# Advancing Mouse-Tracking Research:

# New Solutions for Study Design, Implementation, and Analysis

Pascal J. Kieslich

Inaugural dissertation

submitted in partial fulfillment of the requirements for the degree Doctor of Social Sciences

in the Graduate School of Economic and Social Sciences (GESS) at the University of Mannheim

*Thesis defense:*

December 7, 2018

*Supervisors:*

Prof. Benjamin E. Hilbig, Ph.D.

Prof. Dr. Arndt Bröder

*Dean of the Faculty of Social Sciences:*

Prof. Dr. Michael Diehl

*Academic Director of the CDSS:*

Prof. Dr. Edgar Erdfelder

*Thesis Reviewers:*

Prof. Dr. Arndt Bröder

Prof. Balazs Aczel, Ph.D.

*Defense Committee:*

Prof. Dr. Edgar Erdfelder

Prof. Dr. Arndt Bröder

Prof. Balazs Aczel, Ph.D.

## Contents

## Abstract

This thesis addresses the topic of mouse-tracking, that is, the recording and analysis of mouse movements in computerized experiments. Mouse-tracking is an increasingly popular process tracing method in many psychological disciplines as it allows capturing the temporal development of the relative attraction to and conflict between response alternatives. It thus provides the opportunity to test psychological theories about factors that influence the conflict involved in making a decision, and how this conflict develops over time. So far, researchers have faced a number of difficulties when conducting mouse-tracking studies: There has been no easy-to-use, flexible and open software for creating experiments and no general-purpose package for analysis. Researchers also have had to make many choices regarding the study setup, with no evidence-based guidelines to support their decisions. This thesis aims to provide solutions for these challenges.

First, this thesis introduces free and open-source software packages for creating and analyzing mouse-tracking experiments. The mousetrap plugin enables researchers to implement mouse-tracking in their experiments without programming and, through integration with the experiment builder OpenSesame, offers a graphical user interface that makes it easy to create a variety of experiments and designs. The mousetrap R package provides extensive functionality for processing, analyzing, and visualizing mouse-tracking raw data of all major formats. It implements most of the commonly used preprocessing procedures and mouse-tracking indices, as well as a set of novel visualization and classification procedures for analyzing trajectory shapes.

Second, this thesis presents results from a series of experiments that investigate how the methodological setup influences mouse-tracking data. In separate experiments, I manipulated the design factors starting procedure, mouse sensitivity, and response indication and investigated their impact on trajectory curvature and shape. An additional study investigated the effects of the starting procedure on movement consistency and also included dynamic analyses. While central cognitive effects on trajectory curvature were replicated in all setups, their size varied considerably between some of the setups. In addition, the setup strongly influenced the trajectory shapes and dynamic analyses. Based on this evidence, I discuss implications for interpreting mouse-tracking data and offer preliminary recommendations for conducting mouse-tracking experiments.

In sum, I hope this thesis will contribute to advancing mouse-tracking research and making the method accessible to a broader audience.

# 1 Articles

This cumulative dissertation is based on two published articles, one book chapter in press, and one manuscript under review. They will be discussed in two sections: The first set of two papers presents mousetrap, a collection of software packages that I developed during my dissertation. The second section focuses on the results from a series of experiments that address the influence of methodological choices regarding the study design on mouse-tracking data.

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods, 49*(5), 1652-1667.

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., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2018). *Design factors in mouse-tracking: What makes a difference?* Manuscript submitted for publication.

Scherbaum, S., & Kieslich, P. J. (2018). Stuck at the starting line: How the starting procedure influences mouse-tracking data. *Behavior Research Methods*, *50*(5), 2097-2110.

During my dissertation, I have also worked on a number of research projects that used mousetrap for running mouse-tracking experiments, analyzing mouse-tracking data, or both. I will refer to them in this dissertation since they represent applications of the software.

Kieslich, P. J., & Hilbig, B. E. (2014). Cognitive conflict in social dilemmas: An analysis of response dynamics. *Judgment and Decision Making, 9*(6), 510-522.

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*(6), 918-926.

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*(1), 23-32.

Heck, D. W., Erdfelder, E., & Kieslich, P. J. (in press). Generalized processing tree models: Jointly modeling discrete and continuous variables. *Psychometrika*.

Horwitz, R., Brockhaus, S., Henninger, F., Kieslich, P. J., Schierholz, M., Keusch, F., & Kreuter, F. (in press). Learning from mouse movements: Improving questionnaire and respondents' user experience through passive data collection. In P. C. Beatty, A. Wilmot, D. Collins, L. Kaye, J. L. Padilla, & G. Willis (Eds.), *Advances in Questionnaire Design, Development, Evaluation and Testing*. New York, NY: Wiley.

One further article presents software that can be used in combination with mousetrap for creating experiments in which participants interact.

Henninger, F., Kieslich, P. J., & Hilbig, B. E. (2017). Psynteract: A flexible, cross-platform, open framework for interactive experiments. *Behavior Research Methods, 49*(5), 1605-1614.

One additional book chapter extends the analysis methods available in mousetrap.

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.

# 2 Introduction

This thesis is dedicated to the topic of mouse-tracking, which is the recording and analysis of mouse cursor movements in computerized experiments (Spivey & Dale, 2006; Spivey, Grosjean, & Knoblich, 2005). As a fairly novel technique in psychological research, it complements the toolbox of existing process tracing methods like think-aloud protocols, information boards or eye-tracking (see Schulte-Mecklenbeck et al., 2017, for an overview of process tracing methods). Going beyond these previous methods, mouse-tracking aims to provide a more direct measure of how an individual's preference for different response alternatives develops over time. As such, mouse-tracking allows for decomposing the cognitive processes underlying outcome-based measures (i.e., the final decision and total response time), does not require the preference development process to be verbalizable (as in think-aloud protocols) and aims to exclude the influence of other processes like information acquisition behavior (which are the main focus in eye-tracking and information boards; see Koop & Johnson, 2011, for a more detailed discussion).

In the following, I will first give an introduction to the mouse-tracking method and its previous applications in psychological research. I will then outline the main motivation for and the contributions of my dissertation. My first goal was to develop flexible, open-source software for creating mouse-tracking experiments, and processing and analyzing the resulting data, thus enabling researchers to easily create and analyze even complex mouse-tracking experiments on their own. Second, the thesis systematically investigated how differences in the methodological setup can influence mouse-tracking data, as a basis for interpreting data from previous mouse-tracking studies that utilized different setups and for formulating guidelines regarding the design of future mouse-tracking studies.

## 2.1 Mouse-tracking

Mouse-tracking is becoming an increasingly popular method in psychological research and has been used to investigate cognitive processes and test psychological theories in a wide range of psychological disciplines (see reviews by Freeman, 2018; Stillman, Shen, & Ferguson, 2018). In the most basic mouse-tracking paradigm, participants make a decision between two alternatives that are presented as buttons in the top left and top right corners of a computer screen (see Figure 1). As they indicate their decision by moving the cursor to either of the buttons, its position is continuously recorded, resulting in a mouse trajectory for every trial (i.e., a time series of x and y coordinates). Mouse-tracking is based on the assumption that cognitive processes are continuously revealed in motor responses (Freeman, Dale, & Farmer, 2011; Spivey & Dale, 2006), building on studies reporting a close link between cognitive and motor processes on a neuronal level (see review by Song & Nakayama, 2009). More specifically, the assumption is made that mouse movements are driven by the relative activations of the response options over the course of the decision process, in that the mouse cursor tends to be moved more toward the option with the higher activation at any point in time (Spivey, Dale, Knoblich, & Grosjean, 2010). Accordingly, mouse-tracking can be used to infer the degree of conflict between the response options during the decision process, with trajectories that deviate more towards the ultimately non-chosen option indicating greater amounts of conflict.

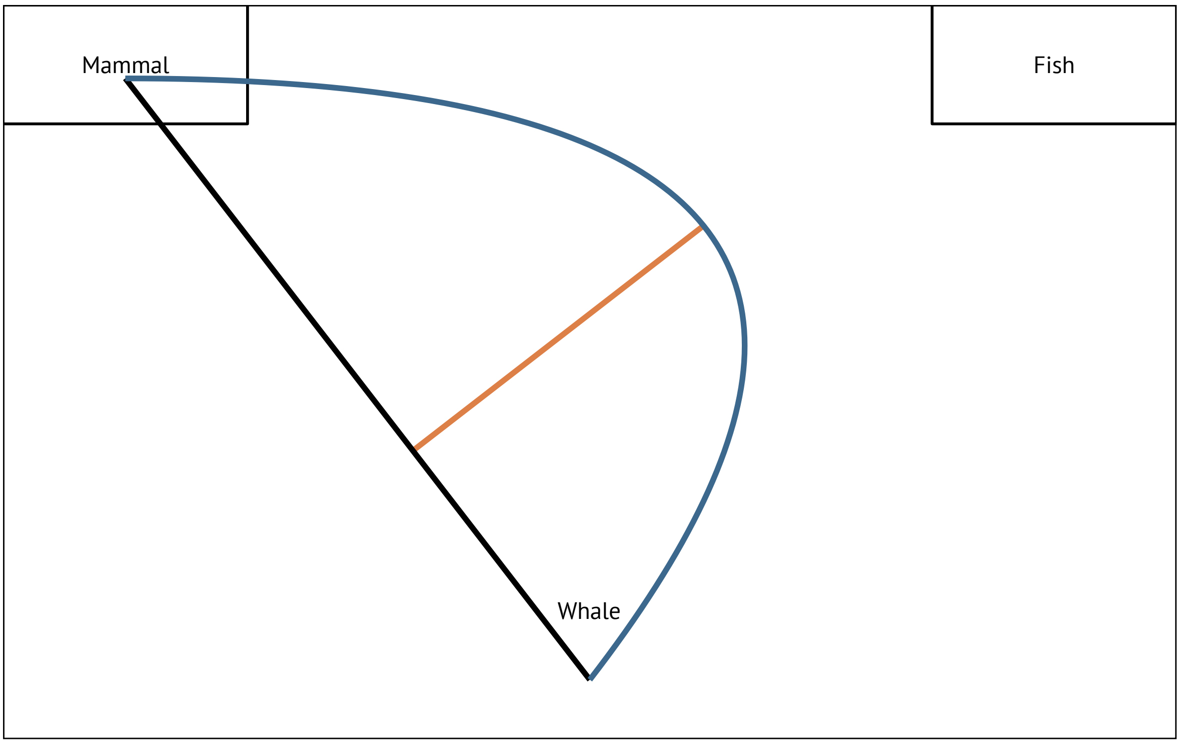


*Figure 1.* Setup of the experiment by Dale, Kehoe, and Spivey (2007), including a simulated cursor trajectory (in blue). The trial is initiated by clicking on a start button in the bottom center of the screen (not shown) after which the name of the to-be-classified animal is presented and mouse-tracking starts. Participants indicate their classification decision by clicking on one of two response buttons. For the example trajectory, the maximum absolute deviation (MAD) is depicted (in orange) as the maximum deviation from the direct path between its start and end point.

Mouse-tracking has thus provided two major opportunities for testing psychological theories (Freeman, 2018; Stillman et al., 2018): First, by providing indices of the overall amount of conflict between response options, mouse-tracking allows for testing the effects of individual differences and contextual factors that are theorized to influence the conflict involved in making a decision. Second, mouse-tracking enables researchers to assess the development and resolution of this conflict over the course of the decision process, making it possible to test temporal predictions of theories about how decisions and judgments unfold.

To illustrate a typical mouse-tracking study, consider the following experiment by Dale, Kehoe, and Spivey (2007), which I will use as an example throughout this dissertation. The authors conducted an experiment in which, at the start of each trial, participants were presented with two animal categories as the response alternatives on their screen (e.g., mammal and fish). Next, participants clicked on a start button and the name of an animal appeared, which participants had to assign to the correct category by clicking on the corresponding button (see Figure 1 for the visual setup). The experiment included names of typical animals (e.g., dog for mammal) and atypical animals (e.g., whale), the latter of which share features both with the correct (mammal) and incorrect category (fish). The authors' main hypothesis regarding the decision process was that atypical exemplars would activate both response options to some degree, whereas typical animals would mostly activate the correct category. Given the assumptions above, this typicality effect should be visible in the mouse trajectories, in that trajectories for atypical animals should deviate more towards the incorrect response option. This example study incorporates many features of a typical mouse-tracking experiment. It includes simple stimuli (i.e., single words), which participants have to assign to one of two alternatives in a forced-choice design, where one option represents the correct response. The spatial setup of the study is also quite common, with the two response buttons presented in the top left and right screen corners and a start button in the bottom center of the screen, that participants click to start the trial screen (in order to align the initial cursor position across trials). Finally, mouse-tracking studies often manipulate a within-participants factor (e.g., in this study, the typicality of the stimulus) with a directed hypothesis regarding its influence on mouse trajectories.

While mouse-tracking was first applied in the domain of language processing (Spivey et al., 2005; Dale et al., 2007), it has since spread to many other areas of psychological research, and studies using the method have covered a broad range of psychological topics. To provide a selection of subfields, research using mouse-tracking spans studies on social cognition (e.g., Freeman & Ambady, 2009; Freeman, Ambady, Rule, & Johnson, 2008), action control (e.g., Scherbaum, Dshemuchadse, Fischer, & Goschke, 2010), numerical cognition (see review by Faulkenberry, Witte, & Hartmann, 2018), perception (e.g., 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), judgmental biases (e.g., Szaszi, Palfi, Szollosi, Kieslich, & Aczel, 2018; Travers, Rolison, & Feeney, 2016), and self-control (e.g., Stillman, Medvedev, & Ferguson, 2017; Sullivan, Hutcherson, Harris, & Rangel, 2015). These and many further studies have recently been summarized in two reviews on mouse-tracking (Freeman, 2018; Stillman et al., 2018).

### 2.2 Software

Researchers aiming to conduct a mouse-tracking study cannot usually rely on standard experimental software packages without further modifications, as most do not support the continuous recording of cursor movements out of the box. In addition, mouse-tracking raw data is more complex than data collected in standard psychological experiments, and requires a number of preprocessing operations, which are usually not covered by general-purpose statistical software. Therefore, researchers have to rely on specific approaches for creating mouse-tracking experiments and analyzing the resulting data. In the following, I will give a brief overview of the two different approaches that researchers have previously used in this regard. Overcoming the limitations of these approaches, I will outline how the software I developed in my dissertation provides a third option.

As a first approach, the pioneers of mouse-tracking, and several research groups since, built their own mouse-tracking implementation from scratch in the experimental software or programming environment of their choice (e.g., Koop & Johnson, 2011; Scherbaum et al., 2010; Spivey et al., 2005). This approach allowed them to tailor experiments to their specific needs, but also requires thorough knowledge of the experimental software as well as programming skills. In addition, researchers afterwards had to write their own scripts for preprocessing the collected tracking data – requiring yet more advanced programming skills and considerable effort. In both cases, these individual implementations were often limited to a specific paradigm and analysis approach, and could not easily be adapted to other projects.

As a second approach, researchers have used *MouseTracker* (Freeman & Ambady, 2010), a software package dedicated specifically to creating mouse-tracking experiments. It provides a graphical user interface for designing the mouse-tracking screen, while the experimental procedure and material is specified in a csv-file. MouseTracker has played an important role in making the mouse-tracking method accessible to researchers, allowing them to quickly build simple mouse-tracking experiments without programming. However, MouseTracker also imposes significant constraints on the studies created with it, as it only offers a limited set of options for fine-tuning experiments and lacks a scripting language that would allow further customization and the implementation of complex experimental designs. For data processing and analysis, MouseTracker supports researchers by visualizing the trajectory data, and providing a built-in set of basic preprocessing steps and standard trial-level indicators. However as with the study design, the fixed and limited set of analysis options renders MouseTracker relatively inflexible. It cannot perform statistical tests, thus requiring that the preprocessed data be imported into another software package for further analysis and statistical inference. In addition, the software only processes and analyzes data that it collected itself, restricting its potential as a general-purpose analysis tool. Finally, while MouseTracker is available free of charge, its source code is not openly available for inspection and extension, and it can only be used on Windows systems.

In sum, researchers conducting mouse-tracking studies have, for the most part, been limited to a choice between (a) programming their entire experiment and analysis from scratch, and, (b) using the stand-alone software MouseTracker, which makes conducting basic mouse-tracking studies easy, but cannot easily accommodate more complex tasks and analyses. To address these limitations, together with colleagues I have taken a new approach during my dissertation by extending general-purpose experimental and statistical software to implement mouse-tracking data collection and analysis (Kieslich & Henninger, 2017; Kieslich, Henninger, Wulff, Haslbeck, & Schulte-Mecklenbeck, in press). This combines the ease-of-use provided by existing software that researchers are already familiar with, and the flexibility of open-source tools that can be extended and adapted.

For creating mouse-tracking experiments, we have developed the *mousetrap plugin* (Kieslich & Henninger, 2017) for the experiment builder *OpenSesame* (Mathôt, Schreij, & Theeuwes, 2012). It allows researchers to easily create mouse-tracking experiments via a graphical user interface that does not require programming skills. Through the integration with a general-purpose experiment software, it is very flexible and supports a diverse range of designs. Complex tasks as well as advanced features like on-the-fly stimulus generation and feedback can be realized using the underlying programming language Python. OpenSesame also allows for the combination of mouse-tracking with its other features and additional plugins, including the integration of external hardware (such as eye-tracking).

For processing, analyzing, and visualizing mouse-tracking data, we have developed the *mousetrap* package (Kieslich et al., in press) for the statistical programming language *R* (R Core Team, 2018). This package covers the entire process from importing and preprocessing mouse-tracking raw data to the computation of many established measures, as well as the visualization of individual and aggregate mouse trajectories. Mousetrap supports raw data in a variety of formats, including data collected using MouseTracker and the mousetrap plugin for OpenSesame. Since the mousetrap package is integrated into the statistical programming language R, researchers can draw upon the many available packages for descriptive statistics, inferential tests, and general visualizations, so that they can perform the complete data preparation and analysis process in one environment.

All software I present is open-source and available free of charge for all major platforms (Windows, Linux, and MacOS). This is of great importance in light of the current focus on replication and open science (e.g., Asendorpf et al., 2013; Munafò et al., 2017; Nosek et al., 2015). It allows researchers to share their material and experiments with other researchers, who can directly run them in their own labs, making it easy to perform both direct replications and extensions of previous experiments. Similarly, researchers can easily share their raw data alongside the scripts used for preprocessing and analysis, which other researchers can immediately reproduce themselves, providing full transparency of the analysis process as well as enabling new analyses of previously published data. Finally, since the complete source code of the software is open-source, advanced users can audit and adapt every implemented feature, be it a specific preprocessing operation or the computation of particular mouse-tracking indices.

### 2.3 Design factors

Given the relative novelty of the mouse-tracking method, no standards for conducting mouse-tracking experiments have been established so far. As a result, the methodological setup has varied considerably in previous mouse-tracking studies. For example, in some studies participants indicated their response by clicking on the corresponding button (e.g., Dale et al., 2007; Freeman et al., 2008; Koop & Johnson, 2013) while in other studies they would simply move the cursor onto the button without clicking (e.g., Dshemuchadse et al., 2013; Huette & McMurray, 2010). Similarly, some studies have left the mouse sensitivity settings at the system defaults (resulting in a medium cursor speed, e.g., Freeman, 2014; Kieslich & Hilbig, 2014; Szaszi et al., 2018), while others deliberately reduced cursor speed and disabled acceleration (Dshemuchadse et al., 2013; Frisch, Dshemuchadse, Görner, Goschke, & Scherbaum, 2015; Scherbaum et al., 2010). The starting procedure has also varied between studies, with some studies giving no instructions to participants regarding the initiation of the mouse movement (e.g., Dale et al., 2007; Kieslich & Hilbig, 2014; Koop, 2013), some explicitly instructing participants to start moving early in the trial (e.g., Freeman & Ambady, 2009; Papesh & Goldinger, 2012; Yu, Wang, Wang, & Bastin, 2012), and others enforcing an initial movement by hiding the critical stimulus until participants moved the cursor upwards (e.g., Frisch et al., 2015; Huette & McMurray, 2010; Scherbaum et al., 2010).

The previous examples illustrate a number of design factors that have varied across previous mouse-tracking studies, the influence of which has not been investigated so far. This also means that researchers creating mouse-tracking experiments face a number of design choices without having empirical data available that could guide their decisions. Nevertheless, some researchers have provided recommendations regarding the basic setup of mouse-tracking studies (Fischer & Hartmann, 2014; Hehman, Stolier, & Freeman, 2015). They suggest that researchers should employ a starting procedure that increases the likelihood that participants start their mouse movement early in the trial, to ensure that cognitive processing takes place during the movement and not before. Regarding the mouse sensitivity settings, Fischer and Hartmann (2014) recommend to reduce cursor speed and turn off acceleration to better capture cognitive effects in the mouse trajectories, as participants have to move the hand smoothly across a greater distance. However, Hehman et al. (2015) note that “these approaches have not been empirically validated, and instead are derived from our previous experience” (p. 388). In other words, while some basic recommendations for the setup of studies exist, they are based largely on theoretical considerations and “lab-lore” without a systematic empirical foundation.

It is, in my view, of critical importance to investigate and understand the impact of the methodological setup on mouse-tracking studies. If design factors influence mouse-tracking data, this may impact the theoretical conclusions that can be drawn from mouse-tracking studies. For example, as will be discussed later, many studies have used the shape of the mouse trajectories to determine which theoretical model accounts best for a particular cognitive process; if this shape does not only depend on the cognitive process, but also on the methodological setup, this considerably limits the conclusions that can be drawn solely based on the shape. Furthermore, future studies could draw on this knowledge to optimize their paradigm, so that the mouse-tracking procedure becomes maximally informative regarding the processes it is intended to measure. Therefore, I conducted a series of experiments as part of my dissertation to systematically investigate how design factors influence mouse-tracking data. Specifically, using the previously described paradigm by Dale et al. (2007) across a series of experiments, we investigated the impact of the starting procedure, mouse sensitivity settings, and response indication on common mouse-tracking measures (Kieslich, Schoemann, Grage, Hepp, & Scherbaum, 2018). In another study, we investigated the influence of the starting procedure on both typical mouse-tracking measures as well as more complex dynamic analyses, using a Simon task (Scherbaum & Kieslich, 2018). Before turning to these studies of the specific methodological setup, I will outline how mouse-tracking experiments can be created in general.

## 3 Mouse-Tracking Software

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods, 49*(5), 1652-1667.

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.

In this section, I will give an overview of the software I developed as part of my dissertation. This includes the mousetrap plugin for OpenSesame, which is used to create mouse-tracking experiments, and the mousetrap R package for analyzing mouse-tracking data. Kieslich and Henninger (2017) present a first version of the mousetrap plugin along with a technical validation, an example mouse-tracking experiment (replicating Experiment 1 by Dale et al., 2007), and a short demonstration of a simple analysis using the mousetrap R package. Kieslich et al. (in press) provide an updated version of the mousetrap plugin as well as a more in-depth tutorial covering the mousetrap R package. In the following, I will briefly summarize both.

### 3.1 Mousetrap plugin for OpenSesame

The first software I present is dedicated to the creation of mouse-tracking experiments by extending the open-source software OpenSesame. *OpenSesame* is a general-purpose experiment builder that can be used for creating experiments via a graphical user interface (Mathôt et al., 2012). Different functionality is implemented in OpenSesame via different items that are responsible for stimulus display, response collection, and other features. A complete experiment is created by combining the different items and arranging them in a temporal order. Complex procedures can be implemented by including Python code at any point in the experiment. OpenSesame can be obtained for free for all major platforms from https://osdoc.cogsci.nl/, where it is also documented in depth. It can also be extended by third-party plugins to implement additional functionality.

OpenSesame by default collects key presses and mouse clicks. We have extended it by implementing mouse-tracking functionality through the *mousetrap* plugin (Kieslich & Henninger, 2017). Mousetrap integrates into OpenSesame's graphical interface, providing items that allow researchers to track participants' mouse movements; it also implements many additional features that are commonly used in mouse-tracking experiments. Mousetrap can be obtained along with extensive documentation and examples from https://github.com/pascalkieslich/mousetrap-os.

Mousetrap offers two options for implementing mouse-tracking, reflecting the two major ways researchers create visual stimuli in OpenSesame (Kieslich & Henninger, 2017). The easiest option is to create the stimulus display via OpenSesame's graphical user interface (Figure 2) and use a *mousetrap_response* item to track participants' mouse movements (Figure 3). This way, researchers can create a mouse-tracking experiment without writing any code. Alternatively, a *mousetrap_form* item can be employed to create the stimulus display via a simple script syntax and implement mouse-tracking within the same item.

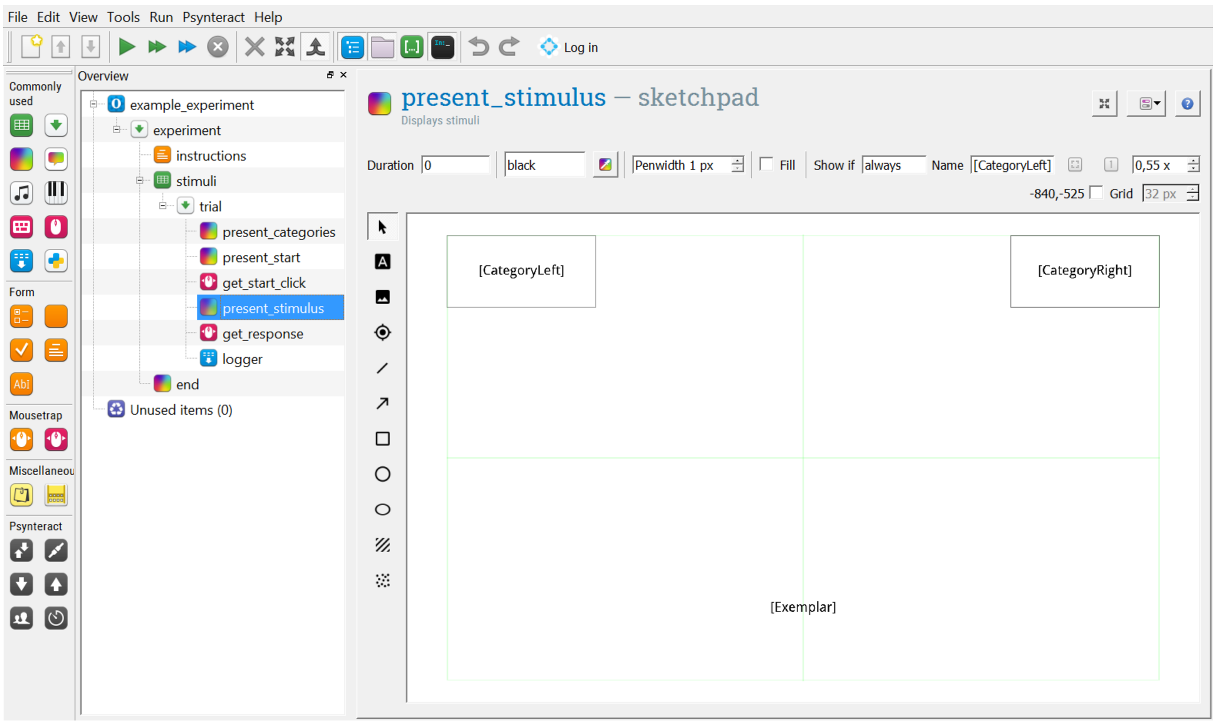


*Figure 2.* User interface of OpenSesame. The leftmost panel shows the item toolbar containing the items that can be included in an experiment (including the mousetrap items). Next to it, the overview area represents the study's structure. The rightmost panel shows the interface of the sketchpad item that is used to create the main stimulus display (resulting in the setup displayed in Figure 1). Variable names are placed in square brackets, so that their values will be substituted when the experiment is run.

