Design factors in mouse-tracking: What makes a difference?

Pascal J. Kieslich[1,2] · Martin Schoemann[3] · Tobias Grage[3] · Johanna Hepp[4] · Stefan Scherbaum[3]

Abstract

Investigating cognitive processes by analyzing mouse movements has become a popular method in many psychological disciplines. When creating mouse-tracking experiments, researchers face many design choices—for example, whether participants indicate responses by clicking a button or just by entering the button area. Hitherto, numerous different settings have been employed, but little is known about how these methodological differences affect mouse-tracking data. We systematically investigated the influences of three central design factors, using a classic mouse-tracking paradigm in which participants classified typical and atypical exemplars. In separate experiments, we manipulated the response indication, mouse sensitivity, and starting procedure. The core finding that mouse movements deviate more toward the nonchosen option for atypical exemplars was replicated in all conditions. However, the size of this effect varied. Specifically, it was larger when participants indicated responses via click and when they were instructed to initialize the movement early. Trajectory shapes also differed between setups. For example, a dynamic start led to mostly curved trajectories, responses via click led to a mix of straight and "change-of-mind" trajectories, and responses via touch led to mostly straight trajectories. Moreover, the distribution of curvature indices was classified as bimodal in some setups and as unimodal in others. Because trajectory curvature and shape are frequently used to make inferences about psychological theories, such as differentiating between dynamic and dual-system models, this study shows that the specific design must be carefully considered when drawing theoretical inferences. All methodological designs and analyses were implemented using open-source software and are available from https://osf.io/xdp7a/.

Keywords Mouse-tracking · Cognitive processes · Experimental design · Decision-making · Response dynamics

Over the past decade, mouse-tracking—the recording and analysis of computer mouse movements—has become an important addition to the toolbox of experimental psychologists. By recording mouse trajectories during psychological tasks, mouse-tracking has allowed researchers to investigate a range of cognitive processes that unfold in real time while people are making their decisions (Freeman, Dale, & Farmer, 2011; Spivey & Dale, 2006). As such, mouse-tracking has extended the window into cognition that classic reaction time analyses

and newer developments, such as eye-tracking or electroencephalography, have opened (Schulte-Mecklenbeck et al., 2017). Though it is a relatively new method, mouse-tracking has quickly spread across a broad range of psychological fields, as recent reviews have demonstrated (Freeman, 2018; Stillman, Shen, & Ferguson, 2018).

Yet, despite mouse-tracking's newfound glory and widespread application in psychology, no standard exists for the design of mouse-tracking studies (Scherbaum & Kieslich, 2018). As a result, methodological setups have varied considerably between mouse-tracking experiments, but almost nothing is known about the implications of such variation. At the same time, it is probable that the methodological setup impacts how cognitive processes are reflected in mouse trajectories. If so, the curvature and overall shape of cursor trajectories may vary for different setups. Variation in these measures would have far-reaching implications for past and future mouse-tracking experiments, since they form the basis for conclusions about psychological theories in these studies. The previous evidence potentially affected by methodological choices spans inferences about the influences of various psychological factors on decision conflict (e.g., semantic

✉ Pascal J. Kieslich
   kieslich@psychologie.uni-mannheim.de

[1] Department of Psychology, School of Social Sciences, University of Mannheim, L13, 17, 68161 Mannheim, Germany

[2] Mannheimer Zentrum für Europäische Sozialforschung (MZES), University of Mannheim, Mannheim, Germany

[3] Technische Universität Dresden, Dresden, Germany

[4] Department of Psychosomatic Medicine, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany

representations, stereotypes, self-control, or personality differences; see Stillman et al., 2018) and inferences about which theoretical model might best account for the data (e.g., dynamic vs. dual-system models; see Freeman, 2018). For this reason, there is a pressing need to understand how design factors in mouse-tracking affect trajectories, and ultimately inferences about psychological theories.

In an effort to provide insight into the consequences of different design factors in mouse-tracking, we herein report three experiments that assessed the impact of the most central design choices, and discuss how the consequences might influence theorizing in general. We first give an overview of previous mouse-tracking research and the varying methodological setups. Next, we present the three experiments and report the effects that different design factors have on mouse-tracking data. Finally, we discuss implications for the connection of mouse-tracking data and theorizing, and provide recommendations for future mouse-tracking studies.

Mouse-tracking: Basic paradigm and design factors

In typical mouse-tracking experiments, participants decide between two options represented as buttons on a computer screen while their cursor movements are continuously recorded (see Fig. 1 for the basic setup and an exemplary mouse cursor trajectory). These cursor movements are taken as an indicator of the relative activation of response options over the course of the decision-making process, assuming that the more an option is activated, the more the mouse trajectory deviates toward it (Freeman et al., 2011; Spivey & Dale,

2006). Thus, the degree of curvature is used as an indicator of the amount of activation of or attraction to this option.

The recording and analysis of mouse trajectories has offered two major opportunities for testing psychological theories (Freeman, 2018; Stillman et al., 2018): First, it provides fine-grained measures of the amount of *conflict* between response options, thus allowing researchers to test predictions about individual differences and contextual factors that influence the amount of conflict in a specific decision. Second, mouse-tracking allows for assessing the *temporal development* and resolution of this conflict over the course of the decision process, which makes it possible to test theories that make predictions about how decisions and judgments unfold over time. In this regard, a central usage of mouse-tracking has been to differentiate between dynamic and dual-system models (Freeman & Dale, 2013). Dynamic process models predict a continuous competition of the response options that gradually gets resolved over time and should be reflected in continuously curved trajectories in all trials. In contrast, dual-system models predict a mixture of trials with little conflict and trials in which, at first, one option is strongly activated and then a change of mind occurs; this should lead to a mix of straight trajectories and trajectories displaying abrupt shifts in the movement.

Mouse-tracking was first applied in the area of language processing little more than a decade ago (Dale & Duran, 2011; Dale, Kehoe, & Spivey, 2007; Spivey, Grosjean, & Knoblich, 2005) and has since spread to a broad range of psychological disciplines. To date, mouse-tracking has been used to study social cognition (e.g., Freeman & Ambady, 2009; Freeman, Ambady, Rule, & Johnson, 2008; Hehman, Carpinella, Johnson, Leitner, & Freeman, 2014; Johnson, Freeman, & Pauker, 2012), action control (e.g., Scherbaum,
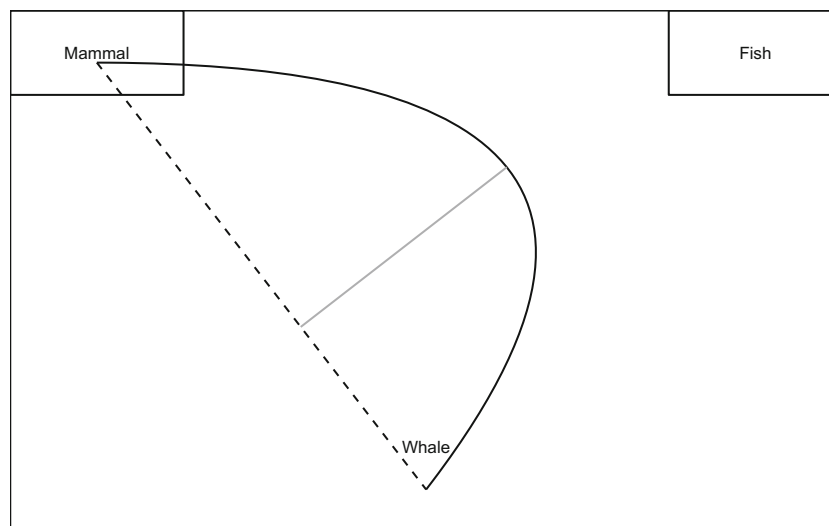
Fig. 1 Setup of the mouse-tracking experiment, including an exemplary cursor trajectory. The trial is initiated by clicking on a start button in the bottom center of the screen (not displayed), after which the name of the to-be-classified animal is presented. Participants indicate their classification decision by clicking on one of two response buttons. For the example trajectory, the maximum absolute deviation (MAD) is depicted (gray line) as the maximum perpendicular deviation of the trajectory from a straight line connecting the start and end points of the trajectory (dashed line)

Dshemuchadse, Fischer, & Goschke, 2010), numerical cognition (see the review by Faulkenberry, Witte, & Hartmann, 2018), political beliefs (e.g., Duran, Nicholson, & Dale, 2017), perception (e.g., Frisch, Dshemuchadse, Görner, Goschke, & Scherbaum, 2015; Huette & McMurray, 2010; Lepora & Pezzulo, 2015), memory (e.g., Koop & Criss, 2016; Papesh & Goldinger, 2012), value-based decision-making (e.g., Dshemuchadse, Scherbaum, & Goschke, 2013; Kieslich & Hilbig, 2014; Koop & Johnson, 2013), self-control (e.g., Stillman, Medvedev, & Ferguson, 2017; Sullivan, Hutcherson, Harris, & Rangel, 2015), and other disciplines. Two recent reviews have summarized the myriad ways in which mouse-tracking has helped advance psychological theory in some of the aforementioned areas (Freeman, 2018; Stillman et al., 2018).

To provide a few examples of mouse-tracking applications in different research areas, we have selected three exemplar studies, devoted to social categorization, preferential choice, and action control. The first example study used mouse-tracking to investigate social categorization (Freeman & Ambady, 2009), asking participants to select which of two adjectives fit with the gender-stereotype for a presented face. The study included two groups of faces: gender-typical faces (e.g., a face with uniquely female features) and gender-atypical faces (e.g., a female face that was to some degree morphed with a male face). For the atypical faces, mouse trajectories were more curved toward the opposite-gender stereotype. For instance, for an atypical female face (as compared to a typical female face), the mouse trajectory was more curved toward a stereotypically male adjective, such as "aggressive", before ultimately selecting the stereotypically female adjective "caring". The authors also analyzed the time course of the trajectories, as well as their shape (via bimodality analysis), finding a continuous and unimodal distribution of curvature values. From this, they concluded that a dynamic process model can best account for the data, assuming simultaneous coactivation of competing stereotypes that gradually gets resolved over time.

The second exemplar study used mouse-tracking to examine participants' preferential choices (Koop & Johnson, 2013, Exp. 1), asking them to indicate which of two pictures they preferred by clicking on the corresponding picture. The picture pairs were created with systematically varying differences in pleasantness as assessed by norming data, and participants' mouse trajectories reflected these differences: The curvature of trajectories systematically increased toward the nonchosen option, the smaller the pleasantness difference between the two pictures became. This study was used as a first step toward validating mouse-tracking in the area of preferential decision-making, to show that trajectory curvature can be used to measure differences in personal preference.

The third exemplary mouse-tracking application comes from the area of action control (Scherbaum et al., 2010): In a mouse-tracking version of the Simon task, participants had to choose the left or right option, depending on the direction of an arrow that was presented on the left versus the right side. Participants' mouse trajectories reflected the typical Simon effect, in that they were more curved toward the nonchosen option if the location of the arrow was incongruent with its direction. This effect was reduced if a trial was preceded by an incongruent trial, the so-called congruency sequence effect. In addition, mouse-tracking allowed for disentangling the temporal development of the different effects. Specifically, temporal analyses of the mouse movement direction revealed that the congruency sequence effect set in after the Simon effect—a finding that allowed for disentangling different theoretical accounts of the cognitive processes underlying action control.

In all three example experiments, participants had to click on a start button in the bottom center of the screen to start the trial (to align the starting position of the cursor across trials), but beyond this, the procedures differed substantially. In the experiment by Koop and Johnson (2013), the stimuli appeared immediately after the click on the start button, and participants could indicate their response by clicking on one of the two buttons. Participants did not receive any specific instructions about how to move the mouse. In the experiment by Freeman and Ambady (2009), participants were explicitly encouraged to initiate their movement early in the trial, and a warning message was displayed if the time for movement initiation exceeded a predefined threshold. In the experiment by Scherbaum et al. (2010), participants had to move the mouse upward at the beginning of the trial for the stimulus to be displayed, and they could indicate their response merely by moving the cursor onto the corresponding button (no click was required). Evidently, these three studies varied considerably with regard to their methodological setup, with three different starting procedures and two different response indication procedures—further hardware- and software-related factors not even considered (e.g., the cursor speed settings or computer screen resolution).

The methodological diversity in these exemplar studies illustrates that researchers face a number of design choices when creating mouse-tracking experiments, for which there are no empirically based recommendations. These choices include the examples mentioned above, such as the variations in starting procedure and type of response indication, but also choices pertaining to the screen layout as well as hard- and software settings. Given the relative novelty of the method, to date there are almost no empirical investigations of how design factors affect mouse-tracking data, but some of their potential implications have previously been discussed. Hehman, Stolier, and Freeman (2015) and Fischer and Hartmann (2014) provided some recommendations for the basic setup of mouse-tracking studies. Both suggested that researchers should implement measures that increase the likelihood of participants initializing their mouse movement early in the

trial, to ensure that cognitive processing takes place while participants are moving the mouse, and not beforehand. Fischer and Hartmann additionally recommended reducing the cursor speed to better capture cognitive effects in the trajectory measures. Importantly, Hehman et al. (2015) noted that "these approaches have not been empirically validated, and instead are derived from our previous experience" (p. 388).

In the, to date, only published empirical investigation of the effects of a specific design factor on mouse-tracking data, Scherbaum and Kieslich (2018) examined two different starting procedures. They compared data from the previously described experiment by Scherbaum et al. (2010) that had used a dynamic starting procedure (i.e., a procedure in which participants have to move the mouse upward for the stimulus to be displayed) to those from a new experiment that replicated the same study using a different starting procedure, in which the stimulus was presented after a short, fixed delay. They found that the cognitive effects on trajectory curvature were comparable for both starting procedures. However, the dynamic starting procedure led to stronger, more distinguishable effects in the temporal analyses of mouse movement direction. While this study provided a first indication that design factors may play an important role in mouse-tracking studies, it focused only on two variations of a single design factor (i.e., the starting procedure). Potential effects of other starting procedures and the broad range of further design factors still remain unexplored.

