Validating mouse-tracking: How design factors influence action dynamics in intertemporal decision making

Martin Schoemann¹ · Malte Lüken¹ · Tobias Grage¹ · Pascal J. Kieslich² · Stefan Scherbaum¹

Published online: 25 January 2019
© The Psychonomic Society, Inc. 2019

Abstract

Mouse-tracking is an increasingly popular process-tracing method. It builds on the assumption that the continuity of cognitive processing leaks into the continuity of mouse movements. Because this assumption is the prerequisite for meaningful reverse inference, it is an important question whether the assumed interaction between continuous processing and movement might be influenced by the methodological setup of the measurement. Here we studied the impacts of three commonly occurring methodological variations on the quality of mouse-tracking measures, and hence, on the reported cognitive effects. We used a mouse-tracking version of a classical intertemporal choice task that had previously been used to examine the dynamics of temporal discounting and the date–delay effect (Dshemuchadse, Scherbaum, & Goshke, 2013). The data from this previous study also served as a benchmark condition in our experimental design. Between studies, we varied the starting procedure. Within the new study, we varied the response procedure and the stimulus position. The starting procedure had the strongest influence on common mouse-tracking measures, and therefore on the cognitive effects. The effects of the response procedure and the stimulus position were weaker and less pronounced. The results suggest that the methodological setup crucially influences the interaction between continuous processing and mouse movement. We conclude that the methodological setup is of high importance for the validity of mouse-tracking as a process-tracing method. Finally, we discuss the need for standardized mouse-tracking setups, for which we provide recommendations, and present two promising lines of research toward obtaining an evidence-based gold standard of mouse-tracking.

Keywords Mouse-tracking · Action dynamics · Process-tracing · Boundary conditions · Intertemporal choice

Decision science has experienced a paradigmatic shift evolving its focus, methods, and approaches from an outcome-based perspective toward a more process-oriented paradigm (Oppenheimer & Kelso, 2015). This process paradigm acknowledges the temporal nature of basic mental processes and, hence, builds theories of choice incorporating perceptual, attentional, memory, and decisional processes. To test these process explanations, process-tracing methods are required. In the last 60 years, decision scientists introduced a variety of process-tracing methods to the field—for example, verbal

protocols (e.g., Ericson & Simon, 1984), eye tracking (e.g., Russo & Rosen, 1975), and most recently, mouse-tracking (e.g., Dale, Kehoe, & Spivey, 2007; Spivey, Grosjean, & Knoblich, 2005) (for an overview, please see Schulte-Mecklenbeck et al., 2017).

Whenever scientists apply such process-tracing methods, they rely on specific prerequisites and core concepts in order to conduct the *reverse inference* (Poldrack, 2006): Reverse inference describes the reasoning by which the presence of a particular cognitive process is inferred from a pattern of neuroimaging or behavioral data (cf. Heit, 2015). One prerequisite of reverse inference stipulates a close mapping between the covert mental process and the overt observable behavior. Given this mapping, reverse inference is used to test process explanations and produce meaningful interpretation of the results. However, this direct interpretation is hampered by the fact that the mapping between mental processes and behavior might be mediated by the (process-tracing) method itself. Validation of the assumed mapping is hence essential. In this study, we concentrate on the validation of mouse-tracking as a process-tracing method by evaluating the interaction between

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13428-018-1179-4>) contains supplementary material, which is available to authorized users.

✉ Martin Schoemann
martin.schoemann@tu-dresden.de

¹ Department of Psychology, Technische Universität Dresden, Dresden, Germany

² Experimental Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany

the mapping and different implementations of the method. Eventually, we derive recommendations for the implementation of the method that we hope will reinforce extensive discussions about standards (i.e., best practices) in mouse-tracking research.

Mouse-tracking as a process-tracing method

Mouse-tracking is a relatively new and convenient process-tracing method (Freeman, 2018; Koop & Johnson, 2011). It offers five advantages: The temporal resolution is high, the risk of distortion tolerable, the technical equipment is cheap and runs in almost every lab facility, most participants are highly familiar with moving a computer mouse, and mouse-tracking algorithms can be implemented in most experimental software with only a few technical skills (Schulte-Mecklenbeck et al., 2017). Additionally, ready-to-use experimental builder and analysis software has been developed that allows researchers to create mouse-tracking experiments without programming and to process, analyze and visualize the resulting data with little technical effort (Freeman & Ambady, 2010; Kieslich & Henninger, 2017; Kieslich, Henninger, Wulff, Haslbeck, & Schulte-Mecklenbeck, *in press*).

Hence, in the last 10 years, mouse-tracking flourished in many fields of psychological research (for a review, see Erb, 2018; Freeman, 2018; Freeman, Dale, & Farmer, 2011), finding applications in studies of phonological and semantic processing (Dale et al., 2007; Dshemuchadse, Grage, & Scherbaum, 2015; Spivey et al., 2005), cognitive control (Dignath, Pfister, Eder, Kiesel, & Kunde, 2014; Incera & McLennan, 2016; Scherbaum, Dshemuchadse, Fischer, & Goschke, 2010; Yamamoto, Incera, & McLennan, 2016), selective attention (Frisch, Dshemuchadse, Görner, Goschke, & Scherbaum, 2015), numerical cognition (Szasz, Palfi, Szollosi, Kieslich, & Aczel, 2018), perceptual choices (Quinton, Volpi, Barca, & Pezzulo, 2014), moral decisions (Koop, 2013), preferential choices (Koop & Johnson, 2013; O'Hara, Dale, Piironen, & Connolly, 2013), lexical decisions (Barca & Pezzulo, 2012, 2015; Lepora & Pezzulo, 2015), and value-based decisions (Calluso, Comitteri, Pezzulo, Lepora, & Tosoni, 2015; Dshemuchadse, Scherbaum, & Goschke, 2013; Kieslich & Hilbig, 2014; Koop & Johnson, 2011; O'Hara, Carey, Kervick, Crowley, & Dabrowski, 2016; Scherbaum, Dshemuchadse, Leiberg, & Goschke, 2013; Scherbaum, Frisch, & Dshemuchadse, 2018a, 2018b; Scherbaum et al., 2016; van Rooij, Favela, Malone, & Richardson, 2013).

By measuring mouse movements, all those studies have advanced their field into a process-oriented paradigm, by revealing participants' cognitive processing over the time course (see Fig. 1). The core concept underlying mouse-tracking is that cognitive processing is continuously leaking into motor (e.g., hand or computer mouse) movements (Spivey, 2007;

Spivey & Dale, 2006). Ideally there would be a one-to-one correspondence (i.e., mapping) between overt behavior and covert cognitive processes. However, this mapping might crucially depend on the applied mouse-tracking paradigm, in such a way that it must generally support the continuous leaking of cognitive processing into the mouse movements.

This continuous leakage in a mouse-tracking paradigm is not trivial: Scherbaum and Kieslich (2018) recently showed that different procedures for starting a trial in a mouse-tracking paradigm resulted in different mouse movements (in the following discussion, this premise is referred to as *PI*). This finding puts in question whether the mapping between overt behavior and covert cognitive processes holds equally for different implementations of a starting procedure. The results of the Scherbaum and Kieslich study give rise to the supposition that mouse-tracking paradigms must be carefully designed and that specific variations can strongly influence the mapping, and hence challenge the validity of the reverse inference. Consequently, this introduces questions about the validity, generalizability, and replicability of mouse-tracking results.

Varieties in the implementation of mouse-tracking

Because mouse-tracking is relatively new, nothing like a gold standard exists for mouse-tracking paradigms, which is reflected in the diversity in methodological setups of mouse-tracking studies over all psychological research. Besides the starting procedure, mouse-tracking paradigms differ—for instance—with respect to the procedure of final response indication and the position of the stimuli—both of which can influence the mapping between cognitive processing and mouse movements in substantial way, as described in the following.

Concerning the response procedure, in most mouse-tracking paradigms participants indicate their responses by clicking with the mouse cursor in specific areas (i.e., response boxes) on the screen (e.g., O'Hara et al., 2016). In contrast to this *click procedure*, in other paradigms participants indicate their responses just by moving the mouse cursor into those areas (e.g., Scherbaum et al., 2010), which we will call a *hover procedure*. Comparing the two procedures, it is rather obvious that the click procedure allows for second thoughts, and hence possibly (discrete) changes of mind (Barca & Pezzulo, 2015; Resulaj, Kiani, Wolpert, & Shadlen, 2009), since the response process is not terminated when the cursor reaches the response box, but only after clicking in it (P2).

Concerning the positions of stimuli, there are paradigms in which stimuli are presented very close to the center of the screen (e.g., Calluso et al., 2015), whereas in other paradigms the stimuli are presented close to or within the response boxes, which are usually positioned at the upper left and right edges of the screen (e.g., Faulkenberry, Cruise, Lavro, & Shaki,

Fig. 1 Simplified illustration of mouse-tracking as a process-tracing method. Cognitive processing (on the left) is depicted as the activation difference between two options as a function of time. The corresponding continuous mouse movement (on the right) is depicted as the recorded mouse cursor position (on the x/y -plane) in a standard mouse-tracking paradigm in which participants have to choose between two options,

2016). Comparing the two implementations of stimulus position, it is rather obvious that a paradigm in which the stimuli are positioned at the upper left and right edges of the screen demands many eye movements. Those eye movements might interfere with cognitive processing (see Orquin & Mueller Loose, 2013) and, in turn, mouse movements (P3). In sum, it appears that at least some of the aforementioned design factors might jeopardize the basis of mouse-tracking, and thus its validity as a process-tracing method.

