/80 to 133/133 is to increase the estimated cloglog -hazard in the early bins (Parameters 32–33). The estimated cloglog -hazard in these bins increases even further when SOA combination is changed to 80/187 (Parameters 29–31). In other words, response occurrence speeds up with increasing P2–T SOA. The remaining interactions between dual-prime (II, CC, CI), SOA, and TIME (Parameters 34–43) mainly reflect a lower cloglog -hazard in Bin 175, especially for SOA combination 80/187.

The first five parameters in Table 4 model the shape of the $\text{logit-}ca(t)$ function in the baseline condition, IC-187/80 in Trial 1,000 (see Fig. 10, row 3, column 4, blue line). Most importantly, compared with condition IC, changing to CC increases the estimated $\text{logit-}ca(t)$ in Bin 275 by .111 units (Parameter 15; $p = .4645$), changing to CI decreases it by 3.678 units (Parameter 19; $p < .0001$), and changing to II decreases it by 2.997 units (Parameter 12; $p < .0001$). These main effects of CC, CI, and II in Bin 275 decrease in magnitude over time (Parameters 12–23).

There is no significant main effect of decreasing the P1–P2 SOA from 187 to 133 (Parameters 28–30), but decreasing it to 80 ms increases the estimated $\text{logit-}ca(t)$ in Bin 200 and decreases the estimated $\text{logit-}ca(t)$ for bins >300 ms (Parameters 24–27). Finally, there are time-invariant interactions between II and SOA combinations (Parameters 31–32), and time-varying interactions between CI and SOA combinations (Parameters 33–39), which all increase the estimated $\text{logit-}ca(t)$ in at least some of the bins.

Summary

Overall, even with a long P1–T SOA, single-prime conditions produced the common finding in response priming experiments: faster and more accurate responses in consistent trials

and slower and less accurate responses in inconsistent trials. Again, when two primes were presented, responses were fastest and most accurate for two consistent primes, and slowest and least accurate for two inconsistent primes. However, under the SOA conditions of Experiment 2, when primes were mixed, we found a clear dominance of the second prime. In particular, CI was almost identical to II, in both response times and error rates. Similarly, IC was almost identical to CC in error rate, but slightly slower. In other words, in both RT and ER the second prime seemed to dominate the response, yet an inconsistent first prime could still slow down response times. This might reflect early response competition due to conflicting prime information in mixed prime conditions.

Again, in order to investigate the temporal dynamics of sequential motor activation, we performed an event history analysis. Altogether, the findings suggest that with prolonged SOAs: (1) The earliest systematic responses were predominantly controlled by the second prime, (2) the slowest responses were controlled by the target, (3) overt responses to the first prime were extremely rare; however, (4) the first prime was able to slow down initial response hazards in mixed prime conditions compared with conditions with identical primes.

General discussion

The goal of the current study was to investigate (a) whether sequential primes initiate immediate sequential response activation or integrate in a buffer before a response is emitted, (b) whether sequential response activation at short SOAs conforms to the rapid-chase criteria, and (c) how the influence of the first prime changes when the SOAs are prolonged so that participants have to safeguard against early errors from inconsistent primes.

Event history analysis provides substantial evidence that sequential primes initiate strictly sequential response activation at short SOAs (Experiment 1). First, we found that earliest responses were exclusively controlled by the first prime irrespective of the identity or onset time of the second prime, that intermediate responses were influenced by the second prime (with the magnitude and timing of this effect depending on the second prime's onset time), and that only late responses were controlled by the actual target. This strongly supports the notion of feedforward and sequential activation, and is in line with previous findings that first responses are exclusively triggered by prime properties, independent of the target, and only later responses are influenced by target properties (Eimer & Schlaghecken, 1998; Grainger et al., 2013; Schmidt & Schmidt, 2010; Schmidt, 2002; Schmidt & Schmidt, 2009; Vath & Schmidt, 2007). Thus, the data adhere to the chase criteria proposed by T. Schmidt (2014): (1) The first prime

rather than the target or subsequent prime signals determine the onset and initial direction of the response; (2) target and second prime influence the response before it is completed; (3) movement kinematics initially depend on characteristics of the first prime only and are independent of all characteristics of target and subsequent prime signals.

Second, as mentioned before (see Multiple-Prime Paradigm section), a simple feedforward-sweep model seems to account very well for response priming effects at short SOAs (up to 100 ms). However, priming effects at longer SOAs are more plausibly carried by the content of a response buffer that carries information from both primes but is dominated by the second one (Grainger et al., 2013). This notion is supported by our findings. When SOAs were long (Experiment 2), we found that early systematic responses were predominantly triggered by the second prime's identity and that later responses were triggered by the target's identity. In contrast to the first experiment, we found that overt responses to the first prime were extremely rare, but we identified an indirect, covert influence of the first prime's identity on motor response activation, as there were signs of response competition due to conflicting prime information in mixed prime conditions. This strongly suggests that information of a first prime was indeed maintained in a memory buffer and could influence the response that is otherwise dominated by the second prime. Future computational models of decision-making (cf. Mattler & Palmer, 2012; Schmidt & Schmidt, 2018; Schubert, Palazova, & Hutt, 2013; Ulrich, Schröter, Leuthold, & Birngruber, 2015; Vorberg et al., 2003) should test whether the observed hazard and conditional accuracy functions can be simulated with or without a memory buffer.

It is important to point to the different insights that can be gained from an ANOVA on mean correct RT, versus an event history analysis. First, in accordance with the conclusions from previous findings (Breitmeyer & Hanif, 2008; Grainger et al., 2013), a second prime dominates the priming effect in mean correct RTs and ER, at least for short interprime intervals. However, the event history analysis showed that the first prime dominated the motor response in the earliest bins (Experiment 1). Thus, in contrast to Breitmeyer and Hanif (2008), this suggests that the second prime does not update and override the effects of the first prime, but that both prime-triggered motor responses are competing in mixed conditions, and under the right SOA setup, even the first prime is able to dominate the motor response.

Second, when SOAs were long (Experiment 2), we found an even clearer dominance of the second prime since participants seemed to safeguard against early errors provoked by the first prime by waiting out the target. Although response accuracy was entirely dominated by the second prime, RT analysis revealed that an inconsistent first prime in IC conditions could still slow down responses compared with consistent-only conditions. The event history analysis confirmed that early hazards were lower in mixed prime

conditions compared with identical prime conditions. We propose that this is due to response competition created by conflicting prime information. Further, this effect increased with prolonged SOAs between primes, again reflecting a reduced dominance of the second prime due to an increase of the first prime's effect. Thus, the first prime can still influence the motor response with long SOAs. Note that these systematic differences between SOA ranges imply that long and short SOAs should not be mixed within the same experiment, since the presence of long SOAs would enforce a strategy of waiting out the target even in trials where the SOA is short (Schmidt, Haberkamp, & Schmidt, 2011).

When we compare SOA combination 187/80 of Experiment 2 with SOA combination 27/80 of Experiment 1, we see that P2 dominated behavior more in the former than in the latter condition. Therefore, in line with Grainger et al. (2013), we propose that the first prime can influence the response only out of a memory buffer in Experiment 2, since prime information seemed to be kept active for a prolonged period of time without activating a response on its own.

Importantly, we designed our lollipop stimulus in such a way as to minimize masking effects (no spatial overlap) and Simon/flanker effects (prime information is presented at both sides of the target). However, it is unclear if active response inhibition was playing a role in the generation of the behavior in Experiment 2. Panis and Schmidt (2016) and Schmidt, Hauch, and Schmidt (2015) showed that a second stimulus can trigger active and selective inhibition of the response triggered by a first stimulus, within about 360 ms. For example, for SOA combinations 133/133 and 187/80 we see that CI has a lower conditional accuracy than II for bins after 225 ms. This might be caused by active inhibition of the first compatible response, creating an even stronger activation of the incompatible response channel in condition CI than in II. Future modeling studies should investigate this issue further.

More generally, the information obtained from an event history analysis can provide strong constraints for computational models of the underlying sensory integration, decision, and cognitive control processes (Panis, Moran, Wolkersdorfer, & Schmidt, 2020). For example, existing models differ in (a) whether sensory integration is perfect (e.g., the drift-diffusion model; Ratcliff & Rouder, 1998) or leaky (e.g., the leaky competing accumulator model of Usher & McClelland, 2001), (b) whether the response criterion is fixed (e.g., Poisson accumulator models; Schmidt & Schmidt, 2018; Schubert et al., 2013; Vorberg et al., 2003) or variable (the urgency gating model of Cisek, Puskas, & El-Murr, 2009), and (c) whether classic computational principles (e.g., the Bayesian reader model of Norris, 2006) or dynamic principles (e.g., the dynamic field theory of Schöner, Spencer, & The DFT Research Group, 2016) are used (e.g., see Carland, Thura, & Cisek, 2019, for a discussion of these issues). Comparing empirical and simulated data from such models using event history analysis will allow

future studies to better select between and validate the different computational models available in the literature.

While behavioral experiments are informative, they allow only indirect inferences about the underlying neural correlates. Ultimately, one wants to complement the behavioral data with physiological data such as EEG, fMRI, single-cell data, and so forth. Note that hazard modeling allows incorporating time-varying explanatory covariates such as heart rate, EEG signal amplitude, and gaze location (Allison, 2010), which is useful for cognitive psychophysiology (Meyer, Osman, Irwin, & Yantis, 1988).

In summary, the current study provides substantial evidence that sequential primes actually initiate sequential response activation, and that this sequence conforms to the chase criteria at short SOAs. However, when SOAs are prolonged participants have to delay their responses, the first prime seems to influence responses out of a memory buffer.

Open practices statement All data, materials and analyses are available from the authors upon request. All independent and dependent variables are identified. Preregistration was not employed.

Acknowledgements Part of this work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projektnummer PA 2947/1-1 (to S.P.).

Funding Information Open Access funding provided by Projekt DEAL.

Compliance with ethical standards

Declarations of interest None.

Author contributions We thank Stefan Hellrigel, Larissa Leist, Stefanie Scülfert, and Thomas Wegner for their contributions in designing the paradigm.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology*, 13, 61–98. <https://doi.org/10.2307/270718>
- Allison, P. D. (2010). *Survival analysis using SAS: A practical guide* (2nd ed.). Cary, NC: SAS Institute Inc.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.
- Breitmeyer, B. G., & Hanif, W. (2008). “Change of mind” within and between nonconscious (masked) and conscious (unmasked) visual processing. *Consciousness and Cognition*, 17, 254–266. <https://doi.org/10.1016/j.concog.2007.08.001>
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, 36(2/3), 96–107. [https://doi.org/10.1016/S0165-0173\(01\)00085-6](https://doi.org/10.1016/S0165-0173(01)00085-6)
- Burle, B., Vidal, F., Tandonnet, C., & Hasbroucq, T. (2004). Physiological evidence for response inhibition in choice reaction time tasks. *Brain and Cognition*, 56(2), 153–164. <https://doi.org/10.1016/j.bandc.2004.06.004>
- Carland, M. A., Thura, D., & Cisek, P. (2019). The urge to decide and act: Implications for brain function and dysfunction. *The Neuroscientist*, 25(5), 491–511. <https://doi.org/10.1177/1073858419841553>
- Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: The urgency-gating model. *Journal of Neuroscience*, 29(37), 11560–11571. <https://doi.org/10.1523/JNEUROSCI.1844-09.2009>
- Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), 369–382. <https://doi.org/10.1177/2F0267658312443651>
- Eimer, M., & Schlaghecken, F. (1998). Effects of masked stimuli on motor activation: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 24(6), 1737–1747. <https://doi.org/10.1037/0096-1523.24.6.1737>
- Eisenhart, C. (1962). Realistic evaluation of the precision and accuracy of instrument calibration systems. In H. H. Ku (Ed.), *Precision measurement and calibration* (pp. 21–48). Washington, DC: National Bureau of Standards.
- Fernández-López, M., Marcet, A., & Perea, M. (2019). Can response congruency effects be obtained in masked priming lexical decision? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(9), 1683–1702. <https://doi.org/10.1037/xlm0000666>
- Ferrand, L., & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked non-word priming. *The Quarterly Journal of Experimental Psychology Section A*, 45(3), 353–372. <https://doi.org/10.1080/2F02724989208250619>
- Grainger, J. E., Scharnowski, F., Schmidt, T., & Herzog, M. H. (2013). Two primes priming: Does feature integration occur before response activation? *Journal of Vision*, 13(8), 19. <https://doi.org/10.1167/13.8.19>
- Jaśkowski, P., Skalska, B., & Verleger, R. (2003). How the self controls its “automatic pilot” when processing subliminal information. *Journal of Cognitive Neuroscience*, 15(6), 911–920. <https://doi.org/10.1162/089892903322370825>
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). *What's new in Psychtoolbox-3?* *Perception*, 36(14), 1–16.
- Klotz, W., & Neumann, O. (1999). Motor activation without conscious discrimination in metacontrast masking. *Journal of Experimental Psychology: Human Perception and Performance*, 25(4), 976–992. <https://doi.org/10.1037/0096-1523.25.4.976>
- Klotz, W., & Wolff, P. (1995). The effect of a masked stimulus on the response to the masking stimulus. *Psychological Research*, 58(2), 92–101. <https://doi.org/10.1007/BF00571098>
- Lamme, V. A.F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11), 571–579. [https://doi.org/10.1016/S0166-2236\(00\)01657-X](https://doi.org/10.1016/S0166-2236(00)01657-X)

- Li, X., Liang, Z., Kleiner, M., & Lu, Z.-L. (2010). RTbox: A device for highly accurate response time measurements. *Behavior Research Methods*, 42, 212–225. <https://doi.org/10.3758/BRM.42.1.212>
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York, NY: Oxford University Press.
- Mattler, U., & Palmer, S. (2012). Time course of free-choice priming effects explained by a simple accumulator model. *Cognition*, 123(3), 347–360. <https://doi.org/10.1016/j.cognition.2012.03.002>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological psychology*, 26(1/3), 3–67. doi: [https://doi.org/10.1016/0301-0511\(88\)90013-0](https://doi.org/10.1016/0301-0511(88)90013-0)
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, 14(2), 247–279. [https://doi.org/10.1016/0010-0285\(82\)90010-X](https://doi.org/10.1016/0010-0285(82)90010-X)
- Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, 113(2), 327. Retrieved from <https://doi.org/10.1037/0033-295X.113.2.327>
- 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 and Performance*, 40(1), 129–144. <https://doi.org/10.1037/a0032949>
- Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). 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. *Attention, Perception, & Psychophysics*, 1–26. Advance online publication. <https://doi.org/10.3758/s13414-019-01897-z>
- 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. https://doi.org/10.1162/jocn_a_00998
- 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. <https://doi.org/10.1080/13506285.2017.1330790>
- Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization and identification of fragmented object outlines. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 661–687. <https://doi.org/10.1037/a0013547>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- 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. <https://doi.org/10.1162/0898929053279513>
- 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/>
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347–356. <https://doi.org/10.1111/1467-9280.00067>
- Rosenbaum, D. A. (1983). The movement precuing technique: Assumptions, applications, and extensions. *Advances in Psychology*, 12, 231–274. [https://doi.org/10.1016/S0166-4115\(08\)61994-9](https://doi.org/10.1016/S0166-4115(08)61994-9)
- Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron*, 20(2), 185–195. <https://doi.org/10.1126/science.281.5380.1188>
- Schmidt, F., Haberkamp, A., & Schmidt, T. (2011). Dos and don'ts in response priming research. *Advances in Cognitive Psychology*, 7, 120–131. <https://doi.org/10.2478/v10053-008-0092-2>
- Schmidt, F., & Schmidt, T. (2010). Feature-based attention to unconscious shapes and colors. *Attention, Perception, & Psychophysics*, 72(6), 1480–1494. <https://doi.org/10.3758/APP.72.6.1480>
- Schmidt, T. (2002). The finger in flight: Real-time motor control by visually masked color stimuli. *Psychological Science*, 13(2), 112–118. <https://doi.org/10.1111/2F1467-9280.00421>
- Schmidt, T. (2014). *Behavioral criteria of feedforward processing in rapid-chase theory: Some formal considerations*. Retrieved from <https://arxiv.org/abs/1405.5795>
- Schmidt, T., Haberkamp, A., Veltkamp, G. M., Weber, A., Seydell-Greenwald, A., & Schmidt, F. (2011). Visual processing in rapid-chase systems: Image processing, attention, and awareness. *Frontiers in Psychology*, 2, 169. <https://doi.org/10.3389/fpsyg.2011.00169>
- Schmidt, T., Hauch, V., & Schmidt, F. (2015). Mask-triggered thrust reversal in the negative compatibility effect. *Attention, Perception, & Psychophysics*, 77(7), 2377–2398. <https://doi.org/10.3758/s13414-015-0923-4>
- Schmidt, T., Niehaus, S., & Nagel, A. (2006). Primes and targets in rapid chases: Tracing sequential waves of motor activation. *Behavioral Neuroscience*, 120(5), 1005–1016. <https://doi.org/10.1037/0735-7044.120.5.1005>
- Schmidt, T., & Schmidt, F. (2009). Processing of natural images is feedforward: A simple behavioral test. *Attention, Perception, & Psychophysics*, 71(3), 594–606. <https://doi.org/10.3758/APP.71.3.594>
- Schmidt, T., & Schmidt, F. (2018). *An accumulator model for primes and targets with independent response activation rates: Basic equations for average response times*. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1804/1804.08513.pdf>
- Schöner, G., Spencer, J., & The DFT Research Group. (2016). *Dynamic thinking: A primer on dynamic field theory*. New York, NY: Oxford University Press.
- Schubert, T., Palazova, M., & Hutt, A. (2013). The time course of temporal attention effects on nonconscious prime processing. *Attention, Perception, & Psychophysics*, 75(8), 1667–1686. <https://doi.org/10.3758/s13414-013-0515-0>
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.
- Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of small-N designs. *Psychonomic Bulletin and Review*, 25, 2083–2101. <https://doi.org/10.3758/s13423-018-1451-8>
- Ulrich, R., Schröter, H., Leuthold, H., & Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognitive Psychology*, 78, 148–174. <https://doi.org/10.1016/j.cogpsych.2015.02.005>
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3), 550. <https://doi.org/10.1037/0033-295X.108.3.550>
- VanRullen, R., & Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process. *Journal of Cognitive Neuroscience*, 15(2), 209–217. <https://doi.org/10.1162/089892903321208141>
- Vath, N., & Schmidt, T. (2007). Tracing sequential waves of rapid visuomotor activation in lateralized readiness potentials. *Neuroscience*, 145(1), 197–208. <https://doi.org/10.1016/j.neuroscience.2006.11.044>
- Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of the National Academy of Sciences of the United States of America*, 100(10), 6275–6280. <https://doi.org/10.1073/pnas.0931489100>
- Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, 8(2), 227–233. [https://doi.org/10.1016/S0959-4388\(98\)80144-X](https://doi.org/10.1016/S0959-4388(98)80144-X)
- Zuur, A. F., & Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*, 7(6), 636–645. <https://doi.org/10.1111/2041-210X.12577>