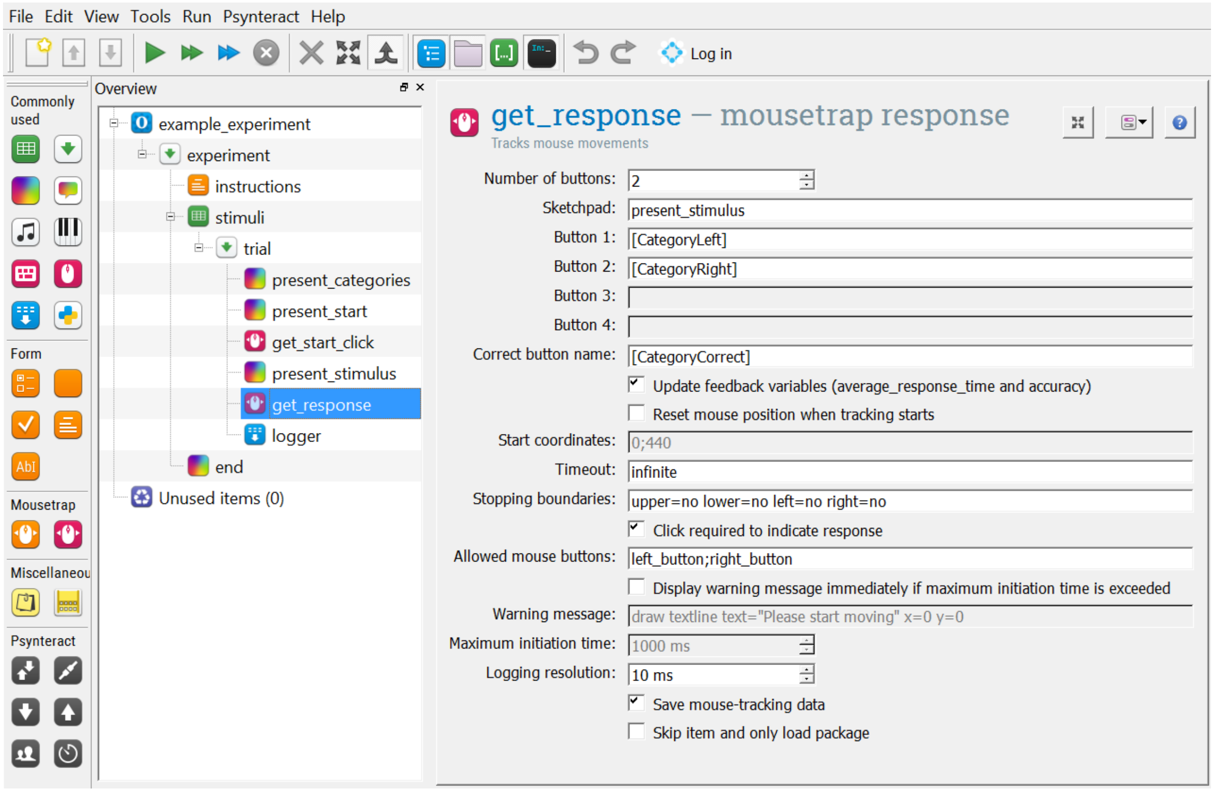


*Figure 3.* User interface of the *mousetrap_response* item (in the right panel). The topmost settings are essential for implementing the mouse-tracking procedure and include the number of buttons and the name of the buttons (and the sketchpad item that displays them, see Figure 2). The settings below provide many options for fine-tuning the mouse-tracking procedure and giving feedback to participants. The mouse-tracking related options include, among other things, resetting the cursor to exact start coordinates at tracking onset, limiting the time participants have to give their answer, and specifying whether participants can indicate their response by clicking or merely moving the cursor onto the button.

The mousetrap plugin implements many options that are commonly used in mouse-tracking studies (Figure 3). For instance, researchers can specify that the cursor position should be reset to specific screen coordinates at tracking onset, which simplifies later analyses. Mousetrap also supports different starting procedures – an issue that will be discussed in more detail in the design factors section. Besides, participants can give their response by clicking or by simply moving the cursor onto one of the buttons. It is also possible to limit the total time participants have for giving a response. Moreover, mousetrap can automatically code the correctness of a response, which can be used to provide feedback to participants or to pay them contingent on their performance. Finally, the plugin automatically computes the time participants took to initiate any movement in a trial and the total response time, both of which can be used during the experiment to display a warning message if a predefined threshold is exceeded, in order to encourage early movement initiation and fast responding.

Through integration with OpenSesame, mousetrap allows researchers to create a vast range of different designs, since the underlying experimental software supports many different types of visual and auditory stimuli and implements many common randomization schemes. Even where the interface reaches its limits, almost any conceivable study can be implemented by including Python code, such as complex feedback and incentivization schemes, or the random creation of visual stimuli for each participant in real-time. Because mousetrap can be combined with other plugins, mouse-tracking can be used in yet more complex experiments. This includes the combination of mouse- and eye-tracking (e.g., Koop & Johnson, 2013, Experiment 3), which can be realized by combining the mousetrap and the PyGaze plugins (Dalmaijer, Mathôt, & Van der Stigchel, 2014), as well as mouse-tracking experiments involving real-time interactions between participants (such as social dilemmas, as in Kieslich & Hilbig, 2014), which have been implemented by combining the mousetrap and the psynteract plugin (Henninger, Kieslich, & Hilbig, 2017).

To ensure that the mousetrap plugin reliably records the cursor position, we conducted a technical validation using external hardware to generate synthetic cursor movements (see Kieslich & Henninger, 2017, Appendix). Virtually all changes in cursor position were captured by the mousetrap plugin and the recorded positions corresponded to their expected values in almost every case. To demonstrate the practical utility of the plugin, we conducted a replication of the previously described experiment by Dale et al. (2007) with 60 participants (see Kieslich & Henninger, 2017, Example experiment). The experiment, raw data, analyses, and results are provided at https://github.com/pascalkieslich/mousetrap-resources. The data from this experiment will be used in the following section to demonstrate mouse-tracking analyses using the mousetrap R package.

### 3.2 Mousetrap R package

To analyze mouse-tracking data, we have developed the *mousetrap* R package (Kieslich et al., in press). *R* is an open-source programming language for statistical analysis (R Core Team, 2018) that has become popular in many scientific disciplines (Tippmann, 2015) and that is freely available from https://www.r-project.org/. Mousetrap can be installed in R with a single command; installation instructions, extensive documentation and examples can be found at http://pascalkieslich.github.io/mousetrap/. Once installed, mousetrap covers the complete process from raw data import and preprocessing to the computation of many established mouse-tracking measures and (if desired) their aggregation, after which standard R functions and packages can be used to perform statistical analyses. Besides, mousetrap provides novel functionality for the advanced visualization of mouse trajectories (cf. Kieslich et al., in press) as well as offering spatial clustering and classification procedures for identifying groups of similar trajectories (cf. Wulff, Haslbeck, Kieslich, Henninger, & Schulte-Mecklenbeck, in press). In the following, I will discuss mousetrap's most important functions. A tutorial for using the package can be found in Kieslich et al. (in press).

Before running any analyses, users have to load the collected data into R. Depending on the data format, they may use one of R's standard functions. For the two software packages discussed earlier (mousetrap plugin for OpenSesame and MouseTracker), we have developed specific functions that automate and thereby simplify this step.[1] After reading the data into R, users need to import the data into the mousetrap package. Mousetrap offers different import functions depending on the type of format that was used to store the collected positions (e.g., long or wide), covering all major formats that are used in existing software packages.

To enable the comparison of trajectories across trials, a set of preprocessing steps are necessary, in particular spatial transformation and resampling (cf. Freeman & Ambady, 2010; Hehman et al., 2015; Kieslich et al., in press). Mousetrap covers most of the preprocessing operations from the literature and provides them as easy to use functions. Spatial transformations include the *remapping* of trajectories so that the overall direction of all trajectories is the same (e.g., in the example experiment all trajectories should end on the left option regardless of which option was

[1] We have also developed the readbulk R package (Kieslich & Henninger, 2016) that allows researchers to merge data files (of any format) of individual participants into one large dataset.

chosen) as well as the *alignment* of trajectories to a constant start position across trials (which is relevant if the cursor position was not reset at tracking onset in the experimental software). Most studies also perform one of several resampling operations with the aim to represent each trajectory with the same, fixed number of positions, regardless of the number of recorded positions for the raw trajectory. This includes *time-normalizing* trajectories so that the time interval between adjacent positions within every resampled trajectory remains constant (Spivey et al., 2005). This procedure has been used in many studies to date, especially when creating plots of aggregate trajectories. Alternatively, trajectories can also be *spatially normalized* to a constant distance between adjacent positions in a trial (Wulff et al., in press). While the former procedure ensures that the temporal development of the raw trajectory is preserved, the latter procedure emphasizes the trajectory shape, which can be useful if this is the main focus of the analysis (Wulff et al., in press). An important feature of mousetrap is that it stores both the original and the different preprocessed trajectories and allows researchers to specify explicitly which of these are used in any particular analysis, making it possible to explore the consequences of different preprocessing steps and to use different types of trajectories in different analyses and visualizations.

The final step in preprocessing is to condense each trajectory into one numeric value that represents a specific property of the trajectory, with the goal of capturing a particular aspect of the cognitive process. Mousetrap calculates a multitude of different mouse-tracking indices (see Kieslich et al., in press, for an overview of the measures). The most frequently used class of indices quantifies the *curvature* of a trajectory, which aims to assess the degree of conflict present in a trial. The idea behind this is that larger deviations towards the non-chosen alternative indicate greater conflict between the two options. A number of different indices have been suggested to quantify curvature; they differ in their exact computation, but are often highly correlated in practice (Stillman et al., 2017). One common measure is the maximum absolute deviation (MAD) illustrated in Figure 1. In addition to curvature, mouse-tracking studies have also assessed the *complexity* of the movement as an indicator of response competition and uncertainty. Indices for complexity are x-flips (the number of directional changes along the horizontal axis) or sample entropy (Hehman et al., 2015). In a similar vein, other studies have tried to infer how often participants change their mind, using measures like x-reversals (the number of crossings of the y axis, cf. Koop & Johnson, 2013) or based on areas of interest (specifying areas of interest around each response button and counting how often participants move between them, cf. Szaszi et al., 2018). In addition, there are a number of indices related to time, velocity, and acceleration, like movement initiation time or maximum velocity (see Freeman & Ambady, 2010; Hehman et al., 2015; Kieslich et al., in press; Koop & Johnson, 2011).

The core analysis in many mouse-tracking studies compares a specific mouse-tracking index between different levels of an independent variable (e.g., different experimental conditions). For instance, the central hypothesis in the example experiment is that mouse trajectories deviate more towards the non-chosen option in atypical than typical trials (in the standard analysis focusing only on correctly answered trials). Consequently, in the replication experiment we tested whether the MAD is larger in atypical than typical trials, which was the case (Kieslich & Henninger, 2017). Many mouse-tracking studies test a hypothesis like this using values that are aggregated per participant and condition. Mousetrap provides basic aggregation functions for this purpose, and general-purpose R functions can then be used for the statistical test (like *t*-tests, ANOVAs, or linear models in a frequentist or Bayesian implementation). Alternatively, analyses can be performed directly on trial-level data using (generalized) linear-mixed models, for which the aggregation step can be skipped.

Instead of analyzing one index value per trial, mouse-tracking studies have also looked at the development of a particular characteristic of the mouse trajectory over time, such as the temporal development of the horizontal position, velocity, or movement angle (cf. Hehman et al., 2015; Scherbaum et al., 2010). This approach provides insights into the temporal dynamics of the cognitive processes, for example, the temporal order in which different psychological factors exert their influence. Mousetrap can also calculate a number of variables for every time point in the trial (such as the velocity, acceleration, and movement angle) and, if desired, can also aggregate these values across trials for different conditions. Afterwards, standard R functions like (generalized) linear-mixed models can be used for further analysis.

Another focus in mouse-tracking analyses has been the qualitative *shape* of the mouse trajectories. The main motivation for this is that different types of movement trajectories would be expected for different types of theoretical models; specifically, researchers are often interested in disentangling whether a dynamic or a dual-system model accounts best for the data (Freeman & Dale, 2013). Dynamic models assume a continuous competition of the response options that is gradually resolved over time and, consequently, would expect continuously curved trajectories in all trials. Dual-system models assume two distinct systems that can drive a decision and therefore predict a mixture of distinct trajectory shapes with either little or no conflict (where both systems agree) or a visible change in preference, when one option is initially favored by system I (which provides a quick evaluation), followed by a *change of mind* driven by system II (whose late evaluation favors the other option and eventually dominates). This would lead to a mixture of relatively straight trajectories and change of mind trajectories that first approach the non-chosen option before moving to the chosen option.

Different graphical and numerical methods have been proposed to assess the degree to which these and additional types of trajectories are present in the data (Freeman & Dale, 2013; Kieslich et al., in press; Wulff et al., in press), the majority of which are implemented in the mousetrap package. The simplest graphical method is to plot the *aggregate trajectories* – usually based on time-normalized trajectories, aggregated separately per experimental condition. As can be seen in Figure 4 for the example study, the aggregate trajectory for atypical trials displays greater attraction to the non-chosen option than the aggregate trajectory for typical trials. Both aggregate trajectories look curved, which would be more in line with a dynamic than a dual-systems model.

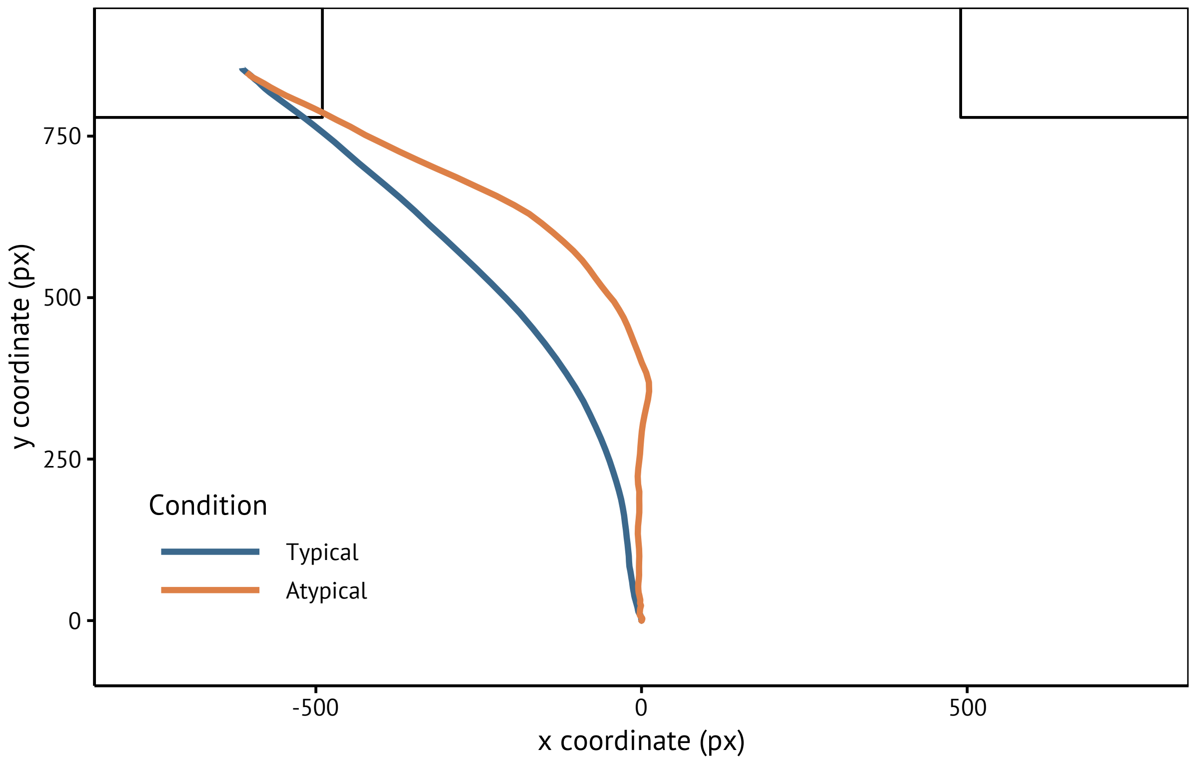


*Figure 4.* Aggregate time-normalized trajectories per typicality condition (including boxes representing response buttons). After excluding incorrectly answered trials, trajectories were first aligned to a common start position, remapped to the left, and then aggregated first within and then across participants.

Of course, aggregate trajectories are not necessarily representative of individual trajectories, which is why analyses and visualizations should also take into account the individual trajectories directly (Kieslich et al., in press; Wulff et al., in press). For this purpose, mousetrap provides *heatmap* functions, which plot individual trajectories and can apply varying degrees of smoothing. As can be seen in Figure 5 (top and middle panel), the individual trajectories in the example experiment are not represented well by the aggregate trajectories, because there appear to be different types of movement trajectories at the trial level. The majority of trajectories move fairly straight from the start button to the chosen option, some trajectories are visibly curved, and several trajectories display extreme curvature or even a discrete change in direction, moving first all the way to the non-chosen alternative and horizontally from there to the chosen option. This latter trajectory in particular might suggest a discrete change of mind. Thus, unlike the aggregate trajectories, many individual trajectories are not in line with the assumption of continuous competition between the response options. Whether this means that they support a dual process model of choice is a question that cannot be definitely answered since, as will be discussed in the next section, the methodological setup of the study also has a strong influence on trajectory shapes. Finally, with regard to the effect of the experimental manipulation, a plot of the difference in densities between conditions (Figure 5, bottom) reveals that change of mind trajectories in particular occur more frequently in the atypical condition.

In order to identify different types of trajectories via numerical methods, previous studies have often relied on *bimodality analyses* of curvature indices (Freeman & Ambady, 2010; Freeman & Dale, 2013). The idea behind this approach is that a mixture of straight and extremely curved (or even discrete) change of mind trajectories should lead to a bimodal distribution of curvature indices, whereas a unimodal distribution would be expected if all trajectories are of the same type, for example, if all trajectories reflected some degree of continuous competition. Currently, there are two predominant bimodality analysis approaches, both of which are implemented in mousetrap. First, researchers can compute the bimodality coefficient (SAS Institute Inc, 1990), which classifies a distribution as bimodal if it exceeds a threshold derived from simulation studies (Pfister, Schwarz, Janczyk, Dale, & Freeman, 2013). Second, researchers can perform an inferential test using Hartigan's dip statistic (Hartigan & Hartigan, 1985), which tests if the distribution deviates significantly from unimodality.

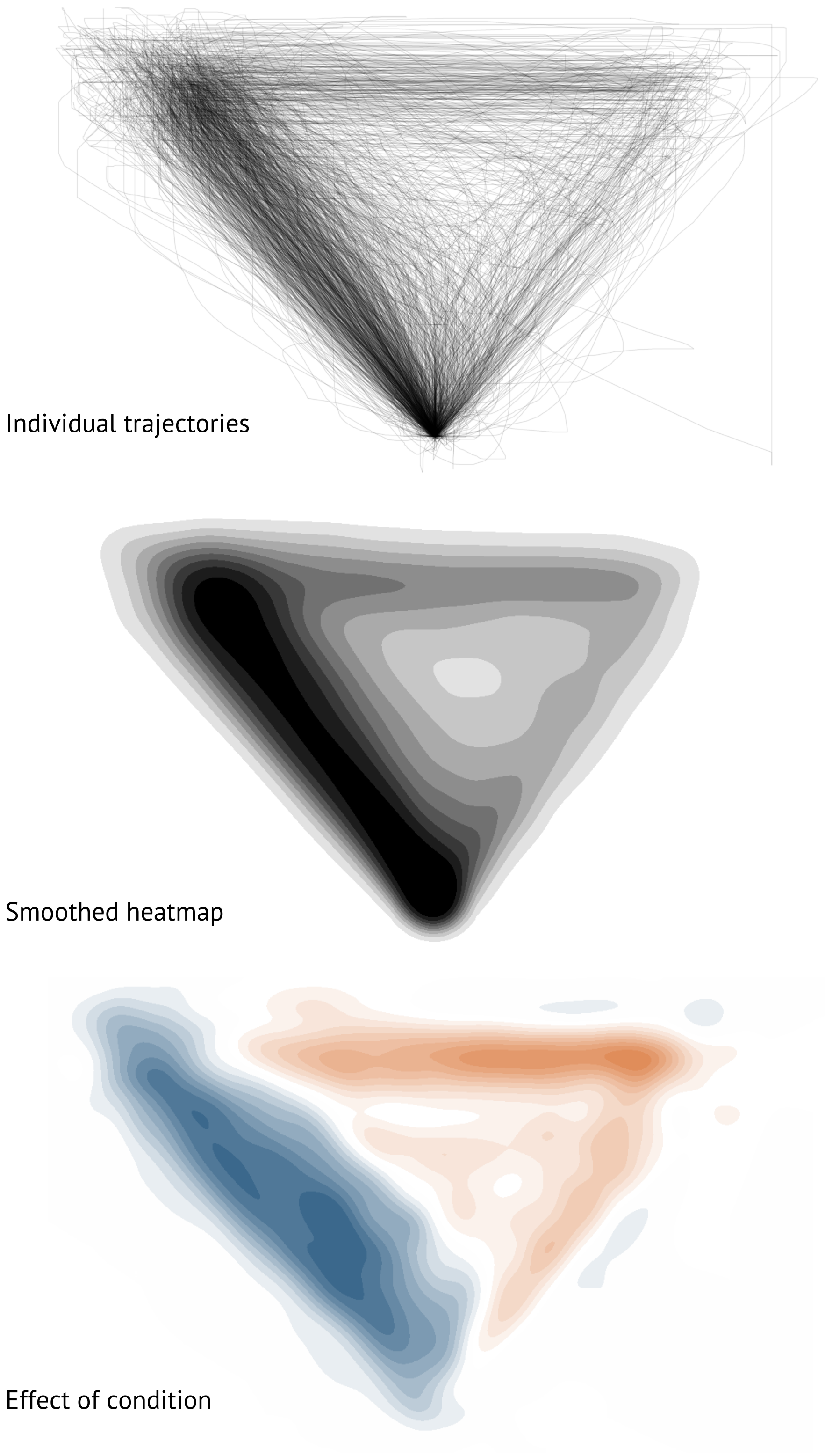


*Figure 5.* Heatmap of the (remapped) individual trajectories (top panel), smoothed heatmap indicating the density of trajectories at every point on the screen (middle panel) and difference of smoothed heatmaps between conditions (bottom panel), where blue indicates higher density in the typical and orange higher density in the atypical condition (white indicates comparable density).

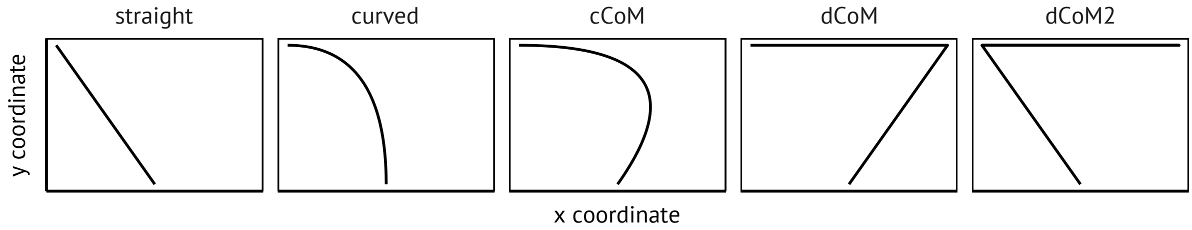


*Figure 6.* Set of prototype trajectories included in the mousetrap R package. These comprise straight and curved trajectories, as well as three types of change of mind trajectories: Curved change of mind (cCoM) trajectories show an initial attraction towards the non-chosen option while maintaining a smooth curvature. Discrete change of mind (dCoM) trajectories first move straight to the non-chosen option and horizontally from there to the chosen option. Double change of mind (dCoM2) trajectories first move to the chosen option, horizontally from there to the non-chosen option and then back to the chosen option.

However, my colleagues and I have recently argued that, to identify different types of trajectories, analyses should not be based on a single numeric value per trial, but should instead take into account the complete trajectory shape (Wulff et al., in press; Wulff, Haslbeck, & Schulte-Mecklenbeck, 2018). Mousetrap provides two approaches that fulfill this requirement (cf. Wulff et al., in press)[2]: First, trajectories can be grouped into a predefined number of *clusters* based on their spatial similarity. Second, trajectories can be assigned to one of several *prototype* trajectories based on their shape, that is, they are assigned to the prototype to which they have the smallest spatial distance. Mousetrap includes a default set of five commonly occurring prototype trajectories (Figure 6) based on a meta-analysis by Wulff et al. (2018).[3] Figure 7 shows the frequency with which trajectories from the example study were assigned to each of these prototypes, and enables a visual assessment of the degree of fit of each trajectory to its assigned prototype.[4] The plots for each

[2] These approaches were developed by Dirk Wulff, Jonas Haslbeck, and Michael Schulte-Mecklenbeck, and contributed by Dirk Wulff to the mousetrap R package with my assistance. They are not discussed in the methodological papers that make up this dissertation (they are introduced in a second book chapter based on our collaboration, cf. Wulff et al., in press). The prototype assignment method is briefly described and used in one of the design factor papers in this dissertation (Kieslich et al., 2018).

[3] Other prototypes are of course possible, and the suitable prototypes often depend on the methodological setup of the study. Mousetrap makes it easy to modify and extend the set of prototypes if needed.

[4] Mousetrap also quantifies the spatial distance of each trajectory from its prototype.

category show that most trajectories are well described by the prototypes, indicating that the prototypes capture the core features of each trajectory concisely.[5] The distribution of classifications corroborates the impression from the heatmaps (Figure 5) that different types of trajectories are present in the data – including straight, curved and change of mind trajectories.

