



Psychometrics of the continuous mind: Measuring cognitive sub-processes via mouse tracking

Stefan Scherbaum¹ · Maja Dshemuchadse²

Published online: 12 November 2019
© The Psychonomic Society, Inc. 2019

Abstract

The measurement of psychological properties often relies on discrete measures, for example, answers in questionnaires or responses in tasks. This focus on discrete measures neglects information that is present in the process leading to an answer or a response. A method to trace such processes is mouse tracking. Mouse tracking promises to open a continuous window onto the processes leading from a stimulus to a response. However, most mouse-tracking studies fall short of the promise to extract dynamic psychometrically valid markers for the different sub-processes, which are intertwined on the way to the final response. Here we used time-continuous multiple regression (TCMR) to extract dynamic markers for the different sub-processes leading to a response. From these markers, we extracted information about the timing, the duration, and the strength of the influence of the different sub-processes. We evaluated these dynamic measures of sub-processes for their psychometric properties, i.e. reliability, which is a basis for their use in the study of individual differences. Furthermore, we applied these dynamic measures in a group-level study to identify differences in the sub-processes of resolving response conflict between groups performing either a Simon or a flanker task. We found specific temporal patterns that match predictions from a conceptual model of these tasks. We concluded that the extracted information from mouse movements could be used as psychometrically valid dynamic measures of psychological properties and their differences across individuals and situations.

A software toolbox to perform the described analyses in Matlab is provided (osf.io/5e3vn).

Keywords Mouse tracking · Process tracing · Psychometrics · Individual differences · Cognitive control

Introduction

The psychological study of differences between individuals and between different situations usually relies on outcome-

based measures of tests or tasks, that is, choices, ratings, or response times. From these outcome-based measures psychologists make inferences to uncover or quantify underlying constructs. Such a construct could be, for example, cognitive control, i.e., the ability to focus on relevant information in the face of distraction, which might in turn be measured in a Stroop task (Stroop, 1935) or a Simon task (Simon, 1969) by response time differences between congruent and incongruent trials. The implicit assumption of this approach is that the outcome measure tells us something about the process that led to the final response. However, the inference from the outcome on the process is based on a single measurement. Much more information about the decision process becomes available when we use process-tracing measures, such as eye tracking or mouse tracking, the tracing of a person's computer mouse movements during the decision (Koop & Johnson, 2011; Spivey & Dale, 2006; Spivey, Grosjean, & Knoblich, 2005). Here, we use time-continuous multiple regression to exploit the full potential of process tracing using mouse tracking, and extract individual process markers from mouse movements. Such markers could be used for both statistical group-level analyses of subtle process-related differences between

Public significance statement Our study shows that we get deeper insight into the difference between persons and situations from looking at how persons give responses in psychological tasks. We measured this “how” by tracking participants’ computer mouse movements while they decided on one of two response boxes on the screen. In contrast to previous studies, we provide here a method that allows for analyzing the movements statistically to estimate the reliability of the found differences.

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13421-019-00981-x>) contains supplementary material, which is available to authorized users.

✉ Stefan Scherbaum
Stefan.Scherbaum@tu-dresden.de

¹ Department of Psychology, Technische Universität Dresden, Zellescher Weg 17, 01062 Dresden, Germany

² Department of Social Sciences, Hochschule Zittau-Görlitz, Brückestraße 1, Görlitz 02826, Germany

conditions and for the study of individual differences in the processes that lead to responses. For all steps of analysis described here we provide a toolbox of Matlab functions for download (Scherbaum, 2017).

Mouse tracking gained momentum in recent years (Dale, Kehoe, & Spivey, 2007; Dshemuchadse, Scherbaum, & Goschke, 2012; Kieslich & Hilbig, 2014; Koop & Johnson, 2011; McKinstry, Dale, & Spivey, 2008; Scherbaum, Dshemuchadse, Fischer, & Goschke, 2010; Spivey & Dale, 2006; Sullivan, Hutcherson, Harris, & Rangel, 2015). While mouse tracking is in the tradition of other reach-tracking methods (e.g., Buetti & Kerzel, 2009; Song & Nakayama, 2009), the ease of implementation and widespread use of computer mice allows for a cheap and easy to implement method of process tracing (Schulte-Mecklenbeck, Kuehberger, & Ranyard, 2011). In a typical mouse-tracking paradigm, participants indicate their response by using a computer mouse. For example, they might have to work on a Simon task (Simon, 1969) and are instructed to respond to the direction of an arrow (left/right pointing) presented on two different positions on the screen (left/right side). Hence, in this task, the relevant information (the direction of the arrow) might interfere with the irrelevant information (the location of the arrow). Normally, participants respond in this task via a left or right key-press, which allows for measuring response times. This yields the so-called Simon-effect: Participants are faster when direction and location of the arrow correspond (so-called congruent trials) than when direction and location of the arrow do not correspond (so-called incongruent trials). When using mouse tracking, participants indicate their response by moving a computer mouse from a starting field in the bottom-center of the screen to pre-defined choice-fields in the upper-left and upper-right corners of the screen. Mouse-tracking studies assume that the choice process continuously leaks into the choice movements of participants, allowing the choice process to be traced within a trial/item while participants move from the starting field to the final choice-field (Spivey & Dale, 2006; but see Fischer & Hartmann, 2014). In the Simon task, this leads to relatively direct movements in congruent trials and movements showing a deflection to the incorrect choice-field in incongruent trials (Scherbaum et al., 2010; Scherbaum, Frisch, Dshemuchadse, Rudolf, & Fischer, 2016).

Using process tracing to investigate the choice process over time should in principle allow for more than only studying deflections – it should allow for studying individual and situational differences in the (sub-)processes leading to a choice. Such differences might show up in the strength, duration, or timing of sub-processes which, in turn, offers new markers for the study of differences between individuals or situations. However, most mouse-tracking studies focus on static measures to quantify mouse movements, for example, the average deflection of a movement to the unchosen alternative or the maximum deviation of the movements (Freeman & Ambady,

2010). By focusing on such static measures, these studies ignore the precise dynamics of sub-processes that might be hidden in mouse movements and gain little more than could be found by the analysis of response-time data. Here we show how to fully gain the advantage from analyzing mouse movements. To analyze the temporal patterns of different sub-processes we use an approach that bears similarities to the methods of analysis applied to neural data from fMRI: A general linear model is applied coding the different trial properties to each time point of the mouse movements (compared to spatial points of the BOLD signal). This procedure results in time-varying beta-weights¹ that indicate which trial properties, and in turn which related potential sub-processes, influence the mouse movement at which point in time to which extent. It hence comprises a full temporal analysis (in contrast to a spatial analysis in FMRI) of all sub-processes tapped by different trial properties. We termed this form of analysis time-continuous multiple regression analysis (TCMR; e.g., Scherbaum, Dshemuchadse, Leiberg, & Goschke, 2013). In the Simon task, this approach allowed us to study the temporal profiles of at least three sub-processes (Scherbaum et al., 2010, 2016): First, the interference from the irrelevant information (the Simon effect), second how this interference changes depending on the congruency of the previous trial (so-called congruency sequence effects), and third how responding is influenced by the response in the previous trial (the so-called response bias).

However, early applications of TCMR (Dshemuchadse et al., 2012; Scherbaum et al., 2010; Sullivan et al., 2015) posed two challenges for the statistical analysis of temporal patterns. The first challenge is that mouse-movement data show a reasonable amount of noise, which makes peak detection (peak strength and timing) based on individual data error prone. This difficulty is typical for many forms of dynamic data, for example, lateralized readiness potentials, and is often solved by statistical methods, for example, analyses based on jack-knifing (Miller, Patterson, & Ulrich, 2001). Such methods, however, come at the cost of restricting statistical analyses of peak data to the group level. Hence, situational differences could be studied on the group level, but the analysis of individual differences is hampered. Furthermore, jack-knifing works on averaged data, and, hence, smearing artefacts can occur due to different peak curves of individual subjects.

The second challenge is that for detecting coherent temporal segments of activity in the beta-weights, one tests these beta-weights across participants (Scherbaum et al., 2010) for every time step of the movement data. This leads, again, to the problem that identified segments are defined at the group level and further statistical analysis is not possible – neither inferential statistics on the group level nor on the individual level. Furthermore, the multiple testing of consecutive time steps poses the problem of how to correct for multiple comparisons,

a problem that until now had to be solved by Monte Carlo simulations determining correction criteria (Dale et al., 2007; Scherbaum, Gottschalk, Dshemuchadse, & Fischer, 2015).

Here, we extend the original approach. The extension rests on the observation that the temporal profiles as reflected in time-varying beta-weights roughly follow a Gaussian shape with their initial positive main component (see Fig. 1). This positive main component is followed by a compensatory negative component. We call this negative component *compensatory* since it is a necessary consequence of the spatial setup forcing participants to reach the response box to give their response. As an example (see Results of Study 1 and Fig. 7), we assume that the correct response box in a Simon task trial is on the right side. Hence, incongruent trials will lead to an initial movement to the (incorrect) left side (the initial effect of irrelevant information). This initial movement will have to be corrected by a strong rightwards movement (the consequence of processing the relevant information) so that the cursor finally reaches the correct response box on the right. In contrast, congruent trials will lead to an initial movement to the (correct) right side. This initial movement will then be followed by a further but relatively weak movement to the right side since the

correct response box is almost at reach already. Since the regressors for the interference effect are coded in a way that a positive component in beta-weights mirrors the initial impact of the irrelevant information, the beta-weights will show an initial positive component (the initial movement to the left or to the right) followed by a negative component (the later movement to the right, which was either large or small). The negative component is hence a direct consequence of the positive component and can be ignored for our purposes.²

We will hence fit Gauss curves to the positive main components of the time-varying beta-weights and use the parameters defining the Gauss curve, i.e. peak time (mean of Gauss curve), peak strength (peak height of Gauss curve), and peak width (SD of Gauss curve) as markers of the dynamic process. In contrast to our original approach, this addition will allow for, first, the extraction of parameters representing the temporal properties of each sub-process for each individual participant and, second, the *statistical comparison* of temporal profiles between different situations.