In this study, we systematically investigated how different design factors influence mouse-tracking data and results. In this regard, we build on the previous work by Scherbaum and Kieslich (2018) but extend their focus—first by extending the application to value-based decision making, and second by studying two additional design factors. We will contrast different implementations of the starting procedure, response procedure, and stimulus position. For each of these design factors, we compared two of the most common approaches. For the starting procedure, we compared (1) a “dynamic starting procedure,” in which participants had to initiate a mouse movement first to trigger stimulus presentation, and (2) a “static starting procedure,” in which a stimulus was presented after a fixed time interval, and participants can freely decide when to initiate their mouse movement. For the response procedure, we compared (1) a “hover response procedure,” in which participants merely had to enter the response box with the mouse cursor to indicate a response, and (2) a “click response procedure,” in which participants additionally had to click in the response box. For the stimulus position, we compared (1) a “centered stimulus position,” in which stimuli were presented close to the horizontal and vertical center of the screen, and (2) an “edged stimulus position,” in which stimuli are presented within or close to the response boxes located at the upper edges of the screen.

To study the influences of the different design factors, we used a standard intertemporal choice task, a value-based decision-making task that is well established in decision science (e.g., Cheng & González-Vallejo, 2017; Dai & Busemeyer, 2014; Franco-Watkins, Mattson, & Jackson, 2015). In this task, participants have to choose between two monetary rewards that

are available at different time points, constituting a soon–small and a late–large option. The classic effect that can be observed in this task is temporal discounting (for an overview, see Frederick, Loewenstein, & O’Donoghue, 2002), the devaluation of a monetary reward with increasing temporal delay. A well-established modulation of temporal discounting is represented by the date–delay effect (DeHart & Odum, 2015; Read, Frederick, Orsel, & Rahman, 2005), which refers to decreased temporal discounting when temporal delays are presented as calendar dates rather than intervals of days. In an earlier study, these effects and their dynamics had been investigated in a specific mouse-tracking setup, using a dynamic starting procedure, a hover response procedure, and a centered stimulus position (Dshemuchadse et al., 2013). This study revealed that a lack of reflection could account for temporal discounting, as indicated by more direct mouse movements when choosing the soon–small option, in contrast to more indirect mouse movements when choosing the late–large option. Furthermore, an analysis of movement dynamics extracted the relative weights of the different sources of information in this paradigm and suggested that a change in the weighting of the monetary reward could be a source of the date–delay effect.

In the present study, we used the original study as a starting point to investigate how far the different design factors influence the consistency of mouse movements within and across participants. Hence, we studied how these design factors affect different mouse-tracking measures, and eventually the original effects. So far, most mouse-tracking studies have relied on discrete measures at the trial level—for instance, calculating initiation times, movement times, movement deviation, or the number of changes in movement direction for statistical analyses (Freeman & Ambady, 2010). *Discrete* measures integrate information over the course of the whole trial, and thus should be more robust against changes in the specific study design (Scherbaum & Kieslich, 2018).¹ In

¹ In this article, the term *discrete mouse-tracking measure* refers to any measure that summarizes (i.e., over time) dynamic characteristics of a continuous mouse movement in one single value.

contrast, *dynamic* measures focus on within-trial continuous mouse movement over time (Dshemuchadse et al., 2013; Scherbaum et al., 2010; Scherbaum et al., 2013; Scherbaum et al., 2018a; Sullivan, Hutcherson, Harris, & Rangel, 2015). Hence, the latter measures should be more prone to changes in the setup or procedure of a mouse-tracking paradigm (Scherbaum & Kieslich, 2018).

In sum, in a mouse-tracking version of a standard intertemporal choice task, we investigated to what extent specific design factors influence the consistency of mouse movement data. Specifically, we investigated three factors: First, we investigated to what degree a static starting procedure decreases data quality, as compared to a dynamic starting procedure. We examined this factor by comparing data from a previous study (Dshemuchadse et al., 2013), which had applied a dynamic start procedure, with data from a new sample of participants who performed the identical task with a static starting procedure. Second, we investigated how the response procedure (hover vs. click) influences data quality, by varying this procedure within our new study. Third, we investigated how the stimulus position (centered vs. edged) influences data quality, by varying this factor within our new study.

Hypotheses

For the comparison between studies (original vs. new), we expected (H1) that cognitive² effects on discrete movement measures should be influenced only slightly by differences in the starting procedure (see P1), whereas (H2) cognitive effects on within-trial continuous movement measures should be larger and more reliable when using the dynamic starting procedure (see P1). Furthermore, we expected (H3) that the consistency of mouse movements within trials, across trials, and across participants would be higher when using the dynamic rather than the static starting procedure (see P1).

For the comparison of design factors within the new study, we expected (H4) that the cognitive effects based on discrete mouse movement measures should be influenced only slightly by variation of the response procedure and the stimulus position (see P1), whereas (H5) cognitive effects on within-trial continuous movement measures should be larger and more reliable when using the hover response procedure and the centered stimulus position, respectively (see P1, P2, and P3). Moreover, we expected (H6) that the consistency of mouse movements within trials, across trials, and across participants would be higher when using the hover response procedure (as compared to the click response procedure), as well as when

using the centered stimulus position (as compared to the edged stimulus position; see P1, P2, and P3).

Method

Ethics statement

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. All data were analyzed anonymously.

Participants

Forty participants (60% female, 40% male; mean age = 23.28 years, $SD = 4.73$) completed the new experiment, conducted at the Technische Universität Dresden, Germany. The experiment lasted 75 min. Participants were uniformly recruited through the department's database system, which is based on ORSEE (Greiner, 2004). Three participants were left-handed, and another four participants did not indicate their handedness. As in the original study (Dshemuchadse et al., 2013), five participants were excluded because they did not exhibit any discounting (k parameter of the fitted hyperbolic function $< .01$) in the different conditions (delay or date), and thus did not execute a sufficient number of soon–small choices. An additional analysis including all participants did not qualitatively change the pattern of results.

In the original experiment, 42 right-handed students (62% female, 38% male; mean age = 22.95 years, $SD = 2.72$) had participated, of whom six participants were excluded from the data analysis; see above.

All participants had normal or corrected-to-normal vision. They received either class credit or €6 payment (€5 in the original study).

Apparatus and stimuli

The apparatus in the new experiment was identical to the apparatus in the original experiment, using the same lab facility. 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 rate). We used Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997) in Matlab 2010b (MathWorks Inc., Natick, MA) as the presentation software, running on a Windows XP SP2 personal computer. Participants performed their responses with a standard computer mouse (Logitech Wheel Mouse USB). The mouse speed was reduced to one-quarter in the systems settings, and nonlinear acceleration was

² In this article, the term *cognitive effect* is used rather conventionally—that is, the effect of a paradigm, a manipulation, or stimuli on cognitive processing. This effect on cognitive processing should be expressed in mouse movements; in turn, effects on mouse movements should reflect cognitive effects (the reverse inference; see Fig. 1).

Fig. 2 Setup of the experiment: Participants had to click with the mouse cursor on a red box at the bottom of the screen. After clicking, response boxes appeared at the upper edge of the screen. In the dynamic starting procedure (i.e., original experiment; top row), participants had to move the cursor upward in order to start stimulus presentation—only after a movement threshold had been reached was the stimulus presented. In the static starting procedure (i.e., new experiment; bottom row), the stimulus was presented 200 ms after participants had clicked in the red box.

Stimuli were presented either close to the vertical and horizontal center of the screen (i.e., centered stimulus position) or within the response boxes at the top left and right edges of the screen (i.e., edged stimulus position). To respond, participants had to move the mouse cursor to the left or the right response box, either just moving into the box (i.e., hover response procedure; top row) or clicking in the box (i.e., click response procedure; bottom row)

switched off. Mouse movement trajectories were sampled with a frequency of 92 Hz and were recorded from the presentation of the complete choice options (including temporal delays and monetary values) until participants had indicated

Fig. 3 Proportions of choices of the LL option as a function of the interval between the LL and SS options, for the original study with dynamic start condition (*dyn*) and the new study with static start condition (*sta*)

Fig. 4 Differences in average deviations [z -score(AD)] between SS and LL choices as simple main effects for each experiment (dyn vs. sta), as well as separately for each condition within the new experiment (h/c, hover/centered; h/e, hover/edged; c/c, click/centered; c/e, click/edged). Note that error bars depict standard deviations. *Significant t test with $p < .05$; n.s., nonsignificant t test with $p \geq .05$

Fig. 5 Average x -coordinates per normalized time step, depending on choice (SS vs. LL), for the original (A) and new (B) experiments. Coordinates were first averaged within and then across participants. Note that confidence bands depict standard errors

their response and the trial ended. As the targets for mouse movements, response boxes were presented at the top left and top right of the screen.

Due to variation of the mouse-tracking design factors, the setup of the present study deviated from that of the original study with respect to the starting and response procedures (see the Procedure section) and stimulus position. Concerning the stimulus position, the choice options were presented either in the vertical center of the left and the right half of the screen

(centered stimulus position, analogous to the original study; see Fig. 2 top) or within the response boxes at the top left and top right edges of the screen (edged stimulus position; see Fig. 2 bottom).