If different types of trajectories are present in the data, the prototype classification itself can also be used as a dependent variable indicating the degree of conflict in a trial (assuming that conflict increases with the degree of curvature and the number of horizontal reversals for the discrete change of mind prototypes, i.e., in Figure 6 the inferred conflict would increase from left to right). In line with the idea that atypical exemplars should produce more conflict, classifications indicating a higher degree of conflict are more frequent in atypical than in typical trials (Figure 7; this can be tested using an ordinal mixed regression at the trial level).

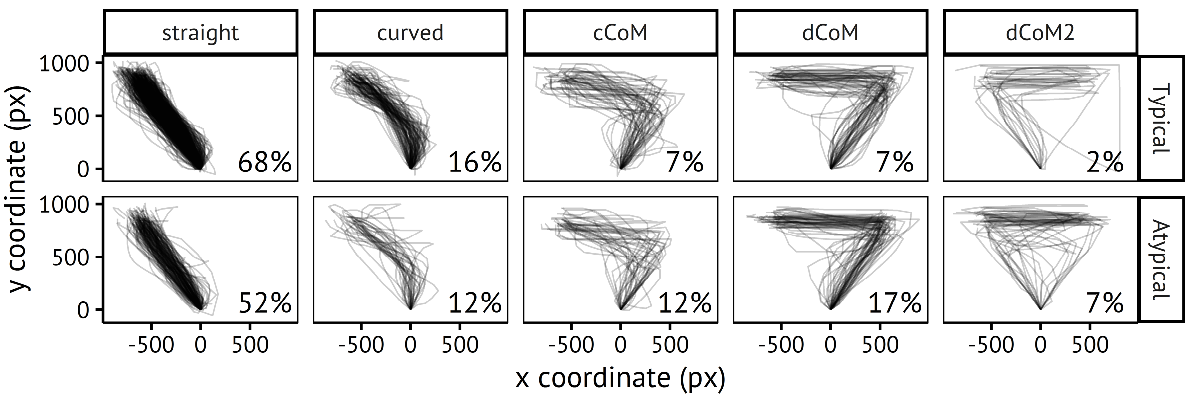


*Figure 7*. Individual trajectories per assigned prototype separately for each typicality condition. For each prototype, the relative frequency of classifications per typicality condition is displayed.

[5] A minority of trajectories assigned to the dCoM and dCoM2 prototypes are not well represented by their prototype. This is likely due to the fact that in these trajectories participants switch horizontally between the options multiple times indicating multiple changes of mind. As a solution, dCoM3 and dCoM4 prototypes can be included that extrapolate the logic of the dCoM2 prototype by including additional horizontal switches. Alternatively, these trajectories could be excluded.

In sum, the mousetrap package implements a multitude of both established and upcoming procedures for analyzing and visualizing mouse-tracking data. The package includes extensive documentation and examples, making these fairly complex analyses accessible to users of all levels of experience with R. A tutorial into using the package can be found in Kieslich et al. (in press). Figure 8 provides an exemplary R script that replicates the main processing steps, analyses, and visualizations that were presented in this section.

### 3.3 Usage

Since their first release in spring 2016, the mousetrap software packages have been used by many researchers from different institutions and research domains. The mousetrap R package has been downloaded 7342 times[6], and the mousetrap plugin more than 4400 times[7]. There are already a number of published research projects that relied on mousetrap for creating experiments, analyzing mouse-tracking data, or both (e.g., Aczel, Szollosi, Palfi, Szaszi, & Kieslich, 2018; Calcagnì, Lombardi, & Sulpizio, 2017; Horwitz et al., in press; Leontyev, Sun, Wolfe, & Yamauchi, 2018; Scherbaum & Kieslich, 2018; Schulz, Speekenbrink, & Krause, 2018; Szaszi et al., 2018).

The increasing popularity of the mousetrap packages has resulted in many questions and discussions around the creation of mouse-tracking experiments and the analysis of the resulting data, which have taken place via email, at workshops and conferences. To offer a central place for support, we have recently set up a forum for questions about mouse-tracking, which is available at http://forum.cogsci.nl/index.php?p=/categories/mousetrap.[8] One reoccurring issue that researchers are concerned with is the question of how to ideally setup a mouse-tracking experiment to ensure that the cognitive processes of interest are reflected in the mouse movements. While some general recommendations for designing mouse-tracking experiments have been discussed (Fischer & Hartmann, 2014; Hehman et al., 2015), this question has not previously been addressed empirically. In the following section, I discuss this very topic and present results from a series of experiments that investigate the impact of different design choices in mouse-tracking studies.

---

[6] Downloads of all CRAN releases as of October 11, 2018.
[7] Downloads of all releases from the Python Package Index as of October 11, 2018. The total number of downloads is likely considerably higher, as direct downloads from GitHub source code releases cannot be tracked.
[8] We are thankful to Sebastiaan Mathôt, who is hosting the mousetrap forum.

```
# Load libraries
library(readbulk)
library(mousetrap)

# Read in and filter raw data (only keep correctly answered trials)
raw_data <- read_opensesame("raw_data")
raw_data <- subset(raw_data, correct==1)

# Import and preprocess trajectories
mt_data <- mt_import_mousetrap(raw_data)
mt_data <- mt_remap_symmetric(mt_data)
mt_data <- mt_align_start(mt_data)

# Compute indices, aggregate and compare MAD values between conditions
mt_data <- mt_measures(mt_data)
agg_mad <- mt_aggregate_per_subject(mt_data, subject_id = "subject_nr",
  use_variables = "MAD", use2_variables = "Condition")
t.test(MAD~Condition, data = agg_mad, paired = TRUE)

# Time-normalize trajectories and plot aggregate trajectories per condition
mt_data <- mt_time_normalize(mt_data)
mt_plot_aggregate(mt_data, use = "tn_trajectories",
  color = "Condition", subject_id = "subject_nr")

# Plot heatmaps (raw, smoothed and difference between conditions)
mt_heatmap(mt_data)
mt_heatmap(mt_data, smooth_radius = 20, n_shades = 10, mean_image = 0.2)
mt_diffmap(mt_data, condition = "Condition",
  smooth_radius = 20, n_shades = 10)

# Spatially normalize trajectories, map them onto prototypes and plot result
mt_data <- mt_spatialize(mt_data)
mt_data <- mt_map(mt_data, prototypes = mt_prototypes, save_as = "data")
mt_plot(mt_data, use = "sp_trajectories",
  facet_col = "prototype_label", facet_row = "Condition", alpha = 0.2)
```

*Figure 8.* Example of a complete analysis script in R, covering the merging of the raw data files, their import, typical preprocessing steps, the computation and aggregation of mouse-tracking indices, the assignment of trajectories to a set of prototypes, and the visualization of mouse trajectories (as in Figure 4, Figure 5, and Figure 7).

## 4 Design Factors in Mouse-Tracking

As previously discussed, mouse-tracking lacks standards for designing and running studies and the methodological setup has varied considerably in previous mouse-tracking studies. To understand the influence of design factors on mouse-tracking data and provide a first step towards an evidence-based standard for conducting mouse-tracking studies, we performed a series of experiments, reported in two articles. In the first article (Kieslich et al., 2018), we investigated how the design factors starting procedure, mouse sensitivity, and type of response indication influence the most frequently used dependent variables in mouse-tracking studies, which are trajectory curvature and shape. In the second article (Scherbaum & Kieslich, 2018), we focused on the influence of the starting procedure on both trajectory curvature and dynamic analyses that investigate the temporal effects of different cognitive factors on movement direction. In the following, I will briefly summarize the main results of each study and then jointly discuss their implications and some preliminary recommendations for conducting future mouse-tracking studies.

### 4.1 Effects of design factors on trajectory curvature

Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2018). *Design factors in mouse-tracking: What makes a difference?* Manuscript submitted for publication.

In this article, we replicated the example experiment presented earlier (Dale et al., 2007; as implemented by Kieslich & Henninger, 2017) while systematically varying a separate design factor in each of three experiments. The analyses focused on how design factors impact trajectory curvature and the size of the central cognitive effect (the typicality effect, i.e., greater curvature for atypical than for typical exemplars). In addition, we examined how design factors influence trajectory shapes, using both traditional bimodality analyses and the prototype assignment method. All experiments, data, preprocessing and analysis code, as well as the results are freely available from https://osf.io/xdp7a/.

In the first experiment, we examined the influence of the *response indication procedure.* We compared a click condition, in which participants had to click on the response button, with a touch condition, in which simply moving the cursor onto the button was enough to indicate a response. The typicality effect was replicated with both starting procedures. However, it was significantly larger in the click condition (the click condition also led to a higher degree of curvature overall). This larger effect was related to the occurrence of more extreme trajectory types in the click condition, specifically more discrete change of mind trajectories. In the touch condition, most trajectories were classified as either straight or curved. In line with this, the distribution of the curvature index exceeded the cut-off for bimodality in the click condition while it was below it in the touch condition.

In the second experiment, we investigated the influence of *mouse sensitivity settings,* comparing a condition, in which these settings were left at the system defaults (medium speed, acceleration enabled, resulting in a relatively fast cursor), with a slow condition, in which cursor speed was reduced and acceleration disabled. The typicality effect was replicated in both the default and slow conditions, and its size did not differ significantly between conditions. The default condition led to greater trajectory curvature on average, which was likely driven by a higher percentage of trajectories with extreme movement patterns (such as discrete changes of mind). In both conditions, bimodality coefficients indicated the presence of bimodality in the distribution of the curvature index.

The third experiment examined four different *starting procedures*. The baseline was a static starting procedure in which the stimulus was presented immediately after the click on the start button and participants did not receive any instructions about movement initiation (as in the preceding two experiments). The static start was compared to three procedures that were used in previous mouse-tracking studies to encourage early movement initialization. These included a restriction of the total time for responding (rtmax), an instruction to initialize movement early in the trial (initmax), and requiring an upwards movement to trigger the stimulus display (dynamic). The typicality effect was replicated with all four starting procedures. However, its size differed significantly between conditions, with the initmax condition leading to the largest effect (the size of the typicality effect did not differ significantly between the rtmax and static condition, nor between the dynamic and static condition). Regarding trajectory shapes, the majority of trajectories in the static and rtmax condition were classified as straight. In the dynamic condition, the majority of trajectories were classified as curved. The initmax condition led to a roughly even split of straight and curved classifications and a considerable increase of extreme trajectory types, specifically more discrete change of mind trajectories. Somewhat contrary to the impression derived from the trajectory classifications, almost all bimodality coefficients in the third experiment were below the cut-off for bimodality, suggesting a unimodal distribution.[9]

While these experiments provide a first insight into how central design factors influence trajectory curvature and shape, other mouse-tracking analyses could not be covered, especially dynamic analyses that focus on the temporal development of trajectories. This is due to the limited number of nineteen trials per participant in the paradigm by Dale et al. (2007), which is not sufficient for these analyses. Therefore, we conducted a separate study to explore effects of the starting procedure on the temporal development of trajectories.

### 4.2 Dynamic effects of the starting procedure

Scherbaum, S., & Kieslich, P. J. (2018). Stuck at the starting line: How the starting procedure influences mouse-tracking data. *Behavior Research Methods*, *50*(5), 2097-2110.

In this article, we investigated effects of the starting procedure in a numeric version of the Simon task (Scherbaum et al., 2010), in which participants have to repeatedly choose the left versus right option depending on the size of a number that is presented on the left versus right side of the screen in three blocks of 256 trials each. In this task, two cognitive effects are expected: The Simon effect predicts that mouse trajectories should deviate more towards the non-chosen option if the number's location on the screen is incongruent with its implied direction. In addition, there should be a congruency sequence effect, in that the Simon effect is attenuated if a trial is preceded by an incongruent trial. The study also explored the influence of these factors on the temporal development of the trajectories. In particular, it investigated whether the congruency sequence effect set in together with or after the Simon effect – a finding that can be used to disentangle different theoretical accounts of the cognitive processes underlying action control.

The original implementation of the Simon task in a mouse-tracking experiment by Scherbaum et al. (2010, Experiment 2) used a dynamic starting procedure, in which participants

[9] One methodological difference from the preceding two experiments was that the cursor speed was reduced in all conditions and the stimulus was presented at a higher point of the screen. This was done to accommodate the dynamic and inimtax starting procedures which require that the stimulus information can be acquired during the upwards movements. These changes may have contributed to overall differences in trajectory shapes between studies.

had to move the mouse upwards for the numerical stimulus to be displayed. The data from this experiment was compared to a new experiment that used an identical setup except for changing the starting procedure to a static start, in which the stimulus was presented after a short, fixed delay and participants could freely decide when to initialize their mouse movement. When looking at trajectory curvature, the Simon and the congruency sequence effect were replicated with both starting procedures. Besides, there were no significant differences in the size of these effects between the two starting procedures, although the dynamic starting procedure led to a greater overall curvature. Bimodality coefficients suggested a bimodal distribution of trajectory curvature in the static condition, but a unimodal distribution in the dynamic condition.

To analyze the temporal development of movements within a trial, time continuous multiple regression analyses were performed. At each time point, these predicted the movement angle (as an indicator of movement towards the chosen vs. non-chosen option) based on the response in the preceding trial, the stimulus location (i.e., the Simon effect), and the congruency sequence effect. In the dynamic condition, the response of the preceding trial influenced the movement angle at the earliest stage of the trial, followed by the location of the stimulus, and the congruency sequence effect thereafter. The temporal order was identical in the static condition, but the effects of each factor on movement angle showed greater temporal overlap and the effect of the stimulus location (Simon effect) was considerably weaker.

We also performed a number of analyses regarding the consistency of movements. Most importantly, participants in the dynamic condition moved the mouse upwards more continuously throughout each trial, while participants in the static condition often stayed at the bottom of the screen for the first half of the trial.

The two starting procedure conditions investigated in Scherbaum and Kieslich (2018) are comparable to the static and dynamic condition investigated in the previously summarized Experiment 3 by Kieslich et al. (2018). However, there are a number of methodological differences between the two studies, which I will discuss in the following. On the one hand, these differences are a chance for exploring the generalizability of starting procedure effects to different tasks, setups, and analyses. On the other hand, they decrease the comparability of the two studies. For this reason, I have performed a set of additional analyses applying the analytic approach from Kieslich et al. (2018) to the data from Scherbaum and Kieslich (2018). I summarize these analyses below.

With regard to the study design, the general methodological setup and the specific implementation of the starting procedures differed (in addition to the task under investigation): Scherbaum and Kieslich (2018) used a touch response procedure, restricted the total time participants had for giving a response, and instructed participants to continuously move upwards once they had initiated a movement. Kieslich et al. (2018) used a click response procedure and neither imposed a total time limit nor gave more specific instructions regarding the movement. When implementing the static starting procedure, Scherbaum and Kieslich (2018) presented the stimulus with a fixed delay of 200 ms while Kieslich et al. (2018) presented the stimulus immediately. When implementing the dynamic starting procedure, Scherbaum and Kieslich (2018) used an upwards movement criterion of 4 px in two consecutive time steps while Kieslich et al. (2018) used an upwards movement criterion of 50 px in total, following Frisch et al. (2015). Besides, Kieslich et al. (2018) followed an exploratory approach with open research questions regarding the influence of the different design factors (outlining possible outcomes depending on previous recommendations and speculations about potential effects of design factors), while Scherbaum and Kieslich (2018) tested specific hypotheses.[10]

Concerning trajectory curvature, the dynamic starting procedure led to an overall higher degree of curvature than the static starting procedure in both studies, while the size of the cognitive effects on trajectory curvature did not differ significantly between the dynamic and static starting procedure. Trajectory curvature was assessed via the average deviation (AD) in Scherbaum and Kieslich (2018) whereas trajectory curvature was assessed using the more common MAD measure in Kieslich et al. (2018). Results in Scherbaum and Kieslich (2018) replicate when using MAD instead of AD in the analysis.

Regarding the shape of the individual trajectories, bimodality coefficients in Scherbaum and Kieslich (2018) were below the threshold for a bimodal distribution for the dynamic starting

[10] The hypotheses in Scherbaum and Kieslich (2018) comprised larger cognitive effects for the dynamic versus the static starting procedure, but specified that the strength of the effect of the starting procedure should depend on the level of analysis. That is, cognitive effects on discrete measures (i.e., overall trajectory curvature) were assumed to be relatively robust and hence only slightly influenced by the starting procedure. In contrast, stronger effects of the starting procedure were expected for continuous within-trial measures (i.e., time continuous multiple regression analyses). This distinction was not made in Kieslich et al. (2018), where the focus was only on discrete measures and on the research question whether starting procedures that induce an early movement initiation generally lead to larger cognitive effects (based on previous recommendations by Hehman et al., 2015, and Fischer & Hartmann, 2014).

procedure while they exceeded it for the static starting procedure. Bimodality coefficients in Kieslich et al. (2018) were all below the threshold for a bimodal distribution except for typical trials in the dynamic condition where the bimodality coefficient slightly exceeded the threshold. However, the exact computation of the bimodality coefficients differed between the two studies. When using the approach by Kieslich et al. (2018) for the data from Scherbaum and Kieslich (2018), bimodality coefficients are still higher for the static condition than for the dynamic condition; however, all bimodality coefficients are below the threshold for a bimodal distribution. This inconsistent pattern of results indicates, in line with earlier discussions, that some caution is advisable when interpreting results from bimodality coefficients and that analyses directly based on trajectory shapes seem preferable. However, while several visualizations of the individual trajectories were included and discussed in Scherbaum and Kieslich (2018), no numeric analysis of trajectory shapes at the trial level was conducted. Therefore, I performed a new analysis, classifying trajectories in Scherbaum and Kieslich (2018) based on the same set of prototypes that were used in Kieslich et al. (2018). The distribution of classifications differed significantly between the two starting procedures: The majority of trajectories in the dynamic condition were classified as curved, while the majority of trajectories in the static condition were classified as straight. This is in line with the results reported for the two starting procedures in Kieslich et al. (2018).

In sum, while the implementation and analysis approaches differed substantially between the two studies, most results regarding the effects of the starting procedures – except for the bimodality analyses – are comparable.

### 4.3 Implications for interpreting mouse-tracking data

This series of studies represents the first systematic investigation of design factors in mouse-tracking. Specifically, we investigated the factors starting procedure, mouse sensitivity, and type of response indication (Kieslich et al., 2018; Scherbaum & Kieslich, 2018), and found them to significantly affect the study results. Thus, the methodological setup needs to be considered when interpreting mouse-tracking data. The implications of our findings are severalfold.

First, all setups that we investigated were able to capture the conflict between response options to at least some extent, and the theoretically predicted cognitive effects on trajectory curvature were replicated in every setup. However, their magnitudes were significantly influenced by several of the design factors, implying that effect sizes are not directly comparable between studies with different methodological setups. Moreover, for studies investigating weaker cognitive effects some setups might not be sensitive enough to detect them at all. Thus, the sensitivity of a setup needs to be taken into account, especially when interpreting null effects. Design factors become even more influential in dynamic analyses (such as time continuous multiple regression) which are more sensitive to inconsistencies in movements within the trial.

In addition, the shape of the individual trajectories varied considerably between studies. Assuming that the setup does not influence cognitive processing[11], this implies that different trajectory shapes can occur for the same cognitive process. Thus, it is not possible to make theoretical inferences about the underlying cognitive process based on the trajectory shape without considering the influence of the design factors.

In my view, a useful way of understanding the influence of design factors is as moderators, that is, in terms of how they change the mapping of the cognitive processes onto mouse movements. The *starting procedure* likely influences the degree to which the early phase of the decision process is reflected in the mouse movement, with starting procedures that explicitly encourage an early movement initiation (i.e., dynamic or initmax) increasing the likelihood that early aspects are captured. This is not the case when using a static starting procedure, where, in extreme cases, the decision process might even be finished before the mouse movement is initiated. A resulting straight trajectory would then not necessarily indicate that there was no response conflict, but merely that it was not captured in the movement. The type of *response indication* is likely to influence the degree to which the attraction of an option is translated into a movement towards that option. Requiring a click to indicate a response allows participants to move all the way to an option and then redirect the movement to the other option (a prototypical change of mind trajectory). In contrast, responding just by moving the cursor onto the response button reduces the likelihood of these extreme movements (as participants would have to move below the button if they wanted to

[11] In my view, it is generally plausible to assume that the cognitive processes themselves are not affected by most of the discussed design factors, especially if they concern a peripheral aspect, such as the speed of the cursor, or whether the response is given via click or touch. However, this assumption cannot be tested directly based on the mouse-tracking data and, hence, a change in the cognitive process cannot be ruled out. One possible test for this could be to look at analyses at the choice level, such as the correctness of participants' responses. Correctness did not differ significantly between the different response indication and mouse sensitivity settings (Experiments 1 and 2 of Kieslich et al., 2018). However, correctness was affected by some of the starting procedures (in Experiment 3), specifically those that induce some amount of time pressure (see more details in the manuscript). Thus, based on the correctness data in Kieslich et al. (2018), I would conclude that the design factors did not significantly affect the cognitive processes, except when they induced some amount of time pressure.

be able to correct their response; some participants might even refrain from initializing an early cursor movement as they might be afraid of giving an unwanted response by accident). The *mouse sensitivity* influences how movements of the hand are translated into the cursor movement, with default settings exaggerating small movements and hence producing more extreme movements compared to a mouse cursor with reduced speed and disabled acceleration. Thus, a study with a static starting procedure, click response mode, and default mouse sensitivity settings is more likely to produce a mix of straight and change of mind trajectories than a study with a dynamic starting procedure, touch response mode, and reduced cursor speed (and disabled acceleration). In other words, if a mix of straight and change of mind trajectories was observed in the latter setup, this would be more convincing evidence for a dual-system model on a process level than if they were observed in the former setup.

Finally, the studies demonstrate the usefulness of methods for visualizing and analyzing the trajectory shape at the trial level (Kieslich et al., in press; Wulff et al., in press). They enable unpacking the effect of a certain manipulation on mouse trajectory curvature. That is, they show whether greater curvature is caused by all trajectories being more curved in one of the conditions, or whether a certain condition leads to the more frequent occurrence of extreme trajectory types, such as discrete changes of mind. With regard to the different analytic methods for identifying trajectory types, the newly proposed prototype assignment method and traditional bimodality analyses did not always agree. It seems that the bimodality coefficient is less sensitive to detecting different types of trajectories, even when visual inspection and the trajectory classification suggest they are present at the trial level.

### 4.4 Implications for designing mouse-tracking studies

The presented studies also have implications for the design of future mouse-tracking studies. In line with previous recommendations (Fischer & Hartmann, 2014; Hehman et al., 2015), it generally seems advisable to use an initmax or dynamic starting procedure that encourages an early movement initiation, to ensure that early stages of the decision process are captured in the movement. Using one of these starting procedures and thereby enforcing a movement initiation as soon as or even before the decision-critical stimulus is presented, has direct implications with regard to other design factors: The cursor speed should be reduced and the stimulus presented a considerable distance above the start button to ensure that participants can acquire the stimulus information while moving the mouse upwards.

However, the suitability of the initmax and dynamic starting procedures also depends on the type of task under investigation. Both are probably best suited for tasks with simple stimuli that can be solved quickly (e.g., the Simon task), for which the time period of the upwards movement is sufficient to acquire the stimulus information and complete the task. Conversely, they might be challenging to implement in tasks involving more complex stimuli or difficult decisions (such as decisions under risk or decisions in social dilemmas) as participants may not be able to acquire all stimulus information and reach their decision before finishing their upwards movement (i.e., they might already arrive at one of the buttons or the top of the screen). If this is the case, participants may stop moving the cursor either already when the stimulus is presented or when they hit the upper screen boundary, disrupting the continuous mapping of the cognitive process onto the mouse movement. For those tasks, a static starting procedure might be an adequate fallback option, as it gives participants the opportunity to initialize their movement only after acquiring the stimulus information – risking that participants might arrive at their decision before starting their movement in some trials.

With regard to the response indication mode, the click condition led to considerably larger cognitive effects than the touch condition. This alone would speak in favor of using a click procedure, but the click procedure was also related to a more frequent occurrence of extreme trajectories, like discrete change of mind trajectories. This, in turn, suggests that researchers might face a trade-off between larger effects which are due to the occurrence of more extreme trajectory types and smaller effects with a more homogeneous trajectory distribution. However, the touch procedure resulted in a very large number of straight trajectories in the current study, suggesting that participants might have made their decision before initiating a movement in some trials to avoid accidentally selecting an unwanted option. This might be particularly relevant in the current study, as it implemented the touch condition in combination with a static starting procedure and default mouse sensitivity settings (i.e., a relatively high cursor speed). Given these considerations, it might make sense to set the response indication mode depending on the starting procedure. That is, if an initmax or dynamic starting procedure is used to ensure that participants start moving early in all trials, a touch response mode could be used to achieve a more homogenous distribution of trajectories and perform traditional mouse-tracking analyses via curvature indices. If a static starting procedure is used, the response mode click might be preferable and analyses should be based on prototypes. However, as the response indication mode was only tested for a static starting procedure in the current studies, these recommendations are somewhat speculative and further research is needed to answer this question.

With regard to mouse sensitivity, the preferred settings likely depend on the starting procedure used. In the current studies, the effect of mouse sensitivity was investigated with a static starting procedure and did not significantly affect the mouse-tracking data (with regard to the strength of the typicality effect). This might indicate that, for a static starting procedure, the mouse sensitivity settings are not that consequential (although an extremely fast cursor should be avoided to prevent chaotic cursor movements, cf. Freeman & Ambady, 2010). However, as discussed above, for an initmax and dynamic starting procedure, it seems advisable to reduce cursor speed and disable acceleration.

As this summary shows, recommendations regarding the design of future mouse-tracking studies are not straightforward. As discussed, the ideal mouse-tracking setup depends on the type of task under investigation and the type of analysis that will be conducted. As a rule of thumb, for very simple tasks, researchers should use an initmax or dynamic starting procedure, reduce the cursor speed, disable acceleration, and use a touch response mode to achieve a more homogenous trajectory distribution. Conversely, if the task under investigation is more complex, researchers may use a static starting procedure and a click response mode (and the cursor speed is not as critical). However, in this setup dynamic analyses (e.g., time continuous multiple regression) may not be possible as the consistency of the movement over the course of the trial may not be sufficient. These recommendations are, of course, preliminary, and require more empirical studies that investigate the influence of design factors in different psychological tasks.

# 5 Discussion

In this dissertation, I first presented the mousetrap plugin for OpenSesame, which allows researchers to easily create mouse-tracking experiments via a graphical user interface and to run them on laboratory computers supporting all major platforms. The mousetrap R package enables users to process, analyze, and visualize mouse-tracking data of all major formats. It implements most of the commonly used preprocessing procedures and mouse-tracking indices, along with a set of novel visualization and classification procedures for analyzing trajectory shapes. All software is open-source and freely available, facilitating open and transparent research practices and the effortless replication of experiments, which has become a topic of critical importance in psychological science and beyond (Asendorpf et al., 2013; Munafò et al., 2017; Nosek et al., 2015). In addition, this dissertation presented results from a first systematic investigation of central design factors in mouse-tracking studies. Results showed that the methodological setup had a considerable influence on trajectory curvature and shape, and should therefore be taken into account when interpreting mouse-tracking data. They also provide a first empirical foundation for informed decisions about future study designs, and I derived some preliminary recommendations.