The goal of the present study was therefore to systematically investigate the influence of a set of design factors that commonly vary between mouse-tracking studies. For this purpose, we used different variations of a classic and simple mouse-tracking paradigm that we describe in the following section. In addition, we provide researchers with open-source implementations for all the methodological setups and analyses that we report. These can be run using free and cross-platform software (Kieslich & Henninger, 2017) and are available online at https://osf.io/xdp7a/. For the analyses of our mouse-tracking data, we focused on a set of traditional mouse-tracking analyses (Freeman & Ambady, 2010), as well as a number of recently proposed graphical and spatial analysis approaches (Wulff, Haslbeck, Kieslich, Henninger, & Schulte-Mecklenbeck, in press). We provide all raw data and analysis codes, so that researchers can both replicate our analyses and apply them to their own data.

Overview of the experiments

For the purpose of assessing the effects of methodological differences in mouse-tracking studies, we conducted three experiments and manipulated a central design factor in each experiment, while keeping the overall paradigm constant

across experiments. We chose a classic paradigm for semantic categorization that was first published by Dale et al. (2007) and marked one of the early mouse-tracking applications. In this experiment, participants classify exemplars as belonging to one of two categories. Specifically, they are presented with the name of an animal that is either typical for its response category (e.g., a lion for mammal) or atypical, in that it shares features with both the correct and the competing category (e.g., a whale, sharing features with both the correct category mammal and the incorrect category fish). The experiment by Dale and colleagues is in many ways representative of a typical mouse-tracking experiment: Participants repeatedly choose between two options, the stimuli are simple and relatively quick to process, and a central factor (typicality) is varied between trials with a directed hypothesis regarding its influence on mouse-tracking data. The central cognitive effect of interest in this experiment is what we will henceforth call the *typicality effect*. It denotes that mouse trajectories deviate more toward the nonchosen option for atypical than for typical exemplars (Dale et al., 2007). As stated above, one design factor was varied between participants in each of the three experiments, implementing the most common variations of this design factor. An overview of all manipulations is given in Table 1. The different design factors will be introduced in detail before we report each experiment.

For analyzing mouse trajectories, we focus on the most frequently used analysis in previous mouse-tracking studies, which is the analysis of trajectory curvature (Freeman, 2018; Stillman et al., 2018). This analysis aims to quantify the amount of response conflict that was present in a given trial. The idea is that the more a participant tends toward the nonchosen option in a trial, the more the mouse trajectory deviates toward it. To quantify curvature, different indices have been proposed that are highly correlated in practice (Stillman et al., 2018). We will use the maximum absolute deviation (MAD), as an easy-to-interpret and commonly used measure in mouse-tracking (Freeman & Ambady, 2010; Kieslich & Henninger, 2017; Koop & Johnson, 2011). The MAD is defined as the signed maximum deviation of the trajectory from a direct path (straight line, see Fig. 1) connecting the start and end positions of the trajectory (with maximum deviations above the direct line, i.e., in the direction of the nonchosen option, receiving a positive sign, deviations below, a negative sign). In line with the typicality effect, MAD values should be higher for atypical than for typical exemplars, and indeed this finding was observed in a recent replication of the experiment by Dale and colleagues (Kieslich & Henninger, 2017). For the purpose of the present study, we examined whether the typicality effect via MAD could be found in the different methodological setups and whether particular design choices influenced the occurrence and size of the typicality effect.

Table 1. Overview of the experiments, including basic design choices and their manipulations

Experiment	Response Indication	Mouse Sensitivity	Starting Procedure
1	Click vs. touch	Default	Static
2	Click	Default vs. slow	Static
3	Click	Slow	Static vs. rtmax vs. initmax vs. dynamic

Default = 50% cursor speed (of maximum speed) with acceleration enabled; Slow = 30% (Exp. 2)/40% (Exp. 3) cursor speed with acceleration disabled; rtmax = static start with restricted total response time; initmax = static start with restricted initiation time

Aside from using mouse-tracking to assess response conflict via curvature analyses, many researchers have also used mouse-tracking to distinguish between different classes of theoretical models (Freeman, 2018; Stillman et al., 2018). For instance, mouse-tracking has been used to determine whether dual-system or dynamic models are better in accordance with the data in a particular task (Freeman & Dale, 2013). Dual-system models, on the one hand, should produce a mixture of straight trajectories (in which both systems agree) and so-called change-of-mind trajectories, in which the initial response (by System I) favors one of the options (and as a consequence, the cursor approaches that option) that is later overridden (by System II), and the other option is chosen. The latter, change-of-mind response should produce large curvature values, whereas the former should result in small curvature values. Thus, across all trials, a bimodal distribution of small and large curvature values would be expected. Dynamic models, on the other hand, expect that both response options will be simultaneously activated to varying degrees until one of the options is chosen. In this case, a unimodal distribution of continuously varying curvature indices is expected. Therefore, researchers have conducted bimodality analyses of curvature indices at the trial level, expecting to find a bimodal distribution if a dual-system model accounted for the data, and a unimodal distribution for dynamic models (Hehman et al., 2015). On the basis of a bimodality analysis, the original study by Dale et al. (2007) provided support for dynamic models in their mouse-tracking data, as trajectory curvatures were classified as being unimodally distributed.

For the present experiments, we examined whether the distribution of curvature values is affected by the methodological setup of the mouse-tracking study. If methodological choices were to affect bimodality analyses, this would pose a general challenge to mouse-tracking studies, because then the theoretical implications of mouse trajectories would always have to be interpreted relative to the specific design that was employed. That is, if a bimodal distribution were observed in one setup, whereas a unimodal distribution was observed in another, this could imply that the setup directly influences the underlying cognitive process. However, an alternative (and in our view more plausible) interpretation would be that the methodological setup changes how the (unaffected) cognitive process is mapped onto the mouse movements. For example, one factor might ensure a continuous mapping of the complete process, while another factor might only capture parts of the process. In this way, it might miss early stages of the decision process, or lead to a discontinuous mapping in which the mouse movements are only periodically updated.

Instead of performing bimodality analyses, which are basically designed to answer the question of whether there are one or two types of trajectories, it has recently been argued that more fine-grained analyses are needed that can allow for inferences about a variety of different trajectory shapes (Wulff et al., in press; Wulff, Haslbeck, & Schulte-Mecklenbeck, 2019). In this regard, one proposed procedure is the mapping of trajectories onto trajectory prototypes. On the basis of their meta-analysis of mouse- and hand-tracking studies, Wulff et al. (2019) suggested that a set of a few prototypical movement trajectories may account for the majority of trajectories in many mouse-tracking studies. To examine whether the methodological setup promotes the occurrence of different trajectory types, we will supplement our analyses using this recently proposed prototype-mapping method (details on the method are given in the Results section of Exp. 1).

Experiment 1

In the first experiment, we examined the effect that the response indication procedure has on mouse-tracking data. For this analysis, we experimentally varied whether participants had to click on a response button to indicate their response (*click* condition) or whether they could simply move the mouse cursor into the area of the response button, with no click required (*touch* condition). Both the click procedure (e.g., Dale et al., 2007; Freeman et al., 2008; Koop & Johnson, 2013; Spivey et al., 2005) and the touch procedure (e.g., Frisch et al., 2015; Huette & McMurray, 2010; Scherbaum et al., 2010) are commonly employed in the literature. Despite substantial variation in the response indication procedures in previous studies, how the type of response indication affects mouse-tracking data remains an open question.

As a direct methodological consequence of the procedure, the click condition gives participants the opportunity to move the cursor onto a response button, hover there, and then either click on it or decide to move all the way to the other option. Thus, the click condition allows participants to produce extreme mouse trajectories with switches from one option to the other. The occurrence of these movements has recently been demonstrated in a number of empirical studies, and they have served as indicators of changes of mind (Freeman, 2014; Szaszi, Palfi, Szollosi, Kieslich, & Aczel, 2018; Wulff et al., 2019). In the touch condition, changes of mind could theoretically be captured if participants hovered below rather than on top of a response button before switching to the other option, but overall, the touch condition renders the occurrence of these extreme movement types much less likely. As a consequence, larger curvature indices would be expected for the click than for the touch condition, and particularly so for trials in which greater response conflict is expected. This, in turn, would lead to larger effects of the typicality manipulation in the click condition. However, if mouse trajectories were more in line with the assumption of continuously curved mouse trajectories, the response indication procedure should be less relevant for discrete mouse-tracking measures such as curvature indices. In this case, the touch condition might even be better at capturing cognitive effects in mouse movements, as it allows participants to indicate their response more smoothly by removing the additional motor process of clicking.

Method

Procedure and materials The experiment was conducted at the University of Mannheim, Germany. After providing written informed consent and answering demographic questions, participants first worked on an unrelated experiment that was followed by the experiment currently under investigation. Participants received partial course credit for completing the experiment.

The basic setup and procedures followed those of Experiment 1 from Dale et al. (2007). In each trial, participants were asked to classify an animal (presented as a written word, e.g., "whale") as belonging to one of two classes (e.g., "mammal" vs. "fish"). The stimulus material included the same 13 typical and six atypical animals and their corresponding response categories that were used by Dale and colleagues in Experiment 1 (all materials were translated into German).

At the beginning of the experiment, participants were randomly assigned to one of two experimental conditions (response indication via *click* vs. *touch*). Participants received a short set of instructions that explained the task to them, including information about the response indication procedure. Afterward, participants worked on three practice trials, followed by another set of short instructions summarizing the task. Then, participants classified the 19 actual stimuli, which were

presented in random order. At the end of the experiment, participants' handedness was assessed via the Edinburgh Handedness Inventory (EHI; Oldfield, 1971, as implemented by Kieslich & Henninger, 2017).

During each trial, a blank screen was first presented for 1,000 ms, followed by the presentation of the two response categories for 2,000 ms in the top-left and top-right corners of the screen (the order of the categories was randomized at the trial level). Next, a start button appeared in the bottom center of the screen, which participants had to click for the animal stimulus to be immediately presented (see Fig. 1 for the layout of the decision screen). After the click on the start button, the mouse cursor was reset to the exact center of the start button, the stimulus was presented 85 pixels (px) above it, and recording of the mouse movements commenced. Depending on the experimental condition, participants could then indicate their response by clicking on the corresponding response button or by touching it (in this case, the response was immediately recorded as soon as one of the button areas was entered with the cursor).

The experiment was created in OpenSesame (Mathôt, Schreij, & Theeuwes, 2012). Mouse cursor movements were recorded every 10 ms using the mousetrap plugin (Kieslich & Henninger, 2017). The experiment was conducted full-screen at a resolution of $1,680 \times 1,050$ px on laboratory computers running Windows 7. The mouse sensitivity settings were left at the system defaults (cursor speed at 50% of maximum speed, with acceleration enabled).

Participants To determine the desired sample size, we conducted a power analysis using G*Power 3.1.9 (Faul, Erdfelder, Buchner, & Lang, 2009). Across all experiments, we aimed to ensure that the power to detect a typicality effect of medium size ($d_z = 0.5$) was at least .95 (with $\alpha = .05$, two-tailed) within each experimental condition. This resulted in a desired sample size of 54 participants per condition. We therefore recruited a total of 108 participants to complete the experiment (85 female, 23 male; between 18 and 38 years of age, $M = 22.0$ years, $SD = 3.7$). The majority (81 participants) indicated a preference for the right hand (EHI score > 60), and another six participants indicated a preference for the left hand (EHI score < – 60); the remaining 21 participants indicated no strong preference.

Results

We focused on a set of typical analyses that are commonly performed in mouse-tracking studies. As such, we compared trajectory curvatures and trajectory shapes, using both traditional analyses at the aggregate level (Freeman & Ambady, 2010) and newly proposed analysis procedures at the trial level (Wulff et al., in press). Analyses were performed in R (R Core Team, 2018) using the mousetrap R package
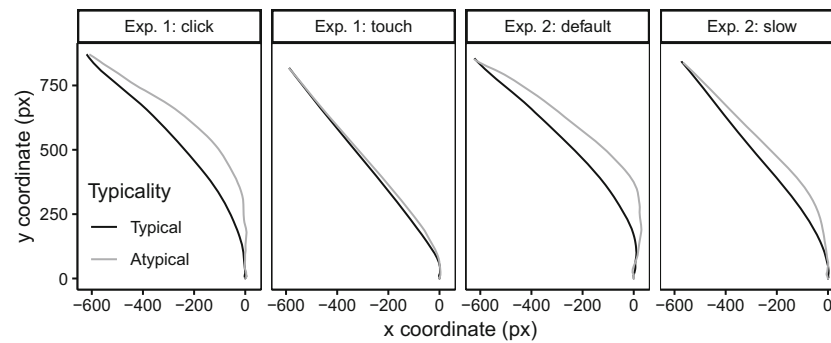
Fig. 2 Aggregate mouse trajectories for Experiments 1 and 2. All individual trajectories were flipped to the left, time-normalized, and aggregated separately per typicality and experimental condition

(Kieslich & Henninger, 2017; Kieslich, Henninger, Wulff, Haslbeck, & Schulte-Mecklenbeck, in press). The raw data, analysis scripts, and results (including the supplementary analyses) for this and all following experiments are openly available from https://osf.io/xdp7a/.

Correctness Before the mouse-tracking analyses, we compared the percentages of correctly answered trials between the two design conditions. The numbers of correctly answered trials did not differ significantly between experimental conditions (93.5% correct answers in the click condition, 93.1% in the touch condition), $\chi^2(1) = 0.16$, $p = .693$.[1] Following Dale et al. (2007), only correctly answered trials were included in the following analyses.

Aggregate trajectory curvature Next we performed a set of analyses focusing on the aggregate trajectory curvature. For this, we flipped all trajectories that ended on the right response option to the left. To visually inspect the shape of the aggregate trajectories, we followed the typical mouse-tracking analysis procedures; that is, we performed time-normalization so that each trajectory would be represented by the same number of temporally equidistant points (101, following Spivey et al., 2005). Then we aggregated the trajectories per typicality condition, first within and then across participants, and separately for the click and touch conditions. The resulting aggregate trajectories are displayed in Fig. 2. As expected, the aggregate trajectories deviated more toward the nonchosen option for atypical than for typical exemplars in both experimental conditions. However, this difference was considerably larger in the click condition.