**Response inhibition in the Negative Compatibility Effect in the
absence of inhibitory stimulus features⁴**

⁴ *Reproduced with permission from De Gruyter Brill under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>).*

Research Article

Thomas Schmidt*, Sven Panis, Maximilian P. Wolkersdorfer, Dirk Vorberg

Response inhibition in the Negative Compatibility Effect in the absence of inhibitory stimulus features

<https://doi.org/10.1515/psych-2022-0012>

received February 9, 2022; accepted May 18, 2022.

Abstract: The Negative Compatibility Effect (NCE) is a reversal in priming effects that can occur when a masked arrow prime is followed by an arrow target at a long stimulus-onset asynchrony (SOA). To test the explanation that the NCE is actually a positive priming effect elicited by mask features associated with the prime-opposed response, we devise masks that always point in the same direction as the prime, eliminating all antiprime features. We find large positive priming effects for arrow primes without masks and for arrow masks without primes. When a neutral mask is introduced, priming effects turn negative at long SOAs. In the critical case where the mask is an arrow in the same direction as the prime, the prime does not add to the positive priming effect from the mask shape, but instead strongly diminishes it and induces response errors even though all stimuli point in the same direction. No such feature-free inhibition is seen when arrows are replaced by color stimuli. We conclude that even though response activation by stimulus features plays a role in the NCE, there is a strong inhibitory component (though perhaps not in all feature domains) that is not based on visual features.

Keywords: Negative Compatibility Effect Response Inhibition Response Priming.

1 Introduction

The time course of activation and inhibition of choice responses can be studied in two closely related paradigms using the same stimuli with slightly different timing. *Response priming* (Klotz & Neumann, 1999; Klotz & Wolff, 1995; Vorberg, Mattler, Heinecke, T. Schmidt, & Schwarzbach, 2003) uses a prime and a subsequent target separated by a stimulus onset asynchrony (SOA) of up to 100 ms. The prime either triggers the same response as the target (*consistent prime*) or the opposite response (*inconsistent prime*). Consistent primes speed up responses to the target while inconsistent primes slow down responses and induce fast errors, and this *response priming effect* increases with SOA (Vorberg et al., 2003). Converging evidence from lateralized readiness potentials (Eimer & Schlaghecken, 1998; Leuthold & Kopp, 1998; Vath & T. Schmidt, 2007; Verleger, Jaśkowski, Aydemir, van der Lubbe, & Groen, 2004), response hazards and response time distributions (Panis & T. Schmidt, 2016; F. Schmidt & T. Schmidt, 2014; Wolkersdorfer, Panis, & T. Schmidt, 2020; Panis & T. Schmidt, 2022), pointing trajectories (Brenner & Smeets, 2004; T. Schmidt, 2002; T. Schmidt & F. Schmidt, 2009), and force profiles (F. Schmidt, Weber, & T. Schmidt, 2014) indicates that response priming is based on strictly sequential response activation, first only by the prime, then additionally by the target, probably by pure feedforward processing (T. Schmidt, Niehaus, & Nagel, 2006). Visual masking may be employed to dissociate the prime's visibility from its ability to activate a response (Vorberg et al., 2003; T. Schmidt & Vorberg, 2006; also see Biafora & T. Schmidt, 2020), but the prime can also be employed unmasked or as a flanker to the target, so that response priming is very closely related to the Eriksen flanker paradigm (Eriksen & Eriksen, 1974; F. Schmidt, Haberkamp, & T. Schmidt, 2011; Schwarz & Mecklinger, 1995).

*Corresponding author: Thomas Schmidt, University of Kaiserslautern, Germany, E-mail: thomas.schmidt@sowi.uni-kl.de
Sven Panis, Maximilian P. Wolkersdorfer, University of Kaiserslautern, Germany
Dirk Vorberg, University of Münster, Germany

When SOAs longer than 100 ms are employed (and especially when a mask intervenes between prime and target), the response priming effect can reverse so that consistent trials become slower and more error-prone than inconsistent trials (*Negative Compatibility Effect*, *NCE*; Eimer & Schlaghecken, 1998; see Lingnau & Vorberg, 2005, for the full time-course of response priming and *NCE*). Eimer and Schlaghecken (1998) used lateralized readiness potentials to discover a sequence of three response activations: an initial activation of the prime-related response at a fixed time after prime onset, followed by a transient activation of the opposite (*antiprime*) response that is responsible for the reversal of the priming effect, and finally another activation of the target-related response. The same sequence can be observed in the time course of response hazards in response time distributions (Panis & T. Schmidt, 2016).

Several theories try to explain the emergence of antiprime activation (for a review, see Sumner, 2007). Eimer and Schlaghecken (1998, 2003) proposed that the prime automatically triggers its own inhibition that becomes noticeable as soon as the mask removes “perceptual evidence” for the prime. Consequently, the antiprime response is disinhibited, and the priming effect is reversed. In contrast, Jaśkowski and Przekoracka-Krawczyk (2005) proposed that it is not the prime that triggers its own inhibition; instead, the mask actively inhibits the premature prime-triggered response (Jaśkowski & Przekoracka-Krawczyk, 2005; Jaśkowski, 2007, 2008, 2009; Jaśkowski, Białuńska, Tomanek, & Verleger, 2008). In contrast to Eimer and Schlaghecken’s proposal, this “emergency break” would not require strong subjective masking of the prime, only a sufficiently strong mask signal. This theory correctly predicts that inhibition is time-locked to the mask, not to the prime (Panis & T. Schmidt, 2016). Finally, T. Schmidt, Hauch, and F. Schmidt (2015) used pointing responses to measure the *NCE* for finger movements towards the target in one of ten directions. They showed that even when prime and target are consistent and indicate the same response direction, slow-starting movements (i.e., from the slowest quartile) tend to start in the opposite direction, indicating massive response activation in the vectorial direction opposite to the primed responses.

However, not all theories of the *NCE* acknowledge a role for response inhibition. Lleras and Enns (2004) argue that the prime and target stimuli employed in Eimer and Schlaghecken’s (1998) original paper were double arrows ($>>$, $<<$) while the mask was a superposition of both these stimuli. When the prime was followed by the mask, this was tantamount to the addition of an arrow pointing in the antiprime direction. According to their *object-updating account*, the *NCE* thus simply reflects positive priming of the antiprime response by the corresponding features in the mask, instead of selective response inhibition (Lleras & Enns, 2004). Similarly, Verleger et al. (2004) proposed the “*active mask hypothesis*” that states that the mask will prime a response of its own if it carries task-relevant features. Figure 1 illustrates this logic for the more complex arrow stimuli employed in our experiment. The left-pointing arrow prime is followed by a metacontrast mask that consists of both arrow primes superimposed, giving it a symmetrical inner contour formed like a razorblade (in this example, the outer contour is neutral). According to Lleras and Enns’s hypothesis, this could be construed as adding features of a right-pointing arrow to the left-pointing prime (Fig. 1a), perhaps by processes of amodal completion (Fig. 1b), or else because the small parts of the cutout that are left uncovered by the prime suggest a pointing direction by themselves.

The goal of this paper is to show that the *NCE* can occur in the absence of *any* visual features eliciting an antiprime response. We essentially replicated an experiment by Dirk Vorberg hitherto published only as a conference presentation (Vorberg, 2005). In our experiment, we presented a prime followed by a metacontrast mask at a fixed interval and finally by a target, while the mask-target SOA was varied (Fig. 2a). We compared the time-course of priming over the SOA in four conditions (Fig. 2b). In the *arrow prime, no mask* condition, no mask was presented, and prime and target were left- or right-pointing arrows that either pointed in the same direction (consistent trials) or in opposite directions (inconsistent trials).¹ The *arrow prime, neutral mask* condition had the same stimuli, but a neutral mask was presented between prime and target. These first two conditions can be compared to examine whether addition of a neutral mask turns positive response priming into an *NCE*. In the *neutral prime, arrow mask* condition, the prime was always neutral, but the cutout in the mask could either point in the same direction as the target (consistent trials) or in the opposite direction (inconsistent trials). Finally, in the crucial *arrow prime = arrow mask* condition, prime and target were both arrows pointing in consistent or inconsistent directions, and the cutout in the mask was also an arrow always pointing in the same direction as the prime. These last two conditions can be compared to test the object-updating hypothesis. Because the *arrow prime = arrow mask* condition contains no visual features at all that could elicit an antiprime response, object updating predicts that only positive response priming can occur that should lead to positive priming

¹ This condition was missing from Vorberg’s (2005) experiment.

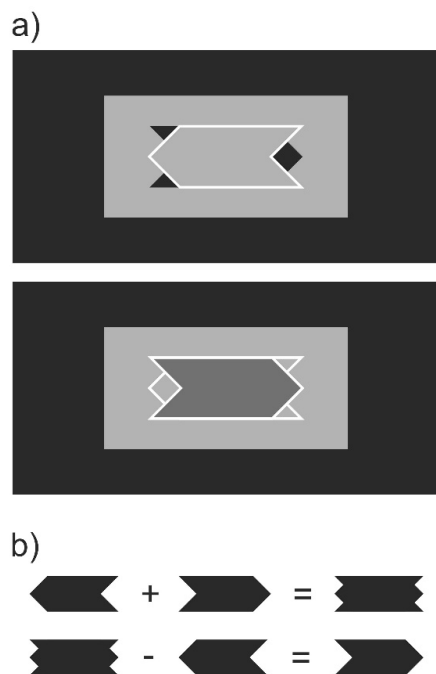


Figure 1: The object-updating argument applied to our stimuli. a) If a left-pointing arrow prime is followed by a mask with a neutral cutout in the center, this could be construed as adding a right-pointing arrow contour to the prime. b) Because the neutral mask contour is the superposition of the two arrow primes (upper panel), presentation of a left-pointing prime might lead the visuomotor system to decompose the neutral mask contour into left- and right-pointing arrow shapes (e.g., by amodal completion), so that the mask would trigger the response opposite to the primed response (lower panel).

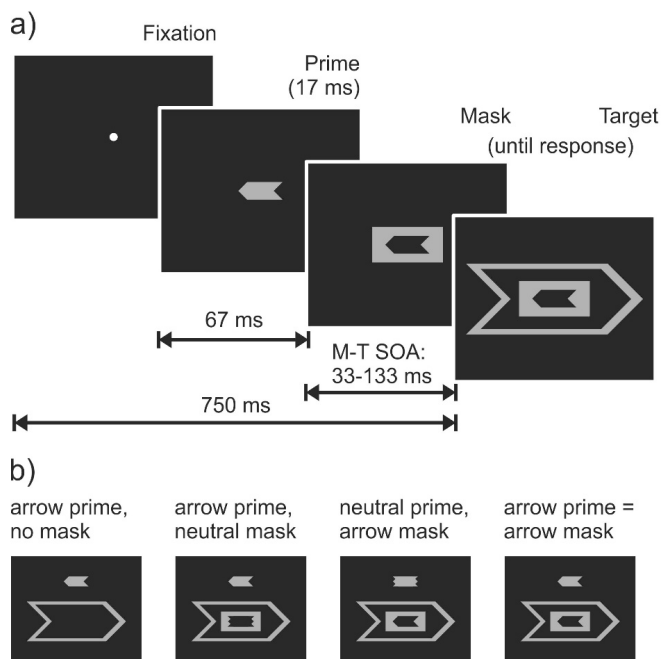


Figure 2: a) Time course of a trial. Note that both mask and target remain on screen until the response is completed. b) The four masking conditions. For better legibility, the prime is drawn above the mask and target. In the experiment, it appeared in the central mask cutout so that their contours were adjacent.

effects at least as large as the one in the *neutral prime, arrow mask* condition. If a reduction of priming is observed instead, we can conclude that an NCE can occur in the absence of features eliciting an antiprime response. We can also conclude that an NCE can occur even if the mask does not remove any perceptual evidence for the prime (as proposed by Eimer and Schlaghecken's self-inhibition account).

2 Methods

Participants. Eight students from Technische Universität Kaiserslautern (four men, three left-handed, age 18-29 years, average age 23.5 years) participated in four 1-h sessions. Their vision was normal or corrected to normal. All participants were naïve to the purpose of the study and received either course credit or a payment of 7 € per hour. Participants gave informed consent and were treated according to the ethical guidelines of the American Psychological Association. After the final session, they were debriefed and received an explanation of the experiment.

Apparatus and Stimuli. The participants were seated in a dimly lit room in front of a color cathode-ray monitor (640 x 480 pixels, retrace rate of 60 Hz) at a viewing distance of approximately 80 cm.

Stimuli. Stimuli were presented in gray (28.4 cd/m²) against a black background (0.03 cd/m²; Fig. 2). Primes were small left- or right-pointing arrows (1.6° x 0.7°) presented at screen center; when the prime was neutral, it was shaped like a left- and a right-pointing prime superimposed. Masks were rectangles (3.3 x 1.4°) with a central cutout shaped like

a neutral prime at the same position as the prime so that the contours were adjacent. When the mask was not neutral, the cutout was shaped like a left- or right-pointing prime. Targets were larger outline arrows ($6.6 \times 2.6^\circ$, line width 0.2°) pointing left or right that shared no contours with the prime or mask.

Procedure. Each trial started with a fixation point (diameter 0.2°) in the center of the screen (Fig. 2a). After a time depending on the mask-target SOA, the prime was presented for 17 ms (one monitor frame). After a prime-mask SOA of 67 ms, either a mask or no mask was presented (depending on condition; see Fig. 2b), followed by the target after a mask-target SOA of either 33, 67, 100, or 133 ms. The time interval between fixation onset and target onset was fixed at 750 ms. Participants were instructed to respond to the direction of the target arrow by pressing a spatially corresponding key on the computer keyboard (“F” or “J”) as quickly and accurately as possible. If the response was incorrect, the message “wrong!” appeared on the screen. If it was slower than 999 ms, the message “too slow” appeared.

Each participant performed four sessions with 21 blocks of 32 trials. The first block of each session was a practice block. All 64 combinations of consistency (depending on condition, between prime and target or mask and target), SOA, masking condition, and target direction occurred randomly and equiprobably over the course of a session.

Data treatment and statistical methods. Practice blocks were not analyzed. Two single sessions were lost due to equipment malfunction.