In the following, I will first discuss future directions with regard to the implementation of mouse-tracking experiments, focusing in particular on limitations of the studies that investigate the influence of the methodological setup. Second, I will discuss future directions regarding mouse-tracking analyses, focusing on additional methods not covered so far and open questions concerning the interpretation of mouse-tracking data.

## 5.1 Implementation and study design

With regard to the implementation of mouse-tracking studies, the mousetrap plugin for OpenSesame covers all current designs and methodological setups (a comprehensive set of example experiments for different setups is provided with Kieslich et al., 2018). Future releases of the plugin will therefore mainly focus on staying up-to-date with OpenSesame, which is continuously developed and improved. Going beyond laboratory studies, an important future direction will be the ability to conduct mouse-tracking experiments online, and we are currently developing a browser-based solution for this in a project led by Felix Henninger (the results of an online study using a development version are published in Horwitz et al., in press). Another direction is the use of alternative tracking methods that more directly record hand movements (using, e.g., a motion capture system, cf., Awasthi, Friedman, & Williams, 2011) or finger movements (e.g., via touch screens, cf., Wirth, Pfister, & Kunde, 2016). The latter approach could, in principle, be implemented in OpenSesame by extending the mousetrap plugin.

A major concern regarding the creation of mouse-tracking experiments remains the many degrees of freedom and open questions regarding the methodological setup. While the current dissertation presents a first set of studies that examine the impact of three central design factors on mouse-tracking data, these studies have a number of limitations. I will address those and directions for future research in the following (a detailed discussion of more specific limitations and open questions concerning specific results is provided in the articles).

Since Scherbaum and Kieslich (2018) compared data from a previous experiment using a dynamic starting procedure (Scherbaum et al., 2010) with a new experiment that employed a static procedure, participants were not randomly assigned to the design factor conditions. Hence, we cannot rule out that there were systematic differences between the participant groups, even though all statistical tests for differences between groups on any of the sample characteristics were non-significant. However, the results of Experiment 3 by Kieslich et al. (2018) largely support the same conclusions based on a randomized experiment (although there are a number of differences between the two studies, as previously discussed). Nevertheless, a replication of Scherbaum and Kieslich (2018) with a randomized assignment to conditions is warranted.

To allow for a clear interpretation of the consequences of each design factor, we manipulated each design factor in a separate experiment. However, this approach would not uncover interactions between design factors, some of which are likely to occur, as noted above. Specifically, it would be interesting to examine combinations of the response indication procedure and different starting procedures. For example, one prediction is that a dynamic or initmax starting procedure reduces the relatively large proportion of straight trajectories that were observed in the touch response condition. The mouse sensitivity settings might also play a more important role for these starting procedures, as a high cursor speed might make it difficult to acquire the stimulus information during the upward movement. While both considerations seem plausible (as discussed in the previous section), they should be tested empirically in future studies.

For each design factor, the studies tested their most common implementations. However, the factors can vary beyond the examined levels. For example, some previous studies have reduced cursor speed even further than we did in the slow condition (e.g., Huette & McMurray, 2010).

Implementations of the initmax and dynamic starting procedures also differed across previous studies, especially with regard to the minimum distance that participants need to move upwards and the time limit for this upwards movement in the initmax condition (a suitable time limit also depends on the complexity of the task, cf. Hehman et al., 2015).

While the current studies covered three important design factors, there are more that should be examined. Several concern the spatial layout of the decision screen, such as the exact stimulus position and the horizontal distance between response buttons. An additional factor is whether participants receive specific instructions regarding mouse movements (e.g., the instruction to continuously move upwards once a movement is initiated, cf., Scherbaum & Kieslich, 2018). Together with different colleagues, I am currently working on projects that address these factors.

The studies reported in this dissertation examined design factors in two different tasks, a Simon task and a semantic categorization task. Both tasks are fairly simple and can be solved relatively quickly. Future studies should examine the effect of design factors in other, more complex tasks. As discussed previously, it is likely that different types of setups are suitable for different types of tasks. An additional question concerns the effect of design factors for different types of stimuli, as several mouse-tracking studies use pictures (e.g., Freeman & Ambady, 2009; Sullivan et al., 2015) or auditory stimuli (e.g., McKinstry, Dale, & Spivey, 2008; Spivey et al., 2005).

The ultimate goal is to provide researchers with a comprehensive set of recommendations for designing mouse-tracking studies, once the empirical evidence base is sufficient. These will likely depend on the type of task that is examined and the analysis that is intended. For this purpose, several colleagues and I are working on conducting additional studies to investigate design factors. Eventually, we want to arrive at a mutually agreed upon standard of how future mouse-tracking studies should be conducted.

### 5.2 Analysis and interpretation

With regard to the analysis of mouse-tracking data, the mousetrap R package covers most of the common preprocessing procedures and established indices from the literature. It also provides different functions for visualizing individual and aggregate trajectories and the development of mouse-tracking variables over time, and covers novel methods for classifying trajectories based on their shape. Nevertheless, some advanced analysis and visualization methods are still missing in the package. For instance, the time continuous multiple regression approach (described in

Scherbaum & Kieslich, 2018) is not yet fully implemented in the package. However, an alternative implementation via linear-mixed models can be realized using mousetrap's data reshaping and aggregation functions in combination with the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). The result of such an analysis is provided in the supplementary material of Scherbaum and Kieslich (2018).[12] Several more recent methods for analyzing mouse-tracking data, such as advanced visualizations via decision landscapes (Zgonnikov, Aleni, Piiroinen, O'Hora, & di Bernardo, 2017) and entropy analyses (Calcagnì et al., 2017) are not yet covered in mousetrap, although I have discussed plans for their integration with the authors.

With many different processing and analysis options conveniently packaged, the main challenge becomes to choose the best approach for a particular research project. This is especially relevant regarding the multitude of different indices, and the different levels of analyses that are available to researchers. I will address these issues in the following.

Many mouse-tracking studies have performed analyses based on aggregate curvature indices and aggregate trajectories. As I have shown in this dissertation, these are not necessarily representative of what is happening at the trial level and it is, therefore, important to visualize and analyze individual trajectories. To assess whether there are different types of trajectories at the trial level, traditional analyses have focused on the bimodality coefficient for the distribution of curvature indices. However, new approaches have been proposed that take into account the complete shape of the trajectory and assign trajectories to different prototypes (Wulff et al., in press). Importantly, the bimodality coefficient and the prototype classifications sometimes differ in their conclusions (as in one of the design factor studies reported in this thesis, Kieslich et al., 2018), and it seems that the bimodality coefficient is generally less sensitive to detecting different types of trajectories. Regardless of the analysis method used, it is also always advisable to plot heatmaps of the individual trajectories to get a visual impression of the data.

One issue which is not completely resolved so far is the best solution in case of a substantial number of different trajectory types present in the data. For now, I would recommend using the prototype classification itself as the dependent variable in the analysis, for example, as an ordinal variable if different amounts of conflict can be clearly assigned to the different prototypes. Generalized processing trees are another recently proposed method (Heck, Erdfelder, & Kieslich, in

[12] The supplementary material is published online together with article or can be downloaded using this link.

press) that can be useful in this regard. This method jointly models choices and continuous variables (such as a trajectory curvature index like MAD) and can hence explain a bimodal distribution of trajectory curvature through different underlying cognitive processes. By jointly modeling choices and MAD, it also takes into account both correct and incorrect answers – whereas previous mouse-tracking studies typically excluded incorrectly answered trials. This method is provided in another R package[13], which can be used in combination with mousetrap. We have previously applied it to mouse-tracking (Heck et al., in press) using data from the example experiment presented in this dissertation (Kieslich & Henninger, 2017).

The variety of mouse-tracking indices reflects the manifold opportunities mouse-tracking provides for testing various research questions. At the same time, the interpretation of many of these indices is still open to debate and the conceptual differences between the measures need to be better understood. In general, studies with theoretically founded and targeted manipulations are needed that validate the interpretation of specific indices. In this regard, Koop and Johnson (2013) provide a useful first step for validating trajectory curvature by demonstrating that a priori differences in the pleasantness of visual stimuli (based on pleasantness norms) systematically influenced the degree of trajectory curvature in preferential decisions. In a similar vein, I am currently working on a project together with Bence Palfi, Barnabas Szaszi, Dirk Wulff, and Balazs Aczel that compares different mouse-tracking indices for quantifying the number of changes of mind in a trial, assesses the degree to which they are sensitive to experimental manipulations that are expected to induce changes of mind, and quantifies their agreement with human raters.

Relatedly, the results from the design factor studies have shown that there is still a need to better understand how cognitive processes in general and the preference development in particular are mapped onto the mouse movement. It seems that the original assumption of a completely continuous mapping of the cognitive process onto the mouse movement responses (Freeman et al., 2011; Spivey & Dale, 2006) depends on the task and methodological setup of the study, and thus cannot be assumed a priori. Future mouse-tracking research therefore should focus more on building and validating explicit models of how the preference development is translated into the cursor movement, accounting for the influence of the setup. In this regard, it might be fruitful to build on promising models that have been proposed for this purpose in the past (e.g., dynamic

[13] The R package is called gpt and can be obtained from GitHub (https://github.com/danheck/gpt).

neural field models, cf. Frisch et al., 2015; or bounded-accumulation models, cf. Resulaj, Kiani, Wolpert, & Shadlen, 2009). This is especially relevant if one wants to draw inferences about the process model underlying a decision (e.g., whether a dual system model is actually supported by the data). This being said, if one is mainly interested in assessing the overall level of conflict that was present in a decision and is somewhat agnostic about the specific process model, one can still rely on trajectory curvature to test theoretical predictions.[14]

In sum, while there is an increasing awareness about many issues that need to be addressed when analyzing mouse-tracking data, a consensus about what constitutes the best study design and analysis approach has yet to be reached. Therefore, when using mouse-tracking to test psychological theories researchers should be transparent in reporting the assumptions they make regarding mouse movements, the methodological setup, data processing, and the analyses they performed. The latter is optimally implemented if researchers share their data, preprocessing, and analysis code. Besides, researchers may often explore different analyses and measures when using mouse-tracking to test theoretical predictions, also to ensure that different approaches arrive at similar conclusions. However, once they have settled on a particular approach to test a theory, additional preregistered studies should be conducted to meet the requirements of strictly confirmatory hypothesis testing (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012).

### 5.3 Conclusions

Mouse-tracking is a promising new method for assessing cognitive processes and testing psychological theories. With the software developed in this dissertation I want to make the method accessible to researchers from different disciplines and levels of technical experience. The presented software should enable researchers to create even complex experiments with ease and to use state-of-the-art methods in their analyses. I also hope to have raised awareness regarding many open questions concerning study design and analysis, and to have contributed some initial steps and recommendations towards resolving them. I am convinced that mouse-tracking will prove to be a useful addition to the toolbox of process tracing methods and will provide novel insights into different cognitive processes. I am looking forward to further advancing the method together with many enthusiastic colleagues.

[14] However, researchers have to ensure that the methodological setup is sensitive enough for detecting conflict in a given task. As discussed, for the current paradigms all investigated setups seemed to be sufficiently sensitive.

## 6 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*(6), 918–926. https://doi.org/10.1037/xlm0000484

Asendorpf, J. B., Conner, M., Fruyt, F. D., Houwer, J. D., Denissen, J. J. A., Fiedler, K., … Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*(2), 108–119. https://doi.org/10.1002/per.1919

Awasthi, B., Friedman, J., & Williams, M. A. (2011). Faster, stronger, lateralized: Low spatial frequency information supports face processing. *Neuropsychologia*, *49*(13), 3583–3590. https://doi.org/10.1016/j.neuropsychologia.2011.08.027

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Calcagnì, A., Lombardi, L., & Sulpizio, S. (2017). Analyzing spatial data from mouse tracker methodology: An entropic approach. *Behavior Research Methods*, *49*(6), 2012–2030. https://doi.org/10.3758/s13428-016-0839-5

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15–28. https://doi.org/10.3758/BF03195938

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, *46*(4), 913–921. https://doi.org/10.3758/s13428-013-0422-2

Dshemuchadse, M., Scherbaum, S., & Goschke, T. (2013). How decisions emerge: Action dynamics in intertemporal decision making. *Journal of Experimental Psychology: General*, *142*(1), 93–100. https://doi.org/10.1037/a0028499

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*(2), 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*(1), 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*(5), 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*(10), 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*(1), 226–241. https://doi.org/10.3758/BRM.42.1.226

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*(4), 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*(1), 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*(4), 377–388. https://doi.org/10.1007/s10339-015-0666-0

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, *13*(1), 70–84. https://doi.org/10.1214/aos/1176346577

Heck, D. W., Erdfelder, E., & Kieslich, P. J. (in press). Generalized processing tree models: Jointly modeling discrete and continuous variables. *Psychometrika*. https://doi.org/10.1007/s11336-018-9622-0

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, *18*(3), 384–401. https://doi.org/10.1177/1368430214538325

Henninger, F., Kieslich, P. J., & Hilbig, B. E. (2017). Psynteract: A flexible, cross-platform, open framework for interactive experiments. *Behavior Research Methods*, *49*(5), 1605–1614. https://doi.org/10.3758/s13428-016-0801-6

Horwitz, R., Brockhaus, S., Henninger, F., Kieslich, P. J., Schierholz, M., Keusch, F., & Kreuter, F. (in press). Learning from mouse movements: Improving questionnaire and respondents' user experience through passive data collection. In P. C. Bearry, A. Wilmot, D. Collins, L. Kaye, J. L. Padilla, & G. Willis (Eds.), *Advances in Questionnaire Design, Development, Evaluation and Testing*. New York, NY: Wiley.

Huette, S., & McMurray, B. (2010). Continuous dynamics of color categorization. *Psychonomic Bulletin & Review*, *17*(3), 348–354. https://doi.org/10.3758/PBR.17.3.348

Kieslich, P. J., & Henninger, F. (2016). Readbulk: An R package for reading and combining multiple data files. https://doi.org/10.5281/zenodo.596649

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods*, *49*(5), 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*(6), 510–522.

Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2018). *Design factors in mouse-tracking: What makes a difference?* Manuscript submitted for publication.

Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judgment and Decision Making*, *8*(5), 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*(5), 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*(8), 750–758.

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, *67*(4), 151–185. https://doi.org/10.1016/j.cogpsych.2013.09.001

Leontyev, A., Sun, S., Wolfe, M., & Yamauchi, T. (2018). Augmented go/no-go task: Mouse cursor motion measures improve ADHD symptom assessment in healthy college students. *Frontiers in Psychology*, *9*, 496. https://doi.org/10.3389/fpsyg.2018.00496

Lepora, N. F., & Pezzulo, G. (2015). Embodied choice: How action influences perceptual decision making. *PLOS Computational Biology*, *11*(4), 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*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*(1), 22–24. https://doi.org/10.1111/j.1467-9280.2008.02041.x

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., du Sert, N. P., … Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. https://doi.org/10.1038/s41562-016-0021

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin & Review*, *19*(5), 906–913. https://doi.org/10.3758/s13423-012-0281-3

Pfister, R., Schwarz, K. A., Janczyk, M., Dale, R., & Freeman, J. B. (2013). Good things peak in pairs: A note on the bimodality coefficient. *Frontiers in Psychology*, *4*, 700. https://doi.org/10.3389/fpsyg.2013.00700

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/

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263–266. https://doi.org/10.1038/nature08275

SAS Institute Inc. (1990). *SAS/STAT User's Guide* (4th ed.). Cary, NC: Author.

Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: The temporal dynamics of action selection. *Cognition*, *115*(3), 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*(5), 2097–2110. https://doi.org/10.3758/s13428-017-0977-4

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*(5), 442–450. https://doi.org/10.1177/0963721417708229

Schulz, E., Speekenbrink, M., & Krause, A. (2018). A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, *85*, 1–16. https://doi.org/10.1016/j.jmp.2018.03.001

Song, J.-H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, *13*(8), 360–366. https://doi.org/10.1016/j.tics.2009.04.009

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, *15*(5), 207–211. https://doi.org/10.1111/j.1467-8721.2006.00437.x

Spivey, M. J., Dale, R., Knoblich, G., & Grosjean, M. (2010). Do curved reaching movements emerge from competing perceptions? A reply to van der Wel et al. (2009). *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 251–254. https://doi.org/10.1037/a0017170

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10393–10398. 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*(9), 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*(6), 531–543. https://doi.org/10.1016/j.tics.2018.03.012

Sullivan, N., Hutcherson, C., Harris, A., & Rangel, A. (2015). Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. *Psychological Science*, *26*(2), 122–134. 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*(1), 23–32.

Tippmann, S. (2015). Programming tools: Adventures with R. *Nature News*, *517*(7532), 109. https://doi.org/10.1038/517109a

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

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Wirth, R., Pfister, R., & Kunde, W. (2016). Asymmetric transfer effects between cognitive and affective task disturbances. *Cognition and Emotion*, *30*(3), 399–416. https://doi.org/10.1080/02699931.2015.1009002

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. (2018). *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*(3), 289–306. https://doi.org/10.1521/soco.2012.30.3.289

Zgonnikov, A., Aleni, A., Piiroinen, P. T., O'Hora, D., & di Bernardo, M. (2017). Decision landscapes: Visualizing mouse-tracking data. *Royal Society Open Science*, *4*(11), 170482. https://doi.org/10.1098/rsos.170482

**Acknowledgments**

This dissertation would not have been possible without the support from a number of people to whom I would like to extend my deepest gratitude.

My girlfriend Johanna Hepp for always supporting me throughout my studies and dissertation (and beyond) with her advice, optimism, patience, and love, for providing extensive and extremely helpful feedback to this thesis and my papers, and for being my perfect match – both personally and academically.

Felix Henninger for being my partner in crime over all those dissertation years, whom I have spent countless days and nights with discussing ideas and building software, and who has contributed tremendously to this thesis through his feedback and valuable suggestions.

Arndt Bröder for giving me the freedom to follow my research interests and motivating me to pursue mouse-tracking as a dissertation project in its own right, for providing so much helpful feedback and advice and always giving me the confidence that I was headed in the right direction.

Benni Hilbig for sparking my interest in science in the first place, for many inspiring discussions and for drawing my attention to the mouse-tracking method all those years ago.

Dirk Wulff, Michael Schulte-Mecklenbeck, and Jonas Haslbeck for contributing their exciting new methods to the mousetrap package along with a fresh new look at mouse-tracking data.

Stefan Scherbaum, Martin Schoemann, and Tobias Grage for joining my quest in the search for the Holy Grail of mouse-tracking (the perfect mouse-tracking design).

Balazs Aczel, Bence Palfi, Barnabas Szaszi, and Aba Szollosi for motivating me to start the mousetrap project through their suggestions (and a phone call), and for many enlightening discussions and productive and fun work sessions concerning future developments and applications of mouse-tracking.

Daniel Heck and Edgar Erdfelder for introducing me to a completely new way of jointly modeling choice and mouse-tracking data.

Sophie Scharf, Jana Berkessel, and Chloe Günther for providing encouragement and feedback in the final stages of my dissertation.

Mila Rüdiger, Monika Wiegelmann, Rebecca Heck, Lara Többen, Florian Storz, and Anja Humbs for their help with data collection, software testing, and manuscript checking.

All colleagues and student research assistants at the Experimental Psychology Lab for providing such a great working environment, both academically and socially.

The Graduate School of Economic and Social Sciences (funded by the German Research Foundation) for providing support to me in my dissertation through its very helpful and fast administration team, great events for academic exchange, and several awards and grants to fund my research.

My parents Conny Meye-Kieslich and Thomas Kieslich and my brother Kimon Kieslich for always supporting me throughout my dissertation and beyond, and for enabling me to follow this path in the first place.

## Statement of Originality

I affirm in lieu of oath that the following statements are to the best of my knowledge true and complete.

1. I hereby affirm that the presented doctoral dissertation with the title *Advancing Mouse-Tracking Research: New Solutions for Study Design, Implementation, and Analysis* is my own work.

2. I did not seek unauthorized assistance of a third party and I have employed no other sources or means except the ones listed. I clearly marked any quotations derived from the works of others.

3. I did not yet present this doctoral dissertation or parts of it at any other higher education institution in Germany or abroad.

4. I hereby confirm the accuracy of the affirmation above.

5. I am aware of the significance of this affirmation and the legal consequences in case of untrue or incomplete statements.

Mannheim, October 2018 Pascal J. Kieslich

## Co-Authors' Statements

- Felix Henninger
- Dirk U. Wulff
- Jonas M. B. Haslbeck
- Michael Schulte-Mecklenbeck
- Martin Schoemann
- Tobias Grage
- Johanna Hepp
- Stefan Scherbaum

**Co-Author: Felix Henninger**

It is hereby confirmed that the articles entitled

Mousetrap: An integrated, open-source mouse-tracking package

and

Mouse-tracking: A practical guide to implementation and analysis

included in the thesis *Advancing Mouse-Tracking Research: New Solutions for Study Design, Implementation and Analysis* were primarily conceived and written by Pascal J. Kieslich, Ph. D. candidate at the Center for Doctoral Studies in Social and Behavioral Sciences of the Graduate School of Economic and Social Sciences, University of Mannheim. I sign this statement to the effect that Pascal J. Kieslich is credited as the primary source of ideas and the main author of the articles.

October 2018 Felix Henninger

**Co-Author: Dirk U. Wulff**

It is hereby confirmed that the article entitled

Mouse-tracking: A practical guide to implementation and analysis

included in the thesis *Advancing Mouse-Tracking Research: New Solutions for Study Design, Implementation and Analysis* was primarily conceived and written by Pascal J. Kieslich, Ph. D. candidate at the Center for Doctoral Studies in Social and Behavioral Sciences of the Graduate School of Economic and Social Sciences, University of Mannheim. I sign this statement to the effect that Pascal J. Kieslich is credited as the primary source of ideas and the main author of the article.

October 2018 Dirk U. Wulff

**Co-Author: Jonas M. B. Haslbeck**

It is hereby confirmed that the article entitled

Mouse-tracking: A practical guide to implementation and analysis

included in the thesis *Advancing Mouse-Tracking Research: New Solutions for Study Design, Implementation and Analysis* was primarily conceived and written by Pascal J. Kieslich, Ph. D. candidate at the Center for Doctoral Studies in Social and Behavioral Sciences of the Graduate School of Economic and Social Sciences, University of Mannheim. I sign this statement to the effect that Pascal J. Kieslich is credited as the primary source of ideas and the main author of the article.

October 2018 Jonas M. B. Haslbeck

**Co-Author: Michael Schulte-Mecklenbeck**

It is hereby confirmed that the article entitled

Mouse-tracking: A practical guide to implementation and analysis

included in the thesis *Advancing Mouse-Tracking Research: New Solutions for Study Design, Implementation and Analysis* was primarily conceived and written by Pascal J. Kieslich, Ph. D. candidate at the Center for Doctoral Studies in Social and Behavioral Sciences of the Graduate School of Economic and Social Sciences, University of Mannheim. I sign this statement to the effect that Pascal J. Kieslich is credited as the primary source of ideas and the main author of the article.

October 2018 Michael Schulte-Mecklenbeck

**Co-Author: Martin Schoemann**

It is hereby confirmed that the article entitled

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

included in the thesis *Advancing Mouse-Tracking Research: New Solutions for Study Design, Implementation and Analysis* was primarily conceived and written by Pascal J. Kieslich, Ph. D. candidate at the Center for Doctoral Studies in Social and Behavioral Sciences of the Graduate School of Economic and Social Sciences, University of Mannheim. I sign this statement to the effect that Pascal J. Kieslich is credited as the primary source of ideas and the main author of the article.

October 2018 Martin Schoemann

**Co-Author: Tobias Grage**

It is hereby confirmed that the article entitled

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

included in the thesis *Advancing Mouse-Tracking Research: New Solutions for Study Design, Implementation and Analysis* was primarily conceived and written by Pascal J. Kieslich, Ph. D. candidate at the Center for Doctoral Studies in Social and Behavioral Sciences of the Graduate School of Economic and Social Sciences, University of Mannheim. I sign this statement to the effect that Pascal J. Kieslich is credited as the primary source of ideas and the main author of the article.

October 2018 Tobias Grage

**Co-Author: Johanna Hepp**

It is hereby confirmed that the article entitled

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

included in the thesis *Advancing Mouse-Tracking Research: New Solutions for Study Design, Implementation and Analysis* was primarily conceived and written by Pascal J. Kieslich, Ph. D. candidate at the Center for Doctoral Studies in Social and Behavioral Sciences of the Graduate School of Economic and Social Sciences, University of Mannheim. I sign this statement to the effect that Pascal J. Kieslich is credited as the primary source of ideas and the main author of the article.

October 2018 Johanna Hepp

**Co-Author: Stefan Scherbaum**

It is hereby confirmed that the article entitled

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

included in the thesis *Advancing Mouse-Tracking Research: New Solutions for Study Design, Implementation and Analysis* was primarily conceived and written by Pascal J. Kieslich, Ph. D. candidate at the Center for Doctoral Studies in Social and Behavioral Sciences of the Graduate School of Economic and Social Sciences, University of Mannheim. I sign this statement to the effect that Pascal J. Kieslich is credited as the primary source of ideas and the main author of the article.

I furthermore grant permission to Pascal J. Kieslich to use the article entitled

Stuck at the starting line: How the starting procedure influences mouse-tracking data

of which I am the first author, as part of his doctoral thesis. He contributed significantly to the conceptualization of the study, by analyzing the data for the “Discrete effects” and “Movement consistency” sections, and by writing parts of the manuscript.

October 2018 Stefan Scherbaum

## Copies of Articles

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods, 49*(5), 1652-1667.

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., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2018). *Design factors in mouse-tracking: What makes a difference?* Manuscript submitted for publication.

Scherbaum, S., & Kieslich, P. J. (2018). Stuck at the starting line: How the starting procedure influences mouse-tracking data. *Behavior Research Methods*, *50*(5), 2097-2110.

Behav Res (2017) 49:1652–1667
DOI 10.3758/s13428-017-0900-z

# Mousetrap: An integrated, open-source mouse-tracking package

**Pascal J. Kieslich**[1,2] · **Felix Henninger**[2,3,4]

Published online: 23 June 2017


**Abstract** Mouse-tracking – the analysis of mouse movements in computerized experiments – is becoming increasingly popular in the cognitive sciences. Mouse movements are taken as an indicator of commitment to or conflict between choice options during the decision process. Using mouse-tracking, researchers have gained insight into the temporal development of cognitive processes across a growing number of psychological domains. In the current article, we present software that offers easy and convenient means of recording and analyzing mouse movements in computerized laboratory experiments. In particular, we introduce and demonstrate the mousetrap plugin that adds mouse-tracking to OpenSesame, a popular general-purpose graphical experiment builder. By integrating with this existing experimental software, mousetrap allows for the creation of mouse-tracking studies through a graphical interface, without requiring programming skills. Thus, researchers can benefit from the core features of a validated software package and the many extensions available for it (e.g., the integration with auxiliary hardware such as eye-tracking, or the support of interactive experiments). In addition, the recorded data can be imported directly into the statistical programming language R using the mousetrap package, which greatly facilitates analysis. Mousetrap is cross-platform, open-source and available free of charge from https://github.com/pascalkieslich/mousetrap-os.