Procedure

As in the original experiment, participants were asked to decide on each trial which of two options they preferred: the left

Fig. 6 Original experiment: Aggregated time-continuous beta weights and beta weight contrast for the predictors value ratio and time interval. Curves indicate influence strength on predicting the trajectory angle of mouse movements. Beta weights are shown for the date condition (A) and

the delay condition (B). Date–delay beta weight contrasts are depicted in panel C. Note that the lines above the graphs show temporal segments significantly greater than zero (with at least ten significant time steps in a row); confidence bands depict standard errors

Fig. 7 New experiment: Aggregated time-continuous beta weights and beta weight contrast for the predictors value ratio and time interval. Curves indicate influence strength on predicting the trajectory angle of mouse movements. Beta weights are shown for the date condition (A) and

the delay condition (B). Date–delay beta weight contrasts are depicted in panel C. Note that the lines above the graphs show temporal segments significantly greater than zero (with at least ten significant time steps in a row); confidence bands depict standard errors

(soon–small, *SS*) or the right (late–large, *LL*) option. Participants were instructed to respond to the hypothetical choices as if they were real choices. As in the original study, trials were grouped into mini-blocks of 14 trials (see Fig. 2).

Within each mini-block, the two monetary values remained constant and only the temporal delay of the two options was varied. At the start of each mini-block, the monetary values were presented for 5 s, allowing participants to encode them in advance (see Fig. 2).

Each trial consisted of three stages: the alignment stage, the start stage, and the response stage. In the *alignment* stage, participants had to click on a red box at the bottom of the screen within a deadline of 1.5 s. This served to align the starting area for each trial. After this box had been clicked, the start stage began, and two response boxes were presented in the right and left upper corners of the screen. The procedure of the *start* stage differed between the new experiment, in which we implemented a static starting procedure, and the original experiment, in which we had implemented a dynamic starting procedure. In the static starting procedure, the start stage simply lasted 200 ms (this was as long as the start stage in the Scherbaum & Kieslich, 2018, paradigm, and in this case slightly longer than the average duration of the start stage in the original experiment that had used the dynamic starting procedure: 184 ms), and participants simply had to wait for the start of the response stage. In contrast, in the dynamic starting procedure, participants were required to start the mouse movement upward within a deadline of 1.5 s.

Fig. 8 Mean continuous-movement index for the original and new experiments over all conditions (A), as well as for each condition within the new experiment (B). Note that error bars depict standard errors

Fig. 9 (A & B) Heat maps of log-transformed probabilities for pooled mouse movements along the x -axis over normalized time steps for the original (A) and new (B) experiments over all conditions. Brighter colors indicate a higher probability that trajectories crossed a bin at a specific time step. White curves represent the mean mouse movements aggregated

over all trials and participants. (C & D) Average velocities of mouse movements per normalized time step between experiments (C) and for each condition within the new experiment (D) (h/c, hover/centered; h/e, hover/edged; c/c, click/centered; c/e, click/edged). Note that confidence bands depict standard errors

Specifically, the response stage only started after participants had moved the mouse upward by at least four pixels in each of

Fig. 10 Mean bimodality indexes for the original and new experiments, over all conditions (A) as well as for each condition within the new experiment (B). Note that error bars depict standard errors

two consecutive time steps. Usually, this procedure is applied in order to force participants to be already moving when they enter the decision process, to assure that they do not decide first and only then execute the final movement (Scherbaum et al., 2010). In the *response* stage, the remaining stimuli (i.e., the temporal delays) were presented beneath the monetary values, either as a delay in days (e.g., “3 Tage”) or as a German calendar date (e.g., “6. Mai”). The procedure of the response stage also differed between the new experiment, in which we implemented both a click and a hover response procedure, and the original experiment, in which we had implemented a hover response procedure only. Participants in the hover response procedure only had to move the cursor into one of the responses boxes, whereas participants in the clicking response condition additionally had to click in the respective response box. Usually, the hover response procedure is applied in order to force participants to finish the evaluation of their preference before they reach the response boxes (Scherbaum et al., 2010). In both conditions, participants were instructed to respond as quickly and informed as possible and to move the mouse continuously upward once they had initialized their movement.

Fig. 11 Mean movement initiation times for the original and new experiments over all conditions (A), as well as for each condition within the new experiment (B). Note that error bars depict standard errors; for the original experiment, the movement initiation time was 0 by definition

The trial ended after the indication of a response within a deadline of 2 s. If participants missed the respective deadline in one of the three stages, the next trial started with presentation of the red start box. Response times were measured as the duration of the response stage, reflecting the interval between the onset of the stimuli and the indication of the response. Intertrial intervals were measured as the duration of the alignment stage, reflecting the interval between a response in a previous trial and the click on the red box at the bottom of the screen in the current trial.

Fig. 12 Mean movement initiation positions (y-coordinate) for the original and new experiments over all conditions (A), as well as for each condition within the new experiment (B). Note that error bars depict standard errors. A lower value on the y-axis/coordinate corresponds to a more advanced movement initiation position

After onscreen instructions and demonstration by the experimenter, participants practiced on 28 trials (14 trials with feedback and no deadline for any stage of a trial, and 14 trials with feedback and deadlines). Furthermore, preceding each block in which changes of the experimental setup occurred (e.g., stimulus position, date/delay condition), participants practiced ten additional trials in order to get used to the alteration.

Design

In the new experiment, we used the same set of trials as in the original experiment. We orthogonally varied the temporal interval between the options (1, 2, 3, 5, 7, 10, and 14 days), the monetary value of the SS option as a percentage of the monetary value of the LL option (20%, 50%, 70%, 80%, 88%, 93%, 97%, and 99%), and the framing of the time (delay vs. date). The framing of the time was varied between blocks; order was balanced between participants; the percentage of the monetary value of the option was varied between mini-blocks; and the temporal interval between the options was varied across trials. Additionally, we orthogonally varied the temporal delay of the SS option (0 and 7 days) and the monetary value of the LL option (€19.68 and €20.32). Both variations were introduced in order to control for specific effects (e.g., delay duration effect and magnitude effect; Dai & Busemeyer, 2014; Green, Myerson, & McFadden, 1997) and to collect more data without repeating identical trials, which could have led to memory effects. As in the original experiment, and due to the focus of the new experiment, these factors were ignored in the following analyses.

In sum, both the new and the original experiment consisted of two blocks (date and delay), with their order balanced across participants, and contained 224 trials (16 mini-blocks) per block in randomized order. In the new experiment, we additionally varied the response procedure between participants (randomly assigned) and the stimulus position within participants. Hence, for the stimulus position, the two blocks (date and delay) were divided into two subblocks, order balanced across participants.

Data preprocessing and statistical analyses

We excluded erroneous trials (original: 3.6%; new: 2.3%)—that is, trials in which participants missed any of the introduced deadlines (see the Procedure section). Additionally, we also excluded trials with response times less than 300 ms (original study: 5.4%; new study: 3.6%). Mouse trajectories were (1) truncated after the first data sample in the response box of the ultimately chosen option (to enhance comparison of the trajectories between the hover and click response procedures); (2) mirrored, so that all trajectories would end in the left response box; (3) horizontally aligned for a common

starting position (the horizontal middle position of the screen corresponded to 0 pixels, and values increased toward the right); and (4) time-normalized into 100 equal time steps (following the original study).

To investigate the unique influence of each design factor's variation, we computed two analyses for each dependent variable: First, we compared data between experiments (original vs. new); second, we compared data within the new experiment between the different design factor conditions. Through the combination of these analyses, we evaluated the contribution of each design factor to the pattern of results. This logic of the analyses also justified the sample size for the new experiment. Additional *a priori* power analyses (power = .95) for replication of the cognitive effects on discrete mouse measures (see the Results section; $g = 0.7$) revealed that a sample size of $N = 24$ would have been sufficient (Faul, Erdfelder, Buchner, & Lang, 2009). Thus, with the sample size in the new experiment ($N = 35$), even small to medium effects—such as the date–delay effect—were appropriately powered (power = .8).

The data from the original experiment had been used in a previous publication (Dshemuchadse et al., 2013). The data were reanalyzed completely for the present article. Due to improvements in algorithms for the analysis of mouse movement data in recent years, the results of the original experiment might exhibit minor variations in comparison to the results reported in the original publication.

Data preprocessing and aggregation were performed in Matlab 2015a (the Mathworks Inc.). Statistical analyses were performed in Matlab and JASP 0.8.6 (JASP Team, 2018). The primary data from both experiments and the analysis scripts are publicly available at the Open Science Framework (<https://osf.io/3w6cr>).

Results

Comparison of groups

Since our analysis built on two independent groups of participants from different experiments (original vs. new), we first checked for differences between the groups, other than the starting procedure, response procedure, and stimulus position. The groups of participants showed significant differences in neither age, $t(69) = -0.46$, $p = .65$, $g = -0.11$, 95% bootstrapped CI of g (bCI_g) $[-0.55, 0.39]$, nor sex (original: 21 female, 15 male; new: 20 female, 15 male). All other descriptive variables also showed no significant differences (all $ps > .120$; see Appendix A), with the exception that the participants in the new experiment reported feeling more rested and less tired, 95% CIs $[-0.98, -0.02]$, $[-0.01, 0.93]$, respectively.

Following Scherbaum and Kieslich (2018), we analyzed the intertrial interval as a general measure for task speed,

which was defined by the time interval between reaching the response box on the current trial and clicking on the start box for the next trial. We found a significant difference between groups, $t(69) = 2.33$, $p < .05$, $g = 0.55$, 95% bCI_g $[0.12, 1.00]$, indicating that the participants in the original experiment were generally slower ($M = 0.92$ s, $SD = 0.26$ s) than the participants in the new experiment ($M = 0.85$ s, $SD = 0.18$ s). Contrasting the difference in intertrial interval as a general measure of task speed, we found the reverse pattern for response times, $t(69) = -3.23$, $p < .05$, $g = -0.76$, 95% bCI_g $[-1.32, -0.28]$, indicating that the participants in the original experiment responded generally faster ($M = 0.88$ s, $SD = 0.27$) than the participants in the new experiment ($M = 1.00$ s, $SD = 0.31$).