Dependent variables were response time and error rate, which can both be analyzed for response priming effects. We also looked at conditional accuracy functions, which plot response accuracy as a function of physical time (Panis & Hermens, 2014; Panis, Moran, Wolkersdorfer, & T. Schmidt, 2020).² For response times, repeated-measures analysis of variance (ANOVA) was performed on the trimmed means after error trials (6.96 %) and outliers (responses faster than 100 or slower than 999 ms; 0.17 %) had been removed. Error rates were arcsine-transformed to meet ANOVA requirements. ANOVAs were performed with factors of consistency (*C*), SOA (*S*), and sometimes masking condition (*M*). For clarity, all tests are reported with Huynh-Feldt-corrected *p* values and the original degrees of freedom, and effects are specified by subscripts to the *F* values (e.g., F_{CS} for an interaction of consistency and SOA). Throughout the paper, we report all ANOVA effects significant at $p \leq .05$, with the understanding that *p* values between .01 and .05 should be regarded with caution. We may mention *p* values between .05 and .10 if important to the argument.

In multifactor repeated-measures designs, statistical power is difficult to predict because too many terms are unknown. Instead, we control measurement precision at the level of individual participants in single sessions. We calculate precision as s/\sqrt{r} (Eisenhart, 1969), where *s* is a single participant’s standard deviation in a given cell of the design, and *r* is the number of repeated measures per cell and subject. With $r = 80$ in the current experiment, we expect a precision of about 6.7 ms in response times (assuming individual SDs around 60 ms) and at most 2.8 percentage points in error rates (assuming the theoretical binomial maximum SD of .50). Precision thus exceeds our previous recommendations for response priming studies ($r = 60$; F. Schmidt, Haberkamp, & T. Schmidt, 2011).

In addition, we used two benchmark datasets that measure response priming effects in the color domain (red vs. green) and the shape domain (circle vs. square). These datasets independently vary prime and target strength (in terms of color or luminance contrast), prime-target SOA, and prime-target consistency. This results in sixteen priming functions per domain, one for each combination of prime and target strength; they are based on eight observers and 60 repetitions per observer and condition, which implies a precision of 7.7 ms in response times. For the shape domain, all sixteen SOA \times Consistency interactions have $11.3 \leq F(3, 21) \leq 78.8$, $.614 \leq \eta_p^2 \leq .916$, and noncentrality parameters between 33.8 and 236.4. All have $p < .001$ and an observed power $\geq .997$ (the critical noncentrality parameter to achieve a power of .95 at $\alpha = .05$ is about 21).

3 Results

Two-factorial analyses (with factors SOA and Consistency) for each masking condition are reported in the order in which they appear in Fig. 3 (left to right). These analyses confirm that there are large priming effects in all masking conditions that either increase with mask-target SOA (*neutral prime, arrow mask*) or slightly decrease with prime-target SOA (*arrow-*

² Average response times can conceal vital information about the time-course of an effect. Therefore, we used Event History Analysis to calculated response hazards, which lead to the same conclusions as the average response times presented here.

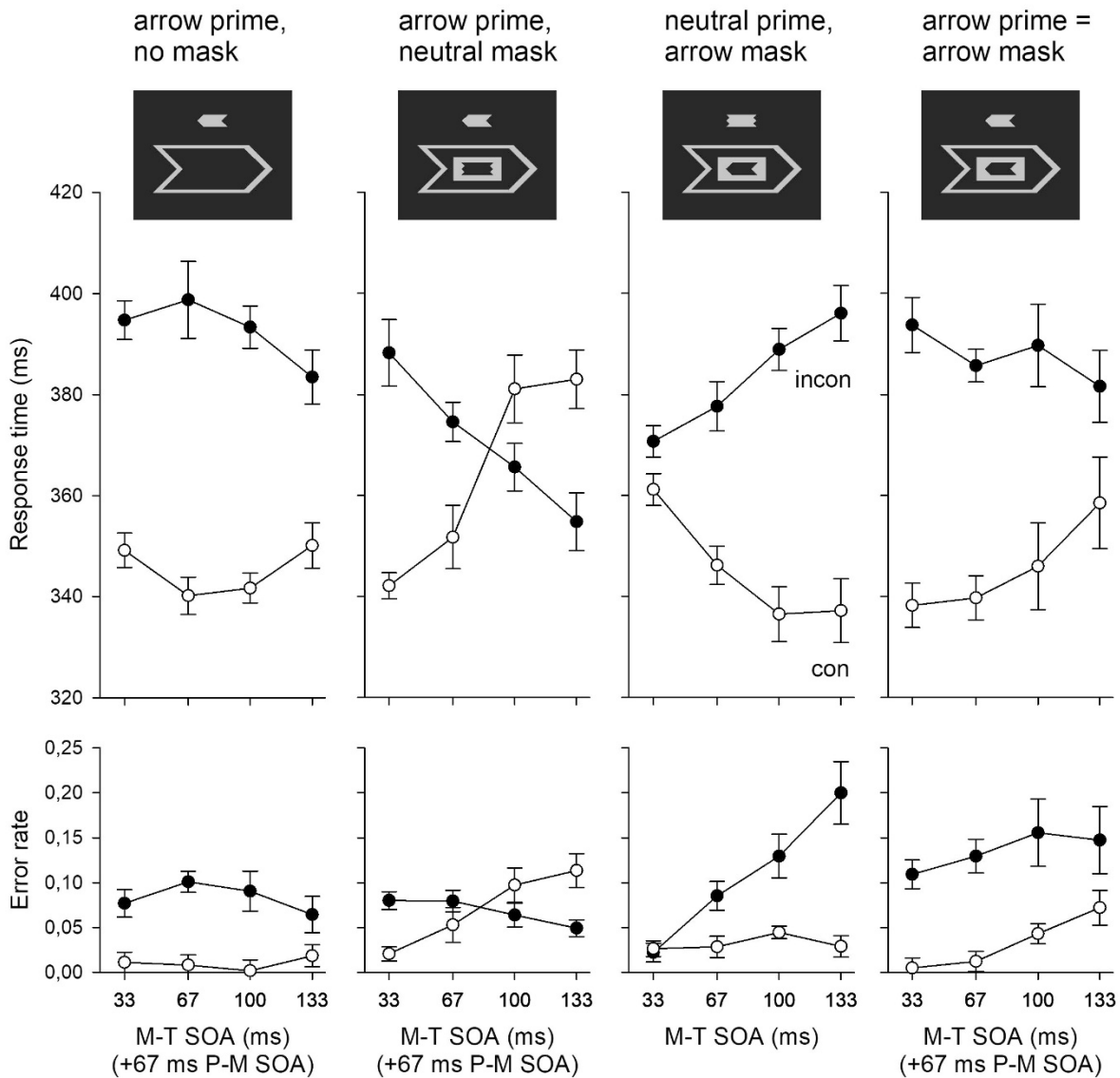


Figure 3: Response times (upper panels) and error rates (lower panels) in the four masking conditions. Because the SOA axis specifies the mask-target SOA, a constant 67 ms must be added in some conditions to yield the effective prime-target SOA. Error bars are corrected for intersubject variance (Bakeman & McArthur, 1996).

prime, no mask; *arrow prime = arrow mask*). The crucial exception is the *arrow prime, neutral mask* condition, which starts with a large priming effect at the shortest SOA that reverses with increasing SOA. In all analyses, remember that Figure 3 plots the data as functions of mask-target SOA and that in some conditions a constant 67 ms must be added to yield the effective prime-target SOA.

In each masking condition, responses were faster in consistent than in inconsistent trials, except for the *arrow prime, neutral mask* condition where the crossed interaction eliminated the main effect; $F_c(1, 7) = 47.10, 0.81, 44.98$, and 16.95 ; $p < .001, .399, < .001, .004$. In each pattern, there was a significant interaction of Consistency and SOA, $F_{cs}(3, 21) = 4.93, 22.51, 13.10$, and 3.80 ; $p = .010, < .001, .002, .039$. This indicates that the *arrow prime, neutral mask* condition is not simply devoid of priming but that the SOA functions for consistent and inconsistent trials are shifted past each other to reverse the effect. The main effect of SOA was nonsignificant in all patterns, $F_s(3, 21) = 1.43, 2.44, 0.80$, and 1.16 ; $p = .265, .133, .506, .359$. The error rates give a similar picture. In each pattern, more errors occurred in inconsistent than in consistent trials, except for the *arrow prime, neutral mask* condition where the crossed interaction

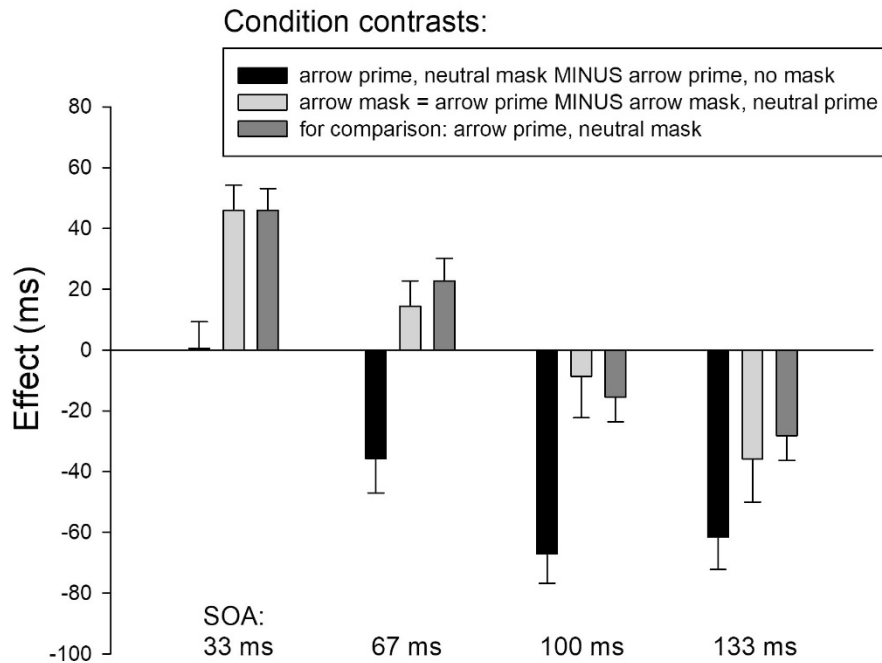


Figure 4: Differences in priming effects at the different mask-target SOAs. Black bars: contrast between arrow prime, neutral mask and arrow prime, no mask conditions. Light grey bars: contrast between arrow prime = arrow mask and neutral prime, arrow mask conditions. Dark grey bars: priming effects in the arrow prime, neutral mask condition are shown for comparison.

eliminated the main effect just as it did in the response times; $F_c(1, 7) = 22.89, 0.03, 21.01, \text{ and } 15.04$; $p < .001, = .869, = .003, = .006$. The interaction of Consistency and SOA was significant in the *arrow prime, neutral mask* and the *neutral prime, arrow mask* conditions only, $F_{cxs}(3, 21) = 2.72, 7.29, 9.41, \text{ and } 3.28$; $p = .117, = .002, < .001, = .053$. The main effect of SOA was significant in all conditions except *arrow prime, no mask*, $F_s(3, 21) = 0.70, 3.71, 16.03, \text{ and } 8.23$; $p = .517, = .028, < .001, = .001$. Note that both response times and error rates show comparable effects. In the following, we contrast pairs of masking conditions to capture the full magnitude of the NCE and to answer the question whether response inhibition can occur without inhibitory stimulus features.

How large is the NCE? To answer this question, we compare the *arrow prime, no mask* condition with the *arrow prime, neutral mask* condition, which only differ in whether a neutral mask is present (Fig. 3, black bars in Fig. 4). The *arrow prime, no mask* condition shows the picture expected for response priming with long prime-target SOAs (100-200 ms): The priming effect is already at a maximum at the shortest mask-target SOA and gently tapers off at longer SOAs. The addition of a neutral mask changes this pattern drastically: the priming effect at the shortest SOA is still comparable to the one in the *arrow prime, no mask* condition, but then it reverses with SOA until consistent responses are actually slower and more error-prone than inconsistent ones. This comparison shows that the NCE is actually much larger than the manifest 30-ms reversal: It is as large as the reversal *plus* the large positive effect it eliminated, and thus larger than the positive response priming effect itself. A repeated-measures ANOVA with factors of Masking Condition (contrasting the two conditions), Consistency, and SOA indicated that on average, responses were slower in inconsistent than in consistent trials, $F_c(1, 7) = 19.20, p = .003$, and that this priming effect increased with SOA, $F_{cxs}(3, 21) = 13.94, p < .001$. Masking Condition interacted with Consistency, $F_{mxc}(1, 7) = 37.79, p < .001$, and with SOA, $F_{mcs}(3, 21) = 3.11, p = .048$, and generated a triple interaction, $F_{mxcxs}(3, 21) = 27.81, p < .001$, all reflecting the dramatic changes in priming effects in the presence of the neutral mask. All other effects were nonsignificant. The error rates yielded a similar picture. On average, more errors occurred in inconsistent than in consistent trials, $F_c(1, 7) = 14.57, p = .007$, but this priming effect decreased from positive to slightly negative values because of the reversal in the *arrow prime, neutral mask* condition, $F_{cxs}(3, 21) = 7.19, p = .002$, where error rates were also higher overall, $F_m(1, 7) = 13.25, p = .008$, and tended to increase a bit faster with SOA, $F_{mcs}(3, 21) = 2.94, p = .058$. Importantly, there was a marked interaction of Consistency and Masking Condition, $F_{mxc}(1,$

7) = 14.29, $p = .007$, and also a triple interaction, $F_{MxCxS}(3, 21) = 4.36$, $p = .039$, reflecting the reversal of the priming effect when a neutral mask was present.

We also looked at the conditional-accuracy functions, which plot response accuracy in consistent and inconsistent trials as a function of response time bins (Fig. 5; Panis & Hermens, 2014). As shown previously, positive response priming in the *arrow prime, no mask* condition was characterized by the fact that early responses go in the direction of the prime: almost all of them are correct in consistent trials but incorrect in inconsistent trials. As response times get longer, consistent trials maintain their high level of accuracy while accuracy in inconsistent trials increases until it reaches the same level as in consistent trials (Panis & T. Schmidt, 2016). This pattern of early errors in inconsistent trials only illustrates the strictly sequential response control by primes and targets (T. Schmidt et al., 2006). The *arrow prime, neutral mask* condition, on the other hand, is surprisingly characterized by *late* errors in *consistent* trials. The consistent trials start at almost perfect accuracy in the fastest responses and then experience a dip in accuracy for RTs around 300 ms that becomes progressively larger with SOA (Panis & T. Schmidt, 2016). This process continues until at the longest SOAs, consistent and inconsistent trials have completely changed their accustomed roles: while accuracy in inconsistent trials is uniformly high across the entire RT distribution, consistent trials start from only 40 % accuracy in the fastest responses and then gradually reach almost perfect accuracy as responses get slower. This pattern is indicative of strong response inhibition directed against the primed response (Panis & T. Schmidt, 2016; T. Schmidt, Hauch, & F. Schmidt, 2015).

Is there an NCE in the absence of inhibitory stimulus features? To answer this question, we compare response times in the *neutral prime, arrow mask* condition with the *arrow prime = arrow mask* condition, which only differ in whether the prime is neutral or has the same direction as the mask (Fig. 3). It is noteworthy that contrasting the priming effects between these two conditions is very similar to the time course of priming in the *arrow prime, neutral mask* condition (Fig. 4, grey bars). The *neutral prime, arrow mask* condition shows the picture expected for response priming if the mask played the role of the prime and the neutral prime itself did not activate a response. Consistent with this expectation, the priming effect is small at the shortest mask-target SOA and increases with SOA to finally reach the full size seen in the *arrow prime, no mask* condition at a prime-target SOA of 133 ms. To evaluate the *arrow mask = arrow prime* condition, assume for a moment that no NCE can occur because there are no stimulus features directed against the primed response. Because we already know that both the arrow prime and the arrow mask can generate a positive priming effect by themselves, we would predict that adding an arrow mask in the same direction as the arrow prime could only make the priming effect more positive. Instead, the arrow mask *diminishes* the priming effect at longer SOAs and provokes errors in consistent trials. A repeated-measures ANOVA with factors of Masking Condition (contrasting the two conditions), Consistency, and SOA indicated that on average, responses were slower in inconsistent than in consistent trials, $F_c(1, 7) = 29.34$, $p = .001$. However, there was a triple interaction of Consistency with SOA and Masking Condition, $F_{MxCxS}(3, 21) = 24.04$, $p < .001$, indicating the marked difference in time-courses of the priming effects. Even though we saw above that both patterns had significant Consistency x SOA interactions when analyzed on their own, these two interactions virtually cancelled each other in the three-factorial analysis, $F_{cxs}(3, 21) = 1.33$, $p = .297$. All other effects were nonsignificant. Error rates gave a similar picture. On average, more errors occurred in inconsistent than in consistent trials, $F_c(1, 7) = 27.15$, $p = .001$, and error rates increased with SOA, $F_s(3, 21) = 20.59$, $p < .001$. However, there was a marked triple interaction, $F_{MxCxS}(3, 21) = 21.56$, $p < .001$, showing that priming effects increased in the *neutral prime, arrow mask* condition but decreased in the *arrow prime = arrow mask* condition. Conditional accuracy functions (Fig. 5) showed that errors were restricted to fast responses in inconsistent trials, where they increased strongly with SOA (Panis & T. Schmidt, 2016).