**Keywords** Mouse-tracking · Experimental design · Software · Response dynamics · Process tracing · OpenSesame · Python

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

1 Experimental Psychology, School of Social Sciences, University of Mannheim, Schloss Ehrenhof Ost, D-68131 Mannheim, Germany

2 Center for Doctoral Studies in Social and Behavioral Sciences, University of Mannheim, Mannheim, Germany

3 Cognitive Psychology Lab, University of Koblenz-Landau, Landau, Germany

4 Max Planck Institute for Research on Collective Goods, Bonn, Germany

## Introduction

Mouse-tracking – the recording and analysis of mouse movements in computerized experiments – is becoming an increasingly popular method of studying the development of cognitive processes over time. In mouse-tracking experiments, participants typically choose between different response options represented by buttons on a screen, and the position of the mouse cursor is continuously recorded while participants move towards and finally settle on one of the alternatives (Freeman & Ambady, 2010). Based on the theoretical assumption that cognitive processing is continuously revealed in motor responses (Spivey & Dale, 2006), mouse movements are taken as indicators of commitment to or conflict between choice options during the decision process (Freeman, Dale, & Farmer, 2011).

Mouse-tracking was first introduced as a paradigm in the cognitive sciences by Spivey, Grosjean, and Knoblich (2005). In their study on language processing, participants received auditory instructions to click on one of two objects (e.g., "click the candle"). A picture of the target object was presented together with a picture of a distractor that was either phonologically similar (e.g., "candy") or dissimilar (e.g., "dice"). Participants' mouse movements were more curved towards the distractor if it was phonologically similar than if it was dissimilar, suggesting a parallel processing of auditory input that activated competing representations.

Following Spivey et al. (2005), mouse-tracking has been used to gain insight into the temporal development of cognitive processes in a growing number of psychological domains, such as social cognition, decision making, and learning (for a review, see Freeman et al., 2011). More recently, researchers have extended the initial paradigm, combining mouse-tracking with more advanced methods. For example, mouse-tracking has been used in conjunction with eye-tracking to study the dynamic interplay of information acquisition and preference development in decision making under risk (Koop & Johnson, 2013). In an experiment with real-time interactions between participants, mouse-tracking uncovered different degrees of cognitive conflict associated with cooperating versus defecting in social dilemmas (Kieslich & Hilbig, 2014). As these examples show, an increasing number of researchers with different backgrounds and demands are using mouse-tracking to study cognitive processes. As a tool, mouse-tracking is increasingly combined with other methods to build complex paradigms and to integrate data across sources, leading to a richer understanding of cognition.

So far, many researchers conducting mouse-tracking studies have built their own experiments manually in code (e.g., Koop & Johnson, 2013; Scherbaum, Dshemuchadse, Fischer, & Goschke, 2010). These custom implementations were often one-off solutions tailored to a specific paradigm, and accompanied by custom analysis code to handle the resulting data specifically and exclusively. Researchers have spent considerable effort and technical expertise building these codebases.

As an alternative, other researchers have used *MouseTracker* (Freeman & Ambady, 2010), a stand-alone program for mouse-tracking data collection and analysis. Its ability to build simple experiments relatively quickly and design the mouse-tracking screen via a graphical user interface, as well as its integrated analysis tools have made mouse-tracking studies accessible to a broader range of researchers. However, researchers choosing MouseTracker lose the flexibility that general-purpose experimental software provides, in particular the ability to implement complex experimental designs within a single tool (involving, e.g., individually generated stimulus material, real-time communication between participants, and/or the inclusion of additional devices for data collection). In addition, many experimental software packages provide a graphical user interface not only for the design of single trials but of the entirety of the experimental procedure. Finally, most experimental software offers a scripting language so that its built-in features can be customized and extended. Although MouseTracker is free of charge (as citation-ware), the source code is not openly available and thereby not open to extensions and customization, limiting its features to those provided by the original authors. Moreover, MouseTracker is only available for the Windows operating system.

Going beyond custom implementations and stand-alone software solutions, there is a third option, namely providing modular components that extend existing experimental software. By building on the user-friendliness and flexibility of these existing tools, complex and highly customized experiments can be created easily, often without resorting to code. By using established open data formats for storage of mouse trajectories alongside all other data, preprocessing and statistical analyses are possible in common analysis frameworks such as R (R Core Team, 2016).

In this article, we present the free and open-source software *mousetrap* that offers users an easy and convenient way of recording mouse movements. Specifically, we introduce a plugin that adds mouse-tracking to *OpenSesame* (Mathôt, Schreij, & Theeuwes, 2012), a general-purpose graphical experiment builder. Together, these offer an intuitive, graphical user interface for creating mouse-tracking experiments that requires little to no further programming. Users can thus not only draw upon the extensive built-in functionality of OpenSesame for designing stimuli and controlling the experimental procedure, but also on additional plugins that extend it further, adding for example eye-tracking functionality (using PyGaze; Dalmaijer, Mathôt, & Van der Stigchel, 2014) and real-time interaction between participants (using Psynteract; Henninger, Kieslich, & Hilbig, in press). Yet further customization is possible through Python inline scripts. Like OpenSesame, mousetrap is available across all major platforms (Windows, Linux, and Mac).

In summary, mousetrap provides a flexible, extensible, open mouse-tracking implementation that integrates seamlessly with the graphical experiment builder OpenSesame and can be included by drag-and-drop in any kind of experiment. Its open data format allows users to analyze the data with a software of their choice. In particular, the recorded data can be imported directly into the statistical programming language R using the mousetrap package (Kieslich, Wulff, Henninger, Haslbeck, & Schulte-Mecklenbeck, 2016), which allows users to process, analyze, and visualize the collected mouse-tracking data.

In the following, we provide a brief introduction to mousetrap in combination with OpenSesame, and demonstrate how a mouse-tracking experiment can be created, what the resulting data look like, and how they can be processed and analyzed. In doing so, we create an experiment based on a classic mouse-tracking study by Dale, Kehoe, and Spivey (2007). In this study, participants' mouse movements are recorded while they classify exemplars (specifically: animals) into one of two categories; for example, a participant might be asked to classify a cat as mammal or reptile. The central independent variable in this paradigm is the typicality of the exemplar for its category: Exemplars are either typical members of their category, as above, or they are atypical (e.g., a whale), in that that they share both features with the correct (mammal) and a competing category (fish). The central hypothesis tested in this paradigm is that there should be more conflict between response options when classifying an atypical exemplar, and that mouse movements should therefore deviate more towards the competing category for atypical as compared to typical exemplars.

## Building a mouse-tracking experiment

In the following, we provide a brief tutorial for building a mouse-tracking experiment with mousetrap, demonstrating the plugin's major features as we do so. Our final result will be a simplified version of Experiment 1 by Dale et al. (2007). This study incorporates many features of a typical mouse-tracking study: participants are presented with simple stimuli (here only a single word) in a forced-choice design with two response alternatives (one of which represents the correct response). Besides, a within-participants factor (typicality) is manipulated with a directed hypothesis regarding its influence on mouse movements.

### Plugin installation and overview

Mousetrap depends on OpenSesame (version ≥ 3.1.0), which is available free of charge for all major operating systems from http://osdoc.cogsci.nl/, where it is also documented in depth. Mousetrap itself is available from GitHub (https://github.com/pascalkieslich/mousetrap-os), and is added to OpenSesame as a plugin.[1] The plugin includes built-in offline help and documentation for all features. Additional online resources are available from the GitHub repository, which offers extensive documentation and several example experiments, including the one built in the following (https://github.com/pascalkieslich/mousetrap-os#examples).

OpenSesame provides a graphical user interface through which users can create a wide range of experiments without programming. The building blocks of OpenSesame experiments are different *items*, from which an entire experiment can be assembled by drag-and-drop. For example, one might use a sketchpad item to present a visual stimulus, a keyboard_response or mouse_response item to record key presses or mouse clicks in response to the stimulus, and a logger item to write the collected data into a log file. Where desired, Python code can be included in an experiment using inline_script items to add further functionality. All of these items can be organized into sequences to run multiple items in direct succession and loops to repeat the same items multiple times (with variations). In a typical mouse-tracking experiment, a loop may contain the list of different stimuli that are presented in different trials, while a sequence contains all the items that are needed for each trial.

The items provided by the mousetrap plugin allow users to include mouse-tracking in any experiment using the same drag-and-drop operations and with the same ease. As OpenSesame provides two different ways of building displays, the mousetrap plugin contains two corresponding items: the mousetrap_response and the mousetrap_form item. Both provide comparable mouse-tracking functionality, but differ in the way the stimulus display is designed.

The *mousetrap_response* item tracks mouse movements while the stimulus display is provided by another item – typically by a sketchpad item that offers a graphical user interface for stimulus design. The mousetrap_response item then monitors the cursor position and registers button clicks.

In comparison, the *mousetrap_form* item extends the built-in OpenSesame form_base item to provide both a visual display as well as mouse-tracking. The visual content (e.g. text, images, and buttons) can be specified directly from within the item using a simple syntax and positioned on a user-defined grid.

Both the mousetrap_response and the mousetrap_form can be used without writing Python code. For even more flexibility, both items provide corresponding Python classes which can be accessed directly from code. Examples as well as documentation for these are provided online.

### Creating a mouse-tracking trial

Figure 1 shows the structure of our example experiment. In the beginning of the experiment, a *form_text_display* item labelled "instructions" is included to explain the task to participants. Next, a *loop* item called "stimuli" is added, which repeats the same sequence of items in each trial while varying the exemplars and response categories in random order (this data, along with additional metadata, is entered in the loop in tabular format – see bottom right of Fig. 1, where each row corresponds to one stimulus and the associated response options).

A simple way to create a mouse-tracking trial via the graphical user interface is to use a sketchpad item to create the visual stimulus display and a subsequent mousetrap_response item to track the mouse movements while the sketchpad is presented. Before creating the individual items, the overall experiment resolution should be set to match the resolution that will be used during data collection, because sketchpad items run at a fixed resolution and do not scale with the display size. As mouse-tracking experiments are normally run in full-screen mode, the experiment resolution will typically correspond to the display resolution of the computers on which the experiment will be conducted.

The trial *sequence* itself begins with a *form_text_display* item that contains a start button in the lower part of the screen, as is typical for mouse-tracking experiments (Freeman & Ambady, 2010). Participants start the stimulus presentation by clicking on this button, which also ensures that the start position of the cursor is comparable across trials. Using a form_text_display item is the most basic way of implementing

[1] Information on installing the plugin is provided at https://github.com/pascalkieslich/mousetrap-os#installation

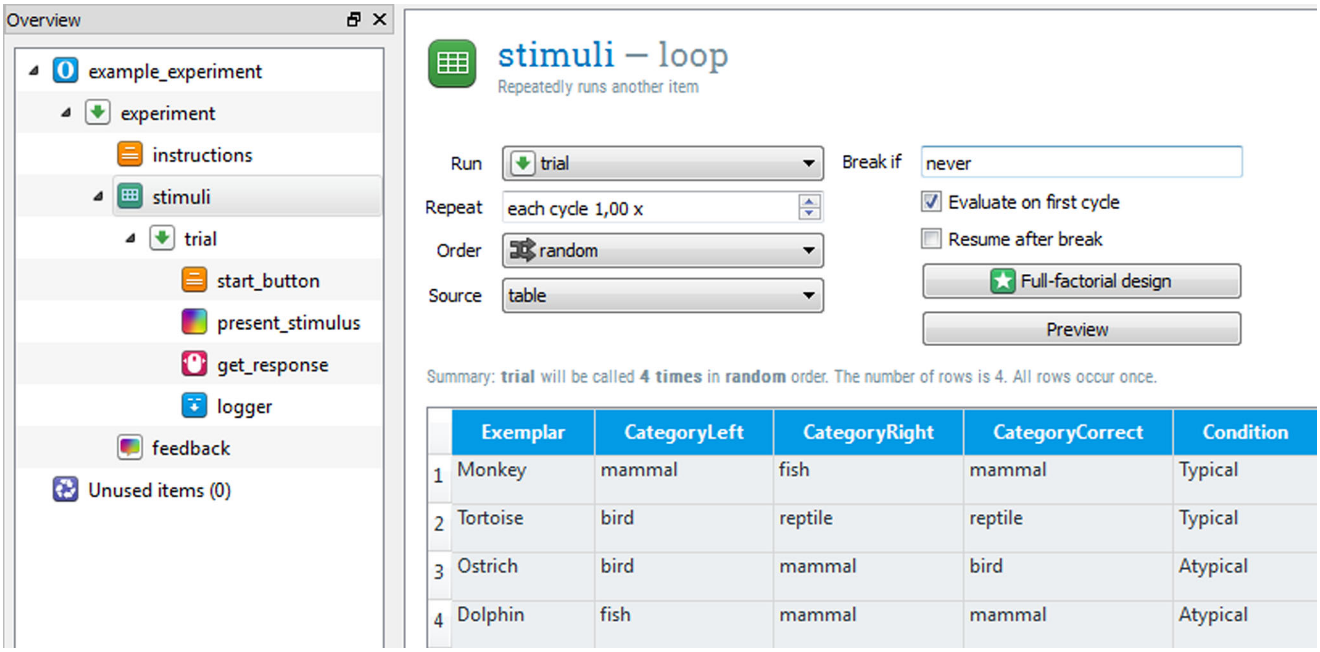


**Fig. 1** Structure of the example OpenSesame experiment (left) and settings for the stimuli loop (right). The panel on the left provides an overview of all items in the experiment, organized in a sequential (from top to bottom) and hierarchical (from left to right) display. On the highest (i.e., leftmost) level, the experiment sequence contains the instructions, the stimuli loop that generates the individual trials, and a final feedback screen. The loop contains a trial sequence, which is subdivided into the start button screen, a sketchpad that presents the stimulus, a mousetrap_response item that collects the participant's response and tracks cursor movements, and a logger item to save the data into the logfile. On the right, the details of the loop are visible. The design options at the top configure the loop such that each stimulus is presented once in random order, and the table at the bottom contains the actual stimulus data for four trials, namely the exemplar and response categories to be shown on screen, the correct response, and the experimental condition for inclusion in the dataset

a start screen because it provides a ready-made layout including some adaptable instruction text and a centered button which can be used to start the trial. Further customization of the start screen is possible, for example, by instead using an additional sketchpad – mousetrap_response combination (as was done in the experiment reported below; see also the online example experiment without forms).

The start item is followed by a *sketchpad* that defines the actual stimulus (Fig. 2). In the most general terms, a typical mouse-tracking task involves the presentation of a stimulus (e.g., a name or picture of an object), and several buttons. In the current study, the buttons correspond to different categories, and the participant's task is to indicate which category the presented exemplar (i.e., the name of the animal as text) belongs to by clicking on the corresponding button.

The most important part of the mouse-tracking screen is the exemplar that is to be categorized. It is added to the sketchpad using a *textline element* which allows for creating formatted text. To vary the presented text in each trial and insert the data from the loop (cf. Fig. 1), the corresponding variable name can be added in square brackets.

Creating button-like elements on a sketchpad item consists of two steps. First, the borders of the buttons are drawn using *rect elements*. Next, the button labels are inserted using *textline elements* (again using the variable names from the loop in square brackets). When designing the buttons, a symmetrical layout is desirable in most cases. Importantly, all buttons should have the same distance from the starting position of the mouse. Typically, the buttons are placed in the corners of the screen so that participants can easily reach them without risking overshooting the button, yet the distance between buttons is maximal.

As the tracking of mouse movements should start immediately when the sketchpad is presented, the duration of the sketchpad is set to 0 and a *mousetrap_response* item is inserted directly after the sketchpad in the trial sequence (see Fig. 1, where the mousetrap_response item is labelled "get_response"). Because the mousetrap_response item is separated from the stimulus display, the number of buttons[2] as well as their location and internal name need to be provided (see Fig. 3). In our case, and indeed for the majority of experiments, the buttons correspond to the rectangles added to the sketchpad earlier. Thus, the appropriate values for x and y coordinates as well as width and height can be copied from the element script, which can be accessed by double-clicking on its border (Fig. 2). In addition to the coordinates, each

[2] The mousetrap_response item supports up to four buttons. More can be added by using the mousetrap_form item or by defining buttons in Python code.

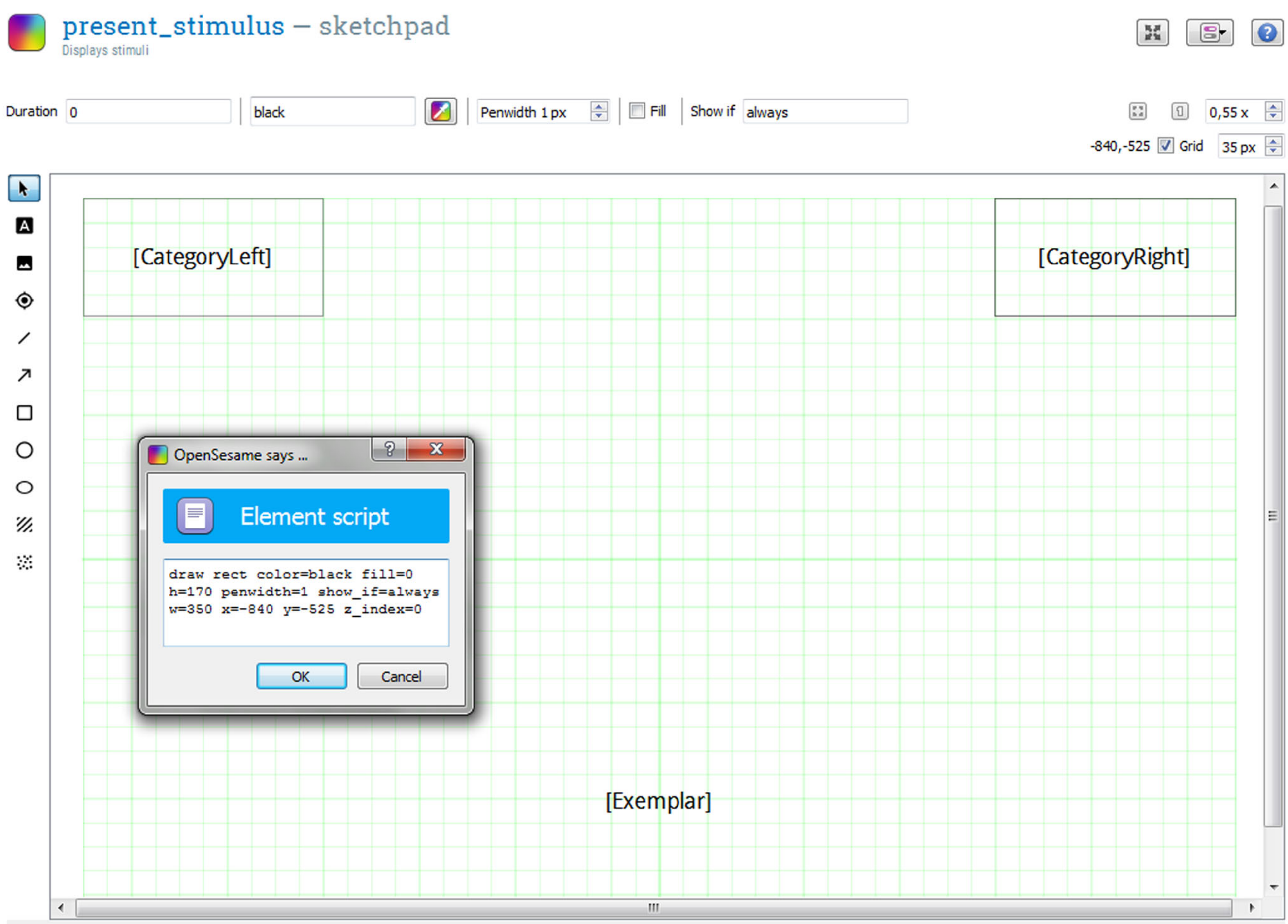


**Fig. 2** Exemplary sketchpad item containing two buttons and a stimulus. The drawing tools used to create the stimulus are shown on the left: The button labels and the stimulus are created using textline elements. As they vary for each trial, the experimental variables defined in the stimuli loop (cf. Fig. 1) are used by enclosing the variable name in square brackets, so that their values will be substituted when the experiment runs. The button borders are drawn using rect elements. The underlying element script for each button can be accessed by double-clicking on the respective rectangle: the script corresponding to the left button is shown in the pop-up window. The *x*, *y*, *w*, and *h* arguments define the left and top coordinates of the rectangle and its width and height. They can be copied and pasted into the mousetrap_response item (cf. Fig. 3) to define the buttons

button receives a name argument that will be saved as response when a participant clicks within the area of the corresponding button. We recommend using the text content of the button for this purpose (e.g., `name =[CategoryLeft]` for the left button in Fig. 2).

When the response options are named, a correct response can be defined by adding the corresponding button's name in the respective field. OpenSesame will then automatically code the correctness of the response (as 1 or 0) in the variable labelled *correct*, which is included in the data for later analysis and can also be used to provide feedback during the study. As with the labels, the correct response in each trial is determined based on the variables specified in the loop (variable *CategoryCorrect*, cf. Fig. 1).

In addition to logging the correctness of a single response, OpenSesame's global feedback variables (e.g., the overall accuracy) can be updated automatically by selecting the corresponding option, which makes it easy to, for example, pay participants contingent on their performance. In the current experiment, participants are provided with feedback on their performance on the last screen of the experiment through this mechanism.

The cursor position is recorded as long as the mousetrap_response item is active. The interval in which the positions are recorded is specified under *logging resolution* in the item settings (see Fig. 3). By default, recording takes place every 10 ms (corresponding to a 100-Hz sampling rate). The actual resolution may differ depending on the performance of the hardware (but has proven to be very robust in our studies, see example experiment below and software validation in the Appendix).

Finally, a *logger* item is inserted at the end of the trial sequence (see Fig. 1). This item writes the current state of all variables to the participant's log file, which will later be used

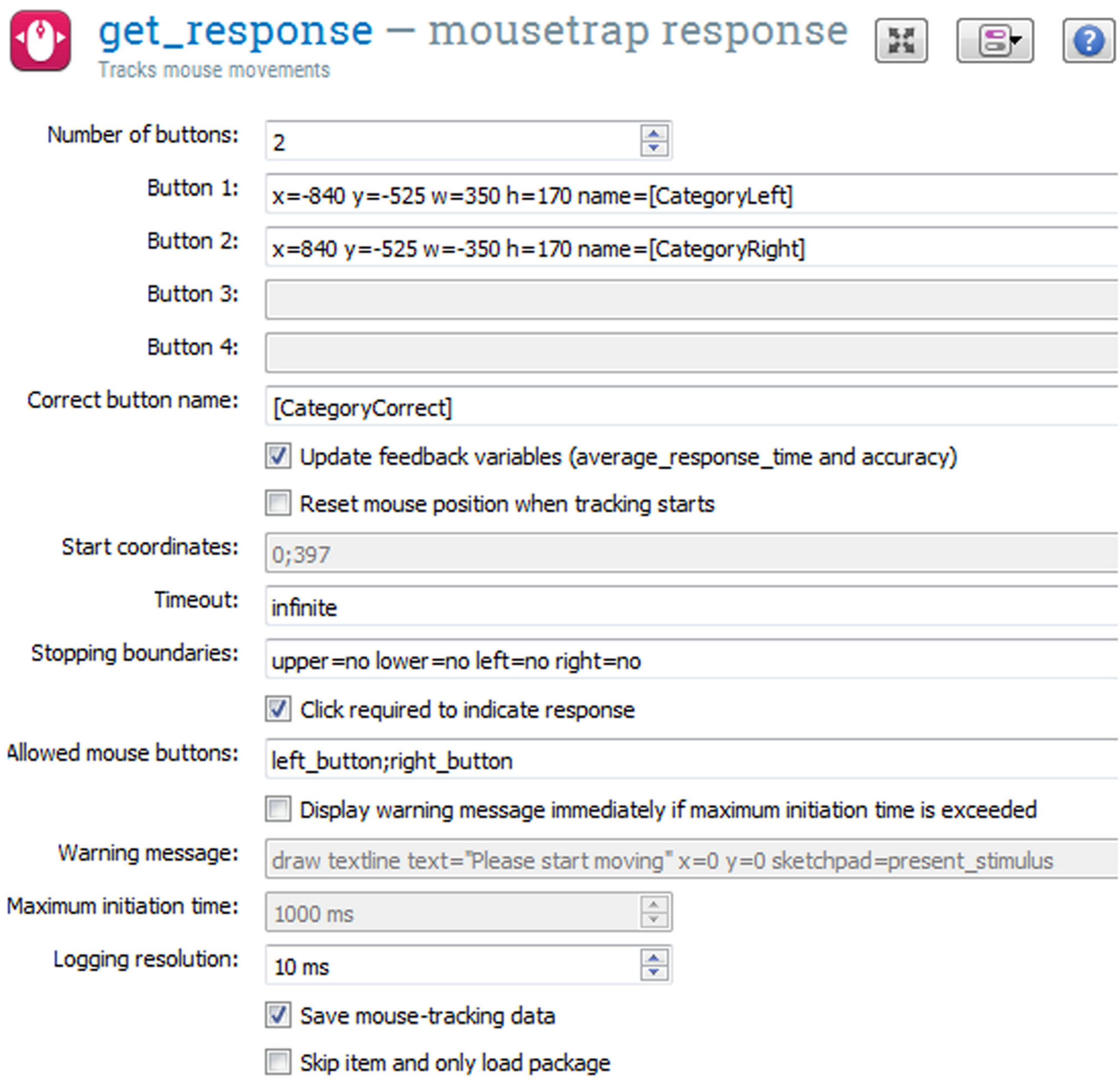


**Fig. 3** Settings of the mousetrap_response item: The topmost settings define the number of buttons used, as well as their position (using the arguments from the rect element script, cf. Fig. 2) and internal name (the button label that was defined in the stimuli loop, cf. Fig. 1). The correct answer can be specified in the Correct button name option to make use of OpenSesame's feedback capabilities. If desired, the mouse cursor can be reset to exact start coordinates at tracking onset. Optionally, a timeout (in ms) can be specified to restrict the time participants have to give their answer. The boundary setting can be used to terminate data collection if the cursor crosses a specified vertical or horizontal boundary on the screen. Additional options concern the possibility to restrict the mouse buttons available for responding, the immediate display of a warning if cursor movement is not initiated within a given interval, and the adjustment of the logging resolution, that is, the interval between subsequent recordings of the cursor position

in the analysis. The variable inspector can be used to monitor the current state of the variables in the experiment if it is run from within OpenSesame. The central mouse-tracking data recorded through mousetrap items is stored in variables starting with *timestamps*, *xpos*, and *ypos*.