To statistically test these differences in aggregate trajectory curvature, we computed the MAD for each trajectory (see Fig. 1; Freeman & Ambady, 2010; Kieslich & Henninger, 2017; Koop & Johnson, 2011). Following common procedures in mouse-tracking studies, the MAD values were aggregated per

typicality condition separately for each participant. The mean MAD values for atypical and typical exemplars are reported in Table 2, separately for all experimental conditions and experiments.

A repeated measures analysis of variance (ANOVA) using the aggregated MAD values per participant with the within-subjects factor typicality (atypical vs. typical) and the between-subjects factor response indication procedure (click vs. touch) revealed a significant main effect of typicality, $F(1, 106) = 25.96$, $p < .001$, $\eta_p^2 = .20$, 90% CI [0.09, 0.30]. In both the click and touch conditions, MAD values were significantly higher for atypical than for typical exemplars (Table 2). In addition, we observed a significant main effect of the response indication procedure, $F(1, 106) = 46.88$, $p < .001$, $\eta_p^2 = .31$, 90% CI [0.19, 0.41], with higher MAD values in the click condition. Finally, there was a significant interaction of typicality and response indication procedure, $F(1, 106) = 12.30$, $p < .001$, $\eta_p^2 = .10$, 90% CI [0.03, 0.20], with a larger typicality effect in the click ($d_z = 0.61$) than in the touch ($d_z = 0.36$) condition (Fig. 3).

Distribution of trajectory shapes To examine the influence of the response indication procedure on trajectory shapes, we first analyzed the bimodality of the distribution of the MAD values at the trial level, following the typical analysis procedures reported in previous mouse-tracking studies (Freeman & Ambady, 2009, 2010; Freeman & Dale, 2013; Spivey et al., 2005). That is, we standardized the MAD values per participant and then computed the bimodality coefficient separately for atypical and typical trials in each experimental condition. As can be seen in Table 2, both bimodality coefficients in the click condition were larger than .555, which—on the basis of simulation studies—is used as a cutoff for assuming a bimodal distribution (see Freeman & Ambady, 2010; Freeman & Dale, 2013). However, in the touch condition, both bimodality coefficients were smaller than .555, which is taken as evidence for a unimodal distribution.

Instead of performing bimodality analyses that are designed to answer the question of whether there are one or two different types of trajectories, it has recently been argued that mouse-tracking researchers should perform more fine-

---

[1] This result was replicated in a generalized linear mixed model at the trial level, using a binomial link function and including a random intercept per participant (see the complete analyses online).

Table 2. Descriptive statistics for maximum absolute deviation (MAD) values (in pixels) per typicality condition, and paired t test results (for a comparison of the atypical and typical conditions)

Experiment	Condition	N	Typical			Atypical			t Test		
			M	SD	BC	M	SD	BC	t	p	d_z
1	Click	53	142.7	111.5	.632	287.5	237.0	.641	4.43	< .001	0.61
1	Touch	55	52.2	78.3	.442	79.0	91.6	.500	2.69	.009	0.36
2	Default	59	157.8	158.8	.558	283.4	225.7	.576	5.49	< .001	0.71
2	Slow	59	73.4	113.1	.573	150.9	141.2	.593	4.52	< .001	0.59
3	Static	59	185.2	134.4	.520	269.7	172.7	.548	4.18	< .001	0.54
3	Rtmax	60	189.8	150.8	.536	301.5	197.8	.501	4.32	< .001	0.56
3	Initmax	66	304.8	140.7	.510	470.9	203.2	.473	7.39	< .001	0.91
3	Dynamic	60	297.0	111.6	.560	364.1	154.0	.508	3.95	< .001	0.51

MAD values were first aggregated per participant and typicality condition. BC = bimodality coefficient, based on the per-participant standardized MAD values

grained analyses to make inferences about the presence or absence of different types of trajectory shapes (Wulff et al., in press; Wulff et al., 2019). One such analysis is a graphical approach that plots a (smoothed) heat map of all trajectories separately for each experimental condition (Kieslich et al., in press). The resulting plots for each experimental condition are displayed in Fig. 4. For the click condition, the plot indeed suggests a mix of primarily straight trajectories and a number of triangular trajectories that first move to the nonchosen option and then horizontally head to the chosen option. In contrast, the latter type of trajectories seems to be almost absent in the touch condition, which consists of mostly straight and slightly curved trajectories.

To quantify and statistically test for differences in the frequency of trajectory types between conditions, a recently proposed procedure has been mapping of the trajectories onto trajectory prototypes (Wulff et al., in press). The prototypes used in the present study are depicted in Fig. 5. They are based on the prototype trajectories proposed in the meta-analysis by Wulff et al. (2019). They include *straight* trajectories that move directly from the start button to the chosen option, *curved* trajectories, continuous change-of-mind (*cCoM*) trajectories that exhibit a curved attraction toward the nonchosen option, discrete change-of-mind (*dCoM*) trajectories that first move straight to the nonchosen option and from there move horizontally to the chosen option, and double change-of-mind (*dCoM2*) trajectories that first move straight to the chosen option and then horizontally switch back and forth between the nonchosen and chosen options.[2]

To assign each trajectory to a prototype, the following analysis steps are employed (following Wulff et al., in press): First, trajectories are spatially normalized so that each trajectory is described by 20 points, ensuring that the spatial distance between adjacent points remains constant across the trajectory (in contrast to time-normalization, it is desirable to use fewer points to put an emphasis on the main shape of the trajectory). Then, the trajectory prototypes are spatially transformed so that their start and end points match the mean start and end points of the trajectories (separately per experimental condition). Next, the spatial distance between each prototype and trajectory is computed (using the Euclidian distance) and each trajectory is assigned to the prototype with the smallest distance.

The results of the prototype assignments and the relative frequency of each prototype classification are displayed in
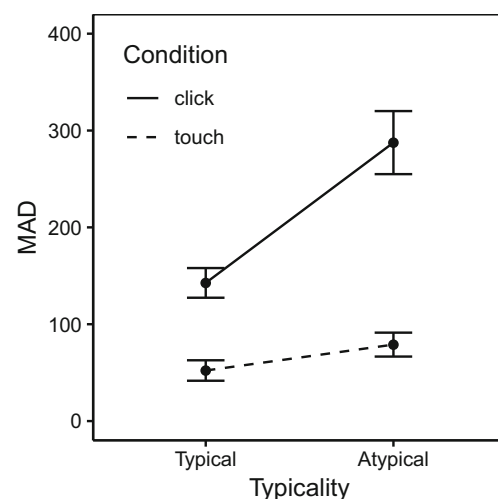
Fig. 3 Mean of maximum absolute deviation values (MAD, in pixels) for Experiment 1, shown separately per typicality and experimental condition. Error bars indicate one *SEM*

---

[2] Other types of trajectories are possible, of course. However, we argue, on the basis of the plots of individual trajectories per assigned prototype (Figs. 6 and 12), that this set of prototypes seems to describe the vast majority of trajectories in the present experiments well. See footnote 14 for a reanalysis of Experiment 3 in which a subset of trajectories seemed to be accounted for better by additional prototypes.
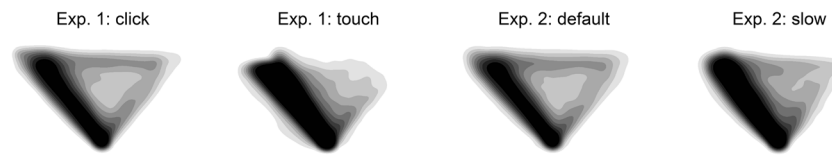
Fig. 4 Smoothed heat maps of the individual trajectories in Experiments 1 and 2, shown separately per experimental condition. Darker colors indicate higher density

Fig. 6. The majority of individual trajectories seem to be explained well by the present set of prototypes. The relative frequencies of prototypes differed significantly between the experimental conditions, $\chi^2(4) = 83.79$, $p < .001$. Whereas most trajectories were classified as straight in the touch condition, there were relatively fewer straight classifications in the click condition. In addition, a considerably greater percentage of trajectories were classified as discrete changes of mind (both dCoM and dCoM2) in the click condition.

To test whether the different trajectory types explained the larger typicality effects on curvature in the click condition, we performed an ordinal mixed regression at the trial level. Assuming that more extreme deviations of a prototype in the direction of the nonchosen option indicated greater amounts of response conflict, we treated the assigned prototype as an ordinal variable (straight < curved < cCoM < dCoM < dCoM2). We included a random intercept per participant, as well as the effect-coded predictors typicality (atypical = .5, typical = − .5), experimental condition (click = .5, touch = − .5), and their interaction. Atypical trials led to a significantly higher probability of more extreme trajectories ($z = 4.23$, $p < .001$), as did the click condition ($z = 5.70$, $p < .001$). Especially in the click condition, atypical trials also led to more extreme trajectories, as indicated by a significant interaction ($z = 2.39$, $p = .017$). The relative frequencies of prototype classifications per typicality and experimental condition are displayed in Table 3.

Discussion

In the first experiment, we examined the influence of the response indication procedure on mouse-tracking data. When participants indicated their response by clicking on the corresponding option, the typicality effect was significantly larger than when they could indicate their response by simply moving the cursor onto the button. This larger effect was related to more extreme trajectory movements at the trial level—specifically, more so-called discrete change-of-mind trajectories, which first move straight to the nonchosen option before heading horizontally to the chosen option. In the touch condition, on the contrary, the majority of trials was either straight or curved. This was also reflected in the bimodality coefficients, which indicated evidence for bimodality in the click condition and for unimodality in the touch condition.

Importantly, previous mouse-tracking studies have used the shape of trajectories (assessed often via bimodality analyses) to draw inferences about whether single or dual decision processes are at work in a given decision situation (see the reviews by Freeman, 2018; Stillman et al., 2018). This experiment demonstrated that using exactly the same task and simply changing a theoretically peripheral design aspect (the response indication procedure) can lead to either a bimodal distribution of curvature values, which could be interpreted as evidence for a dual-system model, or a unimodal distribution, which could be interpreted as evidence for a dynamic model. Although it is possible that changing a design aspect would affect the underlying decision process, we deem it more plausible that the decision process remained unaffected by peripheral changes in the methodological setup, and instead the design factor influenced the *mapping* of the decision process onto the mouse movement. However, this assumption cannot be tested directly (since the cursor movements always reflect both cognitive and motor components); hence, a change in the cognitive process cannot be ruled out solely on the basis of mouse-tracking data. One possible indicator for changes in the cognitive process could be changes at the choice level—for instance, if participants were to make more errors in one of the design conditions. In the present experiment, correctness did not differ significantly between the conditions, which might serve as a first, tentative indicator that the decision process was unaffected.
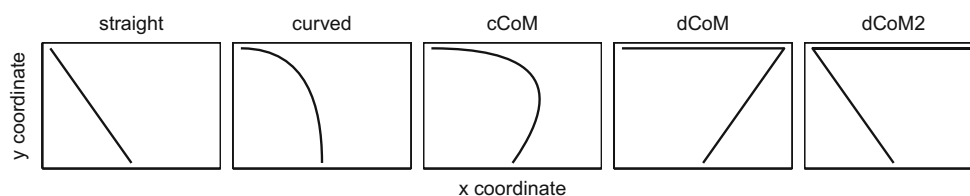
Fig. 5 Set of prototype trajectories used in the present analyses (cCoM = continuous change of mind, dCoM = discrete change of mind, dCoM2 = double change of mind)
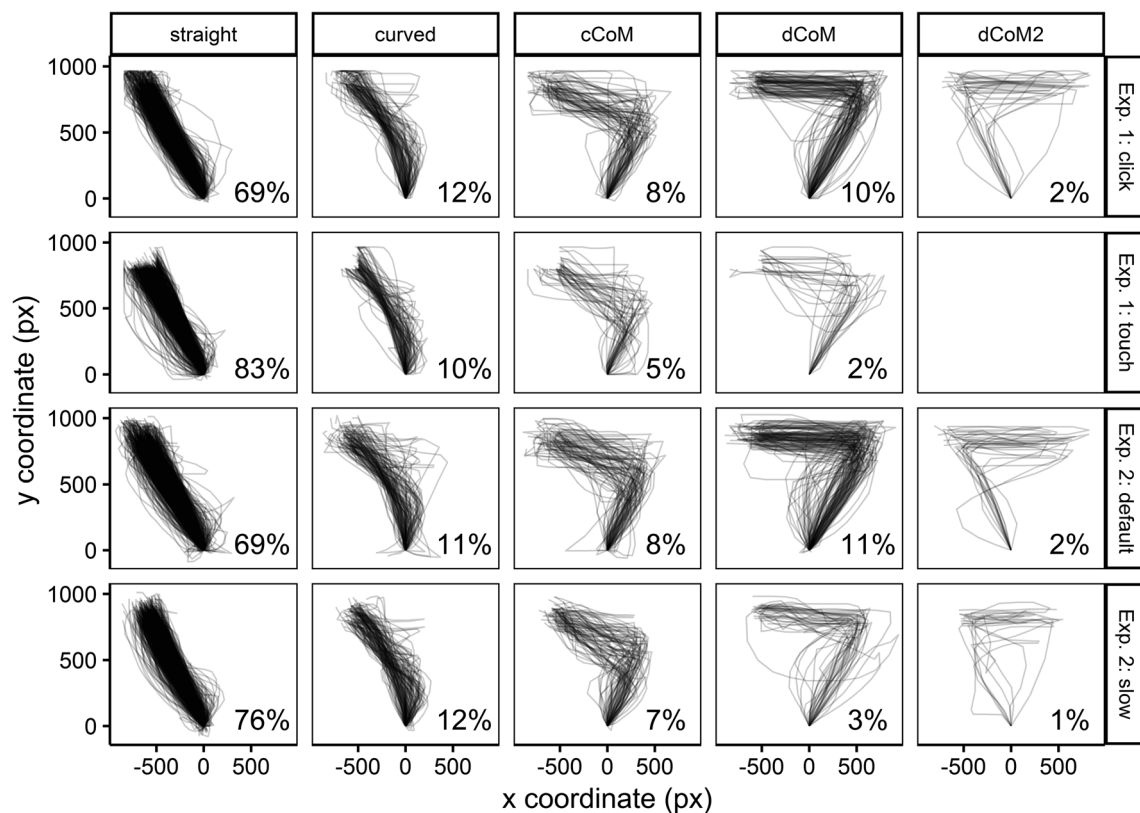
Fig. 6 Individual trajectories per assigned prototype, shown separately for the different experimental conditions in Experiments 1 and 2. For each prototype, the relative frequency of classifications per experimental condition is displayed. Due to rounding, percentages may not sum to exactly 100% for some conditions

In sum, the findings from Experiment 1 strongly suggest that aspects of the study design have to be carefully considered when interpreting mouse-tracking data. This should be particularly important with regard to potential effects of design factors that frequently vary between mouse-tracking studies. Extending the above-presented findings on the effects of the type of response indication, the following experiments are devoted to two further central design factors: the mouse sensitivity settings and starting procedure.