Overall, the *arrow prime = arrow mask* condition was similar to the *arrow prime, no mask* condition. The two conditions can be contrasted as well because they only differ by the presence of the arrow mask. On average, responses were faster in consistent than in inconsistent trials, $F_c(1, 7) = 28.50$, $p = .001$, and this priming effect decreased with SOA, $F_{cxs}(3, 21) = 7.26$, $p = .005$, but there were no significant effects involving the factor Mask Condition (contrasting the two conditions). However, the error rates gave a slightly different picture. On average, more errors occurred in inconsistent than in consistent trials, $F_c(1, 7) = 21.60$, $p = .002$, and at longer SOAs, $F_s(3, 21) = 4.86$, $p = .010$, and the priming effect slightly decreased with SOA, $F_{cxs}(3, 21) = 8.58$, $p = .004$. But more errors occurred in the *arrow prime = arrow mask* condition, $F_M(1, 7) = 15.61$, $p = .006$, and the increase in error rates with SOA was steeper, $F_{MxS}(3, 21) = 5.83$, $p = .005$. Interestingly, those errors also occurred in consistent trials, indicating activation opposite to the primed response. This was confirmed by the conditional accuracy functions (Fig. 5): at long SOAs, consistent trials experienced

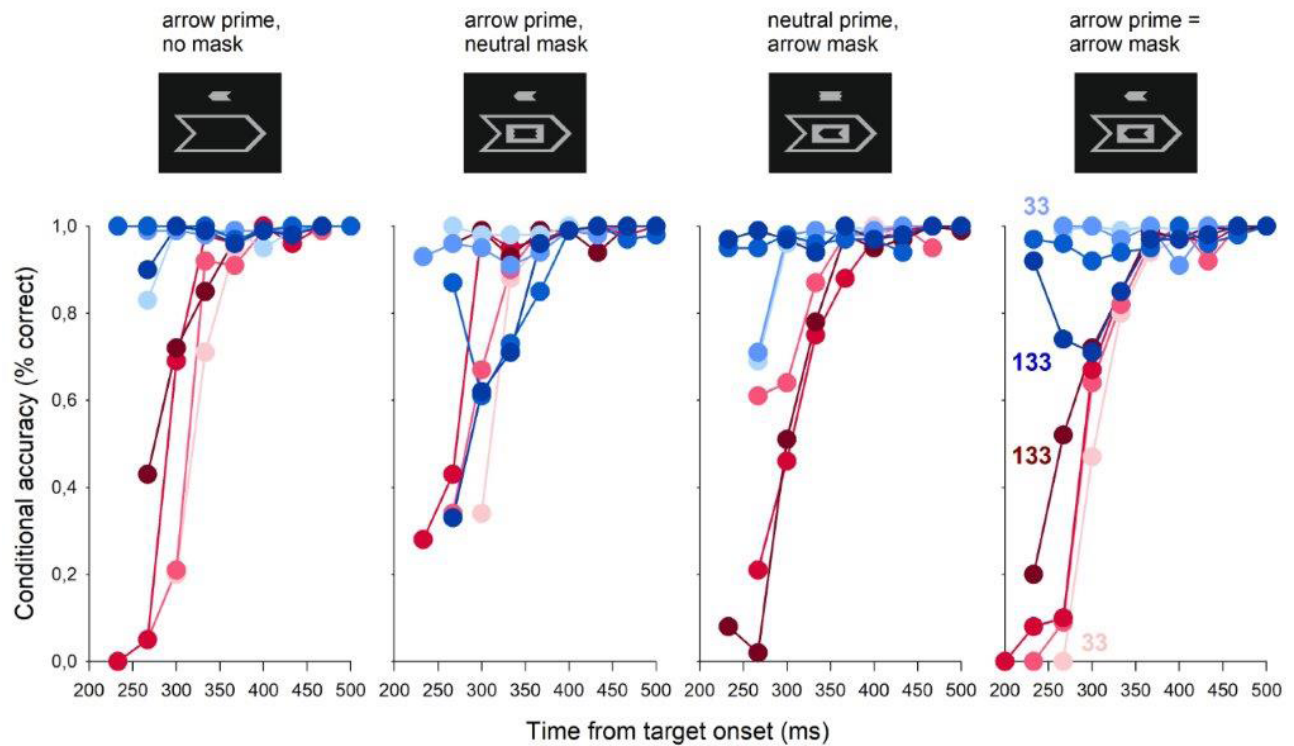


Figure 5: Conditional accuracy functions (accuracy plotted as a function of response time). Blue colors denote consistent conditions, red colors inconsistent conditions. SOA increases from light to dark color shading. Only the RT range from 200 to 600 ms was considered; bin size is 33 ms, and data points are plotted whenever based on a total of at least 15 responses (RTs longer than 500 ms fall short of this criterion).

the characteristic dip in accuracy for response time around 300 ms that seems to be a hallmark of the NCE (Panis & T. Schmidt, 2016). In these responses, participants produced error rates of up to 20 % even though all stimulus features pointed in the same consistent direction.

4 Discussion

In our discussion, we focus on three important theoretical issues. First, the response inhibition processes underlying the NCE are much larger than suggested by the moderate reversal in priming effects in mean response times. Second, the inhibition occurs even if the mask retains all perceptual information about the orientation of the prime. Third, inhibition can occur in the absence of antiprime features, i.e., without any stimulus information that could activate the response opposite to the prime.

The NCE is large -- very large. The first important lesson is that in order to understand the true magnitude of the NCE, it is strongly misleading to only focus on the manifest reversal observable in the *arrow prime, neutral mask* condition. If we define the NCE only in terms of negative-signed priming effects in that one condition, we forget that without a mask we could expect a large *positive* priming effect. The full magnitude of response inhibition is only revealed when the *arrow prime, neutral mask* condition is contrasted with the *arrow prime, no mask* condition. Then it becomes clear that the response inhibition following the mask is even stronger than the response activation following the prime, because the inhibition is not only able to abolish the large response priming effect, but to reverse it.³ From this view, it is obviously not a reversal per se that we are looking for, but *any* reduction in response priming brought on by the mask,

³ In this analysis, we assume that the mask does not diminish the positive response priming effect. This possibility is discussed in F. Schmidt et al. (2011) and awaits systematic evaluation.

even if it fails to lead to a sign reversal (Jaśkowski & Przekoracka-Krawczyk, 2005; Jaśkowski, 2007a, b, 2009; Jaśkowski et al., 2008). Given that response priming can yield large response-time effects (sometimes more than 100 ms), the NCE is probably one of the largest effects known in reaction times of speeded choice responses.

Response inhibition occurs if the mask retains all perceptual evidence of the prime. In the *arrow prime = arrow mask* condition, the mask leads to metacontrast masking of the prime but at the same time retains all the relevant shape information. We can therefore reject Eimer and Schlaghecken's (1998) original proposal that inhibition occurs because the mask suppresses perceptual information about prime shape. Our findings are in line with those by Jaśkowski and Przekoracka-Krawczyk (2005), who compared different types of pattern masks and found that efficient masking of the prime is not a prerequisite of the NCE. In their *mask-triggered inhibition* account, they propose that the mask serves to interrupt processing of the prime and to trigger the inhibition of the primed responses. This inhibition is supposed to be stronger when the mask carries response-relevant features.

In arrow stimuli, antiprime features are not necessary for response inhibition. We have seen that the *arrow prime, no mask* condition generates a positive priming effect that becomes a little weaker with increasing SOA. From this condition, the critical *arrow prime = arrow mask* condition differs only in the presence of the arrow mask. However, the arrow mask itself generates a large, strongly increasing priming effect as seen in the *neutral prime, arrow mask* condition, just as expected if the arrow features in the mask prime a response while the neutral prime does not. The increase of response priming with SOA is also expected because with a neutralized prime, the effective SOA takes place between the mask and the target and is in the range where positive response priming is expected (here, 33-133 ms). Therefore, we would expect that an arrow prime and arrow mask combined should guarantee strong, increasing response priming effects in the *arrow prime = arrow mask* condition.

But instead, we see a pattern very similar to the *arrow prime, no mask* condition, with large priming effects at the shortest SOA and a gradual decline in priming at later SOAs. In other words, the strong response activation previously seen from the arrow mask seems to be missing. From average response times alone, we might conclude that the mask stimulus could be ineffective, ignored, or somehow blocked from response control, but this is clearly not the case: Conditional accuracy functions reveal a marked reduction in response accuracy in consistent (but not in inconsistent) trials for response times around 300 ms, similar to what is seen in the *arrow prime, neutral mask* condition. We interpret this temporary dip in accuracy as a sign of active, selective inhibition of the primed response, which results in a disinhibition of the antiprime response and thus induces a response conflict in consistent trials that leads to increased response times and error rates (Panis & T. Schmidt, 2016). That this inhibition is in fact response-specific was shown by T. Schmidt, Hauch, and F. Schmidt (2015) in a study of pointing movements. Participants pointed to a shape target in one of ten possible directions that was preceded by primes and masks in either the same or the vectorially opposite directions (the actual experiment was more complex; see our paper for full description). Even though arrival times at the target showed only subtle evidence for an NCE, the trajectories showed strong effects of response inhibition. Specifically, we looked at consistent trials where prime and target always specified the same pointing direction. Strikingly, we found that while the faster responses always started out in the vectorial direction of the prime and target, slower responses started out in a direction vectorially *opposite* to both prime and target. This shows not only that the NCE occurs relatively late in the trial (as previously shown by Atas & Cleeremans, 2015, Ocampo & Finkbeiner, 2013, and Verleger et al., 2004) but that it is an actual counteractivation of a response vectorially opposite to the primed one. We dubbed this mechanism underlying the NCE "thrust reversal" (T. Schmidt et al., 2015), in analogy to the common braking mechanism in jetplanes. Thrust reversal may be viewed as an extension of the mask-triggered inhibition hypothesis: Inhibition is not only specific for the primed response (Jaśkowski & Przekoracka-Krawczyk, 2005) but is actually an activation of a vectorially opposite response, a counterthrust opposite in sign but matched in force.

Our findings indicate that the presence of stimulus features mapped to the antiprime response is no necessary condition for response inhibition to occur, as shown in Vorberg's (2005) original conference paper. We can thus reject Lleras and Enns's (2004) hypothesis that reversal of priming effects in arrow stimuli is due to priming by mask features. This is in line with our view that the inhibition is not merely a response to stimulus features, but involves an active, selective response activation specifically directed against the primed response (Panis & T. Schmidt, 2016; T. Schmidt et al., 2015). However, there seem to be limits to this conclusion. In two additional experiments following exactly the same design as the present one, we replaced the arrows by color stimuli (T. Schmidt, 2000), such that the prime (if present)

was a red or green circle, the mask (if present) was an annulus fitting around the circle, and the target was an even larger red or green ring. Neutral stimuli were gray. The two experiments only varied in the color saturations used, and both gave almost identical results.⁴ As with the arrow stimuli, the *color prime, no mask* condition yielded a priming effect that decreased with SOA, while the *color prime, neutral mask* condition abolished and partly reversed this effect, indicating a large NCE. Also, the *neutral prime, color mask* condition showed very large, increasing priming effects, demonstrating that a colored mask alone could activate the response. However, the *color mask = color prime* condition showed no sign of inhibition relative to the *neutral prime, color mask* condition, neither in average response times, nor error rates, nor conditional accuracy functions. If anything, priming effects were even a bit larger, as expected from positive response priming by both prime and mask. Our conclusion is that even though we were able to demonstrate active, selective response inhibition in the absence of inhibitory stimulus features in arrow stimuli, this mechanism may not be in place for all stimulus domains (cf. Verleger et al., 2004). This could either indicate that major principles of response inhibition do not translate well from geometric space to color space, or that in some stimulus domains inhibitory stimulus features are still necessary to reverse response priming. But even with this caveat in mind, the present data provide a proof of concept that negative compatibility effects can occur in the absence of inhibitory stimulus features.

Both the editor and a reviewer asked us to comment on the rather small number of participants we employ here. First, remember that statistical power not only depends on the number of participants, but also on the reliability of the dependent measure, and that this reliability is critically determined by the number of repeated measures per participant and condition (Baker, Vilidaite, Lygo, Smith, Flack, Gouws, & Andrews, 2021; Brysbaert & Stevens, 2018). Therefore, a *small-N design* (Smith & Little, 2018), which is characterized by combining a small number of observers with massively repeated measures, can gain high levels of statistical power (Baker et al., 2021). Indeed, by using noncentrality parameters for the basic Consistency x SOA interactions from previous benchmark data we can predict a power of .95 at an alpha risk of .05 by employing only eight participants with an adequate number of repetitions per cell and subject (here, $r = 80$). We can thus be confident that our previous benchmarks for precision in response priming effects ($r = 60$, F. Schmidt et al., 2011) guarantee excellent statistical power for the effects normally expected for such a study. The great advantage of small-*N* designs is the ability for assessing the homogeneity of effects across participants because each observer's data pattern can be measured with high fidelity. In the present dataset, it turns out that effects are quite uniform across participants and that most participants' data show the overall data pattern.

Open practices statement: All data, materials and analyses are available from the authors upon request. All independent and dependent variables are identified. Preregistration was not employed.

Author note: T.S.: This is the final project I was able to realize with Dirk Vorberg, who passed away in January 2021. We were still able to present this manuscript's results at a small conference, the *Herbstkolloquium Kognitionspsychologie*, in the fall of 2019. The last photograph I have of Dirk shows him attending my presentation of our talk, with my empty seat and coffee mug next to him. This paper is dedicated to my great, great teacher. We also thank Anabel Weis and Laura Wetz for collecting the data. Correspondence may be sent to T.S. (thomas.schmidt@sowi.uni-kl.de).

Financial Support: This research received no specific grant from any funding agency, commercial or nonprofit sectors.

Conflict of Interests Statement: The authors have no conflicts of interest to disclose.

Ethics Statement: This study was carried out in accordance with DGPs guidelines. The protocol was approved by the Ethics committee of the University of Kaiserslautern, Faculty of Social Sciences. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

⁴ We thank Lena Krahel and Judith Geib for devoting their Bachelor theses to these data (sometimes under adverse circumstances).