In the following, we first examine the psychometric properties of the extracted parameters in data that stem

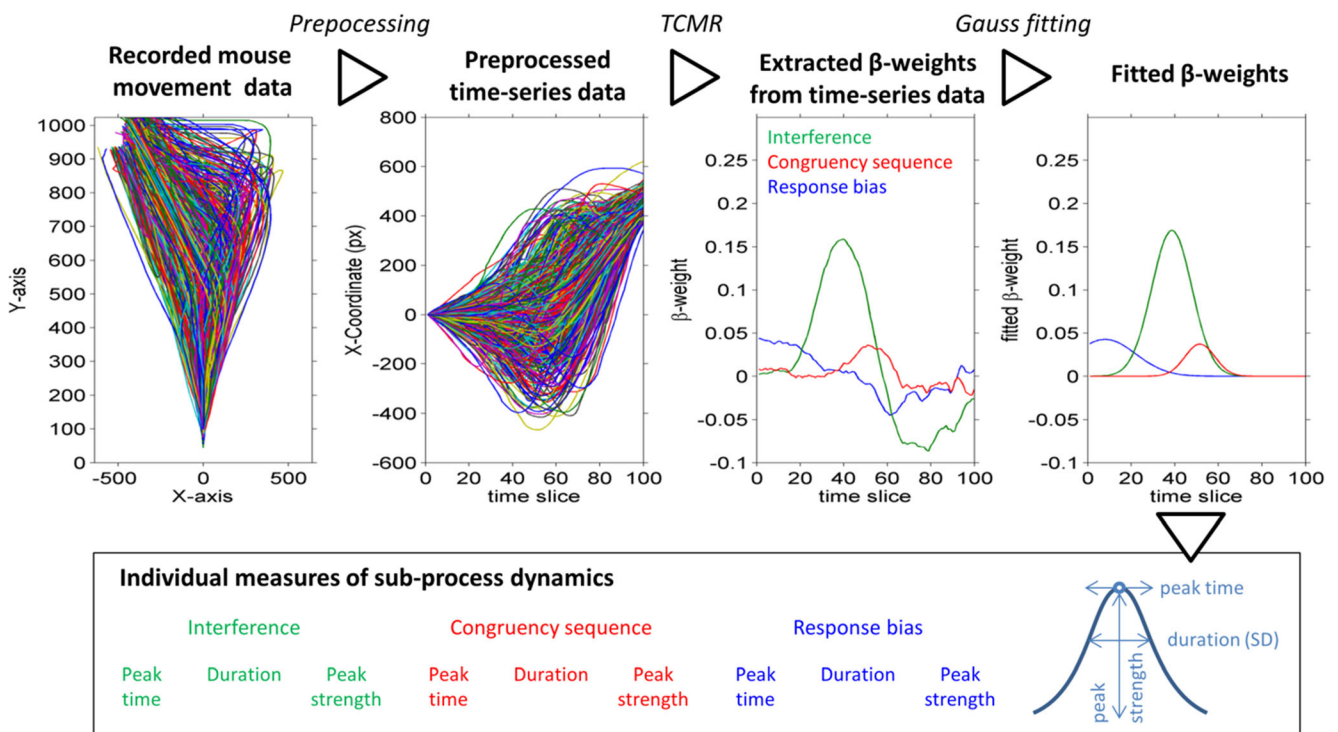


Fig. 1 The processing steps in analyzing mouse-movement data via time-continuous multiple regression (TCMR) and Gaussian fitting, illustrated for the data of one participant from Study 1 (a Simon task). Raw data are pre-processed first, yielding time-normalized time-series data that are transformed into movement angles relative to the X-axis to provide a continuous measure of instantaneous movement tendency. On these angular time-series data, we apply TCMR to gain time-continuous beta-weights representing the influence of different trial properties: For each time-step, multiple regression analyzes the relationship between the trials'

properties and the current mouse-movement angle. These beta-weights hence represent the time-continuous influence of different trial properties (here the three properties are interference, congruency sequence, and response bias; for more information on the properties, see main text) and, in turn, the related sub-processes on the decision process. To quantify the dynamics of these noisy individual beta-weights, Gauss curves are fitted to the positive main components of the raw beta-weights, yielding individual measures of the timing, duration, and strength of each sub-process

from a dynamic version of the Simon task (Scherbaum et al., 2010). The Simon-effect has been used previously to study inter-individual differences in cognitive control, that is, how well a person can shield the response-selection process from interference by irrelevant information of where the stimulus appears. We investigate how reliable the extracted dynamic measures of this interference, congruency sequence effects, and the response bias are for the study of individual differences.

As a second step, we investigate the potential of the method for studying differences between related sub-processes in different situations. We study differences between two cognitive control tasks, i.e., the afore-mentioned Simon task (Simon, 1969) and the flanker task (Eriksen & Eriksen, 1974). In the latter task, participants have to respond to a target stimulus, which is surrounded by distracters that can either indicate the same response – again called congruent trials – or the opposite response – again called incongruent trials. It is an open question how far the cognitive control processes in the Simon task and the flanker task are similar or different.

We provide a complete toolbox of functions for Matlab including all the steps of analysis presented here. Since the toolbox provides not only the TCMR functions, but also further basic pre-processing functions, it could be seen as a Matlab-based complement to similar R-based toolboxes (Kieslich, Wulf, Henninger, Haslbeck, & Schulte-Mecklenbeck, 2017). The article (and the tutorial in the toolbox), in turn, could also be used as a manual on how to perform temporal analyses of mouse movements in Matlab.

Study 1

In the first study we examine the psychometric properties of the extracted parameters, i.e., split-half reliability. We study dynamic markers of the Simon effect and the congruency sequence effects (changes in the Simon effect depending on conflict in the previous trial; Botvinick, Braver, Barch, Carter, & Cohen, 2001; Gratton, Coles, & Donchin, 1992; but see Egner, 2007; Mayr, Awh, & Laurey, 2003). Furthermore, in a previous study (Scherbaum et al., 2010), we had found an early influence of the previous response, so that movements tended initially to the previously chosen direction (response bias). We include this response bias in our analyses.

Whereas in the original study analyses were limited to the group level and stayed descriptive for the dynamics of mouse movements, we now analyze the data using TCMR with Gaussian fitting and analyze split-half reliability of the extracted parameters and their correlation with response time (RT) indicators of the abovementioned sub-processes, namely the interference, congruency sequence, and response bias.

Method

Participants

The data used in this study comprise data from an already published study (Study 2 from Scherbaum et al., 2010) and data newly acquired with the same paradigm. Overall, 72 students (58 female, mean age = 23.36 years, $SD = 3.75$) of Technische Universität Dresden took part in the whole study. Similar selection criteria and procedures were followed in the original and the new study. All participants had normal or corrected-to-normal vision. The study was performed in accordance with the guidelines of the Declaration of Helsinki and of the German Psychological Society. Ethical approval was not required since the study did not involve any risk or discomfort for the participants. All participants were informed about the purpose and the procedure of the study and gave written informed consent prior to the experiment. They received class credit or 5 € payment.

Assuming a minimal acceptable correlation of $r = 0.6$ for reliability, the sample size of 72 participants provided a power of 0.99 (Faul, Erdfelder, Lang, & Buchner, 2007).

Apparatus and stimuli

Target stimuli were presented in white on a black background on a 17-in. screen running at a resolution of $1,280 \times 1,024$ pixels (75-Hz refresh frequency). Target stimuli were numbers (1–4: left response; 6–9: right response). They had a width of 6.44° and an eccentricity (center of stimulus to center of screen) of 20.10° . In both studies, response boxes (11.55° in width) were presented at the top left and top right of the screen. As presentation software, we used Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997) in Matlab 2006b (the Mathworks Inc., Natick, MA, USA), running on a Windows XP SP2 personal computer. Responses were carried out by moving a standard computer mouse (Logitech Wheel Mouse USB). Mouse trajectories were sampled with a frequency of 92 Hz and recorded from stimulus presentation until response in each trial.

Procedure and design

Participants were instructed to respond to the direction indicated by the target stimulus by moving a computer mouse into the left or right response box. Each trial consisted of three stages (see Fig. 2): the alignment stage, the start stage, and the response stage. In the alignment stage, participants clicked into a red box (11.55° in width) at the bottom of the screen within a deadline of 1.5 s. This served to align the starting area for each trial. After clicking within this box, the start stage began and two response boxes at the right and left upper corner of the screen were presented. Participants were

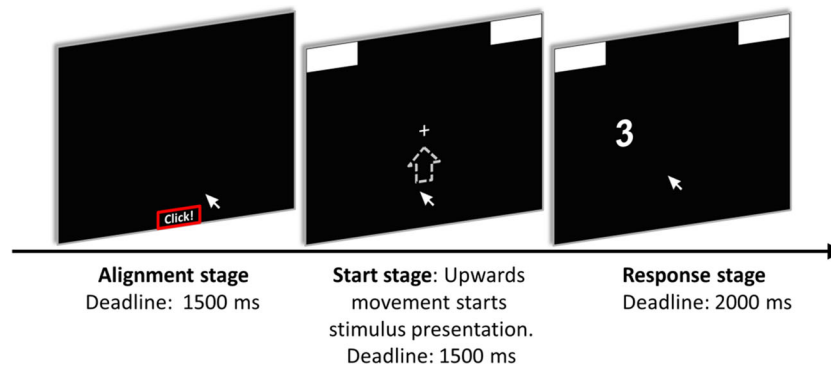


Fig. 2 Setup of the study for number of target stimuli for the three stages of a trial. In the alignment stage participants clicked with the mouse cursor into a red box at the bottom of the screen. This triggered the start stage, in which response boxes appeared at the upper edge of the screen and participants had to move the cursor upwards (as indicated by the dashed upwards arrow, which was not shown to participants) in order to

trigger the next stage. After reaching a movement threshold, the response stage began: the target stimulus (here the number 3, indicating a left response since it is smaller than 5) was presented and participants moved the mouse cursor to the left or the right response box according to direction indicated by the target stimulus

required to start the mouse movement upwards within a deadline of 1.5 s. We chose this procedure forcing participants to be already moving when entering the decision process to assure that they did not decide first and only then execute the final movement (Dshemuchadse et al., 2012; Scherbaum, Fischer, Dshemuchadse, & Goschke, 2011; Scherbaum & Kieslich, 2018). Hence, only after moving at least 4 pixels in each of two consecutive time steps the response stage started: The target stimulus was presented and participants responded by choosing the respective response box. The trial ended after moving the cursor into one of the response boxes within a deadline of 2 s (see Fig. 1). If participants missed the deadline of one of the three stages, the next trial started with the presentation of the red start box. RTs were measured as the duration of the third stage, reflecting the interval between the onset of the target stimulus and reaching the response box with the mouse cursor.

After onscreen instructions and demonstration by the experimenter, participants practiced 40 trials (10 trials with feedback and no deadline for any stage of a trial, 10 trials with feedback and deadline and 20 trials without feedback and with deadline).

The experiment consisted of three blocks and 257 trials per block. We varied the following independent variables: for the current trial, $number_N$ (1–4: left/6–9: right) and $location_N$ (left/right), and for the previous trial, $number_{N-1}$ (1–4/6–9) and $location_{N-1}$ (left/right). This resulted in 16 combinations for the current trial (eight numbers \times two locations) and 16 combinations for the previous trial. The sequence of trials was balanced within each block by pseudo randomization. This resulted in a balanced $Trial_N$ (16) \times $Trial_{N-1}$ (16) \times repetition (3) transition matrix. Concerning congruency of response direction and stimulus location (which leads to the Simon effect and the congruency sequence effects across trials), we hence obtained a balanced sequence of trials with systematically manipulated congruency of direction/location within the

current trial ($congruency_N$) and congruency of direction/location within the previous trial ($congruency_{N-1}$).

Data pre-processing

We excluded erroneous trials, in which participants chose the wrong response box, trials following an error, and trials not fitting the RT outlier criterion of an $RT > 4 SD$ and an $RT < 100$ ms (9.87%, $SD = 8.6\%$). To estimate reliabilities, we used split-half reliability and partitioned the data set into two subsets, i.e., odd and even trials.