Experiment 2

In the second experiment, we focused on the design factor mouse sensitivity. This setting includes both the cursor speed and acceleration, which have varied considerably in previous studies, with some studies leaving the settings at the system defaults (which under Windows 7/8 is medium speed with acceleration enabled; e.g., Kieslich & Hilbig, 2014; Szaszi et al., 2018), and other studies deliberately reducing the cursor speed and disabling cursor acceleration (e.g., Dshemuchadse et al., 2013; Frisch et al., 2015; Scherbaum et al., 2010). For this reason, we compared these two commonly used setups, a *default* condition (medium speed, acceleration enabled) and a

slow condition (reduced speed, acceleration disabled). One challenging aspect of this design factor is that mouse sensitivity settings have rarely been reported explicitly in previous studies (Fischer & Hartmann, 2014).[3]

With regard to the mouse sensitivity settings, Fischer and Hartmann (2014) suggested that reducing the cursor speed and turning off acceleration is preferable for capturing cognitive effects in mouse trajectories. They argued that these settings ensure a linear relationship between hand and cursor movement, such that participants move the hand smoothly across a greater distance. In contrast, under the default settings, small movements of the wrist might already be enough to move the cursor to indicate a response, due to enabled acceleration. Although we, in principle, agree with these recommendations, the actual empirical consequences of different mouse sensitivity settings remain unknown. It might, indeed, be the case that, as Fischer and Hartmann suggest, a slow condition is better for capturing the cognitive effects—which would lead, for example, to a larger typicality effect—but this has never been

---

[3] The authors have to admit that this is also the case for some of their own studies—for example, Kieslich and Hilbig (2014), which used medium speed (50% of maximum speed) with acceleration enabled, and Scherbaum et al. (2010), which used a reduced speed (25%) with acceleration disabled.

Table 3. Percentages of trajectories per assigned prototype in each experimental condition, shown separately for typical/atypical exemplars

Experiment	Condition	Assigned Prototype				
		Straight	Curved	cCoM	dCoM	dCoM2
1	Click	73/60	12/11	7/10	7/16	1/4
1	Touch	84/81	10/12	5/4	2/3	0/0
2	Default	72/61	12/10	6/10	9/16	1/3
2	Slow	80/67	11/15	5/12	3/5	0/2
3	Static	62/49	19/20	12/18	5/10	2/4
3	Rtmax	64/50	15/15	13/18	7/14	1/3
3	Initmax	39/24	33/27	14/17	12/26	2/6
3	Dynamic	28/26	53/45	13/15	6/10	1/4

Due to rounding, percentages may not sum to exactly 100% for some conditions. cCoM = continuous change of mind, dCoM = discrete change of mind, dCoM2 = double change of mind

demonstrated. Alternatively, it could also be that the default condition leads to larger average effects by exaggerating small hand movements—although Freeman and Ambady (2010) noted that extremely high speeds might lead to problems, in that the cursor movements might become ballistic and jerky. By comparing a default and a slow cursor setting in this study, we hoped to provide a first empirical basis for researchers to make an informed decision about which mouse sensitivity setting is most suitable to their research question.

Method

Procedure and materials The general experimental procedure and all materials were identical to those in Experiment 1, and the study was again conducted at the University of Mannheim, Germany. The setup of the default condition in Experiment 2 was identical to the click condition of Experiment 1.[4] For the slow condition, the only change was that cursor acceleration was disabled and the cursor speed was reduced from 50% to 30% (of the maximum speed). Because the mouse sensitivity settings cannot be changed from within OpenSesame, we used a simple program to efficiently change the mouse sensitivity settings in Windows 7, the Mouse Acceleration Toggler.[5]

Participants After providing written informed consent, participants were randomly assigned to the default or the slow

condition and then completed the experiment (which was followed by another experiment). At the end of the study, participants provided demographic information and completed the EHI. They received partial course credit for their participation. On the basis of the power analysis reported in Experiment 1, we intended to ensure a minimum number of 54 participants per experimental condition. A total of 118 participants completed the experiment (88 female, 30 male; between 18 and 35 years of age, $M = 22.7$, $SD = 3.3$). The majority (91 participants) indicated a preference for the right hand, whereas seven participants indicated a preference for the left hand; the remaining 20 participants indicated no strong preference.

Results

The analyses followed those of Experiment 1. In addition, we conducted a manipulation check to examine whether the cursor sensitivity settings affected cursor speed and acceleration.

Correctness The numbers of correctly answered trials did not differ significantly between the experimental conditions (93.5% correct answers in the default condition, 94.5% in the slow condition), $\chi^2(1) = 0.95$, $p = .329$.[6] Again, only correctly answered trials were included in the analyses.

Manipulation check To determine whether participants actually moved the cursor faster in the default than in the slow condition, we computed the maximum velocity (in px/ms) and acceleration (in px/ms^2) for every trial. We then averaged the values per participant and compared them between conditions. As expected, the maximum velocity was considerably larger in the default condition ($M = 10.0$, $SD = 2.2$) than in the slow condition ($M = 4.3$, $SD = 1.1$), $t(116) = 18.09$, $p < .001$, $d = 3.33$, 95% CI [2.77, 3.89]. Similarly, the maximum acceleration was also larger in the default ($M = 0.54$, $SD = 0.12$) than in the slow condition ($M = 0.22$, $SD = 0.06$), $t(116) = 18.67$, $p < .001$, $d = 3.44$, 95% CI [2.86, 4.00].

Aggregate trajectory curvature To get a first impression of the effect of the cursor sensitivity manipulation on trajectory curvature, we inspected the aggregate time-normalized trajectories (Fig. 2). In both experimental conditions, the aggregate trajectories deviated more toward the nonchosen option for atypical than for typical exemplars. The size of this difference seemed to be slightly larger in the default than in the slow condition.

A repeated measures ANOVA using the aggregated MAD values per participant, with the within-subjects factor

---

[4] One minor change was that the cursor was not reset to the exact center of the start button ($x = 0.0$ px, $y = 85.0$ px) after participants clicked on it. However, empirically, the average start position was close to this center (x: $M = -0.2$ px, $SD = 45.6$ px; y: $M = 103.6$ px, $SD = 32.5$ px), and the start positions of all trajectories were aligned statistically during preprocessing.
[5] The program can be obtained for free from http://skwire.dcmembers.com/fp/?page=mat. The settings for the default condition were "accel = on speed = 10", and the settings for the slow condition were "accel = off speed = 6".

[6] This result was replicated in a generalized linear mixed model at the trial level, using a binomial link function and including a random intercept per participant (see the complete analyses online).
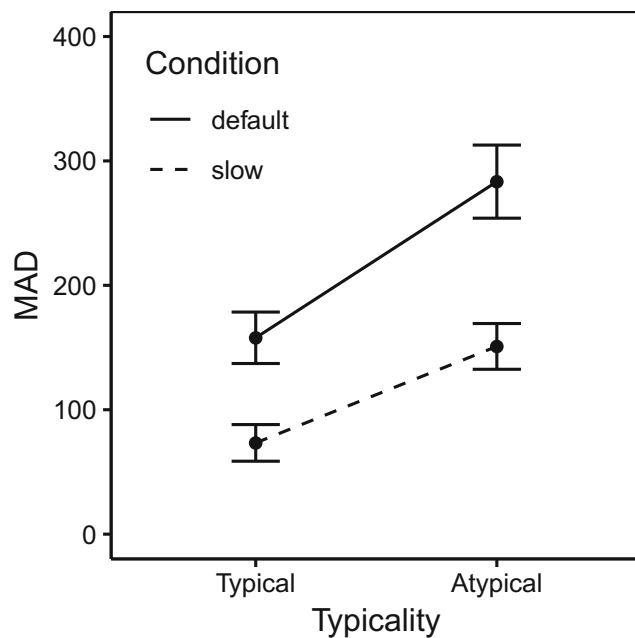
Fig. 7 Mean of maximum absolute deviation values (MAD, in pixels) for Experiment 2, shown separately per typicality and experimental condition. Error bars indicate one *SEM*

typicality (atypical vs. typical) and the between-subjects factor mouse sensitivity settings (default vs. slow), revealed a significant main effect of typicality, $F(1, 116) = 50.49$, $p < .001$, $\eta_p^2 = .30$, 90% CI [0.19, 0.40], with higher MAD values for atypical than for typical exemplars. The effect of typicality was significant in both conditions (Table 2). In addition, we found a significant main effect of the mouse sensitivity settings, $F(1, 116) = 16.37$, $p < .001$, $\eta_p^2 = .12$, 90% CI [0.04, 0.22], with higher MAD values in the default condition. There was, however, no significant interaction of typicality and cursor sensitivity, $F(1, 116) = 2.82$, $p = .096$, $\eta_p^2 = .02$, 90% CI [0.00, 0.09], although the typicality effect was descriptively slightly larger in the default ($d_z = 0.71$) than in the slow ($d_z = 0.59$) condition (Fig. 7).

Distribution of trajectory shapes As a first analysis of trajectory shapes, we again computed the bimodality coefficients for the per-participant standardized MAD values, separately for each typicality and experimental condition. In both the default and the slow condition, the coefficients were larger than .555, indicating a bimodal distribution (Table 2). The smoothed heat maps (Fig. 4) indicated that in both conditions there were a considerable number of straight trajectories, but also a number of change-of-mind trajectories in which the cursor was moved all the way to the nonchosen option before moving to the chosen option. However, the latter type of trajectories seemed to occur less frequently in the slow condition.

To quantify and statistically test for differences in the frequency of trajectory types between conditions, we mapped trajectories on the set of prototypes used in Experiment 1.

The majority of the individual trajectories was again explained well by the prototypes (Fig. 6). The relative frequencies of prototypes differed significantly between experimental conditions, $\chi^2(4) = 49.66$, $p < .001$. The main difference between conditions was that there were relatively more straight trajectories in the slow condition, and more dCoM trajectories in the default condition.

We predicted the trajectory type in an ordinal mixed regression, including a random intercept per participant and the predictors typicality (atypical = .5, typical = − .5), experimental condition (default = .5, slow = − .5), and their interaction. Atypical trials led to a significantly higher probability of more extreme trajectories ($z = 7.13$, $p < .001$), as did the default condition ($z = 2.80$, $p = .005$). The interaction between typicality and condition was not significant ($z = − 0.24$, $p = .809$). The relative frequencies of prototype classifications per typicality condition are provided in Table 3.

Stability of effects across experiments The default condition in Experiment 2 was virtually identical with the click condition in Experiment 1. Therefore, we performed a set of analyses comparing these two conditions in order to replicate the typicality effect and examine its stability.

First, we performed a repeated measures ANOVA on MAD values that were averaged per participant. We included the within-subjects factor typicality and the between-subjects factor experiment. As was theoretically expected, a significant main effect of typicality emerged, $F(1, 110) = 47.34$, $p < .001$, $\eta_p^2 = .30$, 90% CI [0.19, 0.40], with larger values in atypical trials (Table 2). With regard to the differences between experiments, we found neither a main effect of experiment, $F(1, 110) = 0.03$, $p = .856$, $\eta_p^2 = .00$, 90% CI [0.00, 0.02], nor an interaction between experiment and typicality, $F(1, 110) = 0.24$, $p = .624$, $\eta_p^2 = .00$, 90% CI [0.00, 0.04].

The relative frequencies of the classified prototypes did not differ significantly between experiments, $\chi^2(4) = 1.01$, $p = .908$. In an ordinal mixed regression predicting the assigned prototype, with typicality (atypical = .5, typical = − .5), experiment (Exp. 1 = .5, Exp. 2 = − .5), and their interaction, we observed a significant effect of the typicality predictor ($z = 6.85$, $p < .001$). However, there was no significant effect of experiment ($z = 0.10$, $p = .923$), nor was there a significant interaction between typicality and experiment ($z = 0.37$, $p = .713$).

Discussion

In this experiment, we examined the influence of the mouse sensitivity settings on mouse-tracking data. We compared a condition in which these settings were left at the system default under Windows 7 (50% of maximum speed, acceleration enabled) with a slow condition in which the acceleration was disabled and cursor speed reduced (to 30%). The default

condition generally led to greater trajectory curvature (on average), which seemed to be driven by a higher percentage of trajectories with extreme movement patterns. However, there was no significant difference in the sizes of the typicality effect between conditions.

The higher percentage of more extreme movement patterns in the default condition is probably related in particular to the activated acceleration setting, which amplified even small movements toward one of the options. Interestingly, although there was a relatively higher occurrence of these extreme movement patterns for atypical than for typical exemplars, this did not lead to a significant interaction between mouse sensitivity settings and typicality, because there was also an increase in more extreme movement types for atypical exemplars in the slow condition. At the same time, we also found no evidence that a slow condition increases cognitive effects in mouse-tracking data, as had been argued by Fischer and Hartmann (2014).

The second experiment also provided the possibility for an internal replication of the typicality effect across experiments, as the default condition was virtually identical to the click condition of Experiment 1. Across all analyses, there were no significant differences between the experiments, pointing to the stability of mouse-tracking findings across studies—if the methodological setup is held constant.