References

- Atas, A., & Cleeremans, A. (2015). The temporal dynamic of automatic inhibition of irrelevant actions. *Journal of Experimental Psychology: Human Perception & Performance*, 41, 289-305. <http://dx.doi.org/10.1037/a0038654>
- Bakeman, R., & McArthur, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavior Research Methods, Instruments, & Computers*, 28 (4), 584-589. <https://psycnet.apa.org/doi/10.1037/1082-989X.2.4.357>
- Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, 26, 295-314. <http://dx.doi.org/10.1037/met0000337>
- Biafora, M., & Schmidt, T. (2020). Induced dissociations: Opposite time-courses of priming and masking induced by custom-made mask-contrast functions. *Attention, Perception, & Psychophysics*, 82, 1333-1354. <https://doi.org/10.3758/s13414-019-01822-4>
- Brenner, E., & Smeets, J. B. J. (2004). Colour vision can contribute to fast corrections of arm movements. *Experimental Brain Research*, 158, 302-307. <https://doi.org/10.1152/jn.00644.2003>
- Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1, Art. 9. <https://doi.org/10.1152/jn.00644.2003>
- Eimer, M., & Schlaghecken, F. (1998). Effects of masked stimuli on motor activation: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1737-1747. <https://doi.org/10.1037/0096-1523.24.6.1737>
- Eimer, M., & Schlaghecken, F. (2003). Response facilitation and inhibition in subliminal priming. *Biological Psychology*, 64, 7-26. [https://doi.org/10.1016/S0301-0511\(03\)00100-5](https://doi.org/10.1016/S0301-0511(03)00100-5)
- Eisenhart, C. (1962). Realistic evaluation of the precision and accuracy of instrument calibration systems. In H. H. Ku (Ed.), *Precision Measurement and Calibration* (1969, pp. 21-48). Washington, D.C.: National Bureau of Standards.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143-149.
- Jaśkowski, P. (2007). The effect of nonmasking distractors on the priming of motor responses. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 456-468. DOI: 10.1037/0096-1523.33.2.456
- Jaśkowski, P. (2008). The negative compatibility effect with nonmasking flankers: A case for mask-triggered inhibition hypothesis. *Consciousness & Cognition*, 17, 765-777. <https://doi.org/10.1016/j.concog.2007.12.002>
- Jaśkowski, P. (2009). Negative compatibility effect: the object-updating hypothesis revisited. *Experimental Brain Research*, 193, 157-160. <https://doi.org/10.1007/s00221-008-1700-6>
- Jaśkowski, P., Białuńska, A., Tomanek, M., & Verleger, R. (2008). Mask- and distractor-triggered inhibitory processes in the priming of motor responses: An EEG study. *Psychophysiology*, 45, 70-85. <https://doi.org/10.1111/j.1469-8986.2007.00595.x>
- Jaśkowski, P., & Przekoracka-Krawczyk, A. (2005). On the role of mask structure in subliminal priming. *Acta Neurobiologiae Experimentalis*, 65, 409-417.
- Klotz, W., & Neumann, O. (1999). Motor activation without conscious discrimination in metacontrast masking. *Journal of Experimental Psychology: Human Perception and Performance*, 25(4), 976-992. <https://doi.org/10.1037/0096-1523.25.4.976>
- Klotz, W., & Wolff, P. (1995). The effect of a masked stimulus on the response to the masking stimulus. *Psychological Research*, 58(2), 92-101. <https://doi.org/10.1007/BF00571098>
- Leuthold, H., & Kopp, B. (1998). Mechanisms of priming by masked stimuli: Inferences from event-related brain potentials. *Psychological Science*, 9, 263-269. <https://doi.org/10.1111%2F1467-9280.00053>
- Lingnau, A., & Vorberg, D. (2005). The time course of response inhibition in masked priming. *Perception & Psychophysics*, 67, 545-557. <https://doi.org/10.3758/BF03193330>
- Lleras, A., & Enns, J. T. (2004). Negative compatibility or object updating? A cautionary tale of mask-dependent priming. *Journal of Experimental Psychology: General*, 133, 475-493. <https://psycnet.apa.org/doi/10.1037/0096-3445.133.4.475>
- Ocampo, B., & Finkbeiner, M. (2013). The negative compatibility effect with relevant masks: A case for automatic motor inhibition. *Frontiers in Psychology*, 4, Article 822. <https://doi.org/10.3389/fpsyg.2013.00822>
- 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., Moran, R., Wolkersdorfer, M., & Schmidt, T. (2020). 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. *Attention, Perception & Psychophysics*, 82, 689-714. doi.org/10.3758/s13414-019-01897-z
- 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. https://doi.org/10.1162/jocn_a_00998
- Panis, S., & Schmidt, T. (2020, May 25). What is causing “inhibition of return” in spatial cueing tasks? Temporally disentangling multiple cue-triggered effects on multiple time scales using response history and conditional accuracy analyses. Preprint. <https://doi.org/10.31234/osf.io/udpvs>
- Panis, S., & Schmidt, T. (2022). When does “inhibition of return” occur in spatial cueing tasks? Temporally disentangling multiple cue-triggered effects using response history and conditional accuracy analyses. *Open Psychology*, 4, 84-114.

- Schmidt, T. (2000). Visual perception without awareness: Priming responses by color. In T. Metzinger (Ed.), *Neural correlates of consciousness* (pp. 157-179). Cambridge: MIT Press. <https://doi.org/10.7551/mitpress/4928.003.0014>
- Schmidt, T. (2002). The finger in flight: Real-time motor control by visually masked color stimuli. *Psychological Science*, 13, 112-118. <https://doi.org/10.1111%2F1467-9280.00421>
- Schmidt, F., Haberkamp, A., & Schmidt, T. (2011). Dos and don'ts in response priming research. *Advances in Cognitive Psychology*, 7, 120-131. <https://doi.org/10.2478/v10053-008-0092-2>
- Schmidt, T., Hauch, V., & Schmidt, F. (2015). Mask-triggered thrust reversal in the Negative Compatibility Effect. *Attention, Perception, & Psychophysics*, 77, 2377-2398. <https://doi.org/10.3758/s13414-015-0923-4>
- Schmidt, T., Niehaus, S., & Nagel, A. (2006). Primes and targets in rapid chases: Tracing sequential waves of motor activation. *Behavioral Neuroscience*, 120(5), 1005-1016. <https://doi.org/10.1037/0735-7044.120.5.1005>
- Schmidt, F., & Schmidt, T. (2014). Rapid processing of closure and viewpoint-invariant symmetry: Behavioral criteria for feedforward processing. *Psychological Research*, 78, 37-54. <https://doi.org/10.1007/s00426-013-0478-8>
- Schmidt, T., & Schmidt, F. (2009). Processing of natural images is feedforward: A simple behavioral test. *Attention, Perception, & Psychophysics*, 71(3), 594-606. doi:10.3758/APP.71.3.594
- Schmidt, T., & Vorberg, D. (2006). Criteria for unconscious cognition: Three types of dissociation. *Perception & Psychophysics*, 68(3), 489-504. <https://doi.org/10.3758/BF03193692>
- Schmidt, F., Weber, A., & Schmidt, T. (2014). Activation of response force by self-splitting objects: Where are the limits of feedforward Gestalt processing? *Journal of Vision*, 14, 1-16. doi: 10.1167/14.9.20
- Schwarz, W., & Mecklinger, A. (1995). Relationship between flanker identifiability and compatibility effect. *Perception & Psychophysics*, 57, 1045-1052. <https://doi.org/10.3758/BF03205463>
- Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review*, 25(6), 2083-2101. <https://doi.org/10.3758/s13423-018-1451-8>
- Sumner, P. (2008). Mask-induced priming and the negative compatibility effect. *Experimental Psychology*, 55, 133-141. <https://doi.org/10.1027/1618-3169.55.2.133>
- Vath, N., & Schmidt, T. (2007). Tracing sequential waves of rapid visuomotor activation in lateralized readiness potentials. *Neuroscience*, 145, 197-208. <https://doi.org/10.1016/j.neuroscience.2006.11.044>
- Verleger, R., Jaśkowski, P., Aydemir, A., van der Lubbe, R., & Groen, M. (2004). Qualitative differences between conscious and nonconscious processing? On inverse priming induced by masked arrows. *Journal of Experimental Psychology: General*, 133, 494-515. <https://psycnet.apa.org/doi/10.1037/0096-3445.133.4.494>
- Vorberg, D. (2005, April 4-6). *Ist Hemmung oder Bahnung die Grundlage des umgekehrten Priming-Effekts?* [Conference presentation]. 47. Tagung experimentell arbeitender Psychologen, Regensburg, Germany.
- Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of the National Academy of Sciences USA*, 100(10), 6275-6280. <https://doi.org/10.1073/pnas.0931489100>
- Wolkersdorfer, M., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential motor activation in a dual-prime paradigm: Insights from conditional accuracy and hazard functions. *Attention, Perception & Psychophysics*, 82, 2581-2602. <https://doi.org/10.3758/s13414-020-02010-5>

Discussion

The following chapters aim to highlight the advantages of applying advanced distributional analyses. Specifically, which of the findings in the reported articles were possible due to the application of discrete-time EHA, and would have been concealed by traditional methods? In which ways does such a distributional approach advance the field of experimental psychology, in particular regarding the research of visual search behavior, response priming, and visual masking? First, I will highlight the findings presented in the published articles. Next, I will evaluate the method and provide recommendations for the future application in the field. Lastly, I will end with a discussion of ways to further validate the method, and give a preview into potential future enhancements of it.

4 Highlighted Findings

Discrete-time EHA allowed us to study the effects of experimental manipulations in the visual search, the response priming, and the masked response priming paradigm. Collectively, the applications of this method gave new insights into the temporal dynamics and individual differences in performance in these tasks. Crucially, it enabled us to track changes with the passage of time.

4.1 Discrete-time EHA reveals effects that mean performance measures can conceal

First, discrete-time EHA revealed previously concealed features of visual search. As seen in Figure 5, when analyzing mean correct RT, the inclusion of additional

distractors seemingly does not slow down the search process in feature search. This finding led to the classification of this type of search as efficient (Cheal & Lyon, 1992; Liesefeld et al., 2016; Wolfe et al., 2010). However, in Panis, Moran, et al. (2020) we found a systematic temporary effect of set size on early $h(t)$. Specifically, if the target is present, hazards were higher at smaller set sizes compared to larger set sizes. In other words, at least temporarily, the probability of response occurrence in the target present condition decreased with set size. This indicates that a target does not “pop-out” out of a set of distractors. Panis, Moran, et al. (2020) attribute this finding to three factors. First, peripheral targets prolong the search process, as they are more difficult to recognize, due to lower color sensitivity. Wolfe et al. (2010) did not control for the eccentricity on a trial-by-trial basis. Thus, larger set sizes potentially increased the amount of trials with such targets. Next, larger set sizes also confound with stimulus density. This causes more objects to be contained in the receptive field of a single visual neuron. In turn, this might have increased the likelihood of visual crowding in the periphery, further hindering target recognition. Finally, since display presentation lasted until response, it is conceivable that more eye-movements were made in trials with larger set sizes. This would have resulted in varying distances between target location and eye gaze location on a within-trial time basis. If this were the case, gaze-to-target distance should then be seen as a time-varying covariate.

Second, discrete-time EHA gave a more detailed insight into the dynamics of response behavior in a response priming paradigm. In Wolkersdorfer et al. (2020) we employed a dual-prime paradigm, in which a target stimulus was preceded by a prime, which in turn was preceded by a first prime. Previous findings showed that the priming effect in such a sequence of stimuli would be dominated by the second prime

(Breitmeyer & Hanif, 2008; Grainger et al., 2013). Indeed, ANOVA analyses of mean correct RT seemed to corroborate this. However, discrete-time EHA revealed that initial fast responses were in fact driven by the identity of the first prime. Thus, Wolkersdorfer et al. (2020) concluded that a second prime does not override a first prime, as previously assumed (Breitmeyer & Hanif, 2008). Instead, the prime-triggered motor responses compete with each other, with the possibility of the first prime triggering a response by itself.

Third, in T. Schmidt et al. (2022) we investigated response inhibition in the NCE. ANOVAs performed on mean RT and ER, and the contrasting of different masking conditions with them, revealed the true magnitude of the NCE caused by a mask. Moreover, we found response inhibition even with a mask that retains perceptual evidence of the prime, enabling us to reject Eimer and Schlaghecken's (1998) hypothesis that inhibition is the result of the mask suppressing perceptual information about the prime. Crucially, only analyzing conditional accuracy functions revealed that, at least with arrow stimuli, antiprime features are not required to cause response inhibition. If one were to only inspect mean RTs, one might have concluded that an *arrow mask* could be ineffective. This is because, with respect to mean RTs, the condition *arrow prime = arrow mask* behaved very similar to the condition *arrow prime, no mask*, seemingly void of the additional priming effect the mask could trigger by itself, as seen in the *neutral prime, arrow mask* condition. Inspection of the conditional accuracy functions showed, however, a marked reduction in response accuracy in consistent trials (at around 300 ms). This pattern was very similar to the one found in the conditional accuracy function of the *arrow prime, neutral mask* condition. We interpreted this as a sign of active, selective inhibition of the primed response. In line

with Panis and Schmidt (2016), such inhibition would result in the disinhibition of the antiprime response. The resulting response conflict would then cause the observed reduction in accuracy and slow down responses. Moreover, only the application of discrete-time EHA revealed that the NCE is even larger than the response priming effect. Specifically, it is directed against the primed response, with response-specific inhibition not requiring any positive priming features. Previously, the NCE was believed to be only 20 ms large and caused by features of the mask.

To conclude, these examples show the benefits of employing discrete-time EHA in addition to more traditional approaches. While ANOVAs and mean RT and ER analyses are able to reveal certain aspects of the involved processes, they also run the risk of missing and, in the worst case, concealing effects that a discrete-time EHA is able to uncover.

4.2 Discrete-time EHA reveals the temporal dynamics of response behavior

As shown in the previous section, we found a systematic but temporary effect of set size on early $h(t)$ in feature search (Panis, Moran, et al., 2020). Moreover, this effect was present across all three visual search tasks. In addition, we also found a systematic effect of target presence, with hazards rising earlier when a target is present versus when no target is displayed. Importantly, discrete-time EHA revealed that these effects were limited in time, with the ordering of hazards only being present in the left tail of the distribution. Both the effect of set size and target presence persisted longer as task difficulty increased, with effects in feature search lasting until about 500 ms, in conjunction search about 1000 ms, and spatial configuration search about 2000 ms after display onset. From these time points onward, hazard functions were flat without any

systematic effects of set size and target presence. In Panis, Moran, et al. (2020) we argued that these flat hazard functions reflected RT outliers during decision-making and were not related to the respective search processes. Overall, this indicated that search processes are aborted rather early, depending on task difficulty.

Discrete-time EHA further revealed two subsets of subjects that showed differences in their response behavior. One group generally responded slower and showed high accuracy throughout. We argued that these subjects displayed proactive control. Global response inhibition held back responses until enough information about the search was available. Thus, responses occurred later but at high accuracy, compared to the second group. For this second group (fast responders), we identified at least three temporal stages in their respective conditional accuracy functions: (1) Earliest responses showed false alarms (defined as $ca(t) \leq .5$) in the target absent condition, combined with near perfect accuracy in the target present condition, (2) followed by a second stage with near perfect accuracy in the target absent condition, combined with lower accuracy in the target present condition (i.e., an increase in misses), and (3) followed by a third stage, in which high accuracy was achieved in both target presence conditions. The more difficult search tasks, conjunction and spatial configuration, showed an additional fourth stage in slow responses, in which no false alarms were detected, but the miss rate increased with the passage of time. We interpreted these findings as indicative of an early bias towards “target present”, followed by selective inhibition of this bias, due to error-monitoring, resulting in the disinhibition of the “target absent” response. This would lead to a temporary increase in misses, as observed in the second stage. The third stage was seen as the result of information from the actual search process becoming available, resulting in highly

accurate responses. However, as the search process continued without a response in the more difficult tasks, the search was aborted and a tendency to respond “target absent” developed, as seen in the fourth stage. Overall, discrete-time EHA thus enabled us to identify these different stages of the respective search processes, as well as individual differences between fast and slow responding subjects. To our knowledge, these features of visual search were previously unknown and are, as of now, not predicted by cognitive models of visual search.

As shown, discrete-time EHA also revealed temporal dynamics of inhibition in the NCE. In T. Schmidt et al. (2022) we showed a marked reduction in response accuracy in consistent trials after around 300 ms in the *arrow prime = arrow mask* condition, as well as in the *arrow prime, neutral mask* condition. This was in line with the findings of Panis and Schmidt (2016), and appear to be characteristic of the NCE. We also found that, in the condition without a mask but with an arrow prime, early responses were driven by the identity of the prime, while later responses were driven by target information. Through the application of discrete-time EHA Panis and Schmidt (2016) previously revealed the following temporal characteristics of masked priming: (1) an initial PCE for 120 ms, which starts around 320 ms after prime onset and lasts for about 120 ms, (2) followed by an NCE, which is time-locked to the mask at around 360 ms after mask onset, and (3) a final target stage, in which responses are driven by target identity.

Similarly, in Wolkersdorfer et al. (2020) we explored the temporal dynamics of response activation in a dual-prime paradigm. When investigating short SOAs (Experiment 1), we found that conditional accuracy functions were at near-perfect accuracy when the first prime was consistent with the target, and incorrect when the

first prime was inconsistent. Thus, initial responses were exclusively triggered by signals of the first prime. If a first consistent prime was followed by a second consistent prime, responses maintained near perfect accuracy. However, conditional accuracy functions revealed a marked reduction in response accuracy if the second prime was inconsistent. Likewise, if the first prime was inconsistent and followed by an inconsistent second prime, accuracy remained low. If the second prime, however, was consistent, accuracy quickly rose. These findings indicate that intermediate responses reflected conflict between the first and second prime, with responses being triggered by the latter's identity. As expected, later responses had high accuracy, signaling the onset of target information dominating the response. An ordering of hazard rates was present, with hazards decreasing from two consistent, to inconsistent-consistent, consistent-inconsistent, and finally two inconsistent primes. Thus, discrete-time EHA enabled us to support the chase theory of response priming by T. Schmidt et al. (2006): (1) Prime rather than target signals determine the onset and initial direction of the response; (2) target signals influence the response before it is completed; (3) movement kinematics initially depend on prime characteristics only and are independent of all target characteristics; (4) a second prime can influence the response before it is completed, creating conflict if the two prime identities are opposites. These findings reveal that under the right temporal conditions, sequential stimuli result in a strictly sequential response activation. Without the application of discrete-time EHA the only finding would have been that two non-identical primes can lead to intermediary effects, concealing this sequential nature of response activation. While previous studies relied on lateralized readiness potentials (Eimer & Schlaghecken, 1998; Vath & Schmidt, 2007) or the analysis of pointing movements (F. Schmidt & Schmidt, 2010

Schmidt, 2002; T. Schmidt & Schmidt, 2009; Vath & Schmidt, 2007), discrete-time EHA required only the collection of simple behavioral data (RT and accuracy) to achieve this.