Mouse trajectories were aligned for common starting position (horizontal middle position of the screen, 640 pixels). Each trial's movement trajectory was normalized to 100 equal time slices (Spivey et al., 2005) by segmenting each trajectory into 100 equal segments from the first to the last sample of the trajectory using linear interpolation. For analysis of movement dynamics, we focused on the trajectory angle on the XY plane.³ Trajectory angle was calculated as the angle relative to the Y-axis for each difference vector delta-X and delta-Y between two time steps. In other words: For each time slice, we calculated the instantaneous direction of the mouse cursor relative to the y-axis, yielding one value that summarizes the movement on the XY plane. This measure has two advantages over the raw trajectory data. First, it better reflects the instantaneous tendency of the mouse movement since it is based on a differential measure compared to the cumulative effects in raw movement data. Second, it integrates the movement tendency on the XY plane into a single measure. Notably, this procedure also allows for calculating movement velocity. While velocity can also be a valuable source of information, it shows in our experience very similar profiles across conditions in the Simon task, which is why we focus on the trajectory angle in the following. We prepared the temporal analyses described in the next step by introducing temporal correlations between the single data points by convoluting the data over

time with a 10-point Gaussian smoothing window.⁴ Based on this movement angle, we performed TCMR and Gaussian fitting.

Time-continuous multiple regression (TCMR) TCMR follows a procedure of three steps. In the first step, we coded for each participant three predictors for all trials. To better understand this coding step, it is helpful to conceive of the mouse-movement angle as showing positive numbers when the mouse moves to the correct response box and negative numbers when the mouse moves to the incorrect response box. Hence, all predictors will be coded so that when an influence supports the correct response, it will be positive and when an influence supports the incorrect response, it will be negative. The first predictor, *interference*, coded whether the irrelevant location information pointed to the correct or the incorrect response box, which is the Simon effect; the second predictor, *response bias*, indicated whether the previous trials response pointed to the now-correct or the now-incorrect response box; the third predictor, *congruency sequence*, coded whether the current trial's congruency ($congruency_N$) was the same as the previous trials congruency ($congruency_{N-1}$), which represents congruency sequence effects. Hence, it codes how strongly the mouse trajectory would be influenced by *interference* depending on previously induced conflict. To provide comparable beta-weights in the next step, we normalized the predictors to a range -1 and 1. In the third step, we calculated multiple regressions with the normalized predictors on the data from each time slice of the trajectory angle (100 time slices \rightarrow 100 multiple regressions), which had also been normalized for each participant to a range from -1 to 1. This yielded three time-varying beta-weights (three weights \times 100 time slices) for each participant (please see the Appendix for a tutorial of how to run this analysis with the respective Matlab functions). In the original study, we detected significant temporal segments of influence by calculating *t*-tests against zero for each time step of the three time-varying beta-weights. According to Monte Carlo analyses, correction for multiple comparisons in this procedure could be achieved by only accepting segments of more than 10 consecutive significant *t*-tests (Dale et al., 2007; Scherbaum et al., 2015). Here however, we proceeded differently by applying Gaussian fitting.

Gaussian fitting For each time-varying beta-weight of each participant, we fitted a Gauss curve by minimizing the summed squared error for the beta-weight series via a bounded version of the simplex algorithm supported by Matlab (D'Errico, 2012). The parameters of the Gauss curve were its peak time, its duration (the standard deviation), and its peak strength (the height of the Gauss curve at peak time). The algorithm uses estimated parameter bounds and starting values that are based on the grand average of each beta across participants. It first estimates the population peak time and

peak strength from the grand average and consecutively estimates duration by fitting a Gauss curve to the grand average. Based on this initial estimation procedure, it constrains individual peak time to the estimated population peak time \pm 50% of the estimated duration. It constrains individual duration to the estimated population duration \pm 50%. And it constrains individual peak strength to the estimated population peak strength \pm 2.57 SD (99%) of individual peak strengths at the estimated population peak time. R^2 values to estimate fit quality were calculated as correlations of each empirical beta and the fitted Gauss curve. This fit was calculated on 2.57 times the width of the Gauss curve (99% of time points under the Gauss curve).

RT indicators To have a benchmark for the reliability of the parameters from the TCMR analysis, we calculated RT indicators of the three sub-processes of interest. For the *response bias* from the previous trial, we calculated the advantage of repeated responses over alternating responses, which is the contrast $RT_{\text{response-switch}} - RT_{\text{response-repetition}}$. For *interference*, we calculated by how much congruent trials were faster than incongruent trials (the Simon effect), which is the contrast $RT_{\text{incongruent}} - RT_{\text{congruent}}$. For *congruency sequence*, we calculated whether the Simon effect was larger after congruent trials than after incongruent trials (congruency sequence effects indicating conflict adaptation), which is the contrast $Simon_effect_{\text{congruent}N-1} - Simon_effect_{\text{incongruent}N-1}$.

Calculation of statistics All data pre-processing and calculation of statistics were performed in Matlab 2010a (The Mathworks Inc.), using the standard functions of Matlab's Statistics Toolbox.

Results

The analyses of RTs and static mouse measures (average deviation of mouse movements) showed the typical Simon effect and the expected congruency sequence effects as reported in the original publication (Scherbaum et al., 2010). Here, we focus on the results of TCMR and Gaussian fitting with respect to feasibility and reliability.

TCMR and Gaussian fitting

The results of TCMR show the distinct temporal patterns of influences for both sub-sets of data that we created for the analysis of reliability, i.e., odd and even trials. A first peak of *response bias*, followed by the peak of *interference* and then the peak of *congruency sequence* (see Fig. 3).

We applied the classic jack-knifing procedure⁵ as in the original study to compare the results to those of the new Gaussian fitting method. The results can be seen in Table 1.

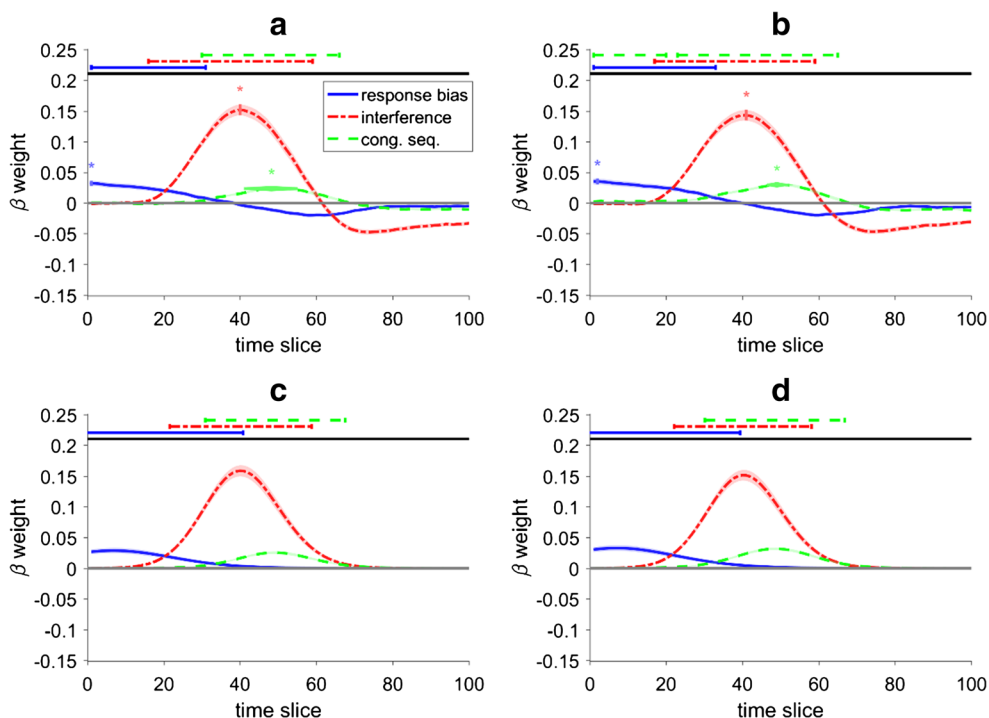


Fig. 3 A and B: Time-continuous beta-weights from time continuous multiple regression for odd trials (A) and even trials (B). Lines above graphs mark significant segments determined by t-test against zero. Curve peaks were determined by a jack-knifing procedure (see main text).

C and D: Reconstructed beta-weights from the Gaussian fitting procedure for odd trials (C) and even trials (D). Lines above graphs indicate the area of 1 SD around the peak of the fitted Gauss curve. In all graphs, shaded areas indicate the standard error of the mean

Fitting quality (R^2) was best for *interference* and *congruency sequence*, and slightly weaker for *response bias*. The spread of parameters did not show any floor or ceiling effects, indicating that the estimation procedure worked correctly (see Fig. 4).

A representative fit to a single subject’s data can be seen in Fig. 5 (subject 2; for graphs of all subjects, please see Supplementary Material).

Reliability of estimated parameters

To check for reliability of the parameters from Gaussian fitting, we calculated split-half reliability for all parameters, which are correlations between odd and even trials (Table 2; for scatter

diagrams please see Supplementary Material). To warrant the assumptions of correlation analysis and avoid outliers driving reliability, we excluded outliers within the parameters ($< >3SD$). Furthermore, we excluded outliers that showed very low values in peak strength ($< 3SD$) since this indicates no peak at all and hence invalid values for the parameters peak time and duration. Notably, this procedure resulted in no exclusions for *response bias* and the peak time and peak strength of *interference*, one exclusion for duration of *interference*, and three exclusions for all parameters of *congruency sequence*. (Notably, including all participants did not change the results qualitatively – see Supplementary Material.)