## Alternative implementation using forms

As mentioned above, mousetrap also provides an alternative way of implementing mouse-tracking via the *mousetrap_form* item. In contrast to the mousetrap_response item, the display is defined directly within the item by using a form. Forms are a general item type which is used throughout OpenSesame. They place content (which is referred to as "widgets" and can include labels, images, buttons and image buttons) on a grid, which allows forms to scale with the display resolution. Forms do not provide a graphical interface, but instead use a simple syntax to define and arrange the content.

In the current example, a mousetrap_form could replace both the "present_stimulus" sketchpad and the "get_response" mousetrap_response item. Assuming a grid with 16 columns and 10 rows, a visual stimulus display similar to Fig. 2 can be created as follows:

```
widget 6 7 4 2 label text="[Exemplar]"
widget 0 0 4 2 button text="[CategoryLeft]"
widget 12 0 4 2 button text="[CategoryRight]"
```

The numbers in the example define the position and extent of each widget on the grid, followed by the type of element and its specific settings. The additional mouse-tracking settings are largely identical to the settings of the mousetrap_response item (see Fig. 3 and the online example experiment demonstrating a mousetrap_form).

## Methodological considerations

With the basic structure of the experiment in place, the stimulus display designed and the mouse-tracking added, the experiment would now be ready to run. However, some additional methodological details should be given consideration. Mouse-tracking studies in the literature differ in many methodological aspects, depending on the implementation and researchers' preferences. We can provide no definitive recommendations, but we aim to cover most common design choices and their implementation using the mousetrap plugin in the following (see also Fischer & Hartmann, 2014; Hehman, Stolier, & Freeman, 2015, for recommendations regarding the setup of mouse-tracking experiments).

### *General display organization*

One general challenge is the design of the information shown during the mouse-tracking task. Because mouse movements should reflect the developing commitment to the choice options rather than information search, the amount of new information that participants need to acquire during tracking should be minimized. At the same time, some information must be withheld until tracking begins, so that participants develop their preferences only during the mouse-tracking task and not before.

To some degree, this also represents a challenge for the current example experiment, where in addition to the name of the exemplar, the information about the two response categories needs to be acquired. Dale et al. (2007) solved this by presenting the response categories for 2,000 ms at the beginning of each trial, even before the start button appeared. The experiment sketched above can be adapted to implement this procedure by including an additional sketchpad in the beginning of the trial sequence that presents only the buttons and their labels for a specified duration. This procedure was also used in the experiment reported below.

Note that other mouse-tracking studies have used an alternative approach by presenting the critical information acoustically (e.g., Spivey et al., 2005). One advantage of this approach is that it prevents any artifacts that might be caused from reading visually presented information. This approach can be implemented easily in OpenSesame, for example, by inserting a sampler item between the sketchpad and the mousetrap_response item.

### *Starting position*

As previously discussed, it is often desirable to have a comparable starting position of the cursor across trials, as is achieved through the start button in our experiment. However, this method only leads to generally comparable, but not identical starting positions across trials. Though the start coordinates can be aligned during the later analysis, the cursor position can also be reset to exact coordinates by the experimental software before the tracking starts. This can be achieved by checking the corresponding option, and the start coordinates can be specified as two integers (indicating pixel values in the sketchpad metric where "0;0" represents the screen center). These values are usually chosen to correspond to the center of the start button, so that the jump in position is minimized (the mousetrap_response item by default uses start coordinates that correspond to the center of the button on a form_text_display item).

### *Movement initialization*

In many mouse-tracking studies, participants are explicitly instructed to initiate their mouse movement within a certain time limit (as described by Hehman et al., 2015) while other studies refrain from giving participants any instructions regarding mouse movement (e.g., Kieslich & Hilbig, 2014; Koop & Johnson, 2013). If such an instruction is given, compliance will typically be monitored and participants may be given feedback. The mousetrap items provide several ways of implementing this. The items automatically compute the *initiation_time* variable that contains the time it took the participant to initialize any mouse movement in the trial. This variable can be used to give feedback to the participant after the task, for example, by conditionally displaying a warning message if the initiation time is above a predefined threshold.

Alternatively, it is also possible to display a warning message while the mouse-tracking task is running. In this case, the time limits and the customized warning message can be specified in the item settings (see Fig. 3). We recommend not using this second option during the actual mouse-tracking task to avoid distracting participants. However, it might be useful in initial practice trials.

Going beyond a mere a priori instruction to initiate movement quickly, some studies have also used a more advanced procedure implementing a dynamic start condition (e.g., Dshemuchadse, Scherbaum, & Goschke, 2013; Frisch, Dshemuchadse, Görner, Goschke, & Scherbaum, 2015). In these studies, the critical stimulus information was presented only after participants crossed an invisible horizontal boundary above the start position, ensuring that movement had already been initiated. A dynamic start condition can be implemented by including an additional sketchpad and mousetrap_response item specifying an upper boundary for tracking in the item settings (see corresponding online example experiment).

In a first attempt to assess the influence of the starting procedure on mouse-tracking measures, Scherbaum and Kieslich (2017) compared data from an experiment using such a dynamic start condition to a condition in which the stimulus was presented after a fixed delay. While results showed that theoretically expected effects on trial-level mouse-tracking measures (i.e., trajectory curvature) were reliably found in both conditions, effects on within-trial continuous measures were stronger and more temporally distinguishable in the dynamic start condition. This was in line with generally more consistent and homogeneous movements in the dynamic start condition.

Another alternative to ensure a quick initialization of mouse movements is to restrict the time participants have for giving their answer. This time limit can be specified (in ms) in the corresponding option in the item settings (Fig. 3).[3]

*Response indication*

An additional methodological factor that varies across mouse-tracking studies is the way participants indicate their response. While many studies require participants to click on the button representing their choice (e.g., Dale et al., 2007; Koop, 2013), in other studies merely entering the area corresponding to the button with the cursor is sufficient (e.g., Dshemuchadse et al., 2013; Scherbaum et al., 2010). Both options are available in mousetrap (see Fig. 3). If a click is required, the (physical) mouse buttons that are accepted as a response indication can be specified. By default, mouse clicks are required and both left and right mouse clicks are accepted.

[3] However, introducing a time limit might also induce time pressure which might lead to other (undesired) effects.

*Counterbalancing presentation order*

A final consideration should be given to potential position effects: So as not to introduce confounds between response alternatives and the position of the corresponding button, the mapping should be varied across trials and / or across participants. This is especially important if the response alternatives stay constant across trials (which is often the case in decision making studies, e.g., Kieslich & Hilbig, 2014; Koop, 2013). In the current study, the position of the correct response and the foil (left vs. right) should be varied. This can be done statically by varying their order across trials (see Fig. 1). To go further, the position of response options can be randomized at run time using OpenSesame's advanced loop operations (as was done in the experiment reported below, see also *shuffle_horiz* online example experiment).

## Data collection

After creating the mouse-tracking experiment, it should be tested on the computers that will later be used to collect the data. We also recommend importing and analyzing self-created test data to check that all relevant independent and dependent variables have been recorded, and to check the logging resolution (see below). When preparing the study for running in the lab, a number of methodological factors need to be considered.

As noted in the previous section, mouse-tracking experiments should be run in full screen mode at the maximum possible screen resolution. The *OpenSesame Run* program, which is included with OpenSesame, can run the experiment without having to open it in the editor, making the starting process more efficient, and hiding the internal structure, conditions, and item names from participants.

In addition, the mouse sensitivity settings of the operating system should be checked and matched across laboratory computers, in particular the speed and acceleration of the cursor relative to the physical mouse (these settings cannot be influenced directly from within OpenSesame). There is currently no single setting applied consistently across studies in the literature, and the settings used in the field are often not reported. Presumably, the settings will often have been left to the operating system defaults (under Windows 7 and 10, medium speed with acceleration) or speed will have been reduced deliberately and acceleration turned off (as recommended by Fischer & Hartmann, 2014).

When preparing the laboratory, it should be ensured that participants have enough desk space to move the mouse. In this regard, we have found it useful to move the keyboard out of the way and design the experiment so that participants can complete the entire experiment by using only the mouse. Additionally, heretofore largely unexplored factors concern the handedness of participants, the hand used for moving the mouse, and their interplay. Some authors go as far as to recommend including only right-handed participants (Hehman et al., 2015). We would recommend assessing the handedness of participants, as well as the hand actually used for moving the mouse in the experiment.

In general, we would like to stress the importance of documenting mouse-tracking studies in sufficient detail, both so that fellow researchers can replicate the experiment and so that potentially differing findings between individual mouse-tracking studies can be traced back to differences in their methodological setup. Ideally, each of the degrees of freedom sketched above should be documented, as well as the specifics of the lab computers (especially screen resolution and mouse sensitivity settings). It is also very useful to provide a screenshot of the actual mouse-tracking task. Finally, to give interested colleagues the opportunity to explore the specific details of the task setup, it is also useful to provide them directly with the experiment files. This is particularly easy if mouse-tracking experiments are created in OpenSesame with the mousetrap plugin, as OpenSesame is freely available for many platforms. OpenSesame also provides the option to automatically save experiments on the Open Science Framework and share them with other researchers.

## Example experiment

Having built and tested the experiment, enterprising colleagues could begin with the data collection immediately. We have done exactly this, and have performed a replication of Experiment 1 by Dale et al. (2007). In doing so, we aimed to assess the technical performance of the plugins (especially with regard to the logging resolution), to demonstrate the structure of the resulting data and how they can be processed and analyzed, and to replicate the original result that atypical exemplars lead to more curved trajectories than typical exemplars. The exact experiment that was used in the study (with German material and instructions) and a simplified but with regard to the task identical version (with English example material and instructions) can be found online at https://github.com/pascalkieslich/mousetrap-resources, as can the raw data and analysis scripts.

### Methods

We used the 13 typical and 6 atypical stimuli from Dale et al.'s Experiment 1 (see Table 1 in Dale et al., 2007) translated to German. Participants first received instructions about their task and completed three practice trials. Thereafter, the 19 stimuli of interest were presented in random order. Participants were not told that their mouse movements were recorded, nor did they receive any specific instructions about moving the mouse.

Each trial began with a blank screen that was presented for 1,000 ms. After that, the two categories were displayed for 2,000 ms in the top left and right screen corners (the order of the categories was randomized at run time), following the procedure of the original study. Next, the start button appeared in the bottom center of the screen, and participants started the trial by clicking on it. Directly thereafter (the cursor position was not reset in this study), the to-be-categorized stimulus word was displayed above the start button and participants could indicate their response by clicking on one of the two categories (see Fig. 2).

The experiment was conducted full screen with a resolution of 1,680 × 1,050 pixels. Laboratory computers were running Windows 7, and mouse settings were left at their default values (acceleration turned on, medium speed). Cursor coordinates were recorded every 10 ms.

The experiment was conducted as the second part in a series of unrelated studies. Before the experiment, we assessed participants' handedness using the Edinburgh Handedness Inventory (EHI; Oldfield, 1971). We used a modified version of the EHI with a five-point rating scale on which participants indicated which hand they preferred to use for ten activities (-100 = exclusively left, −50 = preferably left, 0 = no preference, 50 = preferably right, 100 = exclusively right) and included an additional item for computer mouse usage.

Participants were recruited from a local student participant pool at the University of Mannheim, Germany, and paid for their participation (the payment was variable and depended on other studies in the same session). Participants were randomly assigned to either an implementation of the study using the mousetrap plugin in OpenSesame ($N = 60$, 39 female, mean age = 22.2 years, $SD = 3.5$ years) or another implementation (a development version of an online mouse-tracking data collection tool) not included in the current article. Participants' mean handedness scores based on the original EHI items indicated a preference for the right hand for the majority of participants (50 of 60 participants had scores greater than 60), no strong preference for eight participants (scores between −60 and 60) and preference for the left hand for two participants (below −60). Interestingly, all participants reported using a computer mouse preferably or exclusively with the right hand, as indicated by the newly added item.

**Data preprocessing**

In the following section, we focus on a simple but frequently applied comparison of (aggregate) mouse trajectory curvature.[4] In doing so, we will go through all analysis steps from loading the raw data to the statistical tests in the statistical programming language R (R Core Team, 2016). The complete analysis script is shown in Fig. 4.

The libraries required for the following analyses can be installed from CRAN using the following command: `install.packages(c("readbulk","mousetrap"))`. Thereafter, both libraries are loaded using `library(readbulk)` and `library(mousetrap)` respectively. We will only touch upon the most basic features of both; additional library-level documentation can be accessed with the command `package?mousetrap` (or online at http://pascalkieslich.github.io/mousetrap/), and help for specific functions is available by prepending a question mark to any given command, as in `?mt_import_mousetrap`.

OpenSesame produces an individual comma-separated (CSV) data file for each participant. Because there is a single logger item in the experiment that is repeated with each trial, every line corresponds to a trial. Different variables are spread across different columns. For our purposes, the most important columns are those containing the mouse-tracking data, namely the columns beginning with *timestamps*, *xpos*, and *ypos*. These columns contain the interval since the start of the experiment in milliseconds, and the x and y coordinates of the cursor at each of these time points. The position coordinates are given in pixels, whereby the value 0 for both x and y coordinates corresponds to the center of the screen and values increase as the mouse moves toward the bottom right.

As a first step after opening R (or RStudio), the current working directory should be changed to the location where the raw data is stored (either using *setwd* or via the user interface in RStudio). To read the data of all participants into R, we suggest the *readbulk* R package (Kieslich & Henninger, 2016), which can read and combine data from multiple CSV files into a single dataset. Readbulk provides a specialized function for OpenSesame data (*read_opensesame*). Assuming that the raw data is stored in the subfolder "raw_data" of the working directory, we can combine all individual files into a single data.frame using `read_opensesame("raw_data")`.

Next, the raw data are filtered so that only the trials of interest are retained. Specifically, all trials from the practice phase are excluded. Besides, we determined which trials were solved correctly using the *correct* variable, which was automatically set by the mousetrap_response item. The accuracy in the current study was 88.9% for atypical and 95.4% for typical trials – results comparable to those in the original study. Following Dale et al., only the correctly completed trials were kept for the analyses.

For preprocessing and analyzing mouse-tracking data, we have developed the *mousetrap* R package (Kieslich et al., 2016). A detailed description of the package and its functions is provided elsewhere (Kieslich, Wulff, Henninger, Haslbeck, & Schulte-Mecklenbeck, 2017). In the following, we will focus on the most basic functions needed for the present analyses.

As a precondition for further analysis, the raw data must be represented as a mousetrap data object using the *mt_import_mousetrap* function. This function will automatically select the mouse-tracking data columns from the raw data[5] and transform their contents into a data structure amenable to analysis.

Next, several preprocessing steps ensure that the data can be aggregated within and compared meaningfully between conditions. Trajectories are remapped using *mt_remap_symmetric* which ensures that every trajectory starts at the bottom of the coordinate system and ends in the top left corner (regardless of whether the left or the right response option was chosen). Because the mouse cursor was not reset to a common coordinate at the start of tracking, *mt_align_start* is needed to align all trajectories to the same initial coordinates (0, 0). Trajectories are then typically time-normalized so that each trajectory contains the same number of recorded coordinates regardless of its response time (Spivey et al., 2005). To this end, *mt_time_normalize* computes time-normalized trajectories using a constant (but adjustable) number of time steps of equal length (101 by default, following Spivey et al.).

Several different measures for the curvature of mouse trajectories have been proposed in the literature (Freeman & Ambady, 2010; Koop & Johnson, 2011). One frequently used measure is the maximum absolute deviation (MAD). The MAD represents the maximum perpendicular deviation of the actual trajectory from the idealized trajectory, which is the straight line connecting the trajectories' start and end points.[6] The MAD and many additional trial-level measures can be calculated using the *mt_measures* function.[7] These measures are then typically aggregated per participant for each level of the within-participants factor. For this, *mt_aggregate_per_subject* can be used (see Fig. 4).

[4] These analyses differ from the more elaborate analyses in the original article by Dale et al. (2007), which we have omitted for reasons of brevity. We provide an R script for replication of the original analyses online.

[5] In case that more than one mousetrap item is included in the experiment, the names of the columns need to be provided explicitly using the corresponding arguments.

[6] If this maximum deviation occurs in the direction of the non-chosen option (i.e., "above" the idealized trajectory), it receives a positive sign, otherwise a negative sign.

[7] This function uses the raw trajectories by default to avoid the (unlikely) possibility that relevant spatial information gets lost during time normalization. In the current sample, the MAD values based on the raw trajectories and on the time-normalized trajectories correlate to .9999.

```r
library(readbulk)
raw_data <- read_opensesame("raw_data", verbose = FALSE)
raw_data <- subset(raw_data, Condition!="Example" & correct==1)

library(mousetrap)
mt_data <- mt_import_mousetrap(raw_data)
mt_data <- mt_remap_symmetric(mt_data)
mt_data <- mt_align_start(mt_data)
mt_data <- mt_measures(mt_data)

agg_measures <- mt_aggregate_per_subject(mt_data, subject_id = "subject_nr",
  use_variables = "MAD", use2_variables = "Condition")
t.test(MAD~Condition, data = agg_measures, paired = TRUE)

mt_data <- mt_time_normalize(mt_data)
mt_plot_aggregate(mt_data, use = "tn_trajectories", points = TRUE,
  x = "xpos", y = "ypos", color = "Condition", subject_id = "subject_nr")
```

**Fig. 4** R script for replicating the main data preparation and analysis steps. First, the individual raw data files are merged and read into R. They are then filtered, retaining only correctly solved trials from the actual task. Next, the mouse-tracking data are imported and preprocessed by remapping all trajectories to one side, aligning their start coordinates and computing trial-level summary statistics (such as the maximum absolute deviation, MAD). The MAD values are aggregated per participant and condition, and compared using a paired *t*-test. Finally, the trajectories are time-normalized, aggregated per condition, and visualized

### Data quality check

To check whether the intended logging resolution was actually met, *mt_check_resolution* can be used to compute the achieved interval between logs. Across all recorded mouse positions in all trials that entered the following analyses, 99.4% of the logging intervals were exactly 10 ms, corresponding to the desired logging resolution. An additional 0.5% of intervals were shorter than 10 ms, due to the fact that every click in the experiment leads to an immediate recording of the current cursor position, even outside of the defined logging interval. Finally, 0.1% of logging intervals were greater than 10 ms, of which 76.2% lagged by 1 additional ms only. Overall, the mean timestamp difference was 9.98 ms ($SD$ = 0.43 ms).

A more comprehensive technical validation of the mousetrap plugin is reported in the appendix. Extending a procedure by Freeman and Ambady (2010), we used external hardware (Henninger, 2017) to generate known movement patterns from the start button to one of the response buttons. An analysis of the recorded cursor positions revealed that almost every change in position was captured on the raw coordinate level, and that the recorded positions and derived trial-level measures almost perfectly corresponded to their expected values.

### Results

A quick first visual impression of the effect of the typicality manipulation on mouse movements can be obtained by inspecting the aggregate mouse trajectories. Specifically, *mt_plot_aggregate* can be used to average the time-normalized trajectories per condition (first within and then across participants) and to plot the resulting aggregate trajectories (Fig. 5). In line with the hypothesis by Dale et al., the aggregate response trajectory in the atypical condition showed a greater attraction to the non-chosen option than the trajectory in the typical condition.

To statistically test for differences in curvature, the average MAD values per participant and condition can be compared. In line with the hypothesis and the visual inspection of the aggregate trajectories, the MAD was larger in the atypical ($M$ = 343.8, $SD$ = 218.6) than in the typical condition ($M$ =

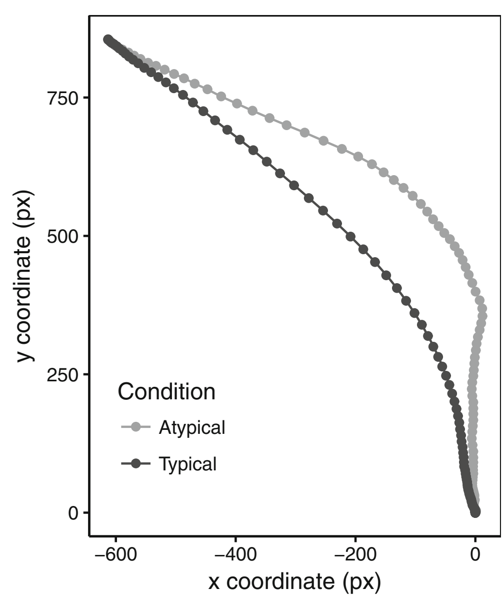


**Fig. 5** Average time-normalized trajectories per experimental condition

172.2, $SD = 110.8$). This difference was significant in a paired $t$-test, $t(59) = 6.73$, $p < .001$, and the standardized difference of $d_z = 0.87$ represented a large effect.

The analyses just described give an initial impression of what mouse-tracking data look like. While we have provided a first simple test of our basic hypothesis, the analysis has barely scratched the surface of what is possible with this data (and what can be realized using the mousetrap package). Specifically, we have skipped a number of important preprocessing and analyses steps that are standard procedure in mouse-tracking studies, such as the inspection of individual trials to detect anomalous or extreme mouse movements (Freeman & Ambady, 2010) and analyses to detect the presence of bimodality (Freeman & Dale, 2013).[8] The original article our study was based upon also contains many more analyses (see online supplementary material for a replication of the analyses by Dale et al., 2007, based on the current dataset).

Several more advanced analyses methods and measures have also been proposed, such as velocity and acceleration profiles, spatial disorder analyses via sample entropy, or the investigation of smooth versus abrupt response competition via distributional analyses (see, e.g., Hehman et al., 2015, for an overview). Many of these methods and measures are implemented in the mousetrap R package, and are described and explained in the package documentation. We discuss elsewhere in detail the methodological possibilities and considerations when processing and analyzing mouse-tracking data, as well as their implementation in mousetrap (Haslbeck, Wulff, Kieslich, Henninger, & Schulte-Mecklenbeck, 2017; Kieslich et al., 2017).

## Discussion

In this article, we presented the free and open-source software *mousetrap* that offers users easy and convenient means of recording mouse movements, and demonstrated how a simple experiment can be built and analyzed. Specifically, we introduced mousetrap as a plugin that adds mouse-tracking to the popular, open-source experiment builder OpenSesame, allowing users to create mouse-tracking experiments via a graphical user interface. To demonstrate the usage of mousetrap, we created and replicated a mouse-tracking experiment by Dale et al. (2007), and analyzed the resulting data using the mousetrap R package. In line with the original hypothesis and results, we found that mouse trajectories displayed greater curvature towards the competing response option for atypical compared to typical exemplars. Naturally, we have only been able to discuss the most salient decisions in the construction of mouse-tracking experiments. However, where possible, we have noted the additional degrees of freedom and design choices, and sketched their implementation.

Mousetrap offers an alternative to the two major ways mouse-tracking studies are currently implemented. First, researchers have built custom code-based implementations of mouse-tracking for specific paradigms. These custom-built experiments can be flexibly tailored to the individual researchers' needs, but their implementation requires extensive programming skills, and paradigms are often cumbersome to adapt to new tasks. Secondly, researchers have relied on MouseTracker (Freeman & Ambady, 2010), a specialized experimental software for building mouse-tracking experiments and analyzing the resulting data. While this software has made mouse-tracking studies accessible to more researchers by providing a visual interface for designing the mouse-tracking screen and recording the mouse movements, it forgoes the flexibility and many useful features of general-purpose experimental software (such as the option to define the structure of the experiment itself via a graphical user interface, or to directly include a scripting language for customization and run time adaptation).

Aiming to combine the advantages while avoiding the disadvantages of both approaches, mousetrap extends the general purpose graphical experiment builder OpenSesame (Mathôt et al., 2012). Thereby, it allows users to easily create mouse-tracking experiments via a graphical interface without requiring programming skills. In addition, it makes available the many useful features of OpenSesame, such as a user-friendly interface for designing the structure of the experiment and implementing advanced randomizations, the support for diverse audiovisual stimuli, an open data format, extensibility via Python scripts, and cross-platform availability.

While mouse-tracking is a frequently used method for assessing response dynamics (Koop & Johnson, 2011), it should be noted that other methods are also available, such as the use of remote controllers (e.g., a Nintendo Wii Remote, cf. Dale, Roche, Snyder, & McCall, 2008) or the direct recording of hand movements (via a handle, e.g., Resulaj, Kiani, Wolpert, & Shadlen, 2009, or using a motion capture system, e.g., Awasthi, Friedman, & Williams, 2011). Another approach that might become more important in future research is the tracking of finger (or pen) movements via touchscreens (e.g., Buc Calderon, Verguts, & Gevers, 2015; Wirth, Pfister, & Kunde, 2016) due to the increasing availability of tablets

[8] A simple bimodality analysis can be conducted by computing bimodality coefficients (BC). Following Freeman and Ambady (2010), we $z$-standardized MAD values per participant and computed the BC separately for the atypical and the typical condition. In both conditions, the BC was higher than the recommended cutoff (.555), $BC_{Typical} = .608$, $BC_{Atypical} = .593$, indicating a bimodal distribution. To analyze whether the difference in MAD between typicality conditions remained significant after excluding outliers, we excluded all trials with $|z_{MAD}| > 1.50$ and repeated the main analyses (for details, see online supplementary material). As in the complete dataset, aggregate MAD was significantly higher in the atypical than in the typical condition, $p < .001$. Note, however, that more advanced and comprehensive alternative analyses are available (Kieslich et al., 2017).

and smartphones. The mousetrap plugin could be extended to implement the latter approach in OpenSesame.

With mousetrap, we hope to make mouse-tracking accessible to researchers from many different fields, and thereby to enable them to gain insights into the dynamics of cognitive processes. Given the fast-paced development of the mouse-tracking method, we hope that our modular and open approach will help users to implement the increasingly complex designs, to combine mouse-tracking with other process tracing methods such as eye-tracking, and to apply the method in fields where only few mouse-tracking studies have been conducted so far, such as behavioral economics with real-time interactive experiments. Similarly, we hope that the open data format and the close link to open analysis tools such as those demonstrated herein will make the manifold methods of analyzing mouse-tracking data widely available.

**Acknowledgments** We thank Anja Humbs for testing a development version of the mousetrap plugin for OpenSesame, Monika Wiegelmann and Mila Rüdiger for collecting the data for the example experiment, and Arndt Bröder and Johanna Hepp for helpful comments on an earlier version of this manuscript. This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences funded by the German Research Foundation

## Appendix

### Software validation

To validate the data collection procedure, we extended the procedure employed by Freeman and Ambady (2010), who simulated and processed artificial mouse trajectories. We used external hardware (Henninger, 2017) to generate two known movement patterns that connected the start and the top left response button: either a diagonal line, or a triangular path leading only upward at first, and then left towards the response button (Fig. 6). The validation experiment was built in OpenSesame (version 3.1.6, using the *legacy* backend[9]) using the mousetrap_response item (version 1.2.1). The screen layout and mouse-tracking settings were identical to the example experiment reported in the main article (cf. Figs. 2 and 3). The study was run on a laboratory terminal with modest hardware (Windows 7 Professional, on an Intel Pentium Dual-Core running at 3 GHz with 4 GB RAM).