The findings from this experiment provide a first empirical insight into the effects of mouse sensitivity settings on mouse-tracking data. Although the default setting increased the occurrence of more extreme mouse movement patterns, these did not affect the strength of cognitive effects reflected in mouse movements, and both settings ultimately produced similar results. However, the investigated conditions only reflect two of the most common settings in the literature, and do not represent an exhaustive sample. Some studies have previously used an even greater reduction in speed (e.g., Huette & McMurray, 2010), and it is possible that this could produce stronger effects than the slow condition herein (or different effects altogether). In addition, acceleration and speed were only varied jointly in this experiment, and consequently their relative impacts on the observed effects are not yet clear. Finally, we could only examine the mouse sensitivity settings for one specific type of response indication and starting procedure. It could be the case that mouse sensitivity settings become more important in other setups—for example, for starting procedures that enforce an early movement initiation—a topic we will return to in the next experiment, which compared different types of starting procedures.

Experiment 3

In this experiment we investigated the influence of the starting procedure on mouse-tracking data. The starting procedure concerns the instructions and settings regarding how participants should initiate their mouse movement and how this relates to the stimulus presentation. A number of starting procedures have been used in previous mouse-tracking studies, and the most common ones will be compared within this experiment. The first and most basic starting procedure we termed *static* start. In this procedure, the stimulus is presented immediately after participants have clicked on the start button, and participants do not receive any instructions how and when to initiate their mouse movements. This procedure has been employed in a number of mouse-tracking studies (e.g., Kieslich & Hilbig, 2014; Koop, 2013; Koop & Johnson, 2013), including the original experiment by Dale et al. (2007).

Other mouse-tracking studies have modified the starting procedure in order to ensure that participants initiate their movements early in the trial, hoping to ensure that the complete decision process will be reflected in the movement (Fischer & Hartmann, 2014; Hehman et al., 2015; Scherbaum & Kieslich, 2018). We implemented and tested three such procedures in the following. A simple method (termed *rtmax*) is to restrict the total time participants have for giving a response in a trial (e.g., Duran et al., 2017; Szaszi et al., 2018). This indirectly also encourages an early movement initiation, because participants have to make their choices quickly. A different, frequently employed procedure (termed *initmax* here) uses a static procedure but explicitly instructs participants to initiate their movement within a certain time limit in each trial, and presents a warning after a trial if participants initiated their movement too slowly (see Hehman et al., 2015, for a discussion; and Freeman & Ambady, 2009, 2011; Papesh & Goldinger, 2012; Stolier & Freeman, 2016; and Yu, Wang, Wang, & Bastin, 2012, for exemplary applications). The fourth starting procedure that we included methodologically ensured a movement initiation even before stimulus presentation. In this *dynamic* starting procedure, participants have to initiate an upward movement for the stimulus to be displayed. This procedure has been employed by a number of mouse-tracking studies (see Scherbaum & Kieslich, 2018, for a discussion; and Dshemuchadse et al., 2013; Frisch et al., 2015; Huette & McMurray, 2010; and Scherbaum et al., 2010, for exemplary applications).

Previous recommendations in the literature have stated that starting procedures that ensure that participants initiate the mouse movement early in the trial should help capture cognitive effects in trajectories (Fischer & Hartmann, 2014; Hehman et al., 2015). Specifically, encouraging participants to start moving the mouse as early as possible may increase the likelihood that important aspects of the decision process are reflected in the movement, such as the initial response tendency, varying activations of the competing options, or changes of mind. If parts of these cognitive processes were already completed before participants even started moving the

mouse, these processes would not be captured in the trajectories. In the extreme case, a decision might already be completed before the movement is initiated, leading to a straight trajectory. A straight trajectory in this case would not necessarily indicate that no conflict was present during the decision process, but rather that it occurred before movement initialization. Applying this reasoning to the present experiment, this implies that all starting procedures that ensure an early movement initiation should lead to a larger typicality effect than a static starting procedure. This should hold in particular for the dynamic and the initmax starting procedures, which directly aim at ensuring an early movement initiation.

So far, only one published study has empirically investigated the influence of different starting procedures on mouse-tracking data (Scherbaum & Kieslich, 2018). This study showed that a dynamic starting procedure did not lead to significantly larger cognitive effects found in aggregate curvature measures than did a static starting procedure in which the stimulus was presented after a fixed, short delay. However, a dynamic starting procedure did lead to larger cognitive effects in temporal analyses that assessed how the cursor movement direction was affected by different factors at a specific time point. Although this study provides a first indication that the starting procedure is an important design aspect of mouse-tracking studies, it only involved a comparison of two conditions across experiments (and, as a consequence, without random assignment). In addition, it only considered two possible starting procedures. Therefore, a study that experimentally compares a larger set of commonly used starting procedures is needed.

Method

Procedure and materials This experiment was again conducted at the University of Mannheim, Germany. After providing written informed consent, participants were randomly assigned to one of four starting conditions and completed the experiment. At the end, participants provided demographic information and answered the EHI. After completing the experiment, participants had the chance to win one of several vouchers for local coffee shops (and other businesses, as well as a voucher for a German soccer league game) or sweets.

The basic setup of all conditions was identical to the click condition in Experiment 1, with the following modifications: The stimulus (the animal name) was now presented 340 px above the center of the start button, the cursor speed was reduced (to 40%), and acceleration was disabled.[7] These changes were introduced in order to ensure that the

participants in the dynamic and initmax conditions could acquire the stimulus information during their upward movement without stopping, which is facilitated if the stimulus is presented at a higher position and if the mouse cursor moves slower. In addition, we increased the number of practice trials to six, so that participants could better acquaint themselves with the more complex starting procedures.

Apart from the starting procedure, which was manipulated between participants (static vs. rtmax vs. initmax vs. dynamic), all experimental conditions were identical. A schematic representation of the trial structure for each starting procedure is presented in Fig. 8. In the static condition, the stimulus was presented immediately after participants clicked on the start button, and participants did not receive any information about movement initiation (as in the previous two experiments). The rtmax condition was identical to this, but participants were told that they would have to provide their answer within 2.5 s; if participants took longer than 2.5 s, the trial was aborted and a reminder to answer within the time limit was presented. The initmax condition was also identical to the static condition, with the addition that participants were told that they would have to initiate an upward movement within 0.6 s; if they exceeded this time limit, a warning message was displayed (after participants had given their response) that reminded them to initialize their upward movement within the time limit.[8] The movement criterion for the dynamic procedure followed the setup by Frisch et al. (2015); that is, participants needed to move the mouse 50 px upward for the stimulus to be presented.[9]

Participants On the basis of the power analysis reported in Experiment 1, we intended to ensure a minimum number of 54 participants per experimental condition. A total of 245 participants completed the experiment and were included in the analysis. The sample comprised 162 women and 83 men, and the participants were between 18 and 50 years old ($M =$ 21.9, $SD =$ 3.3). The majority (172 participants) indicated a preference for the right hand, and 16 participants indicated a preference for the left hand; the remaining 57 participants indicated no strong preference.

Results

The analyses of Experiment 3 mostly followed those of the previous experiments. However, a few additional analyses

---

[7] The cursor sensitivity was again set via the Mouse Acceleration Toggler (specific settings: "accel = off, speed = 8"). However, for six participants, the settings accidentally were not activated (meaning that they remained at the system default), so the data for these participants were discarded.

[8] To calculate the initiation time, we used the same upward movement criterion as in the dynamic condition (an upward movement of 50 px). Regarding the time limit, we initially used a time limit of 0.4 s for a set of six participants. However, the participants reported that they were often not able to initiate their mouse movement within that time period, so we increased the time limit to 0.6 s and discarded the data for these participants.

[9] Unlike Frisch et al. (2015), we did not impose a time limit for participants performing this upward movement (in order to keep the setup as simple and understandable as possible).

Before start click

On start click

After moving 50 px up
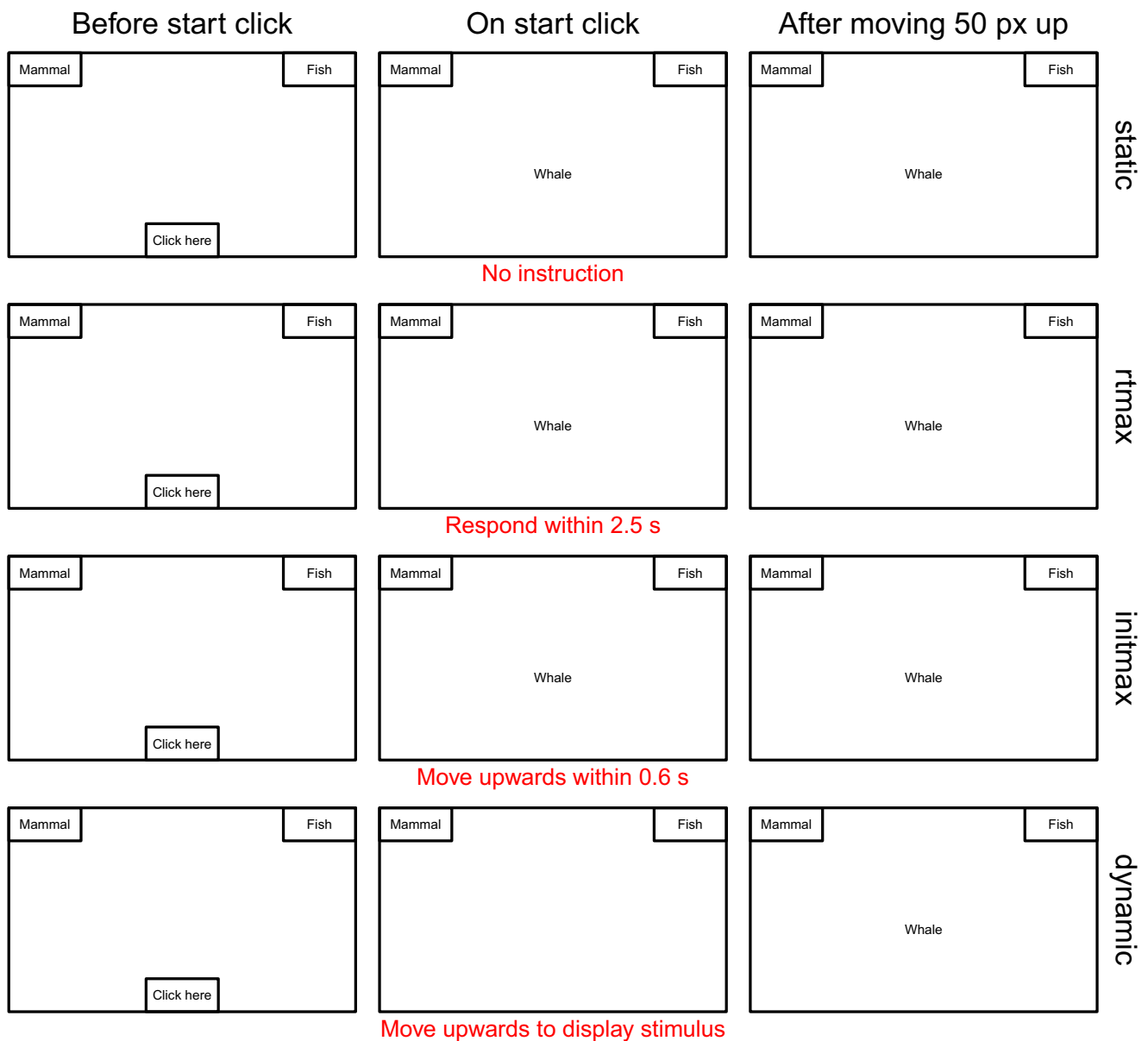
Fig. 8 Schematic representation of the trial structure for each of the starting procedure conditions in Experiment 3, each presented in a separate row. The first column shows the screen containing the start button, which was identical for all conditions. The second column shows the screen that was immediately presented once a participant had clicked the start button. The (shortened) instructions for each starting procedure are shown in red. The last column shows the layout of the screen once the participant had moved the mouse cursor upward for 50 px, which was again identical for all conditions

were conducted as manipulation checks. In addition, since the experimental manipulation now involved more than two conditions, we performed additional contrast analyses to trace back potential effects of the starting procedure to specific conditions. In these analyses, we used dummy-coding in which the static starting procedure served as the baseline condition.

Correctness Across all trials, the numbers of correctly answered trials differed significantly between experimental conditions (static: 94.1%, rtmax: 89.1%, initmax: 89.7%,

dynamic: 93.6%), $\chi^2(3) = 29.93$, $p < .001$. To contrast the effects of the different conditions, we performed a generalized linear mixed model at the trial level using a binomial link function and including a random intercept per participant. The starting condition was included as a predictor, using dummy coding with the static condition serving as the baseline. The dynamic condition did not differ significantly from the static condition ($z = -0.42$, $p = .673$), whereas the initmax condition led to significantly lower performance ($z = -3.12$, $p = .002$). The rtmax condition also led to significantly lower

performance ($z = - 3.53$, $p < .001$); however, when first excluding all trials in the rtmax condition that exceeded the time limit (3.8% of trials, which were counted as incorrect in the previous analysis, as participants did not provide an answer) the performance in the rtmax condition (92.6%) no longer differed significantly from that in the static condition ($z = - 1.18$, $p = .237$). Only correctly answered trials were included in the following analyses.[10]

Manipulation check We analyzed a number of time-related variables as a manipulation check of the starting procedure. For each variable, we first averaged the values per participant and then compared them between conditions. The descriptive statistics of the different variables are displayed in Table 4.