4.3 Discrete-time EHA allows to track performance changes on multiple time scales

In Panis, Moran, et al. (2020), we were able to track performance changes on multiple time scales. Within-trial changes have been presented in the previous section. As the name suggests, this time scale tracks performance within trials, that is, how performance changes with the passage of time (e.g., early vs. late responses). With the inclusion of trial number in the hazard model, we were able to identify how hazard varies on the across-trial (but within-block) time scale. For example, in feature search we found that, with each additional series of ten trials within a block, hazard decreases in the left tail of the distribution and increases in the right tail of the distribution. In other words, responses occur later as trial number increases. A possible explanation would be that in a relatively easy task, such as feature search, fatigue begins to set in towards the end of the block and subjects slow down. Strikingly, the inclusion of block number in the hazard model (enabling tracking on the across-block time scale), revealed that hazard increases in the left tail and decreases in the right tail. In other words, as block number increases, responses occur earlier. This is indicative of a practice effect across blocks. In contrast, the inclusion of trial and block number revealed practice effects in conjunction and spatial configuration search, on both the across-trial and across-block time scales (see Fig. 6 in Panis, Moran, et al., 2020, p. 703).

5 Evaluation of discrete-time EHA

The previous section provided findings that were due to the application of discrete-time EHA. Studies of priming (Wolkersdorfer et al., 2020), visual search (Panis, Moran, et al., 2020; Panis, Schmidt, et al., 2020), and primed masking and the NCE (Panis, Schmidt, et al., 2020; T. Schmidt et al., 2022) make a strong case for the use of discrete-time EHA in the field of cognitive and experimental psychology. First, while traditional analyses focusing on mean performance measures have been shown to conceal effects, discrete-time EHA provides insights into response activation, ranging from the time-locking of effects and response inhibition to changes of effects over time, within and across trials. In other words, the temporal dynamics of response behavior are more accurately captured by this method. Second, the inclusion of conditional accuracy functions allows a deeper look into the decision processes involved. RT and accuracy data are measures of different aspects of the decision process (Mulder & van Maanen, 2013). For example, while hazard functions give the conditional probability of response occurrence, conditional accuracy functions evaluate the performance of these responses. Thus, together they give a more detailed insight into the decision-making process. Third, discrete-time EHA can identify subsets of participants. For example, in Panis, Moran, et al. (2020) we found two distinct groups that showed remarkably different response behavior (fast vs. slow responders).

5.1 Theoretical advantages of discrete-time EHA

To this day, when dealing with time-to-event data, such as RT and accuracy, the dominant method in the field of experimental psychology remains ANOVA. As

detailed in the introduction, this is due to the historically strong influence of serial information processing accounts. The concept of mental operations being separable into distinct consecutive processing stages dates all the way back to Donders (1969) and became even more popular with Sternberg's AFM (1969, 1984, 2011, 2013), continuing on ever since. More recently, distributional methods have become more common in the field, arguing in favor of more advanced or detailed analyses beyond merely investigating means (Balota & Yap, 2011; Ridderinkhof, 2002; van Maanen et al., 2019; VanRullen, 2011). The assumption that cognitive processes or operations can be divided into smaller sub-processes, and that the duration of such a process is reflected in RTs as the cumulative duration of all its sub-processes nonetheless remains popular (e.g., Liesefeld, 2018; Song & Nakayama, 2009). However, the present work and the showcased studies applying discrete-time EHA with conditional accuracy analyses paint a different picture. For example, in Wolkersdorfer et al. (2020) we showed that (1) fast responses were triggered by the first prime's identity, void of any subsequent stimuli's information, (2) intermediate responses showed response conflict between the two primes' information, and (3) late responses reflected the target stimulus' identity. Likewise, different stages of the visual search process were identified in Panis, Moran, et al. (2020): (1) an early bias for target-present, (2) suppression of the target-present response and disinhibition of the target-absent response, resulting in some misses, (3) optimal performance, and (4) an evolving bias towards target-absent responses. Thus, fast, medium, and slow RTs instead relate to different cognitive operations involved in the underlying task, and are not just the cumulative duration of all the sub-processes involved in the task at hand.

Consequently, if one's goal is to understand human behavior, one has to take the passage of time into consideration when analyzing data obtained in such experiments. This is in line with a dynamical systems account (van Gelder, 1995). *Dynamic field theory* (DFT; Schöner G., Spencer, J. P., & the DFT Research Group, 2016) describes cognition as sequential transitions between several states. These states include sensory, motor, and central states. Hazard and conditional accuracy functions enable us to measure the motor output of such a dynamical system. As seen in the presented work, they allow us to investigate how long different motor states last, when transitions to other states occur, and how, if, and when experimental manipulations affect these different states. Through them, we can make inferences about the underlying cognitive processes and better capture the dynamic nature of human behavior. As shown, by simply averaging the motor output over time and analyzing mean RTs and accuracies, we run the risk of missing and concealing effects that the analyses of $h(t)$ and $ca(t)$ reveal. For example, we would have *falsely* concluded that (1) additional distractors do not slow down the search process in feature search (Panis, Moran, et al., 2020), (2) with two sequential primes the priming effect is dominated by the second prime (Wolkersdorfer et al., 2020), and (3) antiprime features are required to cause response inhibition (T. Schmidt et al., 2022). While Donders' pure insertion and Sternberg's AFM rely on unrealistic assumptions (see Introduction, Chapter 1) discrete-time EHA allows to directly measure the time of the onset and the duration of effects, and to track changes across multiple time-scales. In other words, in discrete-time EHA time is treated just as time, resulting in a descriptive approach that operates on actual observed physical time.

5.2 Statistical & methodological advantages of EHA

Compared to the analyses of mean data, EHA also comes with statistical and methodological advantages. As already discussed in the Chapter 2.1, EHA can deal well with right-censored observations, since such observations are not simply discarded when preparing data for this analysis. While the preparation for ANOVAs typically involves some form of censoring in order to mitigate the problem of diminished power due to outliers and the typically right-skewed distribution of RT data, this can often be ill-advised. For example, if the effect of interest occurs in the right-tail of the RT distribution, discarding such trials risks to reduce power, even though preventing this was the original goal of the censoring procedure (Ratcliff, 1993). Generally speaking, EHA retains valuable information that other methods lose due to right-censoring.

In addition, EHA allows the inclusion of a variety of time-varying covariates that could be of interest. When preparing data for EHA, time-to-event data is transformed into time-series data (see Chapter 2.2 for an introduction to this procedure for discrete-time EHA). $h(t)$ gives then the instantaneous likelihood of event occurrence given no previous events, while $ca(t)$ provides the complementary performance measure. In other words, together those measures reveal *what* happens *when*. Physiological measures, such as EEG, can also trace changes over time. When modeling EHA, one can include several of such time-varying covariates (EEG signal amplitude, gaze location, heart rate, galvanic skin response, etc.). This has the potential to generate hazard models of not just behavioral but neural and other physiological event occurrence that go beyond currently existing models of cognitive psychophysiology

(Meyer et al., 1988). Moreover, as shown in Chapter 4, EHA allows to include multiple time scales during modeling. This enables to test how, when, and if (individual) performance changes, for example, within trials, across trials, across blocks, and more.

Finally, the ultimate goal of research in experimental psychology and adjacent fields is to arrive at a precise description of the microscopic dynamics of the neurobehavioral system. For example, we want to connect the behavioral, cognitive, and neural dynamics, that is, the subject's behavior in a given experimental paradigm, the involved cognitive processes, and the underlying neural correlates (ordered from macro to micro). According to Kelso et al. (2013, p. 122) *"it is crucial to first have a precise description of the macroscopic behavior of a system in order to know what to derive"* on the microscopic level. As shown, EHA gives a much more precise description of the macroscopic behavior than traditional measures such as mean RTs. Therefore, one should first study $h(t)$ and $ca(t)$ functions before attempting to fit parametric functions to the data or to generate computational models of human behavior. Chapter 4 highlights how important features of human behavior might otherwise be missed.

5.3 Potential disadvantages of discrete-time EHA and how to address them

Deciding on the optimal bin size is not always a straightforward process. If the chosen bin size is too small, one runs the risk of overinterpreting spurious effects that are based on too few observations. If the chosen bin size is too large, the low temporal resolution can result in missing important features of the distribution of event occurrence. In order to find the best fitting bin size for the data set at hand, one has to explore a few alternatives. However, this crucial decision in the application of discrete-time EHA is not arbitrary. Bin size optimality is dependent on multiple variables. First,

censoring time will influence bin size. It will either be set by the experimental design through the inclusion of an event deadline (i.e., a time after which events such as a button-press response are not recorded and the trial ends), or during analysis (i.e., a time until which all informative events are expected to have occurred); generally, with later censoring time, bin size should increase as well. Second, bin size will also depend on the rarity of event occurrence; more events recorded allows for smaller bin sizes. Third, the number of measurements (either in the number of repeated measures or the number of participants) will affect the optimal bin size; again, more events recorded allows for smaller bin sizes. Forth, the decision may be informed by the experimental design and the hypotheses that are to be tested. For example, if one wants to investigate the effect of different SOAs on an effect of interest (see response priming), intuitively one might want to choose a bin size that reflects the differences in employed SOAs (i.e., if SOAs change in 50 ms steps, choose a bin size of a maximum of 50 ms steps, or other dividers of 50). If effects are time-locked, these bins are more likely to reflect this. Overall, the issue of bin size optimality might not even be a disadvantage. The available information guiding the decision on bin size allows to make an informed tradeoff between temporal resolution and economic data collection. In addition, the process can be adapted to data and research question.

Furthermore, depending on the decision made regarding bin size and the amount of data collected, the person-trial-bin oriented data set might end up very large. This can potentially affect time required for modeling and cause performance issues when working with the data. However, this is a common issue in many areas and with many methods (e.g., EEG, eye-tracking, and other types of data), and not at all unique to discrete-time EHA. Note that time bins do not have to be all of equal size, which can

help with data set size and the issue of response/event availability discussed above (Panis, 2020).

In addition, when applying discrete-time EHA, especially in a small- N design, there are potential sources of noise in the data. Regarding noise within participants, it is crucial to have enough repeated measures, both on the participant level and per condition. This ensures high measurement precision and reduces noise within participants. Noise between participants can have two sources: low measurement precision on the participant level, or differences between participants. The former can again be reduced by sufficient measurements per participant and condition. The latter can actually be informative and reflect systematic differences between participants, as seen in Panis, Moran, et al. (2020). Here, characteristic differences between participants were found in their speed, potentially due to their employment of different response strategies. Assuring high measurement precision and the inclusion of covariates at the participant level when modeling are the recommended strategies to deal with this type of “noise”. Importantly, one should therefore always analyze participants on the individual level first. This allows one to identify such systematic differences between participants and safeguards against the interpretation of spurious effects. If one were to only look at pooled or averaged data instead, effects might otherwise be missed or effects be interpreted that are absent for most individual participants. However, a balance between analyzing individual data patterns and pooled data is crucial. Systematic effects will be present for subsets (e.g., fast and slow responders) or the majority of participants, while patterns due to noise will not. In other words, effects found in pooled data should be present in the majority of individual subjects as well.

Further, if one is not experienced in modeling, inferential statistics for discrete-time EHA can be challenging at first. If data was collected with a large- N design (and without repeated measures), a standard logistic regression on the person-trial-bin oriented data set is sufficient for estimating the parameters of a discrete-time hazard model (see Allison, 2010 for a detailed description). However, if data was collected with a small- N design (and with repeated measures), possible options range from population-averaged methods (such as generalized estimating equations [GEE]), Bayesian methods, to generalized linear mixed models (GLMM; for an overview see Allison, 2010; for a comprehensive tutorial on Bayesian and frequentist approaches, see Panis & Ramsey, 2024). GLMMs were applied in Panis, Moran, et al. (2020) and Wolkersdorfer et al. (2020). In these studies, we employed stepwise techniques to find the best model, both for hazard and conditional accuracy functions. Doing this with complex models, while testing the models on the same data used to estimate them, risks to over- or underfit the data. Thus, the p -values from these models should be treated with caution, especially during model selection. For these reasons, generalizability of analyses and procedures of modeling and model selection are still being discussed and developed (Barr et al., 2013; Cummings, 2012; Matuschek et al., 2017; Zuur & Ieno, 2016). Moreover, models as presented in Wolkersdorfer et al. (2020) can appear rather complex for researchers unfamiliar with the technique. Other potential issues include failures of models to converge, departures from the original analysis plans, ad hoc changes during modeling, and the researcher degrees of freedom in model selection. In order to address this, and to offer an approach that researchers with a different background in data analysis will be more comfortable with, another technique will be discussed in the following.

If one's aim is not to fully describe the shape of the hazard function, but is merely interested in certain properties or parameters of the descriptive function, robust techniques such as bootstrapping or jackknifing become available for inferential analyses (Ulrich & Miller, 2001; Wilcox R., 2011). For example, one might want to investigate divergence and convergence points of $h(t)$, i.e. at which point do hazard estimates of different conditions differ and for how long. Originally developed as a jackknife-based method to investigate differences between two conditions in their respective onset latency of lateralized readiness potentials (LRP; Miller et al., 1998), Ulrich and Miller (2001) generalized it to factorial experiments investigating onset latencies as well.

The main advantage of this method is that it drastically improves the signal-to-noise ratio in relatively noisy EEG data. But let's first revisit the example of a simple response priming experiment introduced in Chapter 2. In Figure 3, we found that hazards for consistent and inconsistent trials diverge after 200 ms have passed, that is, from this time bin on, hazard for consistent trials increases faster than for inconsistent trials. If we know from previous experiments that this is a regular observation that takes place in a certain time window (see, for example, Wolkersdorfer et al., 2020), we can treat such a time window as a region of interest (ROI). We can now apply the same simple, yet efficient, jackknifing procedure proposed by Ulrich and Miller (2001) for such and similar cases. First, one participant is omitted and the average curves for consistent and inconsistent trials in the ROI is computed on the remaining subsample of $N-1$ participants. Second, this computation is repeated until each participant was omitted once, resulting in relatively smooth curves for the N subsamples. Note, each subsample is now treated like a participant. Third, parameters of interest can be

extracted from each subsample. Compared to the noisy data of individual participants, this is now much easier. Forth, these N parameters enter a table and a standard ANOVA can be computed. For example, we can identify the moment of time at which hazards for consistent and inconsistent trials exceed a fixed criterion (e.g., $h(t)=.3$) for each subsample, similar to the onset latency of LRPs in the original work by Ulrich and Miller (2001). These time points are then stored in a table. Crucially, while the mean of the subsample curves and the mean of the individual participants are equal, error variances are artificially reduced by this procedure. Since each participant is included $N-1$ times, error variance is reduced by a factor of $1/(N-1)^2$ (see Ulrich & Miller, 2001 for proofs). For this reason, all F values have to be corrected by this same factor, resulting in $F_c = F/(N-1)^2$. P values are then calculated for this new F_c . Finally, with respect to our example data, this procedure enables us to identify the delay in $h(t)$ between consistent and inconsistent trials. With respect to power, this approach has been shown to perform better than the analysis of individual curves, and deals sufficiently with Type I error (Ulrich & Miller, 2001). This is due to the improved signal-to-noise ratio of subsample scores based on the average of $N-1$ participants compared to scores based on individual participant curves.

Let's consider a slightly different approach to the data analysis by generalizing the procedure to different research goals. For example, one might not be interested in the delay of $h(t)$ increases between conditions, but instead wants to investigate whether hazards significantly differ at a given time bin or in a given time window. Instead of extracting the average time points at which hazards exceed a fixed criterion for each subsample, one can also extract the average hazard at specified time bins or averaged over a time window of interest. This allows to test whether and when hazards are

significantly different between the conditions of interest. Instead of N subsamples of average time, averages for $h(t)$ enter a table for each condition and time bin or window. For our example, a repeated-measures ANOVA computed on this table could answer whether $h(t)$ differ significantly between consistent and inconsistent trials after 200 ms, 300 ms, and 400 ms, respectively. Corrections of F values need to be applied as discussed earlier. Note, if violations of the sphericity assumption occur, common correction methods such as Greenhouse-Geisser and Huynh-Feldt can still be employed (Ulrich & Miller, 2001).