In the following simulations, we ventured to perform a strict test of the software: First, to test the performance of the data collection procedure under heavy load, we simulated rapidly changing cursor coordinates. Specifically, in all simulations, the cursor position was updated at the logging resolution (10 ms) to assess whether data is recorded correctly when the cursor position changes as fast as data are collected. On each update, the cursor moved to the next integer pixel location on its path, that is, both one pixel up and one left for 800 px for the diagonal trajectory, or first one pixel upwards for 800 px and then one left for 800 px for the triangular path. The trial was started by a (simulated) click on a start button, which initiated the display of the response buttons, and ended with a mouse click on the left response button, with pauses of 110 ms before movement initiation and 100 ms between the end of movement and the simulated response. This means that the time between the start and end click was 8,210 ms for the diagonal path and 16,210 ms for the triangular path. Second, we validate the resulting data at the lowest possible level, that is, using the raw trajectory coordinates of each individual (simulated) trial. In scientific practice, standard mouse-tracking analyses will compensate for imperfect measurement to some degree because mouse trajectories are typically time-normalized and analyses are based on aggregate statistics.

For both the diagonal and the triangular path, we simulated 1,000 trials. The resulting data files were read into R and processed and analyzed using the mousetrap R package (Kieslich et al., 2016). All data and analyses scripts can be found at https://github.com/pascalkieslich/mousetrap-os#validation.

To determine the temporal alignment between the external hardware and the data recorded by the mousetrap_response item, we performed several analyses (based on the absolute timestamps recorded in OpenSesame): After the click on the start button, the screen with the response buttons was displayed with an average delay of 6.9 ms ($SD = 0.7$ ms) in both simulations. Mouse-tracking started after an additional delay of 0.7 ms ($SD = 0.5$ ms). This means that, on average, 7.6 ms passed between a click on the start screen and tracking onset on the next screen. Taking this delay into account, the observed tracking durations[10] in both simulations matched the expected value very closely, with an average duration of 8202.9 ms ($SD = 0.9$ ms) for the diagonal simulation, and an average duration of 16203.1 ms ($SD = 0.9$ ms) for the triangular simulation.

Next, we assessed whether the specified logging resolution was met, using the *mt_check_resolution* function to compute the time interval between subsequent recorded cursor positions. In the diagonal simulation, the mean interval was 10.0 ms ($SD = 0.3$ ms) matching the intended logging interval. Specifically, 99.86% of the logging intervals were exactly

[9] OpenSesame provides other backends with superior temporal accuracy. However, we used legacy in our simulations and the example experiment, as it is generally more stable, especially when using forms, which are often used when designing mouse-tracking experiments. More information on general benchmark results for OpenSesame can be found at http://osdoc.cogsci.nl/manual/timing/

[10] Tracking durations can be obtained via the *response_time* variable stored in OpenSesame or by using the *RT* variable computed from the timestamps using the *mt_measures* function of the mousetrap R package. Both approaches lead to identical results.

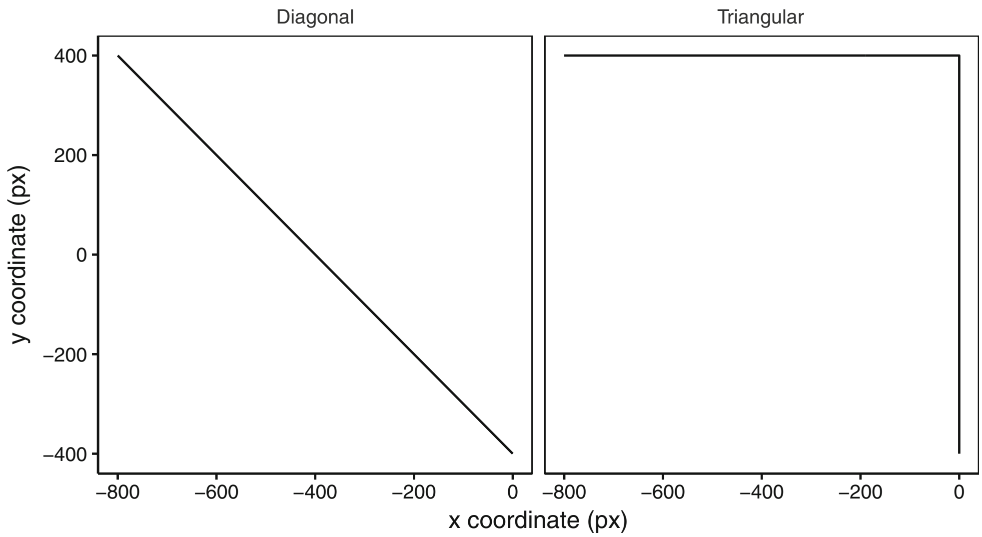


**Fig. 6** Plot of all raw trajectories for each simulation. All trajectories started at the bottom center of the screen and ended at the top left

10 ms, corresponding precisely to the desired logging resolution. An additional 0.12% of intervals were shorter than 10 ms, due to the fact that each click led to an immediate recording of the current cursor position, even before the end of a logging interval (and because logging was not exactly synchronized with simulated cursor movements and clicks). Finally, 0.02% of logging intervals were greater than 10 ms, of which 99.3% lagged by 1 additional ms only. Similar results were obtained in the triangular simulation, in which the mean timestamp difference was 10.0 ms (*SD* = 0.2 ms) and where 99.92% of the logging intervals were exactly 10 ms, 0.06% were shorter, and 0.02% longer (of which 94.4% lagged by 1 ms only).

To gain a first visual impression of the data, all raw trajectories were plotted separately for the two simulations. As can be seen in Fig. 6, all trajectory shapes were perfectly aligned within each simulation and no anomalous positions were recorded.

Missed position changes due to lags in the logging interval can be identified simply by computing the distance between two adjacent cursor positions recorded in each trial. These are expected to be either 0 px for a period where the cursor did not move along the respective dimension or 1 px along one (for the triangular simulation) or both (for the diagonal simulation) dimensions for a period with movement. Any value greater than 1 px indicates a missed change in position. In the diagonal simulation, 99.9995% of the subsequently recorded positions were either 0 px or 1 px apart for both x and y coordinates – the remaining 0.0005% differed by 2 px along either dimension, indicating that a single movement was missed. In the triangular simulation that involved changes in x coordinate only for the first, and y coordinate only for the second half of the trial, for the x coordinates, 99.9949% of the distances were either 0 px or 1 px, and 0.0051% were 2 px indicating that a single movement was missed (in only a single additional case were two changes in position missed). For the y coordinates, 99.9953% of the distances were either 0 px or 1 px, and 0.0047% were 2 px.

To assess the accuracy of the recorded cursor position at each point during the trial, we computed its expected position for each set of recorded coordinates (based on the known path generated by the external hardware, and taking into account the average tracking onset). We then computed Pearson correlations between the observed and the expected position separately for the x and y coordinates. In the diagonal simulation, the correlation was .99999999996 for both x and y coordinates, and the expected and observed position were identical in 99.9995% of cases (and differed by 1 px for the remaining cases). In the triangular simulation, the correlation was .999999993 for the x coordinates, and .999999995 for the y coordinates. For the x coordinates, the observed and expected position were identical in 99.8994% of cases (and differed by 1 px for all remaining cases except one, where it differed by 2 px). For the y coordinates, the observed and expected position were identical in 99.9298% of cases (and differed by 1 px for the remainder).

**Table 1** Expected values, observed mean and standard deviation for selected mouse-tracking measures per simulation

| | Diagonal | | | Triangular | | |
|---|---|---|---|---|---|---|
| | MAD | AUC | AD | MAD | AUC | AD |
| Expected | 0.00 | 0.00 | 0.00 | 565.69 | 320000.00 | 279.01 |
| *M* | 0.00 | 0.00 | 0.00 | 565.69 | 320000.00 | 279.01 |
| *SD* | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 |

*MAD* maximum absolute deviation, *AUC* area under curve, *AD* average deviation.

In the diagonal simulation, *M*s for MAD and AD were $< 9*10^{-14}$ and *SD* for AD was $< 3*10^{-20}$

Finally, we computed a number of mouse-tracking indices based on the raw trajectory data, using the *mt_measures* function. The descriptive statistics for a selection of the measures can be found in Table 1. In line with the expected measures based on the predetermined paths, the maximum absolute deviation (MAD), area under curve (AUC) and average deviation (AD) were 0 for the diagonal simulation and did not vary between trials. For the triangular simulation, the MAD always met the expected value of 565.69 px (which is the height of a right-angled triangle where both legs have a length of 800 px) and the AUC was always 320,000 px$^2$ (which corresponds exactly to the area of the previously described triangle). The AD values were on average also as expected ($M = 279.01$ px) with a minor variation between trials ($SD = 0.02$ px) because the AD takes every logged coordinate value into account and is therefore most sensitive to variations therein.

In sum, with regard to both logging resolution and measured coordinates, the mousetrap plugin for OpenSesame captures the raw mouse trajectory extremely well. It should be noted that the current validation was performed under even stricter conditions than those used in the validation of another software package (Freeman & Ambady, 2010): in the current simulation, the cursor was updated at a higher rate (every 10 ms instead of 30 ms) and more fine-grained analyses were used, focusing on exact raw trajectories instead of averaged data. When applied to actual data, even the remaining minute discrepancies will most often be negligible given that mouse-tracking analyses usually interpolate the raw trajectories to some extent (e.g., through time-normalization) and analyze trial summary statistics such as the measures reported above. Thus, we are confident that our software will perform reliably under most conditions.

## References


Awasthi, B., Friedman, J., & Williams, M. A. (2011). Faster, stronger, lateralized: Low spatial frequency information supports face processing. *Neuropsychologia, 49*(13), 3583–3590. doi:10.1016/j.neuropsychologia.2011.08.027

Buc Calderon, C., Verguts, T., & Gevers, W. (2015). Losing the boundary: Cognition biases action well after action selection. *Journal of Experimental Psychology: General, 144*(4), 737–743. doi:10.1037/xge0000087

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition, 35*(1), 15–28. doi:10.3758/BF03195938

Dale, R., Roche, J., Snyder, K., & McCall, R. (2008). Exploring action dynamics as an index of paired-associate learning. *PLOS ONE, 3*(3), e1728. doi:10.1371/journal.pone.0001728

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods, 46*(4), 913–921. doi:10.3758/s13428-013-0422-2

Dshemuchadse, M., Scherbaum, S., & Goschke, T. (2013). How decisions emerge: Action dynamics in intertemporal decision making. *Journal of Experimental Psychology: General, 142*(1), 93–100. doi:10.1037/a0028499

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

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*(1), 226–241. doi:10.3758/BRM.42.1.226

Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods, 45*(1), 83–97. doi: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. doi: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*(4), 377–388. doi:10.1007/s10339-015-0666-0

Haslbeck, J. M. B., Wulff, D. U., Kieslich, P. J., Henninger, F., & Schulte-Mecklenbeck, M. (2017). *Advanced mouse- and hand-tracking analysis: Detecting and visualizing clusters in movement trajectories*. Manuscript in preparation.

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations, 18*(3), 384–401. doi:10.1177/1368430214538325

Henninger, F. (2017). *The participant on a chip: Flexible, low-cost, high-precision validation of experimental software*. Manuscript in preparation.

Henninger, F., Kieslich, P. J., & Hilbig, B. E. (in press). Psynteract: A flexible, cross-platform, open framework for interactive experiments. *Behavior Research Methods*. doi:10.3758/s13428-016-0801-6

Kieslich, P. J., & Henninger, F. (2016). Readbulk: An R package for reading and combining multiple data files. doi:10.5281/zenodo.596649

Kieslich, P. J., & Hilbig, B. E. (2014). Cognitive conflict in social dilemmas: An analysis of response dynamics. *Judgment and Decision Making, 9*(6), 510–522.

Kieslich, P. J., Wulff, D. U., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2016). Mousetrap: An R package for processing and analyzing mouse-tracking data. doi:10.5281/zenodo.596640

Kieslich, P. J., Wulff, D. U., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2017). *Mouse- and hand-tracking as a window to cognition: A tutorial on implementation, analysis, and visualization*. Manuscript in preparation.

Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judgment and Decision Making, 8*(5), 527–539.

Koop, G. J., & Johnson, J. G. (2011). Response dynamics: A new window on the decision process. *Judgment and Decision Making, 6*(8), 750–758.

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology, 67*(4), 151–185. doi:10.1016/j.cogpsych.2013.09.001

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. doi:10.3758/s13428-011-0168-7

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113. doi:10.1016/0028-3932(71)90067-4

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature, 461*(7261), 263–266. doi:10.1038/nature08275

Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: The temporal dynamics of action selection. *Cognition, 115*(3), 407–416. doi:10.1016/j.cognition.2010.02.004

Scherbaum, S., & Kieslich, P. J. (2017). *Stuck at the starting line: How the starting procedure influences mouse-tracking data*. Manuscript submitted for publication.

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science, 15*(5), 207–211. doi: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 of the United States of America, 102*(29), 10393–10398. doi:10.1073/pnas.0503903102

Wirth, R., Pfister, R., & Kunde, W. (2016). Asymmetric transfer effects between cognitive and affective task disturbances. *Cognition and Emotion, 30*(3), 399–416. doi:10.1080/02699931.2015.1009002

# Mouse-tracking:

# A practical guide to implementation and analysis

Pascal J. Kieslich[a], Felix Henninger[a, b], Dirk U. Wulff[c, d],

Jonas M. B. Haslbeck[e] & Michael Schulte-Mecklenbeck[d, f]

University of Mannheim, Germany[a],

University of Koblenz-Landau, Germany[b],

University of Basel, Switzerland[c],

Max Planck Institute for Human Development, Berlin, Germany[d],

University of Amsterdam, Netherlands[e],

University of Bern, Switzerland[f]

This chapter will be published as: 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.

Correspondence should be addressed to: Pascal J. Kieslich, Experimental Psychology Lab, School of Social Sciences, University of Mannheim, L13, 17, 68161 Mannheim, Germany, E-Mail: kieslich@psychologie.uni-mannheim.de.

This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences funded by the German Research Foundation.

**Mouse-tracking: A practical guide to implementation and analysis**

The motivation behind process tracing is to go beyond the mere observation of a choice as the behavioral outcome and more directly observe the psychological process by collecting additional variables. A central unobserved quantity in choice tasks is the degree to which each alternative received consideration during the choice process, and how commitment to and conflict between options developed over time. Mouse-tracking is based on the assumption that motor movements in a given time interval contain a signal of the cognitive processes during that period (Spivey & Dale, 2006). Specifically, it is assumed that the direction of movement toward or away from alternatives reflects their relative attraction at a given time point during the decision process. To gain access to this information, mouse-tracking records hand movements indirectly by sampling the cursor position of a computer mouse with a high frequency while participants decide between (and move toward) options presented at different locations on the computer screen. Mouse-tracking is an increasingly popular process tracing technique that has been applied to a wide range of questions throughout many fields of psychology (see Chapters 9-10; see also Freeman, Dale, & Farmer, 2011; Stillman, Shen, & Ferguson, 2018).

This chapter provides an introduction to the collection, analysis and visualization of mouse-tracking data using free, open-source software. We show how to create mouse-tracking experiments using the graphical experiment builder OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) in combination with the mousetrap plugin (Kieslich & Henninger, 2017). Analysis and visualization rely on the mousetrap package (Kieslich, Wulff, Henninger, Haslbeck, & Schulte-Mecklenbeck, 2018) for the statistical programming language R (R Core Team, 2016).[1]

To illustrate the method and its implementation in mousetrap, we replicate a mouse-tracking experiment by Dale, Kehoe, and Spivey (2007). In this study, participants classified exemplars (animals) into one of two categories (e.g., mammal or bird) by clicking on the corresponding buttons located at the top-left and top-right of the screen. The independent variable was the typicality of each exemplar for its category. The experiment included typical exemplars (e.g., dog for mammal) as well as atypical ones that shared features both with the correct and the competing

[1] Note that other options for creating mouse-tracking experiments and analyzing mouse-tracking data are available (e.g., MouseTracker, cf., Freeman & Ambady, 2010) and a discussion of the different software packages is provided elsewhere (Kieslich & Henninger, 2017; Kieslich et al., 2018).

category (e.g., a bat, sharing both features with the correct category mammal and the incorrect category bird). Dale et al. (2007) hypothesized that for atypical exemplars, both response options would receive some degree of activation, whereas for the typical exemplars, activation would largely be limited to the correct category. Consequently, for atypical exemplars, the incorrect category should exert a stronger attraction, and mouse movements should deviate more in its direction even if participants finally choose the correct option.[2]

## Creating mouse-tracking experiments

In this section we demonstrate how a mouse-tracking experiment can be created in OpenSesame (Mathôt et al., 2012). OpenSesame is a free, open-source software for creating experiments via a graphical user interface which additionally allows for full customization of studies using Python code.[3] To simplify the creation of mouse-tracking experiments inside this framework, we developed the mousetrap plugin (Kieslich & Henninger, 2017) for OpenSesame. Installation instructions and additional documentation for the plugin are available in its GitHub repository at https://github.com/pascalkieslich/mousetrap-os.

### Creating an experiment

The first step is to start OpenSesame and create a new experiment by clicking on File/New and selecting the default template. Experiments in OpenSesame are assembled from a set of items, for example, a *sketchpad* item for presenting graphical content on the screen, a *keyboard_response* item for collecting key presses, and a *logger* item for writing data into log files. Figure 1 shows the OpenSesame interface with the item toolbar on the left-hand side. To its right, the overview area represents the study's structure, in that the items therein are run sequentially from top to bottom. An experiment is built by dragging and dropping items from the toolbar into the overview area. *Sequences* can be used to run a number of items in succession. In addition, *loop* items can be used to repeatedly run sequences with some degree of variation, for example, trials with varying stimuli (Figure 1, right panel).

[2] The data for this replication were collected by Kieslich and Henninger (2017); the corresponding material, data, analyses, and results are available at https://github.com/pascalkieslich/mousetrap-resources.

[3] OpenSesame can be obtained free of charge from http://osdoc.cogsci.nl/, where a general introduction to the program and detailed documentation are also available.

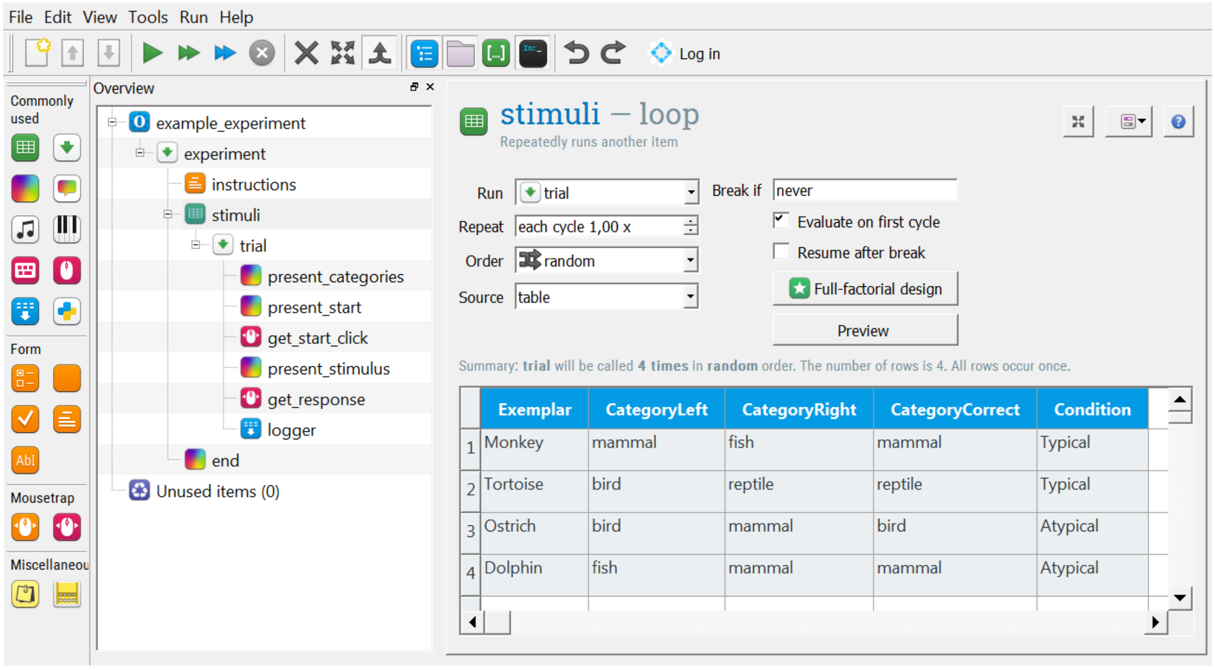


*Figure 1.* User interface of OpenSesame, showing the final state of the tutorial experiment. In the left-most panel, the item toolbar contains the available items, including the mousetrap plugin items visible toward the bottom. The overview area represents the study's structure. The right panel shows the user interface of the stimulus loop containing four exemplary stimuli.

**Setting up the screen.**

Mouse-tracking experiments are typically run in fullscreen mode. Therefore, before adding content to a new experiment, the screen resolution should be adjusted to match that of the computers used for data collection. This is done in the overall experiment settings, which are accessible by clicking on the topmost item in the study overview area ("example_experiment" in Figure 1).

**Creating the study structure.**

The first item in the experiment provides the instructions. For this, we use a *form_text_display* item that presents text and a button to continue the study. It can be added to the study by dragging it from the item toolbar into the overview area (cf., Figure 1).

In the central part of our study, participants will make categorization decisions for different animal exemplars and pairs of response categories. To accommodate this recurring structure, we include a loop item that varies the information presented on each iteration. In the loop options, the stimulus material is represented as a table where rows reflect the different stimuli and columns contain the variables that differ for each stimulus (Figure 1, right panel). In our case, the vital pieces of information are the name of the exemplar and the response categories, which are contained in the columns *Exemplar*, *CategoryLeft* and *CategoryRight*. The additional columns specify the correct response and typicality of each combination; though not presented to participants, they are stored in the dataset and facilitate later analysis. Using the default settings shown in Figure 1, the order of stimuli is randomized, and each stimulus is presented once.

Nested inside the loop, a *sequence* item is used to build each trial. It combines several screen pages as well as the collection of responses and logging of the stimulus and response information.

**Building a mouse-tracking screen.**

The central part of a mouse-tracking experiment is the stimulus display that presents the name of the exemplar and the two response buttons (located in the upper screen corners). We create this display by placing a *sketchpad* item into the trial sequence. In our example, it is named “present_stimulus” (Figure 2).[4] The content of the sketchpad item is added using the visual editor. The available types of elements for creating content are shown in the toolbar to the left of the preview. After selecting an element type, the contents can be drawn inside the preview (to move or edit them afterwards they can be selected using the topmost option in the toolbar). In our example, rectangles (*rect elements*) of equal size represent the response buttons, placed in the top left and right screen corners. Button labels are added in the center of each button using *textline elements*. An additional *textline element* is used to present the name of the to-be-categorized exemplar in the lower part of the screen. By default, the inserted text is presented verbatim. However, one can easily vary content across trials by replacing static text with the appropriate variable names in square brackets (i.e., “[CategoryLeft]” and “[CategoryRight]” for the button labels and “[Exemplar]” for the exemplar name). In every iteration of the loop, OpenSesame will replace the variable name with the variable’s current value. To make sure that the button borders are identifiable in the subsequent *mousetrap_response* item (cf., next section), we must furthermore label the two *rect elements* using the *Name* field (cf., Figure 2 top row). Each button border is labeled using the corresponding variable name (“[CategoryLeft]” and “[CategoryRight]”).

[4] The additional screens that are presented beforehand (“present_categories” and “present_start”) will be described in the section *Design considerations*.

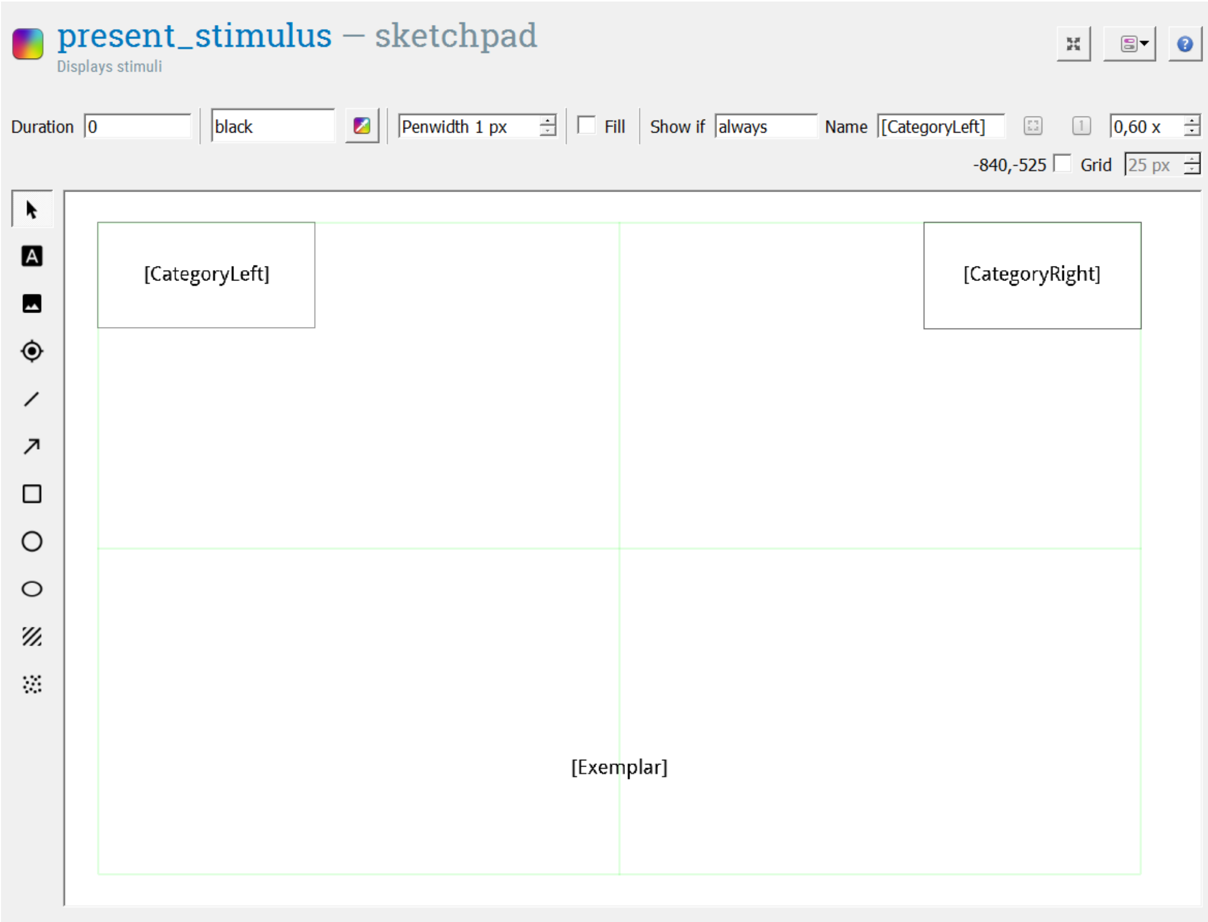


*Figure 2.* Sketchpad item used to create the main stimulus display. The exemplar is displayed using a *textline* element that contains the name of the corresponding variable from the stimulus loop (cf., Figure 1). The two button borders are created using *rect elements*. Each button border is labeled using the *Name* field (see top row) and as label the corresponding values from the stimuli loop are used. The button labels are displayed using *textline elements* that are placed in the center of each button.