As a first variable, we computed the time it took participants to move the mouse upward for 50 px (RT_{init}). Because the participants in the initmax condition were specifically instructed to initiate their movement within 0.6 s, we expected that RT_{init} should be lower in the initmax than in the static condition. The average RT_{init} in the initmax condition was considerably smaller than the instructed time limit (although participants still exceeded the time limit on 12.4% of trials). RT_{init} differed significantly in an ANOVA between the different starting conditions, $F(3, 241) = 13.64$, $p < .001$, $\eta_p^2 = .15$, 90% CI [0.08, 0.21]. Contrast analyses revealed that RT_{init} was significantly smaller in the initmax than in the static condition, $t(241) = - 5.70$, $p < .001$. It was also significantly smaller in the rtmax condition, $t(241) = - 2.05$, $p = .042$, whereas it was not significantly different from the static condition in the dynamic condition, $t(241) = - 0.45$, $p = .651$.[11]

A similar but more traditional mouse-tracking variable is the initiation time—that is, the time in the trial until any movement was initiated. The initiation times also differed significantly between conditions, $F(3, 241) = 22.69$, $p < .001$, $\eta_p^2 = .22$, 90% CI [0.14, 0.29], with a shorter initiation time in the initmax than in the static condition, $t(241) = - 7.78$, $p < .001$. The dynamic and rtmax conditions also led to significantly shorter initiation times than did the static condition, $t(241) = - 4.58$, $p < .001$, and $t(241) = - 2.05$, $p = .042$.[12]

With regard to the total response time (RT) in each trial, we expected that the rtmax condition would lead to shorter RTs. The starting procedure had a significant influence on the RT, $F(3, 241) = 49.61$, $p < .001$, $\eta_p^2 = .38$, 90% CI [0.30, 0.44]. Contrast analyses revealed that the rtmax condition indeed led to shorter RTs than did the static condition, $t(241) = - 4.63$, $p < .001$. RTs were also significantly shorter in the initmax than in the static condition, $t(241) = - 5.14$, $p < .001$. In the dynamic condition, the total RTs were significantly longer than in the static condition, $t(241) = 5.47$, $p < .001$.[13] The overall longer RTs in the dynamic condition make sense, because in this condition the stimulus was only displayed after initiation of the upward movement, and consequently, processing might have started later. Interestingly, if we calculated the RT for the dynamic condition based solely on the part of the trial after the stimulus presentation (which is typically done in studies that use a dynamic starting procedure; e.g., Dshemuchadse et al., 2013; Frisch et al., 2015; Scherbaum et al., 2010), it was on average ($M = 2,021.1$ ms, $SD = 671.1$ ms) quite comparable to the total RT in the static condition.

Aggregate trajectory curvature To get a general impression of the effect of the starting procedure on trajectory curvatures, we inspected the aggregate time-normalized trajectories (Fig. 9). In all experimental conditions, the aggregate trajectories deviated more toward the nonchosen option for atypical than for typical exemplars. The dynamic and initmax conditions generally led to prolonged vertical upward movements as compared to the static and rtmax

Table 4. Means (and *SD*s) of the per-participant aggregated timing variables in Experiment 3, presented separately for each condition (in milliseconds)

Condition	RT_{init}	Initiation Time	RT
static	808.5 (324.1)	508.7 (215.6)	2,110.4 (654.1)
rtmax	650.1 (176.6)	437.3 (160.0)	1,521.6 (183.4)
initmax	377.4 (159.5)	243.1 (142.8)	1,471.7 (248.6)
dynamic	773.4 (752.2)	348.7 (233.2)	2,805.4 (1,199.8)

RT_{init} = Time until cursor was moved 50 pixels upward

---

[10] Initially, we planned to exclude trials exceeding the movement initialization time limit in the initmax condition. We noticed during pilot trials, however, that meeting this criterion was challenging for some participants. Therefore, we decided to slightly oversample the number of participants in this condition, to be able to compensate for participants who would have to be excluded because they did not initiate their movement in time on enough trials. However, as previous studies using an initmax starting procedure (e.g., Freeman & Ambady, 2011) did not exclude trials exceeding the time limit, we eventually decided to follow that procedure. All main results can be replicated when excluding trials in the initmax condition on which the time limit was not met, thereby excluding five participants for whom no correctly answered trials remained in either typicality condition (see the complete analyses online).

[11] Because some trials had extremely large RT_{init} values, we repeated the analyses using median instead of mean values per participant. The general pattern was replicated. However, now the dynamic condition also had a significantly shorter RT_{init} than the static condition (see the complete analyses online).

[12] When repeating the analyses using median instead of mean values per participant, the general pattern was replicated. However, now the rtmax condition did not have a significantly shorter initiation time than the static condition.

[13] The results pattern was comparable when using median instead of mean values per participant.
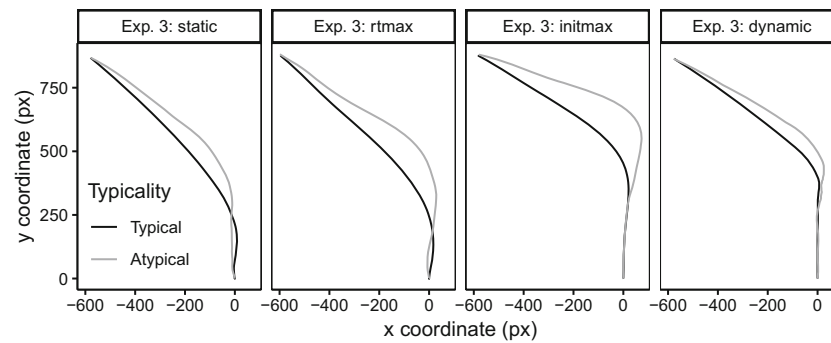
Fig. 9 Aggregate mouse trajectories for Experiment 3. All individual trajectories were flipped to the left, time-normalized, and aggregated separately per typicality and experimental condition

conditions. In addition, the typicality effect especially seemed to be more pronounced in the initmax condition.

A repeated measures ANOVA using the per-participant aggregated MAD values with the within-subjects factor typicality (atypical vs. typical) and the between-subjects factor starting procedure (static vs. rtmax vs. initmax vs. dynamic) revealed a significant main effect of typicality, $F(1, 241) = 97.72$, $p < .001$, $\eta_p^2 = .29$, 90% CI [0.21, 0.36], with higher MAD values for atypical than for typical exemplars. The effect of typicality was significant in all four conditions (Table 2).

In addition, we found a significant main effect of the starting procedure, $F(3, 241) = 18.67$, $p < .001$, $\eta_p^2 = .19$, 90% CI [0.11, 0.25]. Contrast analyses revealed that MAD values were overall significantly higher in the initmax than in the static condition, $t(241) = 6.53$, $p < .001$, as well as in the dynamic than in the static condition, $t(241) = 4.10$, $p < .001$. The MAD values in the rtmax condition did not differ significantly from the static condition, $t(241) = 0.72$, $p = .470$.

There was also a significant interaction between typicality and starting procedure, $F(3, 241) = 4.12$, $p = .007$, $\eta_p^2 = .05$, 90% CI [0.01, 0.09]. As can be seen in Fig. 10, the typicality effect was significantly larger in the initmax than in the static condition, $t(241) = 2.68$, $p = .008$. No significant difference in the size of the typicality effect emerged between the dynamic and static conditions, $t(241) = -0.56$, $p = .576$, nor between the rtmax and static conditions, $t(241) = 0.87$, $p = .383$.

Distribution of trajectory shapes To analyze trajectory shapes, we again computed the bimodality coefficients for the per-participant standardized MAD values separately for each typicality and experimental condition (Table 2). The bimodality coefficients were smaller than .555 for all starting procedures, for both typical and atypical trials, with the exception of the typical trials in the dynamic condition, where the value of .560 was slightly larger than the cutoff.

Smoothed heat maps (Fig. 11) indicated that a considerable number of straight trajectories occurred in the static and rtmax conditions. In the dynamic and the initmax conditions, there were fewer straight trajectories, but instead, many trajectories that moved upward for a distance (longer than the required movement criterion). In all conditions, there also seemed to be a number of change-of-mind trajectories, in which the cursor was moved all the way to nonchosen option before moving to the chosen option.

To quantify and statistically test for differences in the frequencies of trajectory types between conditions, we mapped the trajectories on the same set of prototypes that was used in Experiments 1 and 2. The majority of individual trajectories again seemed to map well onto the set of prototypes
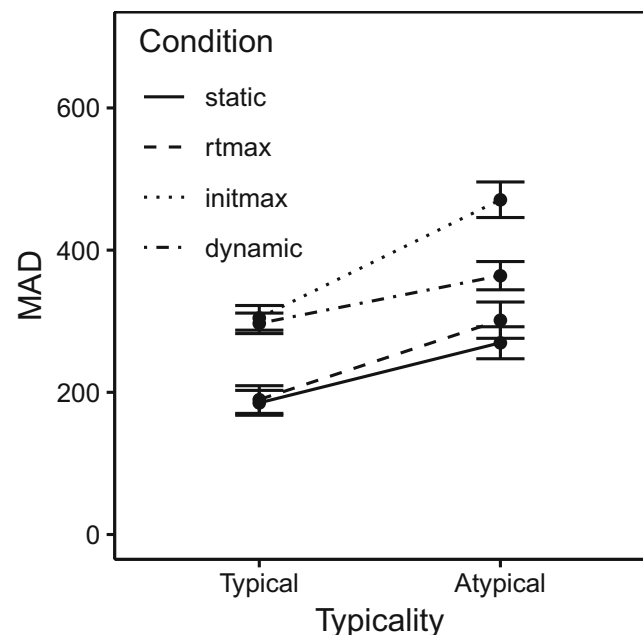
Fig. 10 Mean of maximum absolute deviation values (MAD, in pixels) for Experiment 3, shown separately per typicality and experimental condition. Error bars indicate one *SEM*
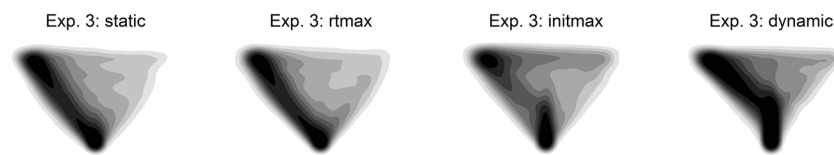
Fig. 11 Smoothed heat maps of the individual trajectories in Experiment 3, shown separately per experimental condition. Darker colors indicate higher density

(Fig. 12).[14] The relative frequencies of prototypes differed significantly between experimental conditions, $\chi^2(12) = 535.73$, $p < .001$. In line with the previous experiments, the majority of trials in the static condition were classified as straight. Similar results were also obtained in the rtmax condition. In the dynamic condition, the majority of trials were classified as curved, whereas the initmax condition led to a roughly even split of straight and curved classifications, as well as a considerable increase of dCoM classifications relative to the other three conditions.

We again predicted the trajectory type in an ordinal mixed regression including a random intercept per participant and the predictors typicality (atypical = .5, typical = − .5) and experimental condition (dummy-coded, with static serving as the baseline condition). Atypical trials led to a significantly higher probability of more extreme trajectories in the static condition ($z = 5.06$, $p < .001$). The rtmax condition did not differ significantly from the static condition, $z = 0.31$, $p = .760$. Both the initmax and the dynamic conditions led to significantly more extreme trajectories than the static condition, $z = 5.99$, $p < .001$, and $z = 4.37$, $p < .001$, respectively. For the initmax condition, more instances of both the curved and change-of-mind trajectory types occurred, whereas for the dynamic condition, there seemed to be especially more curved trajectories (see Fig. 12). With regard to the interaction of typicality and condition, we observed no significant interactions for the rtmax and initmax conditions, $z = 0.79$, $p = .427$, and $z = 1.68$, $p = .094$. The dynamic condition led to a relatively smaller increase in extreme trajectories for atypical trials than the static condition, $z = − 2.19$, $p = .029$.

Discussion

In this experiment, we examined the influences of four different starting procedures on mouse-tracking data. Several

previous mouse-tracking studies have used a static starting procedure in which the stimulus was immediately presented after participants clicked on the start button, and participants did not receive any instructions regarding movement initiation. However, this poses the risk that participants might make their decision before initiating the movement. Therefore, other studies have employed measures to ensure that participants initiate their movement early in the trial, hoping to increase the likelihood that the complete decision process would be reflected in the movement. These methods include restricting the total response time (rtmax), instructing participants to initialize their movement early in the trial (initmax), or requiring an upward movement for the stimulus to be displayed (dynamic).

The results showed that the initmax condition, in which participants were instructed to initialize an upward mouse movement within 0.6 s, led to a significantly larger typicality effect than did the static condition, as measured via MAD. This was accompanied by an increase of change-of-mind trajectories that moved all the way to the nonchosen option before heading to the chosen option. However, participants also made more mistakes in their choices than they did in the static condition. When implementing the initmax condition, a central challenge is to set an adequate time limit for initiation of the mouse movement (see Hehman et al., 2015, for a discussion). With the present setting (move upward 50 px within 0.6 s), we found that participants could not always meet the time limit for initiating their mouse movement. Although a slight increase of the time limit might seem an easy solution for this issue in future studies, it bears the potential of offsetting the above-described effects of this starting procedure. This highlights the need for conducting pilot studies to determine which initiation time threshold works best for the specific task at hand.