Split-half reliability was good for the peak time of *interference* and very good for the peak strength of *response*

Table 1 Results of peak detection via jack-knifing and of Gaussian fitting of beta-weights for odd and even trials

Beta	Jack-knifed peaks		Parameters from Gaussian fitting				
	Peak time (SE)	Peak strength (SE)	Peak time (SE)	Peak strength (SE)	Duration (SE)	R^2 (SE)	
Odd trials	response bias	1 (0)	0.032(0.004)	7.692 (0.797)	0.031(0.005)	16.896(0.629)	0.738(0.032)
	interference	40 (0)	0.152(0.009)	40.156(0.473)	0.171(0.009)	9.496(0.188)	0.954(0.009)
	congruency sequence	48.194 (6.828)	0.024(0.002)	49.216(0.58)	0.029(0.002)	9.354(0.296)	0.791(0.024)
Even trials	response bias	2 (0)	0.035 (0.005)	7.627(0.872)	0.036(0.005)	16.209(0.597)	0.776 (0.066)
	interference	41(0)	0.143 (0.009)	40.115(0.451)	0.163(0.009)	9.185 (0.157)	0.947(0.01)
	congruencysequence	49.014 (2.217)	0.032 (0.002)	48.497(0.6)	0.037 (0.002)	9.353(0.294)	0.802(0.025)

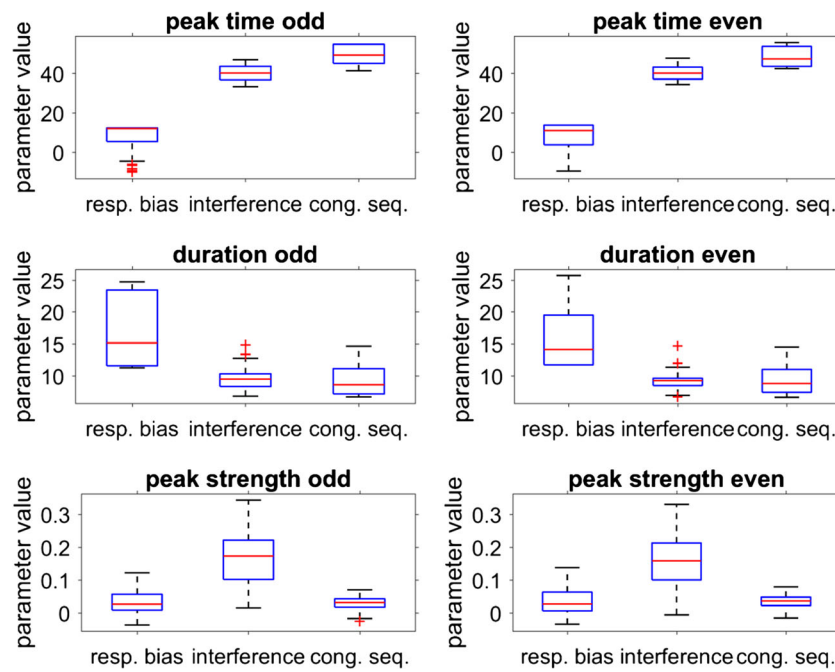


Fig. 4 Spread of parameters from Gaussian fitting for the three betas from TCMR, for odd (left) and even (right) trials. The parameters of the Gauss curve were its peak time, its duration (the standard deviation), and its peak strength (the height of the Gauss curve at peak time)

bias and *interference*. For the peak time of *response bias* and the duration of *response bias* and *interference*, we found lower but significant correlations, while we found only low correlation for the peak strength of *congruency sequence*, marginal correlation for the peak time of *congruency sequence*, and no significant correlation for the duration of *congruency sequence*. To see whether the weak results for *congruency sequence* might stem from a relatively unstable process instead of a weakness in the Gaussian fitting procedure, we checked the benchmark split-half reliabilities of RT measures of *response bias*, *interference*, and *congruency sequence*. As the results in Table 3 indicate, *response bias* and *interference* show fair reliability while *congruency sequence* does not show any correlation.

In summary, this indicates that TCMR combined with Gaussian fitting can in principle produce good reliability for stable sub-processes – namely *response bias* and *interference* – while other sub-processes seem to be unstable on the individual level – namely *congruency sequence*.

We finally pursued two exploratory questions of interest to check the validity of using the extracted parameters for psychometric purposes. The first question asked was what was the number of trials necessary to achieve acceptable levels of reliability. To this end, we analyzed the relationship of trial-number and split-half reliability based on a resampling approach. For a selected number of trials (20, 30, 50, 90, 170, 330), we randomly sampled a sub-set of odd and even trials so that all cells of the design matrix were filled equally. We then

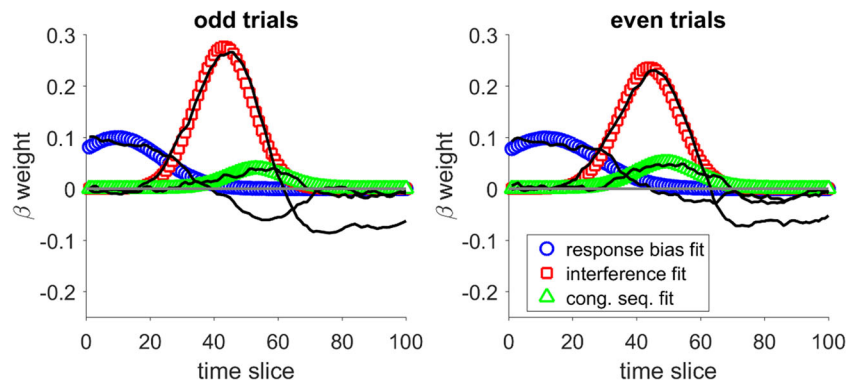


Fig. 5 Data fit for one representative subject (see Supplementary Material for all subjects). Black lines: Time continuous beta-weights from time-continuous multiple regression for odd trials (left) and even trials (right).

Colored markers: reconstructed beta-weights from the Gaussian fitting procedure

Table 2 Correlations (Pearson's r and p values) between odd and even trials and split-half reliability (Spearman-Brown corrected correlations r_c) for the parameters from Gaussian fitting

	Peak time			Duration			Peak strength		
	r	p	r_c	R	p	r_c	r	p	r_c
Response bias	0.58	< .001	0.734	0.603	< .001	0.752	0.96	< .001	0.98
Interference	0.877	< .001	0.934	0.611	< .001	0.759	0.97	< .001	0.985
Congruency Sequence	0.227	0.059	0.37	0.084	0.487	0.155	0.451	< .001	0.622

calculated correlation between the sampled odd and even trials. For each number of trials, this procedure was repeated 50 times, yielding 50 correlation values for each number of trials. Figure 6 shows the resulting curves.

For *response bias* and *interference*, peak strength approaches reliability levels of 0.8 within 90 trials, which is quicker than for the RT measures, while the rise of reliability for the peak duration and peak time follows RT more closely. For *congruency sequence*, it shows the expectable low reliabilities across all numbers of trials, with mentionable reliability for peak strength only appearing with more than 300 trials. The second question asked was whether it was valid to ignore the late negative parts of the analyzed mouse movements – or more clearly, whether these parts only represented compensatory movements that participants had to perform to finally reach the response box. If this was the case, we expected a strong correlation for the relevant positive peak in mouse movements and the later negative peak. For example, in incongruent trials, the stronger the initial deflection to the wrong response box, the stronger the correction needed to be so that the cursor finally reached the correct response box. In contrast, in congruent trials, the stronger the ignition deflection to the correct response box, the weaker the later movement need to be to finally arrive at the correct response box (see Fig. 7, left).

To check this assumption, we calculated the difference in mouse movement angles for congruent and incongruent trials, which shows the same structure as the interference beta-weights (see Fig. 7, middle). We then calculated the area under the curve for the early positive component and the late negative component and correlated these two scores, yielding a good correlation of $r = -.76$, $p < 0.001$ (see Fig. 7, right).

Table 3 Correlations (Pearson's r and p values) between odd and even trials and split-half reliability (Spearman-Brown corrected correlations r_c) for effects in response time (RT)

	RT		
	r	P	r_c
Response bias	0.708	< .001	0.829
Interference	0.752	< .001	0.858
Congruency sequence	0.029	0.808	0.056

This confirms our assumption that the late component indeed represents a compensatory movement necessary to reach the response box and that it does not provide decisive information about the decision process.

Discussion

In Study 1, we investigated the reliability of the parameters extracted from mouse movements via TCMR and Gaussian fitting for three sub-processes in a Simon task: the influence of the previously performed response (response bias), the influence of the location information (interference), and the adaptation of control as reflected in congruency sequence effects. We found that reliability was good for the first two sub-processes, but was overall weak to non-existent for conflict adaptation. Since a similar pattern was present for the respective RT indicators, we concluded that conflict adaptation is a process that is unstable across time within individuals and, hence, when leaving group-level analyses. This finding was not a specific phenomenon for the mouse-movement parameters and fits recent evidence that conflict adaptation might indeed be a temporally fragile construct (Feldman & Freitas, 2016).

Taken together, the extraction of dynamic mouse parameters opens the possibility to study inter-individual differences in markers of sub-processes within a task. It hence fulfils the first aim of the work, enabling future studies to identify relationships between dynamic parameters and individual properties/abilities. In the next study, we show how the extracted dynamic markers could provide insight into differences between cognitive processes in different situations on the group-level.

Study 2

In Study 2, we compared data from a mouse-tracking version of the flanker task (yet unpublished data) with the mouse-tracking data from a Simon task (Scherbaum et al., 2010, Study 1). We aimed to compare the parameters for the previous response, the influence of interference and congruency sequence effects to identify whether these sub-processes work differently in the flanker and the Simon task, similar to

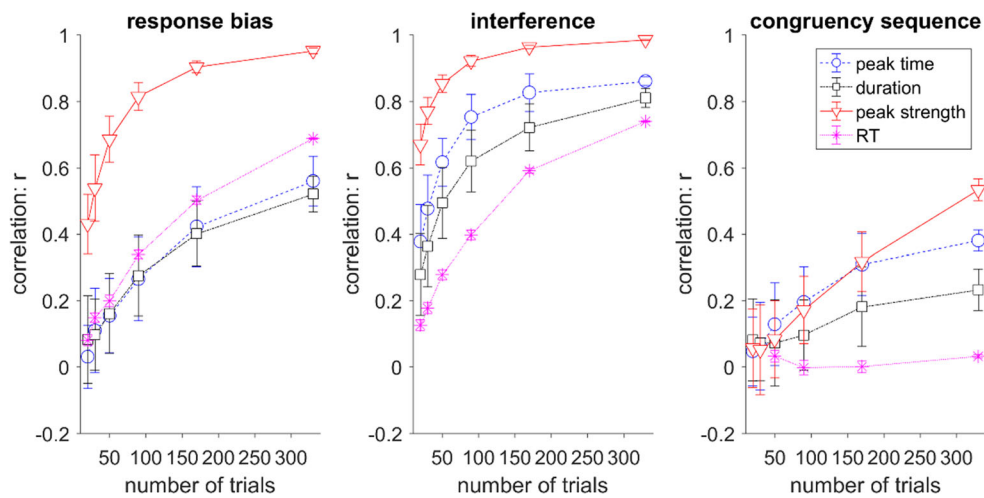


Fig. 6 Resampled split-half reliability (Pearson's r) for each of the three sub-processes, the three parameters, and response time effects as a function of the number of trials

differences that had been studied via distributional analyses between the Stroop and the Simon task (Pratte, Rouder, Morey, & Feng, 2010). The nature of such differences is important on two levels, a theoretical and a measurement level. On a theoretical level, they indicate that different cognitive processes are tackled by different tasks: Though two tasks might superficially tackle the same processes, a deeper process-oriented investigation can provide evidence for distinct processes. In the case of the Simon and the flanker task, both tasks are used widely to study cognitive control in an interchangeable way. Given the different nature of interference in the task, the interchangeable nature of the tasks should not be taken as a given. This different nature is already evident when we look at a simple conceptual model of both tasks (compare Scherbaum et al., 2016), as shown in Fig. 8. In the Simon task, the influence of the distracting information – the location of the arrow – has been proposed to trigger an early automatic response via a fast route. This initial response

impulse decays by itself after the initial peak (Hommel, 1994; Scherbaum et al., 2016; Stürmer, Leuthold, Soetens, Schroter, & Sommer, 2002). Hence, interference comes from the residual activation of this automatic response and the activation of the correct response via a slow semantic route. In contrast, in the flanker task the activation of the incorrect response by the distracting information – the flanker arrows surrounding a central target arrow – takes the same slow semantic route as the correct response indicated by the target arrow. Hence, interference stems from an activation of conflicting responses within the slow semantic route. This interference can only be solved by enhancing the contrast between the relevant information – the target – and the irrelevant information – the distractors (Cohen & Huston, 1994; Scherbaum et al., 2011). The differences in timing of the irrelevant information should lead to a different temporal overlap of the response-selection processes: a small overlap in the Simon task and a larger overlap in the flanker task.