To summarize, questions concerned with bin size optimality are a matter of experimental design and ultimately result in an informed tradeoff between temporal resolution and economic data collection. Large data sets are not unique to the method presented here. Moreover, ensuring high measurement precision is the best strategy to deal with sources of noise, and in case of a small- N design is achieved by sufficient repeated measures. Finally, depending on the research question, there is a wide range of inferential statistics applicable for discrete-time EHA. Rather complex models allow to model the shape of $h(t)$ and $ca(t)$, but model selection can harbor some dangers and difficulties. The jackknifing procedure is a viable alternative. If the goal is not a complete modeling of $h(t)$ and $ca(t)$, but one is rather interested in specific parameters at specific time bins, this method provides a robust test of certain hypotheses, ensuring accessibility to EHA for a wide range of researchers. Jackknifing also allows for the testing of clear a priori hypotheses, can be applied to more than one diagnostic function ($h(t)$, $ca(t)$, $S(t)$, etc.), and allows one to flexibly test for delays and durations of effects. Overall, it is therefore my opinion that the presented advantages of discrete-time EHA outweigh the disadvantages discussed here.

5.4 Comparison with other common distributional methods

How does discrete-time EHA fare in a comparison with other common distributional methods? The first obvious method to compare discrete-time EHA to is continuous-time EHA. At first glance, when treating time as a continuous measure, the high temporal resolution achieved seems sufficient to choose it over discrete-time. However, the latter has some subtle advantages. As mentioned in Chapter 2.2, the continuous-time hazard rate function requires a large amount of data for a good estimate (i.e., ~1,000 trials per subject per condition; see Bloxom, 1984; Luce, 1986; van Zandt, 2000). Depending on the research question, this can be unattainable, for instance due to the increased risk of practice effects or the time required for data acquisition. Discrete-time EHA is a more efficient, yet viable, alternative.

Crucially, statistical modeling of continuous-time EHA requires specialized software, while standard logistic regression software or, in the case of the presented jackknifing procedure, any statistical software capable of simple ANOVAs is sufficient for discrete-time hazard model estimation. Further, modeling is more complex than for discrete-time EHA (which, as the previous sections showed, already is rather complex). Thus, discrete-time EHA can be regarded as a more suitable entry point into these advanced distributional analyses. And while parametric hazard models are available for continuous-time EHA, they can be rather restrictive in the shapes they allow (e.g., Fig. 1 shows an exponential, a Weibull, a gamma, and a lognormal hazard model). A semi-parametric alternative for continuous-time hazard models, the Cox regression, even completely ignores the shape of the hazard function. Another alternative is a piece-wise exponential model. However, this becomes available only if

event times are measured very precisely. During this procedure, time is again divided into intervals, in which the hazard rates are considered to be constant. In other words, it is assumed that RTs are exponentially distributed within these intervals (see Allison, 2010 for an overview of these methods). The application of continuous-time EHA can certainly be desirable under particular circumstances. In my view, however, the complicated and sometimes restrictive nature of its modeling procedures, as well as the large amount of data required for it, make discrete-time EHA the preferred distributional analysis of time-to-event data. The advantages of the latter, from data collection to the acquisition of inferential statistics, more than make up for its lower temporal resolution.

Another common way to investigate distributions of time-to-event data is the creation of quantile and delta plots. RT distributions are partitioned into bins based on quantiles. These quantiles are then plotted against the quantile order. Delta plots can then be created to compare two conditions. Corresponding quantiles are subtracted and the differences of each quantile are then plotted against the average of the two respective quantiles. This enables one to easily identify the ranges of RT in which effects in the cumulative distribution take place, and of which size and direction they are. A big disadvantage of this method is present when participants vary strongly. For example, very fast participants will have small RT quantiles, since many of their responses occur early (i.e., possibly small ranges between quantiles and earlier RT cutoff points), while slow participants will have large quantiles (i.e., possibly large ranges between quantiles and later RT cutoff points). Averaging them will inevitably blur the effect. Let's say effects are indeed time-locked to stimulus onset. Averaging slow and fast participants in this way risks to miss or conceal such an effect and it is

not possible to identify precisely when the effect takes place. In comparison, plotting hazards and conditional accuracy will accurately reveal such time-locked effects (see, for example, Wolkersdorfer et al., 2020).

In Chapter 1, a similar procedure known as Vincentizing was introduced. During this procedure, average RT distributions are generated from the average of their quantiles. It is assumed that this normalizes the RT distributions across participants (Ratcliff, 1979). However, this approach has not been evaluated positively (Rouder & Speckman, 2004). While there are select circumstances for which Vincentizing has been recommended, this is only the case when few observations per participant were made. Moreover, Rouder and Speckman (2004) showed that estimates made with this method are often inconsistent with respect to averaged parameters. Further, error responses are regularly censored before the creation of group RT distributions. Yet, errors can occur outside of the control of the researcher. As discussed in Chapter 2.1, error trials are likely informative and discarding such trials may introduce a sampling bias.

In addition, the hazard is the most diagnostic function. While two hazard functions might cross one or two times, the corresponding cumulative distribution functions might not. Thus, averaging quantiles runs the risk of missing effects and patterns, while the hazard is able to discover them. Finally, the same averaging comes with another drawback. For example, as shown in Panis, Moran, et al. (2020), differences across participants are of high theoretical value. Averaging across them might be possible, however, the subgroups of fast and slow participants would have been missed, and the potential underlying processes involved never further investigated. When applying discrete-time EHA, individual participant inspection is highly

encouraged, and including relevant predictors at the participant level during modeling allows to identify such individual differences.

5.5 Recommendations for the application of discrete-time EHA

If the present work achieved its goal to highlight the advantages of the method of discrete-time EHA and motivates to apply the method, I would like to make some general recommendations regarding the experimental design of RT and other time-to-event data studies. It was shown that discrete-time EHA is a powerful tool, providing us with exactly what we need when investigating time-to-event data: (1) The hazard gives the instantaneous likelihood of event occurrence given no previous events, (2) it is the most diagnostic function describing RT distributions, and (3) the conditional accuracy function allows us to track performance changes over time. Therefore, $h(t)$ and $ca(t)$ functions should become a standard of descriptive statistics. Thorough investigation of them should precede any further analyses, a practice that would reveal the different shapes and qualities of those functions across many different time-to-event data paradigms.

When designing an experiment with the goal of applying discrete-time EHA, the following should be considered. First, a fixed response deadline in each trial is advised. As presented, EHA deals well with right-censored observations. Thus, collecting data beyond a time point after which responses are so slow that other analyses deem them uninformative or meaningless (i.e., they would be trimmed either way), is unnecessary. Response deadlines have the additional advantage that individual distributions will inevitably overlap. This is especially desirable for modeling purposes (Panis & Schmidt, 2016; Wolkersdorfer et al., 2020). For example, a

reasonable response deadline for a simple response priming paradigm (two-button forced-choice) could be 800 ms, since most responses will have occurred way before this, with expected mean RTs between 300-400 ms. Additionally, stimulus presentation should not be determined by RTs. In each trial of a particular condition, present stimuli for exactly the same amount of time and set the trial duration to a fixed time (i.e., the next trial only begins after the response deadline has been reached). This allows for maximal comparability between trials and ensures that averaged trials are in fact a true average. Moreover, this keeps control over the experiment duration in the hands of the researcher.

Second, include as many trials as possible per condition. As discussed, high measurement precision is of utmost importance. More trials allow smaller bin sizes, while still ensuring stable hazard and conditional accuracy estimates. Thus, I generally recommend small- N over large- N designs (see Smith & Little, 2018 for a discussion). In general, more than 100 trials per condition and participant should be achieved. Again, higher measurement precision on the participant level allows modeling from the participant level up. Worries about experiment duration can be neglected at this point. Through response deadlines and fixed trial duration, single sessions can remain short and overall duration stays in the control of the researcher. Alternatively, when feasible, multiple sessions can be planned. Keep in mind, if higher temporal resolution is desired, more trials will necessarily have to be recorded.

Third, if one is interested in time-locking or, in general, the influence of a variable on the shapes of the hazard and conditional accuracy, one should employ sufficient parametric variation of the variable of interest. For example, if one wants to show the time-locked effect in response priming, variation of the SOA is essential. As a general

rule, having at least three parametric variations is advisable, preferably five, a recommendation generalizable to many other experimental designs and statistical methods, since it provides a research design with high internal validity.

6 Outlook

In this present work, I made my case in favor of the analysis of distributional data, in particular through the use of discrete-time EHA. I presented my arguments for a shift from classical methods concerned with mean analysis, described how EHA is performed, showcased its application in three paradigms, discussed its advantages and disadvantages, provided solutions to known issues with the method, compared it with other popular distributional analyses, and ended with recommendations for researchers planning to apply it in the future. It is clear now that RT and accuracy distributions are highly informative. Through EHA, we are able to gain insights into the temporal dynamics of cognitive processes. But where do we go from here?

First, the method as it is right now, can be further developed. For example, in Jubran et al. (2025), we introduce a novel approach in analyzing time-to-event data that utilizes the advantages of discrete-time EHA and generalizes it to trajectory data. Here, we give a concrete example on how to improve measurement precision and how to make the most out of the available data. While discrete-time EHA, as presented in this work, offers insights into the response behavior of subjects, the ongoing process of decision-making is only captured at its final stage, here with a button-press. Jubran et al. (2025) measured hand-movement trajectories in a VR variation of the N-Back task. In our newly developed *Spatiotemporal Survival Analysis* (StSA) we include not just the passage of time but also the movement through space. Thus, this method

makes use of the richness of trajectory data and, as we argue, improves measurement precision at the participant level, capturing the decision-making process from the initiation of movement to its completion. This new method needs further evaluation (e.g., comparing it to other trajectory analyses methods or discrete-time EHA) and application.

Second, more evidence regarding the validity of the method should be collected. For example, earlier I made the recommendation of employing response deadlines. A common concern that researchers have with such a procedure is that it might produce a bias in the data. More specifically, the question arises whether the to-be-expected speed-accuracy tradeoff is troublesome. Indeed, preliminary data suggests that, with shorter deadlines, effects in $h(t)$ decrease in magnitude, while effects in $ca(t)$ increase. However, while such a procedure can lead to a complete erasure of effects in mean RT, hazards remain diagnostic. Instead the observed shift in magnitudes between the two measures (hazard and conditional accuracy) is highly relevant. When no deadline is employed, effects are at a minimum in $ca(t)$, and when the deadline is too short effects are at a minimum in $h(t)$. Thus, depending on the cognitive demand of a given task, we have to identify the sweet spot at which effects are present in both measures. Only then will we get a full picture of the ongoing processes. We should move away from the idea of a tradeoff, which implies that one measure loses and one measure gains. Instead, through the complementary nature of $h(t)$ and $ca(t)$, the analysis of both measures *always* is advantageous.

In order to further validate the method, it simply has to be applied to more experimental data. By exploring the different shapes of hazard and conditional accuracy depending on different experimental paradigms, the field will grow and

learn. Only once we know the possible shapes can we decide on which parametric functions might be suitable to fit to the data and how complex our models need to be (Holden et al., 2009; Townsend & Ashby, 1983; Wickens, 1982).

Finally, applying discrete-time EHA to different paradigms will allow to find more examples of effects concealed by mean RT and accuracy analyses. Effects that are present in mean RT and accuracy will always be present in hazard and conditional accuracy functions as well. However, cases in which effects are only present in $h(t)$ or $ca(t)$, or both, are of high interest to the field. Classical approaches have led to numerous models and theories of human behavior. It is essential to bring them to the test, as some of them are potentially based on false assumptions due to the nature of the performed analyses. In the present work, I have presented just three paradigms, in which classical methods concealed important features of the response behavior of participants (Panis, Moran, et al., 2020; T. Schmidt et al., 2022; Wolkersdorfer et al., 2020). Undoubtedly, there is more to discover out there.

References Frame Text

- Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology*, 13, 61–98. <https://doi.org/10.2307/270718>
- Allison, P. D. (2010). *Survival analysis using SAS: A practical guide, second edition*. SAS Institute Inc.
- Ansorge, U., Kunde, W., & Kiefer, M. (2014). Unconscious vision and executive control: How unconscious processing and conscious action control interact. *Consciousness and Cognition*, 27, 268–287. <https://doi.org/10.1016/j.concog.2014.05.009>
- Austin, P. C. (2017). A tutorial on multilevel survival analysis: Methods, models and applications. *International Statistical Review*, 85(2), 185–203. <https://doi.org/10.1111/insr.12214>
- 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. <https://doi.org/10.1177/0963721411408885>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Biafora, M., & Schmidt, T. (2020). Induced dissociations: Opposite time courses of priming and masking induced by custom-made mask-contrast functions. *Attention, Perception, & Psychophysics*, 82(3), 1333–1354. <https://doi.org/10.3758/s13414-019-01822-4>

- Bloxom, B. (1984). Estimating response time hazard functions: An exposition and extension. *Journal of Mathematical Psychology*, 28(4), 401–420.
[https://doi.org/10.1016/0022-2496\(84\)90008-7](https://doi.org/10.1016/0022-2496(84)90008-7)
- Breitmeyer, B. G., & Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychological Review*, 83(1), 1–36. <https://doi.org/10.1037//0033-295x.83.1.1>
- Breitmeyer, B. G., & Hanif, W. (2008). "Change of Mind" within and between nonconscious (masked) and conscious (unmasked) visual processing. *Consciousness and Cognition*, 17(1), 254–266.
<https://doi.org/10.1016/j.concog.2007.08.001>
- Breitmeyer, B. G., & Ögmen, H. (2007). *Visual masking: Time slices through conscious and unconscious vision* (2. ed., reprinted.). *Oxford psychology series: Vol. 41*. Oxford University Press.
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research. Brain Research Reviews*, 36(2-3), 96–107. [https://doi.org/10.1016/S0165-0173\(01\)00085-6](https://doi.org/10.1016/S0165-0173(01)00085-6)
- Burle, B., Vidal, F., Tandonnet, C., & Hasbroucq, T. (2004). Physiological evidence for response inhibition in choice reaction time tasks. *Brain and Cognition*, 56(2), 153–164. <https://doi.org/10.1016/j.bandc.2004.06.004>
- Cheal, M., & Lyon, D. R. (1992). Attention in visual search: Multiple search classes. *Perception & Psychophysics*, 52(2), 113–138.
<https://doi.org/10.3758/BF03206765>

- Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, 33(1), 269–298.
<https://doi.org/10.1146/annurev.neuro.051508.135409>
- Cummings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), 369–382.
<https://doi.org/10.1177/0267658312443651>
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, 30, 412–431. [https://doi.org/10.1016/0001-6918\(69\)90065-1](https://doi.org/10.1016/0001-6918(69)90065-1)
- Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, 11(5), 14.
<https://doi.org/10.1167/11.5.14>
- Eimer, M., & Schlaghecken, F. (1998). Effects of masked stimuli on motor activation: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 24(6), 1737–1747.
<https://doi.org/10.1037/0096-1523.24.6.1737>
- Eimer, M., & Schlaghecken, F. (2003). Response facilitation and inhibition in subliminal priming. *Biological Psychology*, 64(1-2), 7–26.
[https://doi.org/10.1016/S0301-0511\(03\)00100-5](https://doi.org/10.1016/S0301-0511(03)00100-5)
- 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. <https://doi.org/10.3758/BF03329816>
- Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics*, 25(4), 249–263. <https://doi.org/10.3758/BF03198804>

- Fernández-López, M., Marcet, A., & Perea, M. (2019). Can response congruency effects be obtained in masked priming lexical decision? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(9), 1683–1702.
<https://doi.org/10.1037/xlm0000666>
- Ferrand, L., & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked non-word priming. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 45(3), 353–372.
<https://doi.org/10.1080/02724989208250619>
- Grainger, J. E., Scharnowski, F., Schmidt, T., & Herzog, M. H. (2013). Two primes priming: Does feature integration occur before response activation? *Journal of Vision*, 13(8), 19. <https://doi.org/10.1167/13.8.19>
- Ha, J. C., Kimpo, C. L., & Sackett, G. P. (1997). Multiple-spell, discrete-time survival analysis of developmental data: Object concept in pigtailed macaques. *Developmental Psychology*, 33(6), 1054–1059. <https://doi.org/10.1037/0012-1649.33.6.1054>
- Helmholtz, H. von. (1948). On the rate of transmission of the nerve impulse, 1850. In W. Dennis (Ed.), *Readings in the history of psychology* (pp. 197–198). Appleton-Century-Crofts. <https://doi.org/10.1037/11304-024>
- Helmholtz, H. von. (2021). Über die Methoden, kleinste Zeittheile zu messen, und ihre Anwendung für physiologische Zwecke. In H. Schmidgen (Ed.), *Springer eBook Collection. Hermann von Helmholtz: Versuche zur Fortpflanzungsgeschwindigkeit der Reizung in den Nerven* (pp. 123–135). Springer Spektrum. https://doi.org/10.1007/978-3-662-63833-0_5