A dynamic starting procedure, in which participants had to move the cursor upward 50 px for the stimulus to be displayed, did not significantly influence the typicality effect. However, trajectories were overall classified as being more curved than in all other starting procedures, and the relative occurrence of more extreme trajectory types in atypical versus typical trials was slightly reduced. This indicates that a dynamic starting condition increases the occurrence of trajectories that fit best with the idea of a continuous mapping of the cognitive process onto the movement. However, we cannot rule out that the early part of the trajectory especially might

---

[14] Upon closer inspection of the data, we discovered a small number of trajectories in the initmax and dynamic conditions that were not captured well by the existing prototypes, as they went upward all the way to the top of the screen and then moved either left to the chosen option or first right to the nonchosen option and from there left to the chosen option (or, rarely, even twice back and forth). Comparable results were obtained when including those movement patterns as additional prototypes and repeating the analyses with this extended set (see the complete analyses online).
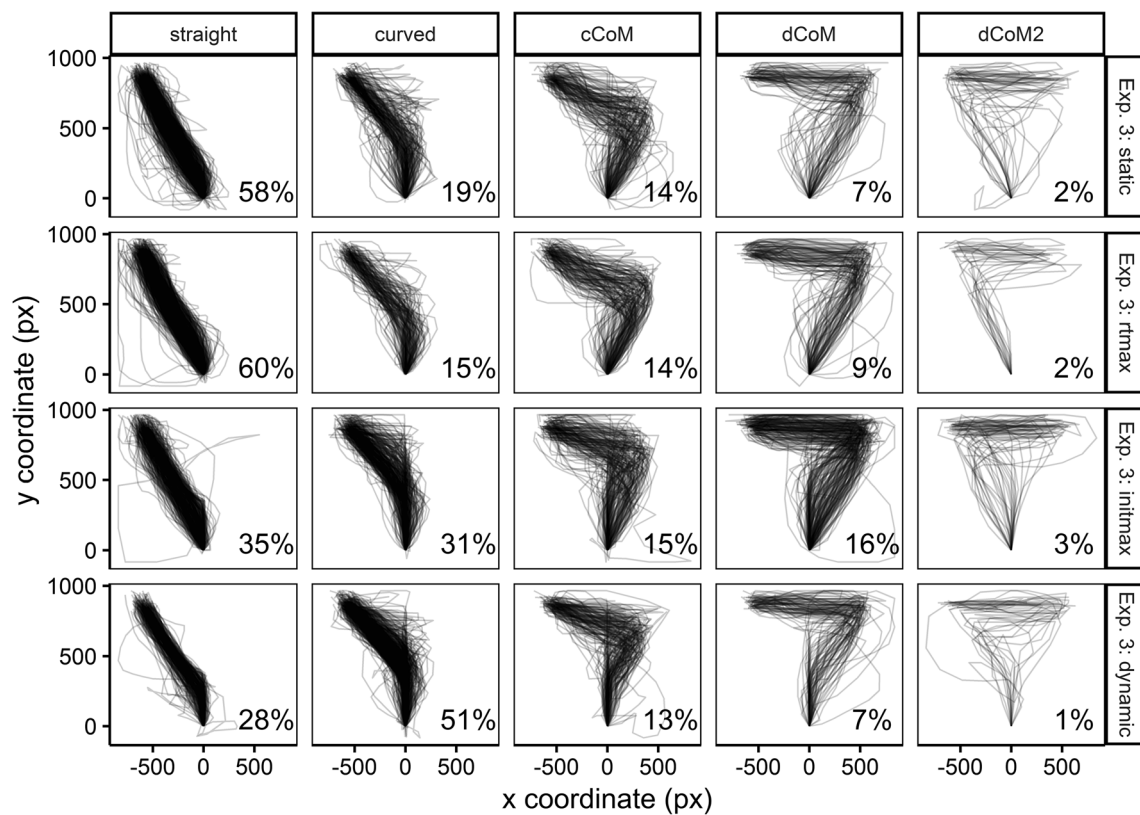
Fig. 12 Individual trajectories per assigned prototype, shown separately for the different experimental conditions in Experiment 3. For each prototype, the relative frequency of classifications per experimental condition is displayed

reflect a purely motoric component (i.e., the vertical movement that needs to be executed for the stimulus to be displayed and a continuation of this movement while the stimulus is processed), and hence, the overall more curved shaped would be driven mainly by this motor process. Nevertheless, as compared to a static start, the dynamic starting procedure methodologically ensures that processing will take place during the movement and, hence, increases the likelihood that these processes will be reflected in the movement.

Combining a dynamic start with a touch instead of a click response, as has been implemented in previous studies (Dshemuchadse et al., 2013; Frisch et al., 2015; Scherbaum et al., 2010), may also lead to a more homogeneous distribution of trajectory curvature. These studies have also often employed a dynamic starting procedure in combination with restrictions regarding the time for initiating the upward movement as well as the time for giving the total response. Future studies should therefore more closely examine the dynamic starting procedure in this setup (see also Scherbaum & Kieslich, 2018).[15]

_____

[15] Older studies used a different movement initiation criterion (moving upward at least 4 px in each of two consecutive time steps; see Scherbaum et al., 2010). However, we think that this does not constitute an important difference, and would argue in favor of the newer criterion (used, e.g., in Frisch et al., 2015), as its definition is more straightforward and easier to implement.

Introducing a total time limit of 2.5 s for giving a response in the rtmax condition did not have a significant effect in any of the trajectory analyses, as compared to the static starting procedure. It is possible that this resulted from the time limit not being strict enough, given that the average response time in the static condition was also shorter than 2.5 s (see Table 4). When setting the time limit, we had intended to encourage participants to start moving earlier without introducing too much overall time pressure (which could alter basic decision processes—e.g., through the introduction of stress). The manipulation generally seemed to be effective, since the average response time in the rtmax condition was more than 0.5 s shorter than that in the static condition, and participants initiated their upward mouse movements earlier in the trial. Still, future studies could explore the use of a stricter total time limit, at the risk of altering the cognitive processes and potentially losing more trials in which participants do not answer within the time limit.

An interesting observation is that the bimodality coefficients in almost all conditions indicated evidence for a unimodal distribution. This was also the case for the static condition, which was quite similar to the slow condition from Experiment 2 (in which the bimodality coefficient indicated bimodality). The main methodological difference for the static condition in the present experiment was that the stimulus was presented at a higher point on the screen. This might explain

the finding, because on average the trajectories headed upward for a longer time in the static condition of Experiment 3 than in the slow condition from Experiment 2 (cf. Figs. 2 and 9). This, in turn, led to higher average MAD values (Table 2), and deviations from this higher baseline would have a smaller effect on the bimodality coefficient. Still, inspection of the heat maps and prototype classifications indicated that different types of trajectories were present in all starting conditions.

In sum, this experiment has shown that the starting procedure has considerable influence on mouse-tracking data, influencing both the size of the cognitive effects reflected in mouse trajectories and their shape. Because both have been used to test psychological theories, this underscores the importance of taking the methodological setup of the study into account when interpreting mouse-tracking data. Specifically, within the same task, a dynamic starting procedure can produce a majority of curved trajectories, whereas a static start leads to a majority of straight trajectories. While the former type of trajectory is usually associated with dynamic process models, the latter type is interpreted instead as belonging to a low-conflict decision in a dual-system model. At the same time, change-of-mind trajectories were present to varying degrees for all starting procedures, which are usually associated with high-conflict decisions in a dual-system model. Assuming that the starting procedure does not really change the underlying decision process, this implies that the mapping of the decision process onto the mouse movement depends on the starting procedure, with the dynamic and initmax starting procedures increasing the likelihood that the decision process, and especially its early stages, is more continuously mapped onto the mouse movement.[16]

General discussion

Over the past decade, mouse-tracking has spread to a multitude of psychological areas and has been used to examine diverse cognitive processes (Freeman, 2018; Stillman et al., 2018). Given the relative novelty of the method, to date no standards for designing and running mouse-tracking experiments have been established, and this has entailed considerable variation in the methodological setups of previous mouse-tracking studies. To improve understanding of the empirical, and ultimately also the theoretical, consequences of methodological differences for mouse-tracking data, in the present study we reported a systematic investigation of three central design factors. In a series of experiments, the design factors response indication, starting procedure, and mouse

___

[16] As was previously discussed, it remains an open question to what degree the higher share of curved trajectories in the dynamic and initmax conditions was driven by motor processes executed at the beginning of the trial to fulfill the upward movement criterion.

sensitivity were varied, while the basic experimental setup remained constant and followed that of a classic mouse-tracking experiment by Dale et al. (2007). In all methodological setups, the postulated typicality effect was replicated, in that mouse trajectories deviated more toward the nonchosen option for atypical than for typical stimuli. However, the size of this effect was influenced by the type of response indication and the starting procedure. In addition, trajectory shapes were influenced by all design factors: In traditional bimodality analyses, some setups have led to the distribution of curvature indices being classified as unimodal, whereas other setups have been classified as bimodal. When mapping individual trajectories onto a set of prespecified prototypes, the relative frequency of the prototypes varied according to the methodological setup, and in many cases this could explain differences in the size of the typicality effect between conditions. The main effects of the design factors on trajectory curvature and shape are summarized in Table 5.

Implications for interpreting mouse-tracking results

The findings of the present study have general implications for interpreting the results from mouse-tracking studies. First, it could be demonstrated that it is possible to find theoretically predicted effects on mouse trajectories in any methodological setup. Thus, in principle, any setup seemed able to capture the conflict between response options to at least some degree. This provides evidence that in the present task, the connection between decision processes and cursor movements was so robust that it prevailed under any type of methodological setup. However, the size of this effect varied considerably between setups. Therefore, it is possible that smaller and less robust effects than the typicality effect we investigated would not be detected with certain methodological setups. As a consequence, before concluding that a certain manipulation does not influence conflict as measured through mouse trajectories, one should consider whether the setup was optimized to detect such effects. More generally, the findings highlight that the comparison of effect sizes between mouse-tracking studies with different methodological setups might be challenging—given that in the present experiments anything from a small (touch condition from Exp. 1) to a large (initmax condition from Exp. 3) effect was obtained for the exact same task and manipulation (see Table 2).

Second, previous mouse-tracking studies have used bimodality analyses of curvature indices to conclude which theoretical model may account for the cognitive process of interest (Freeman & Dale, 2013; Hehman et al., 2015; Stillman et al., 2018). Most often, studies have aimed to differentiate between dynamic and dual-system models, which should lead to unimodal versus bimodal distributions, respectively. The present study demonstrated that, depending on the methodological setup, both unimodal

Table 5. Summary of design factor effects and recommendations for future studies

Design Factor	Influence on Typicality Effect	Influence on Trajectory Shapes	Preliminary Recommendations for Future Studies
Starting procedure	Initmax leads to largest effect. Other conditions are not significantly different.	Static/rtmax: majority straight. Dynamic: majority curved. Initmax: mix of straight and curved, increase of dCoM.	For simple tasks: use dynamic/initmax start to increase chances early aspects of decision process are captured, especially relevant for temporal analyses (Scherbaum & Kieslich, 2018). For complex tasks: use static start (restricting response time had no significant effect, but limit was not very strict).
Mouse sensitivity	No significant influence on typicality effect.	Default settings: increase of dCoM. Reduced speed and disabled acceleration: increase of straight.	For dynamic/initmax: Reduce speed and disable acceleration to ensure decision can be completed during upward movement. For static: settings had no strong effect (avoid very fast cursor nevertheless).
Response indication	Click leads to larger effect than touch.	Click: increase of dCoM. Touch: mostly straight, almost no dCoM.	For dynamic/initmax: use touch to achieve more homogeneous curvature distribution. For static: use click, but consider analysis via prototype classification.

The influences on the size of the typicality effect (maximum absolute deviation in atypical vs. typical trials) and trajectory shapes (relative frequency of prototype classifications) are summarized. Note that the summary and recommendations are preliminary, as not all combinations of each of the design factor conditions were investigated in this study. rtmax = static start with restricted total response time, initmax = static start with restricted initiation time, dCoM = discrete change-of-mind trajectory

and bimodal distributions can be obtained in the very same psychological task. If cognitive processes were not influenced by the setup, this would imply that different shapes can occur for the same cognitive process. This, in turn, would indicate that the mapping of cognitive processes onto mouse movements can vary depending on the methodological setup, and that an interpretation of trajectory shapes needs to consider the methodological conditions under which they were obtained. However, solely on the basis of mouse-tracking data, we cannot rule out that changes in the setup influence cognitive processes directly, and that this might offer an alternative explanation for the findings. For instance, one could argue that changes in the setup induced changes in the movement that had a backward influence on the decision processes via feedback mechanisms (e.g., Lepora & Pezzulo, 2015; see also Aczel, Szollosi, Palfi, Szaszi, & Kieslich, 2018). Future research will need to address this question further using additional indicators of cognitive processes—for example, neurophysiological indices or other process-tracing techniques. For now, we make the tentative (and potentially simplifying) assumption that the setup mostly influences how the cognitive process is mapped onto the movement.

More specifically, the starting procedure may influence the degree to which early aspects of the decision process are reflected in the mouse movement, with an initmax and a dynamic starting procedure increasing the likelihood that early

aspects are captured, whereas this is not guaranteed in studies that use a static start. In extreme cases, the decision process might even be finished before the mouse movement was initiated, in setups with a static start. A resulting straight trajectory would then not necessarily indicate that there was no response conflict, but possibly that such conflict was not captured in the movement. In addition, the response indication probably influences the degree to which the attraction of an option is translated into a movement toward that option. The click condition allowed participants to move all the way to an option (if this was the currently favored option) and then redirect the movement to the other option (a prototypical change-of-mind trajectory), whereas the touch condition reduced the likelihood of these extreme movements. Thus, a study with a static starting procedure and a click response mode is more likely to produce a mix of straight and change-of-mind trajectories than is a study with a dynamic starting procedure and a touch response mode. So, if in the former case mostly curved trajectories were observed, this could be seen as more compelling evidence for a dynamic system model on the process level than if those were observed in the latter case. Conversely, if a mix of straight and change-of-mind trajectories were observed in a setup with a dynamic starting procedure and a touch response mode, this could be seen as more convincing evidence for a dual-process model than if those were observed for a static starting procedure and a click response mode. One recommendation for future studies that aim for a critical test of a certain class of process model

could thus be: "Optimize" the methodological setup in such a way that it increases the likelihood that types of trajectories will be observed that are not expected on the basis of the a-priori-hypothesized process model. For instance, optimize the setup for continuously curved trajectories, if a dual-process model is assumed. However, this is just a preliminary suggestion.

Third, the results demonstrate the usefulness of the newly proposed analysis method for identifying different types of movement trajectories (Wulff et al., in press; Wulff et al., 2019). On the one hand, it allows for unpacking the effect that a certain manipulation has on mouse trajectory curvature. That is, it shows whether higher curvature is caused by all trajectories being more curved in one of the conditions, or whether a certain condition leads to the occurrence of more extreme trajectory types, such as discrete changes of mind. In the present experiments, larger effects on aggregate curvature were often accompanied by a higher share of these types of movements. On the other hand, it offers an alternative way to assess whether or not different types of trajectories are present in the data. In Experiments 1 and 2, the new method generally seemed to agree with the traditional bimodality method, in that conditions with a bimodal distribution also contained a mix of more extreme trajectory types than did conditions with a unimodal distribution. In Experiment 3, bimodality analyses generally suggested unimodal distributions, yet the prototype method still seemed to indicate the presence of different types of trajectories in the data. Future research will need to address the conditions under which these methods agree, and if not, which method offers the more valid interpretation.