In combination, these results corroborated our general subjective intuition that mouse-tracking paradigms are procedurally very demanding for participants, and hence that participants have to take some time to relax during the task—either during the intertrial interval or after starting the trial. This motivated a further analysis of the sum of the intertrial interval and the response time, as a general measure for trial speed. We found no significant difference between groups, $t(69) = -1.16$, $p = .25$, $g = -0.27$, 95% bCI_g $[-0.80, 0.17]$. Taken together, the results suggested that the participants in both experiments completed trials similarly fast, but the participants in the original experiment took more time between trials, whereas the participants in the new experiment took more time within trials.

Cognitive effects

In the next step, we wanted to find out whether our variations in the design factors led to differences in the observed cognitive effects. In this regard, as a manipulation check, we expected that choice effects (i.e., the date–delay effect) would remain unaffected by variations in the design factors. Furthermore, we expected that differences in the starting procedure should only slightly influence discrete cognitive measures (H1) but should have a greater impact on continuous measures, resulting in larger and more reliable effects for the dynamic than for the static starting procedure (H2). Likewise, the variation of the response procedure or stimulus position within the new experiment should only slightly alter discrete measures (H4) but should show larger and more reliable effects for the hover response procedure and centered stimulus position (H5).

Choice effects

We examined choice effects in the new experiment, expecting to find larger temporal discounting in the delay than in the date condition, replicating the results of the original experiment.

To follow the approach of the original experiment with respect to the exclusion of participants (i.e., applying the exact same exclusion criterion; see the Participants section), we calculated indifference points for the two conditions (delay vs. date), which is the estimated value ratio for each interval at which a participant would be indifferent about choosing the SS or the LL option,³ and consecutively fitted a hyperbolic function to these indifference points over time intervals.⁴ From this fitted function, we retrieved the k parameter as an overall measure of discounting (Dshemuchadse et al., 2013).

However, to investigate the date–delay effect we followed a different approach, since k parameters come with several caveats with respect to distribution and sensitivity that might distort comparisons within and between experiments (Myerson, Green, & Warusawitharana, 2001; please see the online supplement). First, we calculated the proportion of LL choices for each interval and derived the area under the curve (AUC) as a summary measure of temporal discounting. For reasons of simplicity, we only calculated and report the results of the difference in AUC between the two conditions (delay vs. date) as a simple main effect for each experiment (original vs. new).

In the original dataset, we revealed a significantly smaller AUC in the delay condition ($M = 0.36$, $SD = 0.16$) than in the date condition ($M = 0.43$, $SD = 0.16$), $t(35) = -3.68$, $p < .001$, $g = -0.41$, 95% bCI_g $[-0.67, -0.20]$. This was also the case for the new dataset, in which we found a significantly smaller AUC in the delay condition ($M = 0.40$, $SD = 0.13$) than in the date condition ($M = 0.47$, $SD = 0.14$), $t(34) = -4.78$, $p < .001$, $g = -0.50$, 95% bCI_g $[-0.74, -0.20]$. Thus, the participants in the new experiment showed more temporal discounting in the delay than in the date condition, replicating the effect from the original experiment (see Fig. 3).

Discrete mouse movement effects

We inspected discrete movement measures by calculating the directness of mouse trajectories toward the response boxes. We expected less direct mouse movements for LL choices, indicating greater reflection on these decisions. Directness

³ The calculation of indifference points was obtained by fitting a logistic regression model to participants' choices for each interval. The fitting of the logistic regression model was performed using the StixBox mathematical toolbox by Anders Holtsberg (www.maths.lth.se/matstat/stixbox/). The fit was based on the model $\log\left[\frac{p}{1-p}\right] = Xb$, where p is the probability that the choice is 1 (SS) and not 0 (LL), X represents the value ratio (SS/LL), and b represents the point estimates for the logistic function. Please see our custom Matlab function findIndifferencePoints.m, available at <https://osf.io/3w6cr/>, for further details and the specific implementation.

⁴ The fitting of the hyperbolic function was performed by applying Matlab's multidimensional unconstrained nonlinear minimization function to the hyperbolic function $1/(1 + kX) = Y$, with X denoting the time interval, Y denoting the indifference point, and k denoting the discounting parameter. Please see our custom Matlab function fitCurve.m, available at <https://osf.io/3w6cr/>, for further details and the specific implementation.

was defined as the average deviation (AD) of trajectories from a notional line between the starting and ending points of the movement. We computed AD for each trial and then z -scored and averaged it per participant according to experiments and/or conditions. For reasons of simplicity, we only calculated and report the results of the difference in AD between SS and LL choices as simple main effects for each experiment (original vs. new), as well as separately for each condition within the new experiment (see Fig. 4); but please see the online supplement for a visual inspection of further, more fine-grained discrete mouse movement effects that were also reported in the original study.

In the original dataset, we reproduced significantly less direct mouse movements for LL than for SS choices, $t(35) = 2.39$, $p = .022$, $g = 0.75$, 95% bCI_g $[0.15, 1.41]$. In the new dataset, the same effect occurred neither over all conditions, $t(36) = 0.81$, $p = .42$, $g = 0.26$, 95% bCI_g $[-0.38, 0.90]$, nor within conditions, all $ts < 0.98$, $ps > .33$, $0.03 < g < 0.40$.

In sum, our analysis of the AD of mouse trajectories did not show the expected robustness of discrete movement effects (controverting H1 and H4). As expected, we reproduced the effect in the reanalysis of the original experiment. In contrast, in the new experiment the effect vanished. A closer examination of the simple main effects (see Fig. 4), as well as a comparison of the raw deviations between the experiments (original vs. new; see Fig. 5), indicated that the static starting procedure accounts for our finding: Fig. 5 illustrates a severe increase of more direct mouse movements when a static starting procedure was used. However, we will elaborate on this mechanism more specifically when turning to the continuous measures of mouse movements later on.

Continuous mouse movement effects

For the statistical analysis of mouse movement dynamics, we performed time-continuous multiple linear regression (TCMR; Scherbaum & Dshemuchadse, 2018) on movement angle on the x/y plane. To dissect the influences of the independent variables on mouse movements within a trial, we applied a three-step procedure: (1) We defined the variables value ratio (eight discrete steps, ranging from 0.2 to 0.99) and time interval (seven discrete steps, ranging from 1 to 14) as predictors for each trial and each participant (see the Design section). The value ratio is given by the monetary value of the SS option as a percentage of the monetary value of the LL option. Time interval is given by the difference of the temporal delays between options. These predictors were normalized to a range of -1 to 1 , in order to extract comparable beta weights. (2) We performed multiple regression with these predictors on trajectory angle (also normalized to a range of -1 to 1) for each time step. This resulted in 100 multiple regressions (from 100 time steps), returning two beta weights for each of the 100 time steps. (3) These beta weights were separately averaged

for each time step and displayed in a time-varying curve of influence strength. We compared curves of influence strength between the date and delay conditions by contrasting the beta weights for each time step and predictor. Furthermore, we computed one-sided Student's *t* tests for each beta weight contrast for each time step. Significant temporal segments were identified with at least ten consecutive significant time steps, to compensate for multiple comparisons (Dshemuchadse et al., 2013; Scherbaum et al., 2010). Curves and contrasts for both experiments are displayed in Fig. 6 (original) and Fig. 7 (new).

With regard to the curves of the beta weights, we found that trajectory angle was significantly predicted by value ratio and time interval in both experiments (see Fig. 6A and B and Fig. 7A and B). However, those curves also illustrate a higher predictive power of value ratio and time interval in the original than in the new experiment, suggesting that mouse movements contain more stimulus information when the dynamic starting procedure is used.

With regard to the beta weight contrast date–delay, we found a significant temporal segment in the original dataset, which was comparable to the one already reported in the original article (Dshemuchadse et al., 2013; see Fig. 6C). Beta weights for the predictor value ratio were significantly higher in the date condition at time steps 49 to 82 (mean times from 429 to 718 ms) than in delay condition. In the new dataset, we found no such temporal segments for the beta weight contrast matching our criterion of significance (Fig. 7C). The same applied for contrasts in the different conditions within the new experiment (please see the [online supplement](#)).

In sum, variation in starting procedure leads to a substantial change in date–delay differences when performing a continuous analysis of mouse movements (supporting H2): Whereas the value ratios in the date condition in the original experiment had a significantly greater impact on predicting mouse movements on the *x/y*-plane, this difference was not significant in the new experiment. However, variations in response procedure and stimulus position made no additional change to the date–delay difference (controverting H5).

Mouse movement consistency

In addition to cognitive effects on the discrete and continuous movement measures, we analyzed the influence of design factors on the consistency of mouse movements within trials, across trials, and across participants. We concentrated on a quantitative evaluation of mouse movement consistency, but will augment the results with some qualitative interpretations that were derived from visual inspection of pooled or mean mouse movements (see also Appendix B).

Consistency within trials was evaluated by calculating the continuous-movement index and inspecting the mean profiles

of movements on the *y*-axis and velocity. Consistency across trials was evaluated by computing the bimodality index and inspecting visual heat maps of pooled movements on the *x/y*-plane. Consistency across participants was evaluated by comparing the distributions of movement initiation times and positions after the stimuli were presented. We expected higher consistency of mouse movements for the dynamic than for the static starting procedure (H3). Within the new experiment, we anticipated the same pattern for the hover response procedure and the centered stimulus position (H6).