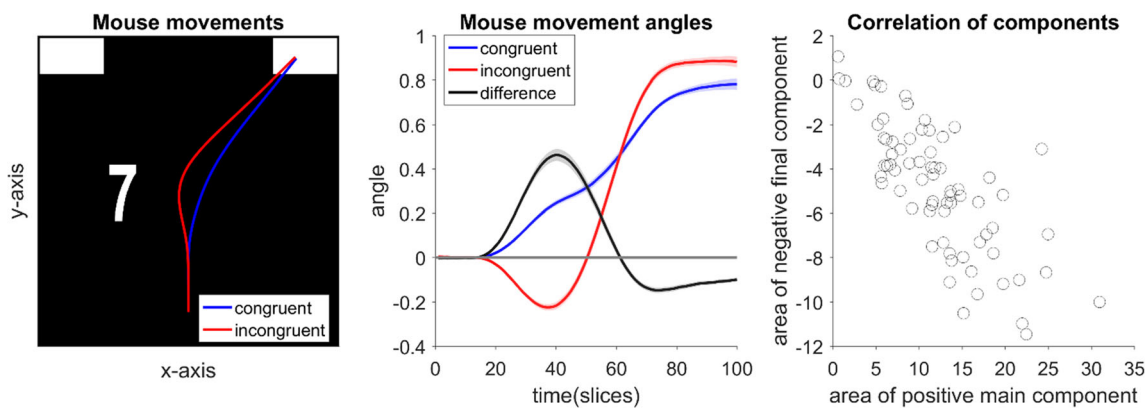


Fig. 7 **Left:** Mouse movements in congruent and incongruent trials. **Middle:** Movement angles for congruent and incongruent trials and the difference between both showing the same temporal signature as the respective beta-weights with a clear early positive component and a

smaller negative component. **Right:** Area under the curve for the each participant's positive component plotted against the area under the curve for the negative component

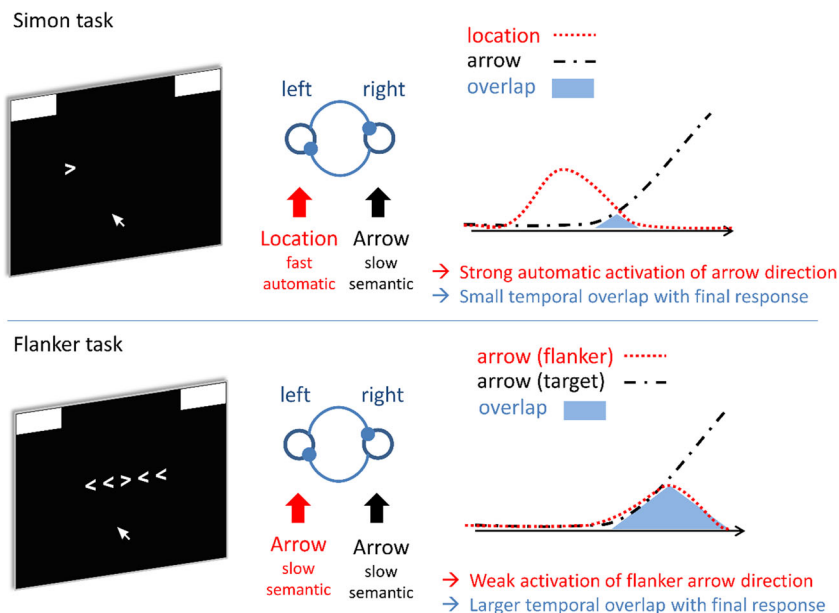


Fig. 8 Conceptual model of the Simon task (**top**) and the flanker task (**bottom**). Irrelevant information (red) and relevant information (black) is fed into two response units inhibiting each other. The decisive difference between tasks lies in the timing of the processing of irrelevant information, which is fast and presumably automatic in the Simon task (the

arrow's location) and slower in the flanker task (the flanker arrows' direction). The resulting timing profiles (right side) show how the differences in the timing of irrelevant information lead to a different temporal overlap, which in turn causes different interference effects in response times

On a measurement level, an important question is whether outcome-based measures correctly inform about these differences between tasks and the respective sub-processes. This boils down to the questions of what differences in RT mean for the sub-processes of interest and whether the correct conclusion could be drawn from such differences.

Based on the reasoning above, we expect interference in the Simon task to affect mouse movements earlier (the automatic activation of a response in the fast route) than interference in the flanker task (the parallel activation of different responses in the same route). This later interference in the flanker tasks in turn leads to a larger temporal overlap of interference with the selection of the final response. This larger overlap in turn leads to more pronounced interference effects in RTs for the flanker task. Hence, looking at RTs, one might conclude that the influence of irrelevant information is stronger in the flanker task than in the Simon task. However, looking at mouse-movement data should show that the influence of the distracting information is similarly strong in both tasks, but shows a different timing between tasks which, in turn, leads to the differences in RT.

For the influence of the previous response, we did not expect any differences as this reflects an intrinsic tendency of response-repetition that should be independent from the stimuli and the task. For conflict adaptation as indicated by congruency sequence effects, we could only speculate: Since conflict adaptation should be related to the experienced interference, differences in the strength of interference should also lead to differences in adaptation. However, the strength of this

difference and the affected parameter (time, duration, strength) are of an explorative nature, especially when considering the low reliability of congruency sequence effects in Study 1.

Methods

Participants

Twenty students (17 female, mean age = 21.1 years) of the Technische Universität Dresden took part in the Simon task (Scherbaum et al., 2010, Study 1) and 20 students (11 female, mean age = 22.1 years) in the flanker task (still unpublished data). Similar selection criteria and procedures were followed for the Simon task and the flanker task samples. All participants were right-handed and had normal or corrected-to-normal vision. The study was performed in accordance with the guidelines of the Declaration of Helsinki and of the German Psychological Society. Ethical approval was not required since the study did not involve any risk or discomfort for the participants. All participants were informed about the purpose and the procedure of the study and gave written informed consent prior to the experiment. They received class credit or 5 € payment.

Given the sample size of 20 subjects per group and assuming a large effect for interference ($d = 0.8$), we calculated a power of 0.8 (Faul et al., 2007) at an alpha-level of 0.05.

Apparatus and stimuli

The Simon task's setup and the flanker task's setup were similar to the Simon task in Study 1, with exception of the imperative stimulus. The Simon task now presented left/right pointing arrows instead of different numbers as stimuli. The flanker task presented a horizontal array of five left/right pointing arrows of which the center arrow was the target and the surrounding four arrows were the distracters. The array was presented in the center of the screen. Each arrow had a visual angle of 1.77° and the whole array spanned 10.62° at 60 cm viewing distance.

Procedure and design

The procedure followed the procedure of the Simon task in Study 1 (see Fig. 9). In the Simon task, the interfering information was the location of the arrow. In the flanker task, the interfering information were the flanking arrows. For a concise description of the design that was common for both tasks, we will subsume both types of interfering information as *distracter* in the following. In both tasks, we varied the following independent variables: for the current trial, $target_N$ (left/right) and $distracter_N$ (left/right), and for the previous trial, $target_{N-1}$ (left/right) and $distracter_{N-1}$ (left/right). This resulted in four combinations for the current trial and four combinations for the previous trial. The sequence of trials was balanced within each block by pseudo randomization resulting in a balanced $Trial_N (4) \times Trial_{N-1} (4) \times trial\ repetition (20)$ transition matrix per block. Each task consisted of two blocks and 320 trials per block.

Data pre-processing

Data pre-processing followed the description from Study 1, with the exception that we did not split the data set into even and odd trials, but analyzed differences between the Simon task and the flanker task. We excluded erroneous trials, in which participants chose the wrong response box, trials following an error, and trials not fitting the RT outlier criterion of an $RT > 4 SD$ and an $RT < 100\ ms$ (6.22%, $SD = 4.6\%$).

Results

RT results

We first checked for differences in the three RT indicators, *response bias* ($RT_{response-switch} - RT_{response-repetition}$), *interference* ($RT_{incongruent} - RT_{congruent}$), and *congruency sequence* ($Simon_effect_{congruentN-1} - Simon_effect_{incongruentN-1}$), as shown in Fig. 10. We performed three independent samples *t*-tests (Bonferroni-corrected alpha-level = .0167). We found a significant difference in the interference effect, $t(38) = 3.906$, $p < .001$, $d = 1.24$. Interference was larger in the flanker task ($M = .109$, $SE = .007$) compared to the Simon task ($M = .074$, $SE = .005$). Furthermore, we found a significant difference for *congruency sequence*, $t(38) = -2.78$, $p < .01$, $d = -.88$. Congruency sequence effects were larger in the Simon task ($M = 0.045$, $SE = 0.004$) than in the flanker task ($M = 0.028$, $SE = 0.005$). There was no significant difference for *response bias*, $t(38) = 0.028$, $p = 0.978$. Hence, congruency sequence effects in RT were inversely related to the effects of interference in RT, which one wouldn't necessarily expect according to conflict-monitoring theory.

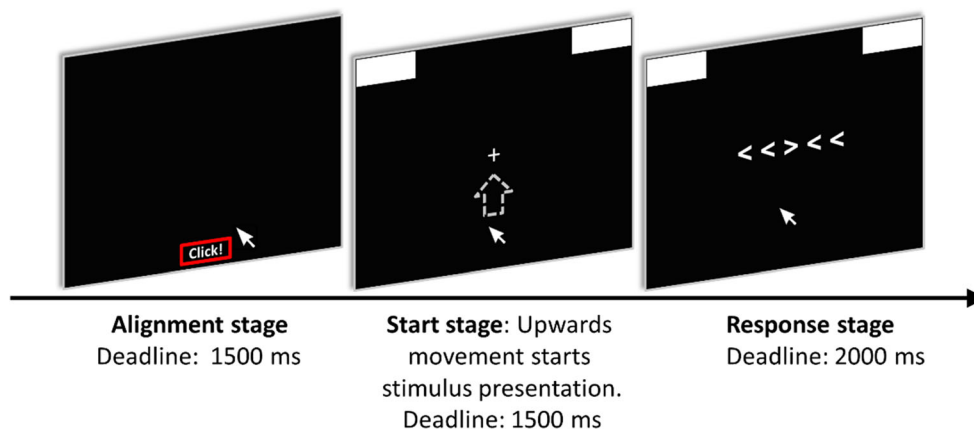


Fig. 9 Setup of the flanker task. In the alignment stage participants clicked with the mouse cursor into a red box at the bottom of the screen. This triggered the start stage, in which response boxes appeared at the upper edge of the screen and participants had to move the cursor upwards (in the sketch indicated by the upwards arrow, which was not

visible to participants) in order to trigger the next stage. After reaching a movement threshold, the response stage began: The target and distracter stimuli were presented and participants moved the mouse cursor to the left or the right response box according to direction indicated by the target stimulus (here the right arrow: right response box)