- Holden, J. G., van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times reveals cognitive dynamics. *Psychological Review*, 116(2), 318–342.
<https://doi.org/10.1037/a0014849>
- Humphreys, G. W. (2016). Feature confirmation in object perception: Feature integration theory 26 years on from the Treisman Bartlett lecture. *Quarterly Journal of Experimental Psychology*, 69(10), 1910–1940.
<https://doi.org/10.1080/17470218.2014.988736>
- Jaśkowski, P. (2007). The effect of nonmasking distractors on the priming of motor responses. *Journal of Experimental Psychology: Human Perception and Performance*, 33(2), 456–468. <https://doi.org/10.1037/0096-1523.33.2.456>
- Jaśkowski, P. (2008). The negative compatibility effect with nonmasking flankers: A case for mask-triggered inhibition hypothesis. *Consciousness and Cognition*, 17(3), 765–777. <https://doi.org/10.1016/j.concog.2007.12.002>
- Jaśkowski, P. (2009). Negative compatibility effect: The object-updating hypothesis revisited. *Experimental Brain Research*, 193(1), 157–160.
<https://doi.org/10.1007/s00221-008-1700-6>
- Jaśkowski, P., Białuńska, A., Tomanek, M., & Verleger, R. (2008). Mask- and distractor-triggered inhibitory processes in the priming of motor responses: An EEG study. *Psychophysiology*, 45(1), 70–85. <https://doi.org/10.1111/j.1469-8986.2007.00595.x>
- Jaśkowski, P., & Przekoracka-Krawczyk, A. (2005). On the role of mask structure in subliminal priming. *Acta Neurobiologiae Experimentalis*, 65, 409–417.
<https://doi.org/10.55782/ane-2005-1569>

- Jubran, O. F., Wolkersdorfer, M. P., Eymann, V., Burkard, N., Czernochowski, D., Herrlich, M., van Leeuwen, C., & Lachmann, T. (2025). Spatiotemporal survival analysis for movement trajectory tracking in virtual reality. *Scientific Reports*, 15(1), 7313. <https://doi.org/10.1038/s41598-025-91471-5>
- Kahneman, D. (1968). Method, findings, and theory in studies of visual masking. *Psychological Bulletin*, 70(6), 404–425. <https://doi.org/10.1037/h0026731>
- Kelso, J. A. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior and brain coordination. *Neural Networks*, 37, 120–131. <https://doi.org/10.1016/j.neunet.2012.09.003>
- Klotz, W., & Neumann, O. (1999). Motor activation without conscious discrimination in metacontrast masking. *Journal of Experimental Psychology: Human Perception and Performance*, 25(4), 976–992. <https://doi.org/10.1037/0096-1523.25.4.976>
- Klotz, W., & Wolff, P. (1995). The effect of a masked stimulus on the response to the masking stimulus. *Psychological Research*, 58(2), 92–101. <https://doi.org/10.1007/BF00571098>
- Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, 38(3), 557–577. <https://doi.org/10.1017/S0048577201990559>
- Kolers, P. A. (1962). Intensity and contour effects in visual masking. *Vision Research*, 2(9-10), 277–294, IN4. [https://doi.org/10.1016/0042-6989\(62\)90037-8](https://doi.org/10.1016/0042-6989(62)90037-8)
- Kunde, W., Kiesel, A., & Hoffmann, J. (2003). Conscious control over the content of unconscious cognition. *Cognition*, 88(2), 223–242. [https://doi.org/10.1016/S0010-0277\(03\)00023-4](https://doi.org/10.1016/S0010-0277(03)00023-4)

- Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11), 571–579.
[https://doi.org/10.1016/S0166-2236\(00\)01657-X](https://doi.org/10.1016/S0166-2236(00)01657-X)
- 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, 765.
<https://doi.org/10.3389/fnins.2018.00765>
- 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.
<https://doi.org/10.1037/xhp0000156>
- Lingnau, A., & Vorberg, D. (2005). The time course of response inhibition in masked priming. *Perception & Psychophysics*, 67(3), 545–557.
<https://doi.org/10.3758/BF03193330>
- Lleras, A., & Enns, J. T. (2004). Negative compatibility or object updating? A cautionary tale of mask-dependent priming. *Journal of Experimental Psychology: General*, 133(4), 475–493. <https://doi.org/10.1037/0096-3445.133.4.475>
- Luce, R. D. (1986). *Response times. Their role in inferring elementary mental organization*. Oxford University Press.
- Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, 15(2), 197–237.
[https://doi.org/10.1016/0010-0285\(83\)90009-9](https://doi.org/10.1016/0010-0285(83)90009-9)

- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86(4), 287–330. <https://doi.org/10.1037/0033-295X.86.4.287>
- Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, 26(1-3), 3–67. [https://doi.org/10.1016/0301-0511\(88\)90013-0](https://doi.org/10.1016/0301-0511(88)90013-0)
- Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, 35(1), 99–115. <https://doi.org/10.1111/1469-8986.3510099>
- 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(5), 1300–1315. <https://doi.org/10.3758/s13423-015-0978-1>
- Mulder, M. J., & van Maanen, L. (2013). Are accuracy and reaction time affected via different processes? *PLoS ONE*, 8(11), e80222. <https://doi.org/10.1371/journal.pone.0080222>
- Neumann, O. (1990). Direct parameter specification and the concept of perception. *Psychological Research*, 52(2-3), 207–215. <https://doi.org/10.1007/BF00877529>
- Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing* (pp. 41–82). Erlbaum.

- Panis, S. (2020). How can we learn what attention is? Response gating via multiple direct routes kept in check by inhibitory control processes. *Open Psychology*, 2(1), 238–279. <https://doi.org/10.1515/psych-2020-0107>
- 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 and Performance*, 40(1), 129–144. <https://doi.org/10.1037/a0032949>
- Panis, S., Moran, R., Wolkersdorfer, M. P., & Schmidt, T. (2020). 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. *Attention, Perception, & Psychophysics*, 82(2), 689–714. <https://doi.org/10.3758/s13414-019-01897-z>
- Panis, S., & Ramsey, R. (2024, December 17). *Event History Analysis for psychological time-to-event data: A tutorial in R with examples in Bayesian and frequentist workflows*. <https://doi.org/10.31234/osf.io/57bh6>
- 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. https://doi.org/10.1162/jocn_a_00998
- Panis, S., Schmidt, F., Wolkersdorfer, M. P., & Schmidt, T. (2020). Analyzing response times and other types of time-to-event data using event history analysis: A tool for mental chronometry and cognitive psychophysiology. *I-Perception*, 11(6), 1–24. <https://doi.org/10.1177/2041669520978673>

- Panis, S., & Wagemans, J. (2009). Time-course contingencies in perceptual organization and identification of fragmented object outlines. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 661–687.
<https://doi.org/10.1037/a0013547>
- Pieters, J. P. (1983). Sternberg's additive factor method and underlying psychological processes: Some theoretical considerations. *Psychological Bulletin*, 93(3), 411–426. <https://doi.org/10.1037/0033-2909.93.3.411>
- 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.
<https://doi.org/10.1162/0898929053279513>
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 86(3), 446–461. <https://doi.org/10.1037/0033-2909.86.3.446>
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114(3), 510–532. <https://doi.org/10.1037/0033-2909.114.3.510>
- Ridderinkhof, K. R. (2002). Activation and suppression in conflict tasks: Empirical clarification through distributional analyses. In W. Prinz & B. Hommel (Eds.), *Common Mechanisms in Perception and Action: Attention and Performance XIX* (pp. 494–519). Oxford University Press.
<https://doi.org/10.1093/oso/9780198510697.003.0024>
- Rosenbaum, D. A. (1983). The movement precuing technique: Assumptions, applications, and extensions. *Advances in Psychology*, 12, 231–274.
[https://doi.org/10.1016/S0166-4115\(08\)61994-9](https://doi.org/10.1016/S0166-4115(08)61994-9)

- Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review*, 18(2), 324–330. <https://doi.org/10.3758/s13423-011-0055-3>
- Rouder, J. N., & Speckman, P. L. (2004). An evaluation of the Vincentizing method of forming group-level response time distributions. *Psychonomic Bulletin & Review*, 11(3), 419–427. <https://doi.org/10.3758/BF03196589>
- Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron*, 20(2), 185–195. [https://doi.org/10.1016/s0896-6273\(00\)80448-1](https://doi.org/10.1016/s0896-6273(00)80448-1)
- Schmidt, F., & Schmidt, T. (2010). Feature-based attention to unconscious shapes and colors. *Attention, Perception, & Psychophysics*, 72(6), 1480–1494. <https://doi.org/10.3758/APP.72.6.1480>
- Schmidt, T. (2002). The finger in flight: Real-time motor control by visually masked color stimuli. *Psychological Science*, 13(2), 112–118. <https://doi.org/10.1111/1467-9280.00421>
- Schmidt, T. (2014, May 22). *Behavioral criteria of feedforward processing in rapid-chase theory: Some formal considerations*. <http://arxiv.org/pdf/1405.5795>
- Schmidt, T., Niehaus, S., & Nagel, A. (2006). Primes and targets in rapid chases: Tracing sequential waves of motor activation. *Behavioral Neuroscience*, 120(5), 1005–1016. <https://doi.org/10.1037/0735-7044.120.5.1005>
- Schmidt, T., Panis, S., Wolkersdorfer, M. P., & Vorberg, D. (2022). Response inhibition in the Negative Compatibility Effect in the absence of inhibitory stimulus features. *Open Psychology*, 4(1), 219–230. <https://doi.org/10.1515/psych-2022-0012>

- Schmidt, T., & Schmidt, F. (2009). Processing of natural images is feedforward: A simple behavioral test. *Attention, Perception, & Psychophysics*, 71(3), 594–606.
<https://doi.org/10.3758/APP.71.3.594>
- Schmidt, T., & Vorberg, D. (2006). Criteria for unconscious cognition: Three types of dissociation. *Perception & Psychophysics*, 68(3), 489–504.
<https://doi.org/10.3758/BF03193692>
- Schöner G., Spencer, J. P., & the DFT Research Group. (2016). *Dynamic thinking. A primer on dynamic field theory*. Oxford University Press.
- Singer, J. D., & Willett, J. B. (1991). Modeling 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.
<https://doi.org/10.1037/0033-2909.110.2.268>
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.
- Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review*, 25(6), 2083–2101.
<https://doi.org/10.3758/s13423-018-1451-8>
- Song, J.-H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, 13(8), 360–366.
<https://doi.org/10.1016/j.tics.2009.04.009>
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276–315. [https://doi.org/10.1016/0001-6918\(69\)90055-9](https://doi.org/10.1016/0001-6918(69)90055-9)

- Sternberg, S. (1984). Stage models of mental processing and the additive-factor method. *Behavioral and Brain Sciences*, 7(1), 82–84.
<https://doi.org/10.1017/S0140525X00026285>
- Sternberg, S. (2011). Modular processes in mind and brain. *Cognitive Neuropsychology*, 28(3-4), 156–208. <https://doi.org/10.1080/02643294.2011.557231>
- Sternberg, S. (2013). The meaning of additive reaction-time effects: Some misconceptions. *Frontiers in Psychology*, 4, 744.
<https://doi.org/10.3389/fpsyg.2013.00744>
- 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. <https://doi.org/10.1111/j.1467-9280.1990.tb00067.x>
- 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. <https://doi.org/10.1037/0033-2909.108.3.551>
- Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge University Press.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, 38(5), 816–827. <https://doi.org/10.1111/1469-8986.3850816>

- van Gelder, T. (1995). What might cognition be, if not computation? *Journal of Philosophy*, 92(7), 345–381. <https://doi.org/10.2307/2941061>
- van Maanen, L., Katsimpokis, D., & van Campen, A. D. (2019). Fast and slow errors: Logistic regression to identify patterns in accuracy-response time relationships. *Behavior Research Methods*, 51(5), 2378–2389. <https://doi.org/10.3758/s13428-018-1110-z>
- van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, 7(3), 424–465. <https://doi.org/10.3758/BF03214357>
- VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, 2, 365. <https://doi.org/10.3389/fpsyg.2011.00365>
- VanRullen, R., & Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process. *Journal of Cognitive Neuroscience*, 15(2), 209–217. <https://doi.org/10.1162/089892903321208141>
- Vath, N., & Schmidt, T. (2007). Tracing sequential waves of rapid visuomotor activation in lateralized readiness potentials. *Neuroscience*, 145(1), 197–208. <https://doi.org/10.1016/j.neuroscience.2006.11.044>
- Verbaten, M. N., Huyben, M. A., & Kemner, C. (1997). Processing capacity and the frontal P3. *International Journal of Psychophysiology*, 25(3), 237–248. [https://doi.org/10.1016/S0167-8760\(96\)00748-9](https://doi.org/10.1016/S0167-8760(96)00748-9)
- Vincent, S. B. (1912). The Functions of the Vibrissae in the Behavior of the White Rat. *Behavioral Monographs*, 1 (5).
- Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of*

the National Academy of Sciences of the United States of America, 100(10), 6275–6280. <https://doi.org/10.1073/pnas.0931489100>

Wenger, M. J., & Gibson, B. S. (2004). Using hazard functions to assess changes in processing capacity in an attentional cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4), 708–719.

<https://doi.org/10.1037/0096-1523.30.4.708>

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, 58(3), 475–482. <https://doi.org/10.1007/BF03395630>

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41(1), 67–85. [https://doi.org/10.1016/0001-6918\(77\)90012-9](https://doi.org/10.1016/0001-6918(77)90012-9)

Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. Freeman.

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, 8(2), 227–233. [https://doi.org/10.1016/S0959-4388\(98\)80144-X](https://doi.org/10.1016/S0959-4388(98)80144-X)

Wilcox R. (2011). *Introduction to robust estimation & hypothesis testing* (3rd ed.). Elsevier.

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. <https://doi.org/10.1037/0022-006X.61.6.952>

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(1), 41–67. <https://doi.org/10.3102/10769986020001041>

- Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202–238. <https://doi.org/10.3758/BF03200774>
- Wolfe, J. M. (2007). Guided Search 4.0: Current progress with a model of visual search. In W. D. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 99–119). Oxford University Press.
- <https://doi.org/10.1093/acprof:oso/9780195189193.003.0008>
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 419–433.
- <https://doi.org/10.1037/0096-1523.15.3.419>
- Wolfe, J. M., Palmer, E. M., & Horowitz, T. S. (2010). Reaction time distributions constrain models of visual search. *Vision Research*, 50(14), 1304–1311.
- <https://doi.org/10.1016/j.visres.2009.11.002>
- Wolkersdorfer, M. P., Panis, S., & Schmidt, T. (2020). Temporal dynamics of sequential motor activation in a dual-prime paradigm: Insights from conditional accuracy and hazard functions. *Attention, Perception & Psychophysics*, 82(5), 2581–2602. <https://doi.org/10.3758/s13414-020-02010-5>
- Zuur, A. F., & Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*, 7(6), 636–645.
- <https://doi.org/10.1111/2041-210X.12577>

Declaration of authorship

Hiermit versichere ich,

- dass ich die vorgelegte Arbeit selbst angefertigt und alle benutzten Hilfsmittel in der Arbeit angegeben habe,
- dass ich diese Dissertation nicht schon als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht, und
- dass weder die gleiche noch eine andere Abhandlung der Dissertation bei einer anderen Universität oder einem anderen Fachbereich der Rheinland-Pfälzischen Technischen Universität Kaiserslautern-Landau veröffentlicht wurde.

06.02.2025, Kaiserslautern

Maximilian Philipp Wolkersdorfer

Curriculum Vitae

Name: Maximilian Philipp Wolkersdorfer
Anschrift: Erwin-Schrödinger-Straße, Geb. 57-566.1, 67633 Kaiserslautern
Email: m.wolkersdorfer@rptu.de

Beruflicher Werdegang

07/2019 –	Wissenschaftlicher Mitarbeiter FG Allgemeine Psychologie FB Sozialwissenschaften RPTU Kaiserslautern (ehemals TU Kaiserslautern)
11/2016 – 06/2019	Fachbereichsstipendium FG Allgemeine Psychologie & FG Cognitive Neuroscience FB Sozialwissenschaften TU Kaiserslautern
11/2016 –	Promotion (Dr. rer. nat.) FB Sozialwissenschaften RPTU Kaiserslautern (ehemals TU Kaiserslautern)
10/2014 – 10/2016	Studium Cognitive Science (M.Sc.) TU Kaiserslautern
10/2010 – 06/2014	Studium Psychologie (B.Sc.) Universität Koblenz-Landau