Continuous-movement index

To examine the consistency within trials, we calculated the continuous-movement index. This was computed for each trial as the correlation of the actual position on the *y*-axis and a hypothetical position on the *y*-axis for a constant and straight mouse movement from the starting point to the response box. A strong correlation would indicate a smooth and constant movement (Scherbaum & Kieslich, 2018). The continuous-movement index was averaged for each participant and first compared between experiments (original vs. new), revealing a significant difference, $t(69) = 5.18$, $p < .001$, $g = 1.22$, 95% bCI_g [0.83, 1.69]: Mouse movements were smoother and more constant in the original experiment ($M = .95$, $SD = .047$) than in the new experiment ($M = .86$, $SD = .097$; the index is correlation [-1, 1]; see Fig. 8A). Furthermore, we compared continuous-movement indexes within the new experiment. An analysis of variance (ANOVA) with the independent variables stimulus position and response procedure showed significant main effects for stimulus position, $F(1, 33) = 14$, $p < .001$, $\eta^2 = .27$, and response procedure, $F(1, 33) = 8.9$, $p = .005$, $\eta^2 = .21$. Mouse movements were smoother and more constant for the edged than for the centered stimulus position, as well as for the click than for the hover response procedure (see Fig. 8B). The interaction Stimulus Position $\times$ Response Procedure was significant, as well, $F(1, 33) = 5.11$, $p = .031$, $\eta^2 = .10$, reflecting a stronger effect of the stimulus position for the hover response procedure (see Fig. 8B).

In sum, our analysis of movement consistency within trials via the continuous movement index provided mixed results with regard to our hypotheses. Mouse movements in the static starting procedure were less smooth and less constant than those in the dynamic starting procedure (supporting H3). Furthermore, the analyses revealed that a click response procedure and an edged stimulus position can partly compensate for the former effect (controverting H6). Our quantitative results were also supported by visual inspection of the mean profiles of movements on the *y*-axis (see Appendix B, Fig. 16) and of velocity (see Figs. 9C and D).

Bimodality index

To examine consistency across trials, we assessed bimodality for distributions of AD per participant. Therefore, we calculated the bimodality index, which uses kurtosis and skewness to estimate the bimodality of a distribution (Freeman & Dale, 2013). The bimodality index ranges from 0 to 1; an index above .555 suggests bimodality—hence, in our terms, lower consistency across trials. A comparison between experiments (original vs. new) did not show significant differences, $t(69) = -1.46$, $p = .15$, $g = -0.34$, 95% bCI_g [-0.82, 0.12], indicating comparable bimodality indexes between the original experiment ($M = .38$, $SD = .13$) and the new experiment ($M = .43$, $SD = .14$; bimodality index [0,1]; see Fig. 10A). Within the new experiment, an ANOVA with the independent variables stimulus position and response procedure revealed significant main effects for response procedure, $F(1, 33) = 4.25$, $p = .047$, $\eta^2 = .11$, and stimulus position, $F(1, 33) = 5.02$, $p = 0.032$, $\eta^2 = .13$. The bimodality index was higher for the click than for the hover response procedure, as well as for the edged than for the centered stimulus position (see Fig. 10B). The interaction was not significant, $F(1, 33) = 0.12$, $p = .73$, $\eta^2 < .01$.

In sum, our analysis of movement consistency across trials via the bimodality index provided mixed results with respect to our hypotheses. In contrast to our expectations, distributions of AD were estimated to be similar and to be distributed rather unimodally, and hence to be equally consistent with both starting procedures (controverting H3). Nevertheless, within the new experiment, consistency decreased for the click response procedure and the edged stimulus position (supporting H6), though in no condition did the bimodality index exceed the threshold of .555 ($.37 < M < .50$, $.12 < SD < .15$). However, our quantitative results were controverted by visual inspection of the pooled profiles of movements on the x/y -plane (see Appendix B, Figs. 13 and 14) and the x -axis (see Figs. 9A and B). The visual inspection suggested that a static starting procedure introduces two sets of sharp and qualitatively different movement profiles: A high rate of mouse movements went straight (i.e., without curvature) toward the chosen option, whereas some mouse movements went straight toward the unchosen option, followed by another straight movement toward the chosen option (i.e., a change of mind; see, e.g., Resulaj et al., 2009). In contrast, the movement profiles with the dynamic starting procedure yielded the full continuum of mouse movements, ranging from straight profiles to changes of mind, with a high frequency of smoothly curved profiles in between. Furthermore, within the new experiment, the inspection suggested, a hover response procedure yielded mainly straight movements toward the chosen option, whereas a click response procedure enhanced discrete changes of mind.

Movement initiation strategies

To examine consistency across participants, we examined both movement initiation times and positions. Analogously to the original experiment, the movement initiation time was given by the time between stimulus presentation and moving at least four pixels in two consecutive time steps. The movement initiation position was given by the mouse position on the y -axis, where participants initiated their movement after stimulus presentation. Since in the original experiment, with the dynamic starting procedure, the movement initiation strategies were highly controlled, here we concentrated on the differences in initiation strategies within the new experiment. However, as for the comparison between experiments (original vs. new), we will check whether the higher-controlled dynamic starting procedure indeed yielded less variance than did the lower-controlled static starting procedure.

Movement initiation time

Due to the differences in starting procedure (dynamic vs. static), a comparison of movement initiation times between experiments (original vs. new) was hardly feasible, since in the dynamic starting procedure the stimulus presentation was triggered by movements, and hence movements had already been initiated at stimulus presentation; in the static starting procedure, stimuli were presented after a fixed 200 ms, and hence movements were not necessarily already initiated at stimulus presentation (please see Appendix B, Fig. 17). Therefore we focused on variation in movement initiation time in the new experiment only. We found a significant initiation time ($M = 0.43$, $SD = 0.21$; see Fig. 11A), $t(34) = 11.73$, $p < .001$, $d = 1.98$, 95% bCI_d [1.54, 2.82]. Additionally, we conducted an ANOVA with the independent variables stimulus position and response procedure. The ANOVA revealed a significant main effect for stimulus position, $F(1, 33) = 20.96$, $p < .001$, $\eta^2 = .37$, indicating higher initiation times with the centered stimulus position than with the edged stimulus position (see Fig. 11B); neither the main effect of response procedure, $F(1, 33) = 0.43$, $p = .52$, nor the interaction, $F(1, 33) = 2.38$, $p = .13$, was significant.

Movement initiation position

To evaluate what participants did during the 200 ms between clicking on the start box and stimulus presentation, we examined movement initiation position. Here the results revealed no different variance in initiation position between the new experiment ($M = 923$ px, $SD = 26$ px) and the original experiment ($M = 896$ px, $SD = 15$ px; see Fig. 12A), Levene's test $F(1, 69) = 3.54$, $p = .06$; but please see Appendix B, Fig. 18. However, as is suggested by the means and also supported statistically, $t(69) = -5.26$, $p < .001$, $g = -1.24$, 95% bCI_g

$[-1.97, -0.69]$, on average, the participants in the static starting procedure supposedly moved the mouse only slightly after clicking the start box, and hence stayed closer to the start box before stimulus presentation than did the participants in the dynamic starting procedure, who had to move to trigger stimulus presentation (see also Figs. 9C and D). To explain the variance in the new experiment, we conducted an ANOVA with the independent variables stimulus position and response procedure. The ANOVA revealed a significant main effect of stimulus position, $F(1, 33) = 4.96, p = .048, \eta^2 = .11$, indicating a more advanced initiation position for the edged than for the centered stimulus position (see Fig. 12B); neither the main effect of response procedure, $F(1, 33) = 1.80, p = .19$, nor the interaction, $F(1, 33) = 0.08, p = .78$, was significant.

In sum, our analysis of movement consistency across participants via movement initiation time and position confirmed our hypotheses. Participants in the lower-controlled static starting procedure exploited their freedom with regard to initiation times (supporting H3) and stuck close to the start box with regard to their average initiation position, though with high variability (see Appendix B, Figs. 17 and 18). The greater variety of initiation strategies in the static starting procedure was partly explained by the stimulus position but could not compensate for the general effect. Thus, with an edged stimulus position, participants initiated their movements earlier and at a more advanced position than in trials with a centered stimulus position (controverting H6).

Discussion

In this study, we investigated how different implementations (i.e., design factors) of the mouse-tracking paradigm influence the mapping between (covert) cognitive processing and (overt) mouse movement. We expected that different design factors would influence the consistency of mouse movements (within trials, between trials, and between participants), and hence the theoretically expected cognitive effects as given by discrete and continuous mouse-tracking measures, whereas choice effects—that is, decision outcomes—should remain unaffected. Overall, we found that design factors significantly influence both the cognitive effects and consistency of mouse movements, though not always in the directions we expected.

Specifically, we investigated the influences of three factors—the starting procedure, the response procedure, and the stimulus position—in a mouse-tracking version of a standard intertemporal choice task. To this end, we compared data from two studies: on the one hand, a previously published study (Dshemuchadse et al., 2013), which used a dynamic starting procedure in combination with a hover response procedure and stimuli that were presented at the center of the screen; on the other hand, a new study, which used a static starting procedure in combination with varying response

procedures (hover vs. click) and stimulus positions (centered vs. edged).