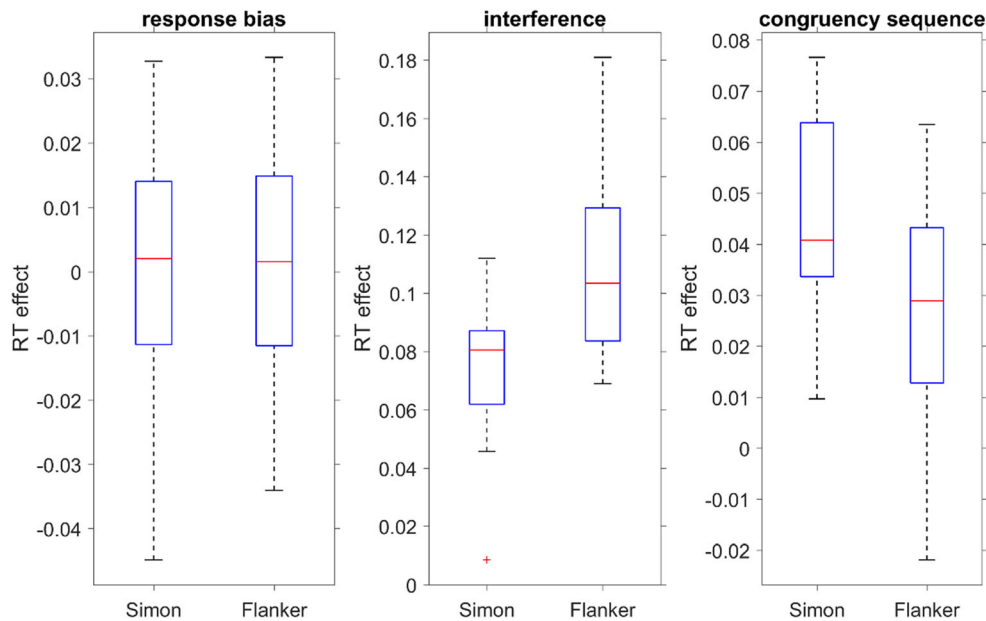


Fig. 10 Response time (RT) effects in the Simon and the flanker tasks. Left: The benefit of response repetitions. Middle: RT slowed by interfering irrelevant information. Right: Modulation of interference by congruency sequence

Mouse results

The results of TCMR for the Simon and the flanker tasks both show the distinct temporal patterns for each influence. A first peak of *response bias*, followed by the peak of *interference* and then the peak of *congruency sequence* (see Fig. 11).

The jack-knifed peak time and strength from TCMR and the parameters extracted by Gauss curve fitting can be seen in Table 4.

Fitting quality (R^2) was best for *interference* and weaker for *response bias* in both tasks. *Congruency sequence* showed good fitting quality in the Simon task, but the worst fit in the flanker task. The spread of parameters did not show any floor or ceiling effects, indicating that the estimation procedure worked correctly (see Fig. 12).

To check for differences in the mouse parameters, we performed nine *t*-tests on the three parameters of the three extracted beta-weights (Bonferroni-corrected alpha level = 0.0056). The results can be found in Table 5.

We found significant differences for all parameters of *interference*. The influence peaked at an earlier time slice in the Simon task ($M = 43.547$, $SE = 0.778$) than in the flanker task ($M = 53.833$, $SE = 0.750$), lasted longer in the Simon task ($M = 9.439$, $SE = 0.164$) than in the flanker task ($M = 8.215$, $SE = 0.166$), and was stronger in the Simon task ($M = 0.208$, $SE = 0.011$) than in the flanker task ($M = 0.159$, $SE = 0.011$). We also found significant differences for the peak of *congruency sequence*, which peaked later in the Simon task ($M = 50.369$, $SE = 0.847$) than in the flanker task ($M = 45.653$, $SE = 1.086$) and was stronger in the Simon task ($M = 0.037$, $SE = 0.004$) than in the flanker task ($M = 0.013$, $SE = 0.003$).

There were no further significant differences, especially, as expected, for *response bias*. Considering the absolute peak times⁶ for the interference effect of the Simon task ($M = 0.253$ s) and the flanker task ($M = 0.367$ s) shows that the difference between these peaks shows a similar magnitude to the interference effect in RT itself and hence was of substantial magnitude.

Finally, we checked for correlations between the mouse parameters and RT indicators with (Bonferroni-corrected alpha level = 0.0056, see Table 6).

The only significant correlation was present for the peak time of *interference*. The later the peak time of this interference, the larger the interference effect in RT.⁷

Discussion

Study two found the expected later peak of interference in the flanker task compared to the Simon task. As theoretically derived above, this indicates that interference in the flanker task can be attributed to a slower sub-process than interference in the Simon task: In the flanker task, interference is caused by parallel competing processing of semantic information (e.g., the direction of arrows in the flanker and the target stimuli) – hence, processing of irrelevant information and correct information competes simultaneously and shows a strong overlap during selection of the final response. In contrast, in the Simon task, interference is caused by an early and presumably automatic activation of the stimulus' position, which then decays and shows only minimal overlap with the selection of the final response by processing semantic stimulus information (the direction of the arrow). Notably, the overall strength of

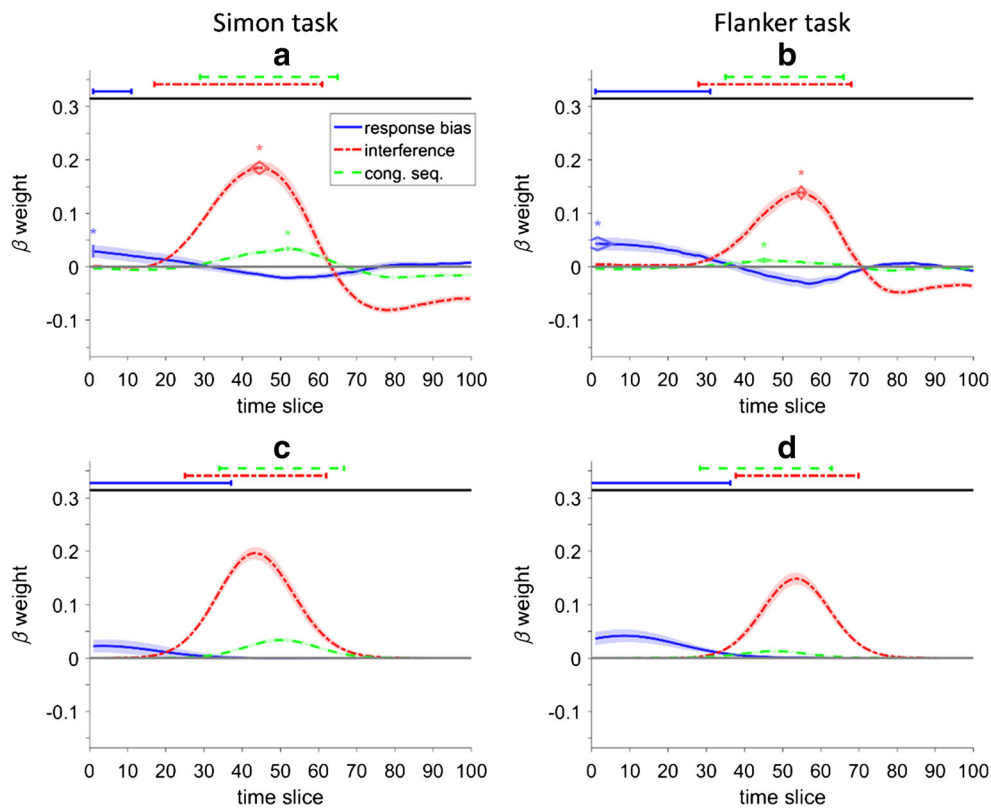


Fig. 11 A and B: Time-continuous beta-weights from time-continuous multiple regression for the Simon task (A) and the flanker task (B). Lines above graphs mark significant segments determined by t-test against zero. Curve peaks were determined by a jack-knifing procedure. **C and D:**

Reconstructed beta-weights from the Gaussian fitting procedure for the Simon task (C) and the flanker task (D). Lines above graphs indicate the area of 1 SD around the peak of the fitted Gauss curve. In all graphs, shaded areas indicate the standard error of the mean

influence by irrelevant semantic information was actually weaker in the flanker task than the strength of influence by the automatic location activation in the Simon task. This finding in mouse movements is important because we also found a difference in interference between tasks in RT. However, this difference alone might have led to the opposite conclusion, since RT indicated a stronger interference in the flanker task than in the Simon task. Considering the mouse movement results, it is not valid to interpret the RT difference as a difference in strength of influence of irrelevant information. It should instead be interpreted as a difference in the overlap of

the two interfering processes (irrelevant information and relevant information). This overlap was larger for the flanker task (later peak) than for the Simon task (earlier peak). Corroborating evidence for this interpretation comes from correlating RT and dynamic markers, which indicates that the stronger interference effect in RT for the flanker task is related to the later peak time of the interference. This was further underlined by a lack of correlation between the interference effect in RT and the peak strength of this influence. Hence, it is not the strength of influence of irrelevant information that causes the difference in RT between the Simon task and the

Table 4 Results of peak detection via jack-knifing and of Gaussian fitting of beta-weights for the Simon and the flanker tasks

Beta	Jack-knifed peaks		Parameters from Gaussian fitting				
	Peak time (SE)	Peak strength (SE)	Peak time (SE)	Peak strength (SE)	Duration(SE)	R ² (SE)	
Simon	response bias	1 (0)	0.029 (0.012)	6.327 (1.199)	0.024 (0.013)	15.718 (0.82)	0.869 (0.047)
	interference	44.5 (2.169)	0.185 (0.012)	43.547 (0.778)	0.208 (0.011)	9.439 (0.164)	0.958 (0.005)
	congruency sequence	52 (0)	0.034 (0.004)	50.369 (0.847)	0.037 (0.004)	8.32 (0.283)	0.888 (0.024)
Flanker	response bias	1.6 (3.995)	0.043 (0.012)	4.751 (1.908)	0.044 (0.014)	16.135 (1.225)	0.8 (0.065)
	interference	54.9 (1.308)	0.138 (0.012)	53.833 (0.75)	0.159 (0.011)	8.215 (0.166)	0.933 (0.008)
	congruencysequence	45.15 (1.556)	0.012 (0.003)	45.653 (1.086)	0.013 (0.003)	8.805 (0.625)	0.712 (0.041)