Implications for designing mouse-tracking studies

The findings have implications for the design of future mouse-tracking studies. In line with previous recommendations (Fischer & Hartmann, 2014; Hehman et al., 2015), a starting procedure that encourages participants to initiate their mouse movement early in the trial led to larger cognitive effects. Interestingly, no larger effects were observed for the dynamic starting procedure, in which participants had to move the mouse upward for the stimulus to be displayed. This, in turn, is in line with previous findings by Scherbaum and Kieslich (2018), who also did not find differences in the cognitive effects on trajectory curvature when comparing a dynamic and a static starting procedure. However, they also showed that a dynamic starting procedure led to larger effects in more fine-grained analyses that investigated the temporal development of within-trial movements in a time-continuous regression framework.

The cursor speed and acceleration settings did not have substantial effects on mouse-tracking data in the present study. However, we only examined the effect of these settings in a static starting procedure. In test runs for the dynamic and initmax starting procedures, we observed that a fast cursor with enabled acceleration made it extremely difficult to acquire the stimulus information during the upward movement. Specifically, the cursor increased in speed so quickly at the beginning of the trial that it was difficult to read the stimulus word before reaching the upper part of the experimental screen. For this reason, we decided to reduce the cursor speed and to turn off acceleration when comparing the different starting procedures—a setup that we recommend to anyone using a starting procedure that encourages early movement initiation (in line with Fischer & Hartmann, 2014; Freeman & Ambady, 2010).

With regard to the response indication mode, a response by click led to considerably larger effects than did a response by touching the button with the cursor. However, this was related to an occurrence of more extreme trajectories. In addition, the distribution of curvature values was classified as bimodal in the click condition and as unimodal in the touch condition. This suggests that researchers might face a trade-off between larger effects that are due to the occurrence of more extreme trajectory types and smaller effects with a more homogeneous trajectory distribution.

Another consideration when making design choices for a mouse-tracking study is the question of which setup is suited for which psychological tasks. Although encouraging early mouse movements through the starting procedure should work well for studies with stimuli that can be processed very quickly and in which decisions are relatively easy, it might provide a challenge for studies involving more complex tasks (such as decisions between monetary lotteries—e.g., Koop & Johnson, 2013—or decisions in the Cognitive Reflection Test—e.g., Travers, Rolison, & Feeney, 2016). In this case, a static starting procedure might be better suited, so that participants can initiate the movement after acquiring the stimulus information. This is at the risk that they might finish processing and arrive at a decision before initiating the mouse movement. This would most likely lead to an increase of straight trajectories (which can also be observed in the present study in conditions that used a static starting procedure).

On the basis of these considerations, we provide a set of preliminary recommendations in Table 5. For simple tasks, we recommend that researchers use an initmax or dynamic starting procedure, disable acceleration, reduce cursor speed, and use a touch response indication procedure in order to achieve a homogeneous trajectory distribution. If the task under investigation is more complex, researchers might use a static starting procedure and collect responses via click. Cursor speed is not as critical in this setup, but an extremely fast cursor should be avoided, nonetheless. In this setup, dynamic analyses may not be possible, because the consistency of the movement over the course of a trial may not be sufficient (Scherbaum & Kieslich, 2018). In addition, the occurrence of different types of trajectories is more likely in this setup, and thus an analysis of trajectory shapes via prototype

classification should be performed (Wulff et al., in press). These recommendations are, of course, preliminary, and require more empirical studies that will investigate different combinations of each of the design factor conditions and different psychological tasks.

Limitations

In the present study we investigated the influence of design factors on mouse-tracking data by replicating a classic mouse-tracking experiment (Dale et al., 2007, Exp. 1) with different methodological setups. To investigate the impacts of these different setups on mouse-tracking data, we had to make a number of choices, each of which entailed certain limitations. Specifically, we selected a specific set of three central design factors and implemented only specific combinations of these. We also decided to implement all manipulations between participants, to avoid carryover effects. Therefore, we could only implement a limited number of conditions for each design factor, to ensure that statistical power was sufficient for each condition. We further only used one mouse-tracking paradigm for all three experiments, to ensure comparability between experiments. Finally, we selected only a subset of all potentially available mouse-tracking analyses for the present purpose, focusing on the most frequently used analyses. In the following, we discuss how each of these choices could pose limitations for the present investigation.

As stated above, we only examined three design factors in total and varied only one of them in each experiment. While this allowed for a clear interpretation of the consequences of each design factor in isolation, it also excluded the possibility of investigating potential interactions between the different design factors—some of which are likely to occur. For example, while it might be the case that a touch response procedure drastically reduces cognitive effects of curvature when used in combination with a static start procedure (as implemented in this study), it could well be that this is not the case when a touch response procedure is combined with a starting procedure that encourages early movement initiation. In addition, for starting procedures that encourage an early movement initiation, a fast cursor speed might lead to problems (as we previously discussed). Consequently, the study of design factors in mouse-tracking warrants further investigation and extension of the present results in order to investigate the effects of each design factor with different combinations of the other factors.

In addition, although we intended to cover the most common implementations of each design factor, we could not cover all possible implementations. With regard to mouse sensitivity, we only compared two commonly used settings (one with default settings—i.e., enabled acceleration and 50% speed—and one with disabled acceleration and 30% speed). Of course, many more speed settings could be

used—both faster and slower speeds—and the effect of speed should be explored independently of acceleration. With regard to the starting procedures, several studies have also employed a static starting procedure with delayed stimulus presentation (e.g., Spivey et al., 2005)—a condition that was not included in the present study. In addition, the dynamic and initmax starting procedures also have been used in combination with a general restriction of the total response time (e.g., Dshemuchadse et al., 2013; Freeman & Ambady, 2011). In this regard, the previously discussed study by Scherbaum and Kieslich (2018) may offer first insight, as it compared a dynamic with a static starting procedure (in which the stimulus presentation was slightly delayed), and the total response time was restricted in both conditions.

We also intended to investigate the most central design factors. However, additional design factors are likely relevant in mouse-tracking studies. One potentially relevant factor is the stimulus position. This is already suggested by the fact that the shapes of the trajectories differed between the slow condition of Experiment 2 and the static condition of Experiment 3, which is likely related to the change of stimulus position between the experiments. Another potentially relevant factor concerns the response button position. Research on these and additional factors is currently under way (Grage, Schoemann, & Scherbaum, 2019; Schoemann, Lüken, Grage, Kieslich, & Scherbaum, 2019). In addition, there are also other factors that are more closely related to the task at hand, such as the stimulus modality (presented as a written word, a spoken word, or a picture). Dale et al. (2007), for example, replicated their experiment both with written words and pictures, the latter generally leading to larger effects.

A further potential limitation of the present investigation is that it only used a paradigm from one content area (semantic categorization) and that the study was implemented in a relatively simplistic methodological setup. As compared to the paradigm by Dale et al. (2007) that we used herein, other studies have included a considerably higher number of practice and actual trials and gave closer instructions to participants with regard to how they should move the mouse (e.g., Dshemuchadse et al., 2013; Scherbaum et al., 2010). Nevertheless, we think that the basic setup of the present study is representative of many mouse-tracking studies and that the effects observed in each experiment were reliable (as indicated by the cross-experiment comparison of two conditions with identical setups that did not differ significantly in any analyses). Still, all of this highlights the need for future studies that can examine the effects of the individual design factors in other mouse-tracking paradigms.

Finally, in our analyses we focused on the influence that design factors have on trajectory curvature and trajectory shape, assessed through the calculation of MAD values, bimodality analysis, and the newly proposed prototype-mapping method. We did this because these are the most

common applications of mouse-tracking (Stillman et al., 2018). Nonetheless, the richness of mouse-tracking data allows for a multitude of further analyses (see Freeman, 2018; Hehman et al., 2015; Scherbaum et al., 2010; Stillman et al., 2018), and it is an interesting question how design factors might also have an influence on each of these other analyses (see Scherbaum & Kieslich, 2018, who performed a number of additional analyses in their comparison of two starting procedures).[17] Because the data for all experiments are freely provided in an open format (along with open-source software for their analysis), interested researchers are invited to use them to explore how the design factors in the present study influence data in their particular analysis of interest.

Conclusion

In this study, we presented a comprehensive investigation of three central design factors in mouse-tracking—response indication, mouse sensitivity, and starting procedure—and of their influence on mouse-tracking data. We demonstrated that these design factors can have considerable impact on trajectory curvature and the shape of individual trajectories. Such differences can, in turn, bias theorizing and lead to premature conclusions about support for or against certain theories—for example, the distinction of dynamic and dual-system accounts. Our results strongly suggest that the specific design of a mouse-tracking study must be carefully considered when interpreting mouse-tracking data with respect to testing theories and when planning mouse-tracking studies. Finally, an extension of the present investigation to further setups and tasks seems imperative—an endeavor that we encourage further researchers to pursue with the help of the experiments, analysis code, and data that we have made available.

---

[17] Scherbaum and Kieslich (2018) analyzed the development of the mouse movement over time using time-continuous multiple regression analyses. We did not replicate these analyses herein, as they require a considerably higher number of trials per participant than the 19 trials available for each participant in the present study.

References

Aczel, B., Szollosi, A., Palfi, B., Szaszi, B., & Kieslich, P. J. (2018). Is action execution part of the decision-making process? An investigation of the embodied choice hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44, 918–926. https://doi.org/10.1037/xlm0000484

Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science*, 35, 983–996. https://doi.org/10.1111/j.1551-6709.2010.01164.x

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

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

Duran, N. D., Nicholson, S. P., & Dale, R. (2017). The hidden appeal and aversion to political conspiracies as revealed in the response dynamics of partisans. *Journal of Experimental Social Psychology*, 73, 268–278. https://doi.org/10.1016/j.jesp.2017.07.008

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., Witte, M., & Hartmann, M. (2018). Tracking the continuous dynamics of numerical processing: A brief review and editorial. *Journal of Numerical Cognition*, 4, 271–285. https://doi.org/10.5964/jnc.v4i2.179

Fischer, M. H., & Hartmann, M. (2014). Pushing forward in embodied cognition: May we mouse the mathematical mind? *Frontiers in Psychology*, 5, 1315. https://doi.org/10.3389/fpsyg.2014.01315

Freeman, J. B. (2014). Abrupt category shifts during real-time person perception. *Psychonomic Bulletin & Review*, 21, 85–92. https://doi.org/10.3758/s13423-013-0470-8

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. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science*, 20, 1183–1188. https://doi.org/10.1111/j.1467-9280.2009.02422.x

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., & Ambady, N. (2011). Hand movements reveal the time-course of shape and pigmentation processing in face categorization. *Psychonomic Bulletin & Review*, 18, 705–712. https://doi.org/10.3758/s13423-011-0097-6

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, 137, 673–690. https://doi.org/10.1037/a0013875

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*, 2, 59. https://doi.org/10.3389/fpsyg.2011.00059

Frisch, S., Dshemuchadse, M., Görner, 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, M., & Scherbaum, S. (2019). *Lost to translation: How design factors of the mouse-tracking procedure impact the inference from action to cognition*. Manuscript submitted for publication.

Hehman, E., Carpinella, C. M., Johnson, K. L., Leitner, J. B., & Freeman, J. B. (2014). Early processing of gendered facial cues predicts the electoral success of female politicians. *Social Psychological and Personality Science*, *5*, 815–824. https://doi.org/10.1177/1948550614534701

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, *18*, 384–401. https://doi.org/10.1177/1368430214538325

Huette, S., & McMurray, B. (2010). Continuous dynamics of color categorization. *Psychonomic Bulletin & Review*, *17*, 348–354. https://doi.org/10.3758/PBR.17.3.348

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, *102*, 116–131. https://doi.org/10.1037/a0025335

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., Henninger, F., Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (in press). 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.

Kieslich, P. J., & Hilbig, B. E. (2014). Cognitive conflict in social dilemmas: An analysis of response dynamics. *Judgment and Decision Making*, *9*, 510–522.

Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judgment and Decision Making*, *8*, 527–539.

Koop, G. J., & Criss, A. H. (2016). The response dynamics of recognition memory: Sensitivity and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 671–685. https://doi.org/10.1037/xlm0000202

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

Lepora, N. F., & Pezzulo, G. (2015). Embodied choice: How action influences perceptual decision making. *PLOS Computational Biology*, *11*, e1004110. https://doi.org/10.1371/journal.pcbi.1004110

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*, 314–324. https://doi.org/10.3758/s13428-011-0168-7

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin & Review*, *19*, 906–913. https://doi.org/10.3758/s13423-012-0281-3

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

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., & 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., Lüken, M., Grage, T., Kieslich, P. J., & Scherbaum, S. (2019). Validating mouse-tracking: How design factors influence action dynamics in intertemporal decision making. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-018-1179-4

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

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). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*, 10393–10398. https://doi.org/10.1073/pnas.0503903102

Stillman, P. E., Medvedev, D., & Ferguson, M. J. (2017). Resisting temptation: Tracking how self-control conflicts are successfully resolved in real time. *Psychological Science*, *28*, 1240–1258. https://doi.org/10.1177/0956797617705386

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences*, *22*, 531–543. https://doi.org/10.1016/j.tics.2018.03.012

Stolier, R. M., & Freeman, J. B. (2016). Neural pattern similarity reveals the inherent intersection of social categories. *Nature Neuroscience*, *19*, 795–797. https://doi.org/10.1038/nn.4296

Sullivan, N., Hutcherson, C., Harris, A., & Rangel, A. (2015). Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. *Psychological Science*, *26*, 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.

Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the Cognitive Reflection Test. *Cognition*, *150*, 109–118. https://doi.org/10.1016/j.cognition.2016.01.015

Wulff, D. U., Haslbeck, J. M. B., Kieslich, P. J., Henninger, F., & Schulte-Mecklenbeck, M. (in press). Mouse-tracking: Detecting types in movement trajectories. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Eds.), *A handbook of process tracing methods*. New York, NY: Routledge.

Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2019). *Measuring the (dis-)continuous mind: What movement trajectories reveal about cognition*. Manuscript in preparation

Yu, Z., Wang, F., Wang, D., & Bastin, M. (2012). Beyond reaction times: Incorporating mouse-tracking measures into the Implicit Association Test to examine its underlying process. *Social Cognition*, *30*, 289–306. https://doi.org/10.1521/soco.2012.30.3.289