With regard to the influence of design factors on the theoretically predicted effects, we split our analysis by examining effects on choice behavior as well as discrete and continuous movement measures. Examining choice behavior, we found the expected effects, such as temporal discounting and the date–delay effect. Interestingly, we found generally stronger temporal discounting in the original than in the new study. Though this is a common finding in the temporal-discounting literature (cf. Lempert & Phelps, 2016), it could potentially be attributed to group differences concerning fatigue, since the participants in the new study reported being less tired and more rested than the participants in the original study (see Appendix A; Berkman, Hutcherson, Livingston, Kahn, & Inzlicht, 2017; Blain, Hollard, & Pessiglione, 2016). The date–delay effect, however, was present for both groups with comparable effect sizes. The occurrence of temporal discounting and the date–delay effect served as a general manipulation check for the present intertemporal choice paradigm and was a prerequisite for investigating the cognitive effects found in both discrete and continuous movement measures. When examining discrete movement measures (i.e., the average deviation), we found that the anticipated cognitive effects were sensitive to the influences of the starting procedure: The originally found effect, showing more direct mouse movements for late–large than for soon–small choices, was not significant when we used a static starting procedure; variations of the response procedure and stimulus position did not introduce further variance. While we had not expected to find a strong effect of the starting procedure on the discrete measures, the finding is in line with previous discussions that, especially for more subtle cognitive effects (as is often the case in value-based decision making), design factors might play a more important role (Scherbaum & Kieslich, 2018). When examining continuous measures in a time-continuous multiple regression analysis, we found that the anticipated movement effects were highly sensitive to influences of the starting procedure: The effect we originally found, showing reliably higher beta weights for the predictor value ratio in the date condition than in the delay condition, vanished when we applied a static starting procedure; again, variations in the response procedure and stimulus position could not compensate for this influence. This finding supported our expectations and substantiates previous results regarding the influence of the starting procedure on the cognitive effects reflected by continuous movement measures (Scherbaum & Kieslich, 2018).

With regard to the influence of design factors on the consistency of mouse movements, we found that the static starting procedure yielded less consistent movements than did the dynamic starting procedure, supporting our expectations and previous research (Scherbaum & Kieslich, 2018). This pattern

was present for the consistency of mouse movements both within trials and across participants. Only for consistency across trials did we not find the expected differences between starting procedures, though on a descriptive level the data pointed in the expected direction. We also found that the additional influences of the response procedure and the stimulus position were less distinct, and largely in contrast to our expectations: We did not find the expected evidence that a hover response procedure and a centered stimulus position would yield more consistent mouse movements than a click response procedure and an edged stimulus position. Instead, we found that a click response procedure and an edged stimulus position yielded more consistent mouse movements. Only for the consistency across trials did we find the expected pattern in favor of the hover response procedure and the centered stimulus position.

Our results suggest that different design factors in a mouse-tracking paradigm—here, specifically the starting procedure, as well as the response procedure and stimulus position—indeed influence the consistency of mouse movements, and thereby the theoretically important effects investigated in such studies. We showed that dynamic effects in value-based decision making—here, intertemporal choice specifically—are highly sensitive to the setup of the mouse-tracking paradigm. Hence, we must assume that each experimenter's decision about the design of a mouse-tracking paradigm might influence the results that she might find and report. Furthermore, we must assume that a comparison of effects over several studies (e.g., in meta-analyses) should take into account differences in the methodological setup.

To exemplify this conclusion in the case of intertemporal choice, we refer to the literature that has also investigated the action dynamics of intertemporal decision making. To our knowledge, only four such studies exist, including our original work (Calluso et al., 2015; Dshemuchadse et al., 2013; O'Hora et al., 2016; Scherbaum et al., 2018a). The most basic results of those studies focused on the comparison of mouse movements between choices of the soon–small and the late–large options. The original article by Dshemuchadse et al. (2013) showed that soon–small choices overall were associated with more direct mouse movements than late–large choices, and that this effect positively correlated with the difficulty of the choice task. Scherbaum et al. (2018a), as well as Calluso et al. (2015), reported the opposite effect concerning choices, but they did not evaluate its connection to difficulty. O'Hora et al. (2016) reported no overall (main) effect for the choice but found that the directness of the mouse movements depended on the subjective evaluation of the choice task, and hence on its difficulty. Our results suggest that those results cannot be compared directly, since in each of these studies another setup of the mouse-tracking procedure was applied: The latter two studies applied a static starting procedure and a click response, whereas our original study applied a dynamic starting procedure and a hover response.

However, besides the fact that we found evidence that a comparison between studies should take methodological differences into account, our results also suggest that the mapping between cognitive processing and mouse movements might vary with differing design factors. Thus, it is a plausible assumption that the validity of the reverse inference from mouse movements to cognitive processing also varies with differing design factors, which makes a comparison between studies even more problematic. To produce (theoretically) valid, reproducible, and comparable results, the mapping should be optimized and held constant.

Toward an evidence-based gold standard for mouse-tracking paradigms

Already, a debate about the boundary conditions and standards for mouse-tracking paradigms is gradually evolving, from discussion at conferences and meetings to articles in the literature (Faulkenberry & Rey, 2014; Fischer & Hartmann, 2014; Scherbaum & Kieslich, 2018; Wulff, Haslbeck, & Schulte-Mecklenbeck, 2018). The discussion has been focused on two main issues: (1) How must a mouse-tracking paradigm be designed in order to continuously capture the ongoing decision process—that is, to ensure a direct mapping of cognitive processing onto motor movements? (2) How must such trajectories be analyzed with the aim to make valid inferences about the underlying cognitive processes? The latter issue is motivated by the notion that effects can easily be driven by only a small subset of trajectories (Wulff et al., 2018). In consequence, the consistency of mouse movements within trials, across trials, and across participants is of superordinate importance for the diagnostic value of trajectories when measuring cognitive processes.

Since the first issue is more analytic (i.e., ad hoc) and the second issue is more synthetic (i.e., post hoc), we should take a new perspective by tackling both issues experimentally. This approach would consist of experimental manipulation of the design factors in mouse-tracking paradigms and examination of their influence on the consistency of mouse movements, which has recently been investigated for the first time (Scherbaum & Kieslich, 2018; but please see Burk, Ingram, Franklin, Shadlen, & Wolpert, 2014, demonstrating a decrease of changes of mind with decreasing vertical distance between choice options). Together with two further studies (Grage, Schoemann, & Scherbaum, 2018; Kieslich, Schoemann, Grage, Hepp, & Scherbaum, 2018), our study has incorporated this approach and the recent results, and eventually will provide a strong argument in favor of evidence-based standards for mouse-tracking paradigms. On the basis of the current evidence, we recommend using dynamic starting and a hover procedure in combination with a centered stimulus position. As such, the recommendations might depend on the relevant task, and hence the measured cognitive process, since

there might be cases in which the methodological implementation must be adjusted (cf. Scherbaum & Kieslich, 2018). Accordingly, we recommend using an edged stimulus presentation when a static starting procedure is required, to facilitate earlier horizontal mouse movement. Certainly, our recommendations constitute only a starting point for future discussions about standards for mouse-tracking paradigms. Hence, we encourage further work in the same direction that can challenge our recommendations.

However, the validation of mouse-tracking as a process-tracing method using the present approach can only provide standards for designing mouse-tracking studies that can produce the most reliable and comparable data. Our approach comes with two caveats: First, we cannot measure the mapping between cognitive processing and mouse movements without interference from design factors. Second, we cannot validate the (theoretical) basis that justifies mouse-tracking as being a process-tracing method. This assumption states that continuous cognitive processing leaks into continuous motor movements (Spivey, 2007; Spivey & Dale, 2006). Whenever we conduct mouse-tracking studies and analyze mouse movement trajectories, we must accept this assumption in order to make the reverse inference.

It has been pointed out that this reasoning is deductively invalid, because different cognitive processes can be responsible for the same observable pattern (Poldrack, 2006, p. 59). To overcome this caveat, *forward inference* (Heit, 2015; Henson, 2006) has been emphasized: This technique turns the direction of inference upside down, by making cognitive processing explicit through instructions, and has recently been studied for eye-tracking paradigms (Schoemann, Schulte-Mecklenbeck, Renkewitz, & Scherbaum, 2018; Schulte-Mecklenbeck, Kühberger, Gagl, & Hutzler, 2017). The forward inference route, from cognitive processing to behavioral patterns, might also be beneficial to approaching the problems raised by mouse-tracking. By taking the forward inference route, one might be able to evaluate to what degree continuous processing leaks into continuous (mouse) movements and what design factors of a mouse-tracking paradigm produce the best fit between processing and movements. Here an appropriate experimental paradigm is still to be designed, but it might be the necessary next step, complementing recent efforts toward the development of an evidence-based gold standard for mouse-tracking paradigms, as well as toward the validation of mouse-tracking as a process-tracing method.

Conclusion

We studied the impact of three commonly occurring methodological variations on the quality of mouse-tracking measures, and hence, on the reported cognitive effects. We found varying effects with varying methodological setups of the mouse-tracking paradigm. In sum, our study—augmented by

previous (Scherbaum & Kieslich, 2018) evidence—suggests that the methodological setup of mouse-tracking studies needs to be taken into account when interpreting mouse-tracking data. The problems concern both the validity of mouse-tracking as a process-tracing method and the principles of reproducible science (i.e., detailed documentation of the setup and procedure; cf. Munafò et al., 2017; Nosek et al., 2015). Both problems can be tackled by understanding how different design choices affect the resulting data, and consequently solved by compiling an evidence-based gold standard for mouse-tracking.