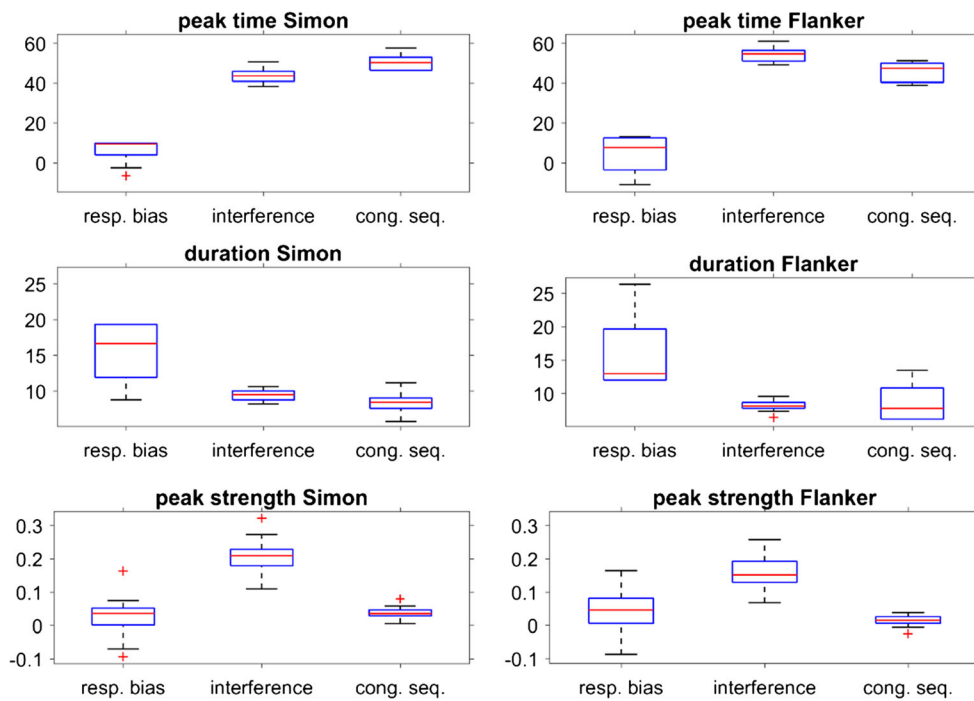


Fig. 12 Spread of parameters from Gaussian fitting for the three betas from TCMR, for the flanker task (left) and the Simon task (right)

flanker task, but the timing of when the irrelevant information is processed.

General discussion

The aim of the current work was to show the potential of extracting individual dynamic parameters that describe the temporal patterns of sub-processes in decision making as they are reflected in mouse movements of participants.

We showed that the extracted parameters from a Simon task (Scherbaum et al., 2010) yield promising reliability for

the peak time and peak strength of two sub-processes, i.e., the influence of the previous response of the irrelevant location information. In comparison, we found similar levels of reliability for the respective RT markers, though with a trend to slightly lower values. However, for the third sub-process, conflict adaptation, we found surprisingly low levels of reliability for mouse parameters, and even worse levels for the RT marker. This finding is startling, as conflict adaptation is one of the hallmarks of cognitive control processes and could be expected to show stable inter-individual differences. However, this observation adds to recent findings of low reliability for conflict adaptation (Feldman & Freitas, 2016) and suggests a heightened level of caution when studying inter-individual differences of this sub-process.

We used the validated parameters to compare the cognitive processes of the Simon task and the flanker task. In RT, we found stronger interference effects in the flanker task than in the Simon task, which might have suggested that the flanker task triggers stronger interference than the Simon task.

Table 5 Independent samples t-tests on mouse parameters

	t	df	p	d
Response bias peak time	-0.729	38.00	0.471 ^{*1}	-0.23
Response bias duration	0.282	38.00	0.779	0.089
Response bias peak strength	1.041	38.00	0.304	0.329
Interference peak time	9.523	38.00	< .001	3.011
Interference duration	-5.258	38.00	< .001	-1.663
Interference peak strength	-3.02	38.00	0.004	-0.955
Cong. Seq. Peak time	-3.423	38.00	0.001	-1.082
Cong. Seq. Duration	0.707	38.00	0.484 ^{*2}	0.224
Cong. Seq. Peak strength	-4.819	38.00	< .001	-1.524

^{*1} Levene's test is significant ($p < .05$), suggesting a violation of the equal variance assumption. Welch's t -test yielded $p = 0.472$

^{*2} Levene's test is significant ($p < .05$), suggesting a violation of the equal variance assumption. Welch's t -test yielded $p = 0.485$

Table 6 Correlation of response time (RT) indicators and mouse parameters (Bonferroni-corrected alpha level = 0.0056)

	Peak time		Peak SD		Peak strength	
	r	p	r	p	r	p
Response bias	.084	.605	.011	.948	.321	.043
Interference	.481	.002	.231	.151	.032	.843
Congruency sequence	.286	.074	.332	.036	.312	.050

However, analysis of the dynamic parameters showed that the strength of interference is larger in the Simon task than in the flanker task, while the peak time of interference is earlier in the Simon task than in the flanker task. The latter could lead to a stronger overlap of the interference influence and response selection and, in turn, lead to the interference effects that we found for RT in the flanker task (Hommel, 1994; Ridderinkhof, van den Wildenberg, Wijnen, & Burle, 2004). Correlation analysis bolstered this interpretation that stronger interference effects in RT are not caused by a generally stronger peak of the interference influence, but by a larger overlap of the interference influence and the response selection process. This indicates that the analysis of temporal dynamics of sub-processes with mouse tracking can provide additional, and even correcting, information to a pure analysis of mean RT or static markers of performance in mouse tracking and motion tracking (Buetti & Kerzel, 2008, 2009). It is noteworthy that mouse tracking is not the only approach that could unveil this information. Properties of and differences between conflict tasks can also be inferred via distribution analyses of RT (e.g., Ridderinkhof, 2002; Ridderinkhof et al., 2004) and physiological measures, for example, the lateralized readiness potential in the EEG (e.g., Stürmer et al., 2002) and subtle muscle movements in EMG (Burle, Possamai, Vidal, Bonnet, & Hasbroucq, 2002). Mouse tracking and the TCMR method fill a gap here in the sense that distribution analysis on the one hand provides a simple but indirect inference of the process from static behavioral markers, and measuring EEG on the other hand provides a methodologically complex though more direct (at least in the case of the lateralized readiness potential) inference from a process-oriented measure. Mouse tracking together with TCMR allows for combining the simplicity of a behavioral measure with the process orientation of physiological measures. It goes without saying that the TCMR method, or more general multiple-regression approaches, could be applied to almost any time-continuous signal and thus also to physiological signals (e.g., Cohen & Cavanagh, 2011; Cohen & Donner, 2013) to gain more information from these signals.

While we and others have shown this gain of information for mouse tracking in principle in previous studies (Dshemuchadse et al., 2012; Frisch, Dshemuchadse, Görner, Goschke, & Scherbaum, 2015; Scherbaum et al., 2016, 2015; Sullivan et al., 2015), the gain of the method provided here – TCMR with Gaussian fitting – is the possibility of an inference-statistical analysis of all extracted parameters rather than the descriptive approach that is usually applied. Hence, the promise of mouse tracking and process tracing in general – an extensive study of the dynamics of cognitive process – is more in reach with the method presented here.

Some limitations concerning the approach presented here need to be considered.

First, it is worth mentioning two approaches that could be seen as a complement to the approach presented here: the analysis of decision spaces (O’Hora, Dale, Piironen, & Connolly, 2013) and the analysis of trajectory prototypes (Wulff, Haslbeck, & Schulte-Mecklenbeck, *in preparation*). In the first approach, one derives attractor landscapes from the recorded mouse movements and compares the landscapes across participants or conditions. This approach is visually intuitive and nicely summarizes the information hidden in mouse movements. Though it lacks the possibility to extract individual parameters, it provides a complement to the detailed analysis of temporal patterns of different influences/sub-process as provided by the method presented here. In the second approach, individual mouse movements are clustered into prototypes to detect whether these movements reflect a continuous decision process (a smooth though deflected movement to a response box) or a discontinuous and presumably step-wise decision process (a disrupted or cornered movement, first to one response box and then to the alternative one). Though this approach considers the whole trajectory, it does not target the temporal dynamics of sub-processes, but is more a method to ensure that mouse movements represent the processes that they are assumed to represent (continuous decision processes). Hence, it represents a methodological complement to our approach that could ensure construct validity especially when the mouse-tracking setup might favor discontinuous processes (Grage, Schoemann, Kieslich, & Scherbaum, *under review*; Scherbaum & Kieslich, 2018; Schoemann, Lüken, Grage, Kieslich, & Scherbaum, *in press*).

Second, the fitted Gauss curves do not fully represent the shape of the extracted beta-weights. The positive peaks of the beta-weights are most often followed by a negative counterpart. We did not consider fitting this part, since it reflects a necessary compensatory movement triggered by the initial (positive) peak, that is, because the start and end point of the movements is fixed by the coordinates of the start box and the response box on the screen. Hence, when an influence leads the movement to divert from its path, this diversion must be compensated for later in the movement so that the cursor finally arrives within the target area. Based on this logic and on our experience, this later part of the movement bears no further information about the specific regressors’ influence compared to the initial main component.

Third, one might ask why we performed single regression analysis for each individual in contrast to a hierarchical model analysis. We have shown several times that hierarchical model analysis provides nearly identical information on the participant-level (Scherbaum et al., 2016; Scherbaum & Kieslich, 2018). Simple regression analysis is a basic tool that could be used and understood by most social scientists and its implementation in Matlab is straightforward. Considering this

simplicity and the comparability of the results, we decided in favor of the simpler method in our toolbox.

Fourth, the extensive study of temporal patterns hidden in mouse movements poses prerequisites on the quality of the acquired data, that is, the homogeneity and continuity of movements (Schoemann et al., *in press*). This quality might depend on many parameters, of which the starting condition is one parameter (Scherbaum & Kieslich, 2018). While many studies use a dynamic starting procedure to encourage participants to move consistently, such a procedure is not applicable to all areas of study, for example when the stimuli are too complex for a swift decision (e.g. Kieslich & Hilbig, 2014). In such setups, certain modifications of the task might be required, for example, sequential presentation of information (Dshemuchadse et al., 2012). Hence, researchers should investigate their data for movement consistency and low movement initiation times to warrant homogeneous data for the analysis of temporal patterns (Scherbaum & Kieslich, 2018).

Fifth, one might ask whether comparing data from two separate samples in Study 2 provides reliable support that the found differences on the congruency effect stem from the differences in the tasks and not from differences in samples. Though we ensured that instructions and setups were completely comparable, the two different samples might still show intrinsic differences. Because of these potential intrinsic differences, it was of utter importance that the predictor *response bias* showed no difference between the two groups, since this predictor does not represent a stimulus-driven influence, but should solely represent participants' intrinsic tendency to repeat their response. The missing difference here hence indicates that the samples were as comparable as possible.

Sixth, and finally, the analyses described here need programming skills to be performed. To ease the use of the method, we provide a toolbox of Matlab functions for downloading. We provide the commented functions and a demonstration script that uses a subset of the data from our Simon task study (Scherbaum et al., 2011), allowing users to learn and explore the method. While our Matlab toolbox is currently the only toolbox offering TCMR, users interested in analyzing mouse movements in general might also refer to alternative open source packages, for example, mouse trap (Kieslich et al., 2017), especially when they prefer the use of R over Matlab.

We hope that the results presented here and the toolbox provided for downloading encourage researchers to enter the study of the temporal dynamics of cognitive sub-process as reflected in mouse movements. While we used cognitive control as an example, the wide range of existing mouse studies, for example, about semantic processing (Dale et al., 2007; Dshemuchadse, Grage, & Scherbaum, 2015; Spivey et al., 2005), value-based decision making (Dshemuchadse et al., 2012; Koop & Johnson, 2013), and moral decision making

(Kieslich & Hilbig, 2014), suggests that the presented method could shed light on the processes in many other areas.