Author note We thank Marie Gotthardt for her support in data collection. This research was partly supported by the German Research Council (DFG grant SFB 940/2 to S.S.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix

Descriptive sample statistics

Table 1 Descriptive sample statistics for the original (Group = 1) and new (Group = 2) experiments

Variable	Group	<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>
Sex (<i>f</i> = 0, <i>m</i> = 1)	1	36	.417	.500	.083
	2	35	.429	.502	.085
Age (years)	1	36	23.000	2.849	0.475
	2	35	23.429	4.840	0.818
ITI (s)	1	36	0.924	0.157	0.026
	2	35	0.855	0.073	0.012
RT (s)	1	36	0.876	0.138	0.023
	2	35	1.004	0.194	0.033
Time of day (h)	1	36	13.412	2.433	0.405
	2	35	12.421	2.889	0.488
Sleep (h)	1	36	7.222	1.105	0.184
	2	35	7.114	1.138	0.192
Rested (1–5)	1	36	3.500	0.941	0.157
	2	35	3.943	0.838	0.142
Motivated (1–5)	1	36	3.806	0.822	0.137
	2	35	4.029	0.707	0.119
Energy (1–5)	1	36	3.250	0.770	0.128
	2	35	3.457	1.010	0.171
Bored (1–5)	1	36	1.556	0.607	0.101
	2	35	1.571	0.608	0.103
Stressed (1–5)	1	36	2.250	1.079	0.180
	2	35	2.086	0.919	0.155
Interested (1–5)	1	36	4.083	0.874	0.146
	2	35	4.171	0.568	0.096
Nervous (1–5)	1	36	1.778	0.832	0.139
	2	35	1.800	0.719	0.122
Aroused (1–5)	1	36	1.861	0.990	0.165
	2	35	2.086	0.887	0.150
Tired (1–5)	1	36	2.417	1.025	0.171
	2	35	2.000	0.767	0.130

Table 2 Results of independent samples *t* tests comparing different sample statistics between the original and the new experiment

Variabel	<i>t</i>	<i>df</i>	<i>p</i>	Cohen's <i>d</i>	95% CI for Cohen's <i>d</i>	
					Lower	Upper
Sex (<i>f</i> = 0, <i>m</i> = 1)	− 0.100	69.000	.921	− 0.024	− 0.489	0.442
Age	− 0.456	69.000	.650 ^a	− 0.108	− 0.574	0.358
ITI	2.334	69.000	.023 ^a	0.554	0.078	1.026
RT	− 3.226	69.000	.002 ^a	− 0.766	− 1.246	− 0.281
Time (h)	1.564	69.000	.122	0.371	− 0.099	0.839
Sleep (h)	0.405	69.000	.686	0.096	− 0.370	0.561
Rested	− 2.092	69.000	.040	− 0.497	− 0.967	− 0.022
Motivated	− 1.225	69.000	.225 ^a	− 0.291	− 0.757	0.178
Energy	− 0.974	69.000	.334	− 0.231	− 0.697	0.237
Bored	− 0.110	69.000	.913	− 0.026	− 0.491	0.439
Stressed	0.690	69.000	.493	0.164	− 0.303	0.629
Interested	− 0.502	69.000	.617	− 0.119	− 0.584	0.347
Nervous	− 0.120	69.000	.905	− 0.029	− 0.494	0.437
Aroused	− 1.006	69.000	.318	− 0.239	− 0.705	0.229
Tired	1.935	69.000	.057 ^a	0.459	− 0.014	0.929

Student's *t* tests were performed. ^a Levene's test was significant ($p < .05$), suggesting a violation of the equal-variance assumption. Applying Welch's *t* test yielded comparable results

Raw and mean mouse movements

Mouse movements on the *x/y*-plane

Fig. 13 Heat maps of log transformed probability for pooled mouse movements along the *x/y* plane for the original (A) and new (B) experiment. Brighter colors indicate higher probability. White curves represent mean mouse movements aggregated over all trials and participants

Fig. 14 Heat maps of log transformed probability for pooled mouse movements along the x/y plane within the new experiment. Movements were separated according to hover (A and B; $n = 16$) and click (C and D; $n = 19$) response procedure as well as centered (A and C) and edged (B and

D) stimulus position. Brighter colors indicate higher probability. White curves represent mean mouse movements aggregated over all trials and participants

Mouse movements on the x -coordinate

Fig. 15 Average x -coordinate of mouse movements per normalized time step between experiments (A) and within the new experiments (B) for each condition (h/c, hover/centered; h/e, hover/edged; c/c, click/centered; c/e, click/edged). Note that confidence bands depict standard errors

Mouse movements on the y-coordinate

Fig. 16 (A and B) Heat maps of log transformed probability for pooled mouse movements along the y-axis over normalized time steps for the original (A) and new (B) experiment. Brighter colors indicate higher probability that trajectories crossed a bin at a specific time step. White curves represent mean mouse movements aggregated over all trials and

participants. (C and D) Average y coordinate of mouse movements per normalized time step between experiments (C) and within the new experiments (D) for each condition (h/c, hover/centered; h/e, hover/edged; c/c, click/centered; c/e, click/edged). Note that confidence bands depict standard errors

Fig. 17 Probability distribution of movement initiation time (in seconds) after stimulus onset for the original experiment (A) and the new experiment overall conditions (B), as well as for each condition within the new

experiment (C–F: C, hover/centered; D, hover/edged; E, click/centered; F, click/edged)

Fig. 18 Probability distribution of movement initiation position (y coordinate in px) after stimulus onset for the original experiment (A) and the new experiment overall conditions (B), as well as, for each

condition within the new experiment (C–F: C, hover/centered; D, hover/edged; E, click/centered; F, click/edged)

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

References

- Barca, L., & Pezzulo, G. (2012). Unfolding visual lexical decision in time. *PLoS ONE*, 7, e35932. <https://doi.org/10.1371/journal.pone.0035932>
- Barca, L., & Pezzulo, G. (2015). Tracking second thoughts: Continuous and discrete revision processes during visual lexical decision. *PLoS ONE*, 10, e116193:1–14. <https://doi.org/10.1371/journal.pone.0116193>
- Berkman, E. T., Hutcherson, C. A., Livingston, J. L., Kahn, L. E., & Inzlicht, M. (2017). Self-control as value-based choice. *Current Directions in Psychological Science*, 26, 422–428. <https://doi.org/10.1177/0963721417704394>
- Blain, B., Hollard, G., & Pessiglione, M. (2016). Neural mechanisms underlying the impact of daylong cognitive work on economic decisions. *Proceedings of the National Academy of Sciences*, 113, 6967–6972. <https://doi.org/10.1073/pnas.1520527113>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436. <https://doi.org/10.1163/156856897X00357>
- Burk, D., Ingram, J. N., Franklin, D. W., Shadlen, M. N., & Wolpert, D. M. (2014). Motor effort alters changes of mind in sensorimotor decision making. *PLoS ONE*, 9, e92681. <https://doi.org/10.1371/journal.pone.0092681>
- Calluso, C., Committeri, G., Pezzulo, G., Lepora, N. F., & Tsoni, A. (2015). Analysis of hand kinematics reveals inter-individual differences in intertemporal decision dynamics. *Experimental Brain Research*, 233, 3597–3611. <https://doi.org/10.1007/s00221-015-4427-1>
- Cheng, J., & González-Vallejo, C. (2017). Action dynamics in intertemporal choice reveal different facets of decision process. *Journal of Behavioral Decision Making*, 30, 107–122. <https://doi.org/10.1002/bdm.1923>
- Dai, J., & Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General*, 143, 1489–1514. <https://doi.org/10.1037/a0035976>
- Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, 35, 15–28. <https://doi.org/10.3758/BF03195938>
- DeHart, W. B., & Odum, A. L. (2015). The effects of the framing of time on delay discounting. *Journal of the Experimental Analysis of Behavior*, 103, 10–21. <https://doi.org/10.1002/jeab.125>
- Dignath, D., Pfister, R., Eder, A. B., Kiesel, A., & Kunde, W. (2014). Something in the way she moves—Movement trajectories reveal dynamics of self-control. *Psychonomic Bulletin & Review*, 21, 809–816. <https://doi.org/10.3758/s13423-013-0517-x>
- Dshemuchadse, M., Grage, T., & Scherbaum, S. (2015). Action dynamics reveal two types of cognitive flexibility in a homonym relatedness judgment task. *Frontiers in Psychology*, 6, 1244. <https://doi.org/10.3389/fpsyg.2015.01244>
- Dshemuchadse, M., Scherbaum, S., & Goschke, T. (2013). How decisions emerge: Action dynamics in intertemporal decision making. *Journal of Experimental Psychology: General*, 142, 93–100. <https://doi.org/10.1037/a0028499>
- Erb, C. D. (2018). The developing mind in action: Measuring manual dynamics in childhood. *Journal of Cognition and Development*, 19, 233–247. <https://doi.org/10.1080/15248372.2018.1454449>
- Ericson, K. A., & Simon, H. A. (1984). Protocol analysis: Verbal reports as data. Cambridge: MIT Press.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Faulkenberry, T. J., Cruise, A., Lavro, D., & Shaki, S. (2016). Response trajectories capture the continuous dynamics of the size congruity effect. *Acta Psychologica*, 163, 114–123. <https://doi.org/10.1016/j.actpsy.2015.11.010>
- Faulkenberry, T. J., & Rey, A. E. (2014). Extending the reach of mousetracking in numerical cognition: A comment on Fischer and Hartmann (2014). *Frontiers in Psychology*, 5, 1436. <https://doi.org/10.1038/35006062>
- Fischer, M. H., & Hartmann, M. (2014). Pushing forward in embodied cognition: May we mouse the mathematical mind? *Frontiers in Psychology*, 5, 1315:1–4. <https://doi.org/10.3389/fpsyg.2014.01315>
- Franco-Watkins, A. M., Mattson, R. E., & Jackson, M. D. (2015). Now or later? Attentional processing and intertemporal choice. *Journal of Behavioral Decision Making*, 29, 206–217. <https://doi.org/10.1002/bdm.1895>
- Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40, 351–401.
- Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science*, 27, 315–323. <https://doi.org/10.1177/0963721417746793>
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42, 226–241. <https://doi.org/10.3758/BRM.42.1.226>
- Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods*, 45, 83–97. <https://doi.org/10.3758/s13428-012-0225-x>
- Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, 35, 59:15–28. <https://doi.org/10.3389/fpsyg.2011.00059>
- 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, 377–388. <https://doi.org/10.1007/s10339-015-0666-0>
- Grage, T., Schoemann, & Scherbaum, S. (2018). *Lost to translation: How design factors of the mouse-tracking procedure impact the inference from action to cognition*. Manuscript submitted for publication.
- Green, L., Myerson, J., & McFadden, E. (1997). Rate of temporal discounting decreases with amount of reward. *Memory & Cognition*, 25, 715–723. <https://doi.org/10.3758/BF03211314>
- Greiner, B. (2004). *The Online Recruitment System ORSEE 2.0—A guide for the Organization of Experiments in Economics* (Working Paper Series in Economics, pp. 1–67). Retrieved from <http://ideas.repec.org/p/kls/series/0010.html>
- Heit, E. (2015). Brain imaging, forward inference, and theories of reasoning. *Frontiers in Human Neuroscience*, 8, 1056:1–5. <https://doi.org/10.3389/fnhum.2014.01056>
- Henson, R. (2006). Forward inference using functional neuroimaging: Dissociations versus associations. *Trends in Cognitive Sciences*, 10, 64–69. <https://doi.org/10.1016/j.tics.2005.12.005>
- Incera, S., & McLennan, C. T. (2016). Mouse tracking reveals that bilinguals behave like experts. *Bilingualism: Language and Cognition*, 19, 610–620. <https://doi.org/10.1017/S1366728915000218>
- JASP Team. (2018). JASP [Computer software]. Retrieved from <https://jasp-stats.org/>
- Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods*, 49, 1652–1667. <https://doi.org/10.3758/s13428-017-0900-z>
- Kieslich, P. J., & Hilbig, B. E. (2014). Cognitive conflict in social dilemmas: An analysis of response dynamics. *Judgment and Decision Making*, 9, 510–522.

- Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2018). *Design factors in mouse-tracking: What makes a difference?* Manuscript submitted for publication.
- Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: A practical guide to implementation and analysis. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Eds.), *A Handbook of Process Tracing Methods*. New York, NY: Routledge.
- Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judgment and Decision Making*, 8, 527–539.
- Koop, G. J., & Johnson, J. G. (2011). Response dynamics: A new window on the decision process. *Judgment and Decision Making*, 6, 750–758.
- Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, 67, 151–185. <https://doi.org/10.1016/j.cogpsych.2013.09.001>
- Lempert, K. M., & Phelps, E. A. (2016). The malleability of intertemporal choice. *Trends in Cognitive Sciences*, 20, 64–74. <https://doi.org/10.1016/j.tics.2015.09.005>
- Lepora, N. F., & Pezzulo, G. (2015). Embodied choice: How action influences perceptual decision making. *PLoS Computational Biology*, 11, e1004110:1–22. <https://doi.org/10.1371/journal.pcbi.1004110>
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie Du Sert, N. ... Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1, 21:1–9. <https://doi.org/10.1038/s41562-016-0021>
- Myerson, J., Green, L., & Warusawitharana, M. (2001). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior*, 76, 235–243. <https://doi.org/10.1901/jeab.2001.76-235>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J. ... Yarkoni, T. (2015). Promoting an open research culture. *Nature*, 548, 1422–1425. <https://doi.org/10.1126/science.aab3847>
- O'Hara, D., Carey, R., Kervick, A., Crowley, D., & Dabrowski, M. (2016). Decisions in motion: Decision dynamics during intertemporal choice reflect subjective evaluation of delayed rewards. *Scientific Reports*, 6, 20740:1–17. <https://doi.org/10.1038/srep20740>
- O'Hara, 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, 2210. <https://doi.org/10.1038/srep02210>
- Oppenheimer, D. M., & Kelso, E. (2015). Information processing as a paradigm for decision making. *Annual Review of Psychology*, 66, 277–294. <https://doi.org/10.1146/annurev-psych-010814-015148>
- Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, 144, 190–206. <https://doi.org/10.1016/j.actpsy.2013.06.003>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10, 437–442. <https://doi.org/10.1163/156856897X00366>
- Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10, 59–63. <https://doi.org/10.1016/j.tics.2005.12.004>
- Quinton, J. C., Volpi, N. C., Barca, L., & Pezzulo, G. (2014). The cat is on the mat. Or is it a dog? Dynamic competition in perceptual decision making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44, 539–551. <https://doi.org/10.1109/TSMC.2013.2279664>
- Read, D., Frederick, S., Orsel, B., & Rahman, J. (2005). Four score and seven years from now: The date/delay effect in temporal discounting. *Management Science*, 51, 1326–1335. <https://doi.org/10.1287/mnsc.1050.0412>
- Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, 461, 263–266. <https://doi.org/10.1038/nature08275>
- van Rooij, M. M. J. W., Favela, L. H., Malone, M., & Richardson, M. J. (2013). Modeling the dynamics of risky choice. *Ecological Psychology*, 25, 293–303. <https://doi.org/10.1080/10407413.2013.810502>
- Russo, J. E., & Rosen, L. D. (1975). An eye fixation analysis of multialternative choice. *Memory & Cognition*, 3, 267–276. <https://doi.org/10.3758/BF03212910>
- Scherbaum, S., & Dshemuchadse, M. (2018). *Psychometrics of the continuous mind: Time continuous multiple regression as a method to exploit the dynamics of computer mouse movements*. Manuscript submitted for publication.
- Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: The temporal dynamics of action selection. *Cognition*, 115, 407–416. <https://doi.org/10.1016/j.cognition.2010.02.004>
- Scherbaum, S., Dshemuchadse, M., Leiberg, S., & Goschke, T. (2013). Harder than expected: Increased conflict in clearly disadvantageous delayed choices in a computer game. *PLoS ONE*, 8, e79310. <https://doi.org/10.1371/journal.pone.0079310>
- Scherbaum, S., Frisch, S., & Dshemuchadse, M. (2018a). A bird in the hand isn't good for long: Action dynamics reveal short-term choice impulses in intertemporal choices. *Experimental Psychology*, 65, 23–31.
- Scherbaum, S., Frisch, S., & Dshemuchadse, M. (2018b). Step by step: Harvesting the dynamics of delay discounting decisions. *Quarterly Journal of Experimental Psychology*, 71, 949–964. <https://doi.org/10.1080/17470218.2017.1307863>
- Scherbaum, S., Frisch, S., Leiberg, S., Lade, S. J., Goschke, T., & Dshemuchadse, M. (2016). Process dynamics in delay discounting decisions: An attractor dynamics approach. *Judgement and Decision Making*, 11, 472–495.
- Scherbaum, S., & Kieslich, P. J. (2018). Stuck at the starting line: How the starting procedure influences mouse-tracking data. *Behavior Research Methods*, 50, 2097–2110. <https://doi.org/10.3758/s13428-017-0977-4>
- Schoemann, M., Schulte-Mecklenbeck, M., Renkewitz, F., & Scherbaum, S. (2018). *Forward inference in risky choice: Mapping gaze and decision processes*. Manuscript submitted for publication.
- Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., & Willemsen, M. C. (2017). Process-tracing methods in decision making: On growing up in the 70s. *Current Directions in Psychological Science*, 26, 442–450. <https://doi.org/10.1177/0963721417708229>
- Schulte-Mecklenbeck, M., Kühberger, A., Gagl, B., & Hutzler, F. (2017). Inducing thought processes: Bringing process measures and cognitive processes closer together. *Journal of Behavioral Decision Making*, 30, 1001–1013. <https://doi.org/10.1002/bdm.2007>
- Spivey, M. J. (2007). *The continuity of mind*. Oxford: Oxford University Press.
- Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, 15, 207–211. <https://doi.org/10.1111/j.1467-8721.2006.00437.x>
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). From the cover: Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, 102, 10393–10398. <https://doi.org/10.1073/pnas.0503903102>
- Sullivan, N. J., Hutcherson, C. A., 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, 122–134. <https://doi.org/10.1177/0956797614559543>
- Szaszi, B., Palfi, B., Szollosi, A., Kieslich, P. J., & Aczel, B. (2018). Thinking dynamics and individual differences: Mouse-tracking analysis of the denominator neglect task. *Judgment and Decision Making*, 13, 23–32. Retrieved from <http://dx.doi.org/10.17818/jdm.17818.pdf>
- Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2018). *Measuring the (dis-)continuous mind: What movement trajectories reveal about cognition*. Manuscript in preparation.
- Yamamoto, N., Incera, S., & McLennan, C. T. (2016). A reverse stroop task with mouse tracking. *Frontiers in Psychology*, 7, 670:1–12. <https://doi.org/10.3389/fpsyg.2016.00670>