Studying processes this way could not only provide more information in empirical studies; recent advances in cognitive modelling ask for more and more data to allow for better model fitting (Turner, Rodriguez, Norcia, McClure, & Steyvers, 2016). The extraction of individual dynamic parameters from mouse tracking and motion tracking in general might provide a way to go beyond qualitative model comparisons (Dshemuchadse et al., 2015; Frisch et al., 2015; Scherbaum et al., 2016) and could be a simple and efficient addition to the pool of data one can fit a model on. Hence, methods for a deeper analysis of mouse movements as presented here allow researchers to use data that have often been ignored before: the way from the start to the end of a decision.

Notes

1. We use the term beta-weight in the following to denote the weights/slopes in linear regression models of the form $y = \alpha + \beta x$.
2. We will, of course, check this assumption in our data, see results of Study 1 and Fig. 7.
3. The movement angle reflects the instantaneous tendency of the movement more precisely as it integrates the movement on the X/Y plane into a single measure. Other potential measures, e.g., the movement on the X-axis, only represent the cumulative previous instantaneous tendencies. While such a cumulative measure yields less noise, it shows a time lag for when exactly an influence affected the movement and, hence, we prefer the instantaneous measure of the movement angle.
4. We apply temporal smoothing for two reasons (similar to spatial smoothing in fMRI analysis; e.g., Mikl et al., 2008): First, it increases the signal-to-noise ratio. Since the movement angle is a differential measure, it shows a higher level of noise than raw movement data; this noise is reduced by smoothing. Second, smoothing improves the validity of statistical tests since it makes error distributions more normal.
5. Jack-knifing represents one of several possible resampling methods: for each dataset d in a group of n datasets, the jack-knife produces a new mean dataset consisting of all datasets in the group, except dataset d . Hence, for dataset 1, the method creates a mean dataset averaging across the data in datasets (2, 3, ..., n). For dataset 2, it creates a mean dataset averaging across the data in datasets (1, 3, 4, ..., n). While this reduces the noise occurring in time-series data, e.g., LRP data, it also reduces the degrees of freedom. Hence, for statistical testing, test parameters have to be adjusted (for further details, see Miller, Patterson, & Ulrich, 2001).

6. *Absolute peak time* (APT) denotes a derived temporal measure that is calculated by applying relative peak times in time slices ts (e.g., time slice 43 of 100) to mean RT of the respective task (RT_{task}): The formula hence is:

$$APT = ts/100 * RT_{task}$$

7. With a sample size of 40 subjects, the correlational results should be taken with care. However, taking the full pattern of correlational results together with a post hoc power analysis (Faul, Erdfelder, Lang, & Buchner, 2007) for the effect of peak time in interference ($r = 0.481$, $\alpha = 0.0058$) yielding a power of 0.75, indicates that the results could be interpreted in favor of the overall interpretation.

Acknowledgements This research was supported by the German Research Council (DFG) (grant SFB 940/2 2016). We thank Diana Schwenke for her support in data collection. We thank Daniel Leising and three reviewers for their very helpful comments on an earlier version of the manuscript.

Open Practices Statement Primary data (csv format) and analysis scripts (Matlab) using the TCMR toolbox are available at the Open Science Framework: <https://osf.io/62gfj>. None of the studies was preregistered. We report achieved power for the used sample size, all data exclusions (if any), and all relevant measures and manipulations in the study.

References

- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.
- Buetti, S., & Kerzel, D. (2008). Time course of the Simon effect in pointing movements for horizontal, vertical, and acoustic stimuli: Evidence for a common mechanism. *Acta Psychologica*, *129*(3), 420–428.
- Buetti, S., & Kerzel, D. (2009). Conflicts during response selection affect response programming: Reactions towards the source of stimulation. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 816–834.
- Burle, B., Possamai, C. A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002). Executive control in the Simon effect: An electromyographic and distributional analysis. *Psychological Research*, *66*(4), 324–336.
- Cohen, J. D., & Huston, T. A. (1994). Progress in the use of interactive models for understanding attention and performance. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 1–19). Cambridge, MA: MIT Press.
- Cohen, M. X., & Cavanagh, J. F. (2011). Single-Trial Regression Elucidates the Role of Prefrontal Theta Oscillations in Response Conflict. *Frontiers in Psychology*, *2*.
- Cohen, M. X., & Donner, T. H. (2013). Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior. *Journal of Neurophysiology*, *110*(12), 2752–2763.
- Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory and Cognition*, *35*(1), 15–28.
- D’Errico, J. (2012). *fminsearchbnd - Bound constrained optimization using fminsearch*.
- Dshemuchadse, M., Grage, T., & Scherbaum, S. (2015). Action dynamics reveal two components of cognitive flexibility in a homonym relatedness judgement task. *Frontiers in Cognition*, *6*, 1244.
- Dshemuchadse, M., Scherbaum, S., & Goschke, T. (2012). How decisions emerge: action dynamics in intertemporal decision making. *Journal of Experimental Psychology: General*, *142*, 151–185.
- Egner, T. (2007). Congruency sequence effects and cognitive control. *Cognitive, Affective & Behavioral Neuroscience*, *7*(4), 380–390.
- 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*(1), 143–149.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.
- Feldman, J. L., & Freitas, A. L. (2016). An Investigation of the Reliability and Self-Regulatory Correlates of Conflict Adaptation. *Experimental Psychology*, *63*(4), 237–247.
- Fischer, M. H., & Hartmann, M. (2014). Pushing forward in embodied cognition: may we mouse the mathematical mind? *Frontiers in Psychology*, *5*.
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241.
- Frisch, S., Dshemuchadse, M., Gömer, M., Goschke, T., & Scherbaum, S. (2015). Unraveling the sub-processes of selective attention: insights from dynamic modeling and continuous behavior. *Cognitive Processing*, *16*(4), 377–388.
- Grage, T., Schoemann, M., Kieslich, P. J., & Scherbaum, S. (under review). Validate mouse-tracking: How design factors influence action dynamics in the Simon task. *Attention, Perception, & Psychophysics*.
- Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*(4), 480–506.
- Hommel, B. (1994). Spontaneous decay of response-code activation. *Psychological Research*, *56*(4), 261–268.
- Kieslich, P. J., & Hilbig, B. E. (2014). Cognitive conflict in social dilemmas: An analysis of response dynamics. *Judgment and Decision Making*, *9*(6), 510–522.
- Kieslich, P. J., Wulf, D. U., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2017). Mousetrap: An R package for processing and analyzing mouse-tracking data. *Zenodo*.
- Koop, G. J., & Johnson, J. G. (2011). Response dynamics: A new window on the decision process. *Judgment and Decision Making*, *6*(8), 750–758.
- Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, *67*(4), 151–185.
- Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience*, *6*(5), 450–452.
- McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*(1), 22–24.
- Mikl, M., Marecek, R., Hlustík, P., Pavlicová, M., Drastich, A., Chlebus, P., et al. (2008). Effects of spatial smoothing on fMRI group inferences. *Magnetic Resonance Imaging*, *26*(4), 490–503.

- Miller, J., Patterson, T., & Ulrich, R. (2001). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, 35(01), 99–115.
- O’Hora, D., Dale, R., Piiroinen, P. T., & Connolly, F. (2013). Local dynamics in decision making: The evolution of preference within and across decisions. *Scientific Reports*, 3.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10(4), 437–442.
- Pratte, M. S., Rouder, J. N., Morey, R. D., & Feng, C. (2010). Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception, & Psychophysics*, 72(7), 2013–2025.
- Ridderinkhof, K. R. (2002). Activation and suppression in conflict tasks: Empirical clarification through distributional analyses. In W. Prinz & B. Hommel (Eds.), *Mechanisms in perception and action, attention and performance* (pp. 494–519). Oxford: Oxford University Press.
- Ridderinkhof, K. R., van den Wildenberg, W. P. M., Wijnen, J., & Burle, B. (2004). Response Inhibition in Conflict Tasks Is Revealed in Delta Plots. In M. I. Posner (Ed.), *Cognitive neuroscience of attention* (pp. 369–377). New York, NY, US: Guilford Press.
- Scherbaum, S. (2017). *TCMR: Time continuous multiple regression toolbox for mouse tracking*. osf.io/5e3vn.
- Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: The temporal dynamics of action selection. *Cognition*, 115(3), 407–416.
- Scherbaum, S., Dshemuchadse, M., Leiberg, S., & Goschke, T. (2013). Harder than expected: increased conflict in clearly disadvantageous intertemporal choices in a computer game. *PLoS ONE*, 8(11), e79310.
- Scherbaum, S., Fischer, R., Dshemuchadse, M., & Goschke, T. (2011). The dynamics of cognitive control: Evidence for within-trial conflict adaptation from frequency-tagged EEG. *Psychophysiology*, 48(5), 591–600.
- Scherbaum, S., Frisch, S., Dshemuchadse, M., Rudolf, M., & Fischer, R. (2016). The test of both worlds: identifying feature binding and control processes in congruency sequence tasks by means of action dynamics. *Psychological Research*, 1–16.
- Scherbaum, S., Gottschalk, C., Dshemuchadse, M., & Fischer, R. (2015). Action dynamics in multitasking: the impact of additional task factors on the execution of the prioritized motor movement. *Frontiers in Cognition*, 6, 934.
- Scherbaum, S., & Kieslich, P. J. (2018). Stuck at the starting line: How the starting procedure influences mouse-tracking data. *Behavioral Research Methods*, 50(5), 2097–2110.
- Schoemann, M., Lüken, M., Grage, T., Kieslich, P. J., & Scherbaum, S. (in press). Validate mouse-tracking: How design factors influence action dynamics in intertemporal decision making. *Behavior Research Methods*.
- Schulte-Mecklenbeck, M., Kuehberger, A., & Ranyard, R. (2011). a handbook of process tracing methods for decision research : A critical review and user’s guide. Hoboken: Taylor and Francis.
- Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, 81(1), 174–176.
- Song, J. H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, 13(8), 360–366.
- Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, 15(5), 207–211.
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, 102(29), 10393–10398.
- Stroop, J. R. (1935). Studies of interference in serial verbal interactions. *Journal of Experimental Psychology*, 18, 643–662.
- Stürmer, B., Leuthold, H., Soetens, E., Schroter, H., & Sommer, W. (2002). Control over location-based response activation in the Simon task: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 28(6), 1345–1363.
- Sullivan, N., Hutcherson, C., Harris, A., & Rangel, A. (2015). Dietary Self-Control Is Related to the Speed With Which Attributes of Healthfulness and Tastiness Are Processed. *Psychological Science*, 26(2), 122–134.
- Turner, B. M., Rodriguez, C. A., Norcia, T. M., McClure, S. M., & Steyvers, M. (2016). Why more is better: Simultaneous modeling of EEG, fMRI, and behavioral data. *NeuroImage*, 128, 96–115.
- Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (in prep.). Measuring the (dis-)continuous mind: What movement trajectories reveal about cognition.

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