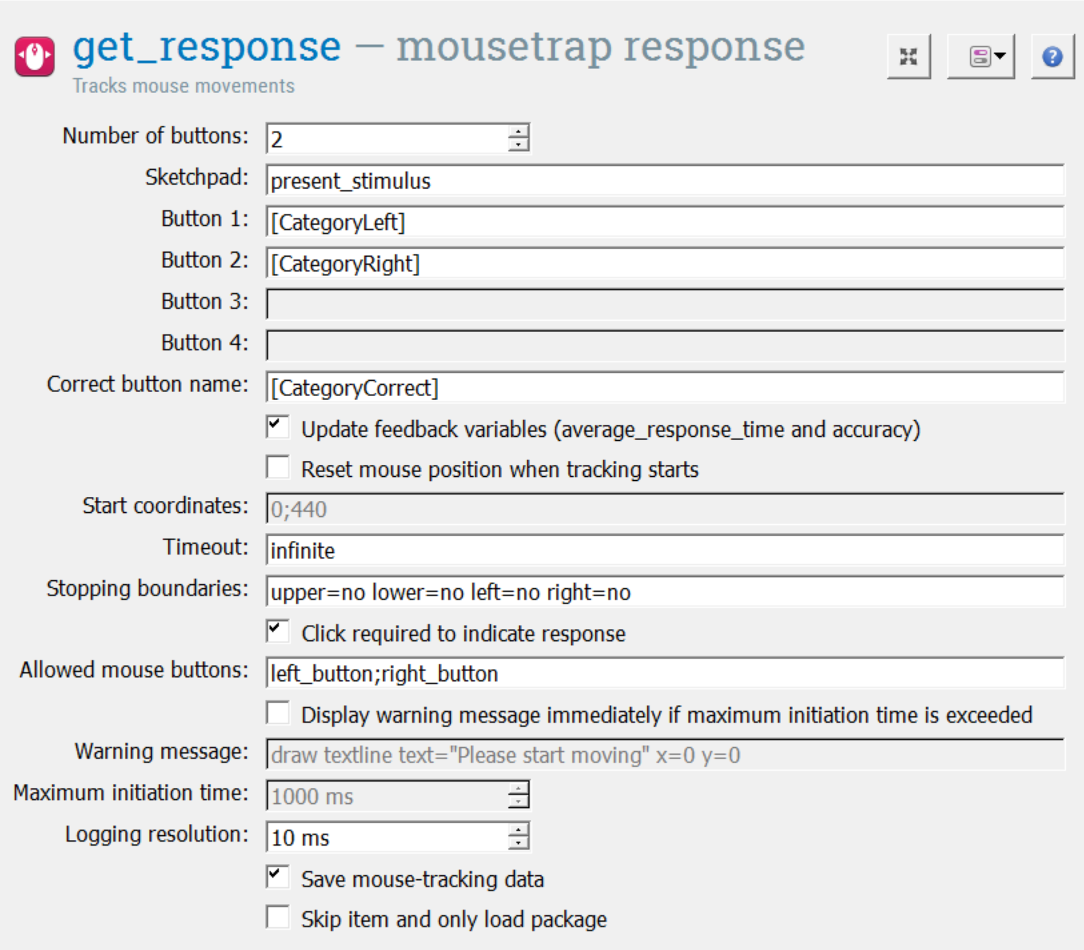


*Figure 3.* Settings of the mousetrap_response item.

**Tracking mouse movements.**

After creating the stimulus presentation, we specify the collection of mouse-tracking data and button clicks using the *mousetrap_response* item, which is inserted directly after the sketchpad item and called "get_response".[5] To start recording cursor positions immediately following stimulus presentation, the duration of the sketchpad is set to 0.

The *mousetrap_response* item records the cursor position at a constant sampling rate (10 ms by default) until the participant clicks on one of the buttons. To register responses, the corresponding buttons need to be defined (Figure 3, upper part): first, the number of buttons is specified. Then, the name of the sketchpad that presents the buttons is entered ("present_stimulus").

[5] The mousetrap plugin includes two items for tracking mouse movements. As an alternative to the *mousetrap_response* item, a *mousetrap_form* item combines stimulus presentation and response collection; its contents are defined using a basic syntax instead of a visual editor. More information is provided in Kieslich and Henninger (2017).

Finally, the buttons are specified via the labels of the button borders used on the sketchpad ("[CategoryLeft]" and "[CategoryRight]"). As a result, if the participant selects the left button, the value of the variable *CategoryLeft* is recorded as their response.

The *mousetrap_response* item also provides additional options (cf., lower part of Figure 3). For example, if the name of the correct button is specified, OpenSesame will automatically create a variable *correct* that is set to 1 or 0 for correct and incorrect answers, respectively (this is useful for analysis, as well as for providing feedback during the study). Additional design options are discussed in the section *Design considerations*.

#### Storing data.

As the final part of the trial sequence, a logger item writes the data from the current trial into a log file. This includes variables pertaining to the study as a whole (e.g., the *subject_nr*), the current values of all variables in the stimuli loop (cf., Figure 1) and the response variables. OpenSesame stores participants' responses in two places – global variables (*response*, *response_time* etc.) that always store the last recorded response and response time in the experiment, and item-specific variables named after the item that collected the response (e.g., *response_get_response* in the current example). The recorded mouse positions and associated timestamps are stored in item-specific variables only, in order to save memory (*xpos_get_response, ypos_get_response* and *timestamps_get_response*).

### Design considerations

When setting up mouse-tracking experiments, researchers are faced with a number of design choices. These include decisions about the starting procedure, the cursor speed and acceleration settings, and the response mode (click or mouse-over). Each of these choices aims to ensure that all cognitive processes relevant to the decision take place while the tracking is active (which is, in many cases, the period between the click on a start button and the selection of one of the response options), so that the process of interest is captured in the trajectories. In the remainder of this section, we discuss available options for a number of important design choices, and their potential impact on the recorded mouse trajectories (see also Fischer & Hartmann, 2014; Hehman, Stolier, & Freeman, 2015; Kieslich & Henninger, 2017; Scherbaum & Kieslich, 2018, for additional discussions about design choices).

**Start button.**

Virtually all mouse-tracking experiments try to enforce a comparable start position of the cursor across trials, thereby ensuring that the cursor is centered horizontally and approximately equidistant to all response options at the beginning of each trial. To achieve this, another screen with a start button can be added prior to the display of the task stimulus. The button ensures that participants have to return to a common area before subsequently initiating mouse movements for a new choice. In the current experiment, this is implemented using a *sketchpad* called "present_start" combined with a *mousetrap_response* item called "get_start_click" (cf., Figure 1). As before, the screen content is assembled in the visual editor and a start button is placed in the lower center of the screen (and labeled "Start"). The name of the start button is entered in the options of the *mousetrap_response* item as the single possible response. As mouse-tracking data prior to the stimulus presentation are not of interest, the option *save mouse-tracking data* can be unchecked for the "get_start_click" item. While the start button ensures that the cursor position at tracking onset is comparable across trials, it does not guarantee that it is identical. If this is desired, one can select "Reset mouse position when tracking starts" and specify coordinates in the "get_response" item (cf., Figure 3).

**Information presentation.**

Another key challenge in designing mouse-tracking studies is the temporal order in which task-relevant information is presented to the participant. On the one hand, the amount of information presented after the onset of tracking should be minimized to ensure that the collected mouse-tracking data reflects the decision processes. On the other hand, the decision-critical information needs to be withheld until tracking begins, to prevent participants from making their decision beforehand. In the current example, these considerations are accommodated by presenting the information about the two response categories for 2000 ms prior to tracking onset, but presenting the to-be-categorized exemplar only after the click on the start button (following the original procedure of Dale et al., 2007). We implemented this procedure by including another *sketchpad* item called "present_categories" at the beginning of the trial that presents only the two response categories, before the start button is made available to participants (cf., Figure 1).

#### Counterbalancing.

Another design factor concerns the assignment of response options to the button positions on the screen. Specifically, in the current study we would like to ensure that the correct answer is not always presented on the same side. One solution for this is counterbalancing the position of the correct answer between stimuli, while keeping their position fixed for all participants (cf., Figure 1). Ideally, however, the position of both response options is drawn anew for each participant and stimulus (this can be achieved in OpenSesame through the advanced randomization operation shuffle horizontal).

#### Starting procedure.

For mouse-tracking to reflect the cognitive processes underlying the choice, movement must occur while the cognitive process is ongoing. It has been shown that the starting procedure has a considerable influence on the obtained trajectories (Scherbaum & Kieslich, 2018).

Many mouse-tracking studies have used a so-called *static starting procedure*, in which the stimulus is shown immediately after participants have clicked on the start button and without any specific measures taken to ensure movement during processing (our tutorial experiment following Dale et al., 2007, is an example for such a setup). While many mouse-tracking studies that use a static starting procedure find theoretically relevant effects in mouse trajectories, this procedure does not exclude the possibility that (in some trials) decision-relevant processes take place before the mouse movement is initiated and therefore are not captured by mouse trajectories.

To ensure that the cognitive processes under investigation do not take place before mouse movement initialization, some studies have modified the starting procedure. One option is the *static starting procedure with delay*, in which a brief lag of, for example, 500 ms, is inserted between clicking the start button and stimulus presentation. Previous studies reported that this often successfully led participants to initialize movement before the stimulus appeared (e.g., Spivey, Grosjean, & Knoblich, 2005). Other studies employ a static starting procedure with immediate stimulus presentation, but explicitly instruct participants to initiate their mouse movement within a certain time limit and display a warning to participants after the trial if the *initiation time* exceeds the threshold. The exact time limit depends on the task (a typical value is 400 ms; see Hehman et al., 2015, p. 388–389, for a discussion).

A more rigorous option, however, is to implement a *dynamic starting* procedure that presents the stimulus only after participants have moved the mouse upwards for a minimum distance (e.g., Scherbaum, Dshemuchadse, Fischer, & Goschke, 2010). The dynamic procedure forces participants to initiate their movement in order to receive the critical information needed to make the choice. It can be implemented by placing an invisible horizontal boundary slightly above the start button that triggers the presentation of the stimulus once it is crossed (cf., Frisch, Dshemuchadse, Görner, Goschke, & Scherbaum, 2015). This procedure has been shown to lead to more consistent movements and larger effects in within-trial temporal analyses (Scherbaum & Kieslich, 2018).[6]

#### Mouse sensitivity.

Another design choice is the computer's mouse sensitivity, in particular the cursor speed and acceleration. One option is to leave these settings to the operating system defaults (under Windows 7 and 10, medium speed with acceleration). However, it is often preferable to reduce mouse speed and switch off mouse acceleration (Fischer & Hartmann, 2014). This is particularly relevant when using a dynamic starting procedure to ensure that participants can read the dynamically presented stimulus information while continuously moving upwards. The mouse sensitivity settings cannot be adjusted directly within OpenSesame, but need to be set in the computer's system preferences.

#### Response mode.

The two main response modes in mouse-tracking studies are clicking on and moving over the response buttons. In the mousetrap plugin, users can switch between the two response modes by checking or unchecking the option *Click required to indicate response*, which is enabled by default (cf., Figure 3).

#### Data collection and testing.

After creating the experiment, it can be run from within OpenSesame for testing or using *OpenSesame Run* for data collection in the laboratory (see Kieslich & Henninger, 2017, for more information on running mouse-tracking experiments). Mouse-tracking studies also usually assess the handedness of participants and the hand participants use for moving the mouse (with some authors recommending only to include right-handed participants, cf., Hehman et al., 2015).

[6] An example experiment implementing this procedure can be found at https://github.com/pascalkieslich/mousetrap-os#examples.

## Analyzing mouse-tracking data

We will now demonstrate the typical steps of a basic mouse-tracking analysis using the data from the replication experiment described above (Kieslich & Henninger, 2017). For this analysis, we will use the *mousetrap* package (Kieslich et al., 2018) in the statistical programming language R (R Core Team, 2016), which facilitates preprocessing, analysis and visualization of mouse-tracking data.[7] Once installed, mousetrap functions can then be made available within an R session by loading the package via:

```
library(mousetrap)
```

A detailed overview of its functionality is provided online at http://pascalkieslich.github.io/mousetrap/ or within R using the command:

```
package?mousetrap
```

In the following, we discuss the most important analysis steps, starting with data import and preprocessing operations, followed by the computation and analysis of common indices, temporal analyses, and visualizations.

### Import

First, the raw data need to be read into R's workspace. OpenSesame stores the data for each participant in a separate csv file. To load all csv files from a directory and combine them into a single dataset, we use the *read_opensesame* function from the *readbulk* package (Kieslich & Henninger, 2016). The following command assumes that all data files can be found in the folder "raw_data" in the working directory and stores the imported data in the dataset "KH2017_raw" (this dataset is available once the mousetrap package has been loaded, so no raw data have to be imported to follow this tutorial):

```
library(readbulk)
KH2017_raw <- read_opensesame("raw_data")
```

[7] R is open-source and freely available from https://www.r-project.org/. We recommend using R in combination with RStudio (available from https://www.rstudio.com/), which greatly facilitates code development and analysis by providing many useful features such as code highlighting, debugging, and tools for data inspection.

Next, the data need to be transformed into a *mousetrap data object* to perform analyses using the mousetrap R package.[8] This results in a mousetrap data object (called "mt_data" in the current analysis), which is described in detail in Information box 1:

```
mt_data <- mt_import_mousetrap(KH2017_raw)
```

Using this two-step procedure of reading and importing the mouse-tracking data, the mousetrap R package can also be used for data collected in other software. An example for reading and importing raw data collected with MouseTracker (Freeman & Ambady, 2010) is given in the documentation of the *read_mt* function, which can be accessed by entering:

```
?read_mt
```

## Preprocessing

### Spatial transformations.

In a typical two-alternative choice design (as implemented in the example experiment, see Figure 2), trajectories end either at the left or the right response option. As the overall spatial direction is irrelevant for most analyses (as opposed to the substantive meaning of the response button, which varies across trials if the position of alternatives is counterbalanced), all trajectories are remapped so that they end on the same side. By default, mousetrap maps the trajectories to the left, implying that trajectories that end on the right-hand side are flipped from right to left:

```
mt_data <- mt_remap_symmetric(mt_data)
```

Similarly, differences in the trajectories' starting points are often not of substantive interest. If the cursor's starting position was not reset to exact coordinates during the experiment (as is the case for the example data set), it can be aligned by shifting the trajectories in preprocessing:

```
mt_data <- mt_align_start(mt_data, start=c(0,0))
```

[8] In case that only one mousetrap item in the experiment collected mouse-tracking data, the *mt_import_mousetrap* function automatically detects the mouse-tracking variables in the raw data. If more than one item stored mouse-tracking data, the variable names have to be set explicitly once using the *xpos_label*, *ypos_label*, and *timestamps_label* arguments when importing data via the *mt_import_mousetrap* function.

*Information box 1*. Working with mousetrap data objects

The mousetrap R package represents mouse-tracking data in a specialized data structure, a mousetrap data object. This allows the package to store and process mouse trajectories efficiently, and to link them to other information collected during the study. All mousetrap analysis functions use mousetrap data objects as input; therefore, the collected data must be imported before processing and analysis. A newly imported mousetrap data object consists of a *data.frame* called *data* containing the trial information (without mouse trajectories) and an *array* called *trajectories* containing the recorded mouse-tracking data.

The mousetrap data object can hold multiple sets of trajectories (e.g., *mt_time_normalize* adds the time-normalized trajectories as *tn_trajectories*). In subsequent analyses, the user can specify via the *use* argument whether an analysis (or visualization) should be performed based on the raw trajectories (`use="trajectories"`, which is used by default in most functions) or another trajectory array (e.g., `use="tn_trajectories"`). Other functions add new *data.frames* to the mousetrap object (e.g., *mt_measures* adds a *data.frame* called *measures* that contains trial-level indices).

The mousetrap package is designed for processing and visualizing trajectories and the computation of indices. For statistical analyses of the computed indices, they can be merged with the other trial data via:

```
results <- merge(mt_data$data, mt_data$measures, by="mt_id")
```

Similarly, mouse trajectories can be transformed into a format required for the statistical analysis using the *mt_export_long* or *mt_export_wide* functions. The resulting data can then be analyzed outside of the mousetrap package using any standard analysis method.

**Resampling.**

The cursor position is typically recorded at a constant sampling rate. The mousetrap plugin in OpenSesame records the mouse position every 10 ms by default (corresponding to a sampling rate of 100 Hz). Due to variation in trial durations, the number of recorded cursor positions may vary considerably across trials. To be able to aggregate trajectories or compare them statistically, one often requires an equal number of coordinates for all trajectories. To achieve this, studies commonly apply time-normalization:

```
mt_data <- mt_time_normalize(mt_data)
```

Time-normalization interpolates trajectories so that each is represented by the same number of positions (101 by default, following Spivey et al., 2005) separated by a (within-trial) constant time interval. Mousetrap stores the time-normalized data as a new set of trajectories within the mousetrap data object (see Information box 1).

Another possibility is to interpolate trajectories so that each is represented by the same number of spatially equidistant positions (using *mt_spatialize*). This processing step facilitates the comparison of trajectory shapes and is instrumental in type-based analyses of trajectories (cf., Chapter 9).

**Data inspection and filtering.**

As a final step prior to analysis, trials are typically screened and filtered based on one or more criteria. If choices can be graded as correct, studies often exclude trials with incorrect responses to ensure a consistent interpretation of curvature across all trials (i.e., that increased curvature always reflects attraction towards the distractor category). The *mt_subset* function can be used to select only correctly answered trials for further analysis (or to apply other filters):

```
mt_data <- mt_subset(mt_data, correct==1)
```

An additional concern in mouse-tracking analysis is whether the data contain movements that are presumably not related to the preference development but to other processes, such as information acquisition or slips of the hand. Information acquisition might, for example, be reflected by directed movements towards a point where information was presented on the screen. Slips of the hand, resulting, for example, from participants placing the mouse device somewhere else in order to avoid a physical obstacle (or in order to more comfortably move it), would lead to erratic movements and result in movements untypical for this context, for example, comparatively large amounts of up and down movements. The challenge is finding precise criteria to differentiate between relevant and irrelevant movements. One possibility is an exploratory approach, for example, visually inspecting all trials by plotting them either in a single figure (using *mt_plot* or *mt_heatmap*, see also top panel of Figure 5 in the section *Trajectory types*) or separately (using *mt_plot_per_trajectory*). If to-be-excluded movement patterns have been specified, separate plots per trajectory might also be provided to human raters who can code whether these are present in a trial. Another possibility is to exclude trials based on a numeric criterion, such as trials exceeding an absolute or relative reaction time or trials containing several flips along the y-axis (which probably indicate large amounts of task-irrelevant tracking data). A more detailed discussion is provided in Kieslich et al. (2018). Especially if exclusion criteria were not defined a priori, the impact of the criterion applied should be reported; additional pre-registered studies might be conducted to validate the chosen criteria and to replicate the results under strictly confirmatory conditions.

### Analysis

To analyze effects of the experimental manipulation, a common first step is the visual inspection of aggregate time-normalized mouse trajectories. Mousetrap provides the *mt_plot_aggregate* function, which, if used as below, aggregates the time-normalized trajectories for each condition first within and then across participants and plots the result:

```
mt_plot_aggregate(mt_data, use="tn_trajectories",
  x="xpos", y="ypos", color="Condition",
  subject_id="subject_nr")
```

As can be seen in Figure 4, the aggregate mouse trajectory in the current study is more curved towards the non-chosen option for atypical than for typical exemplars – consistent with the hypothesis by Dale et al. (2007). Whether the aggregate trajectories are an adequate summary of the trial-level trajectories is discussed in the section *Trajectory types*.

A wide range of analysis methods are available for mouse-tracking data (for overviews, see Hehman et al., 2015; Kieslich et al., 2018). They can roughly be categorized into analyses that focus on the temporal development of a certain characteristic over the course of a trial (such as x-position, velocity or movement direction, see section *Temporal analyses*) and those that summarize a particular characteristic of each trajectory by computing one index value per trial. Many common indices can be computed using the *mt_measures* function:

```
mt_data <- mt_measures(mt_data)
```

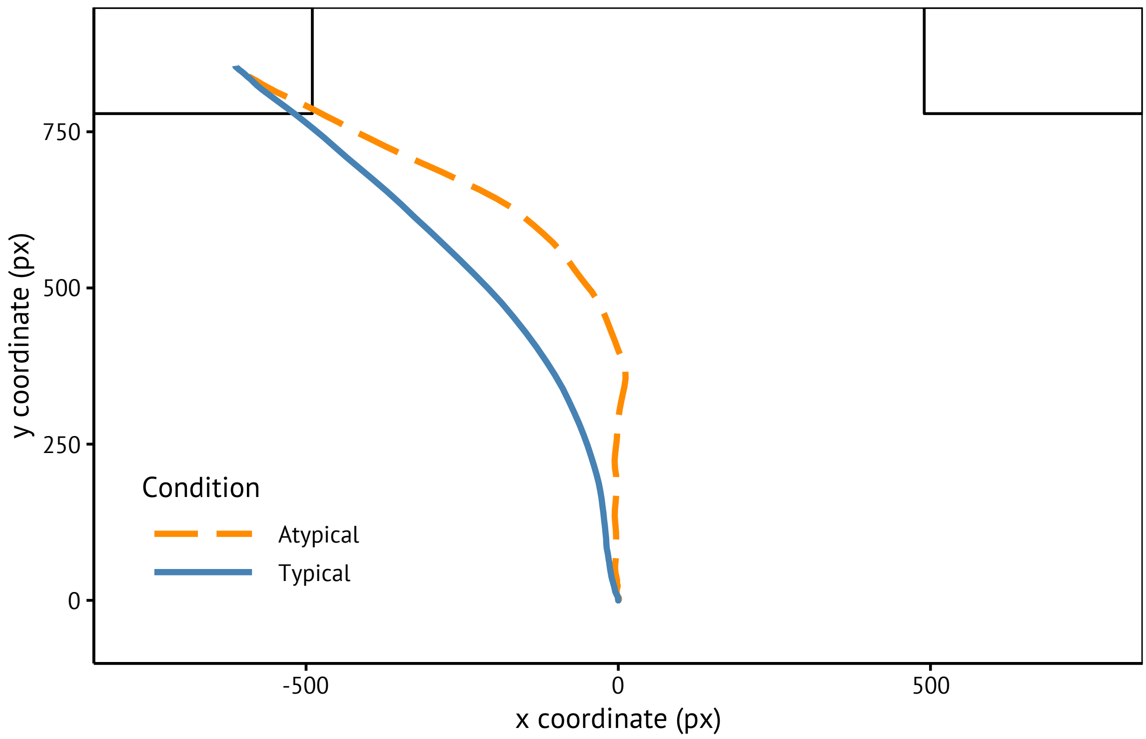


*Figure 4.* Aggregate time-normalized trajectories per typicality condition. Trajectories were first aligned to a common start position, remapped to the left, and finally aggregated first within and then across participants. Boxes representing the response buttons were added for clarity.

An overview of the different indices is given in Table 1 and further information about working with the computed indices is provided in Information box 1. Different types of indices and their interpretation will be discussed in the following.

**Curvature.**

The curvature of the response trajectory is used to assess the degree of its attraction towards the non-chosen option. It is assumed to be driven by the difference in activation between the non-chosen and the chosen option – in that a smaller difference in activations leads to a stronger curvature (Spivey, Dale, Knoblich, & Grosjean, 2010). A number of different indices have been suggested to quantify curvature (cf., Table 1). Their exact computation differs, but they are often highly correlated in practice (see Kieslich et al., 2018; Stillman et al., 2018).

*Table 1.* Selected mouse-tracking measures, their variable name (in brackets, as used by the *mt_measures* function of the mousetrap R package) and definition.

| **Type** | **Measure** | **Definition** |
|---|---|---|
| Curvature | Maximum absolute deviation (MAD) | Signed maximum absolute deviation of observed trajectory from direct path |
| | Maximum deviation above (MD_above) | Maximum deviation above direct path |
| | Average deviation (AD) | Average deviation of observed trajectory from direct path |
| | Area under curve (AUC) | Geometric area between observed trajectory and direct path |
| Complexity | x-flips (xpos_flips) | Number of directional changes along x-axis |
| | x-reversals (xpos_reversals) | Number of crossings of y-axis |
| | Sample entropy (sample_entropy) | Degree of unpredictability of movement along x-axis |
| Time | Response time (RT) | Time until response is given |
| | Initiation time (initiation_time) | Time until first movement is initiated |
| | Idle time (idle_time) | Total time without movement across trial |
| Derivatives | Total distance (total_dist) | Euclidean distance traveled by trajectory |
| | Max velocity (vel_max) | Maximum movement velocity |
| | Max acceleration (acc_max) | Maximum movement acceleration |

*Note*: The direct path refers to the straight line connecting the start and end point of the observed trajectory. Deviations/areas above the direct path receive a positive sign and deviations/areas below receive a negative sign. For all derivative measures, it is assumed that movements across both x and y dimensions are taken into account (derivatives have to be calculated using *mt_calculate_derivatives* before calling *mt_measures*). Sample entropy is computed using *mt_sample_entropy* (based on the time-normalized x-positions by default).

In the following, we focus on a frequently used index known as the *(signed) maximum absolute deviation* (MAD). To compute the MAD, imagine an idealized, direct line between the start and end point of the trajectory, and that lines perpendicular to this idealized line are drawn to connect it with every point on the original trajectory. The value of the MAD is defined as the length of the longest of these lines. The sign of the MAD is positive if the deviation is largest above the direct path (in the direction of the non-chosen alternative) and negative if the point of strongest deviation occurs below.

To assess whether the MAD differs between experimental conditions, mouse-tracking studies often aggregate the MAD values across trials per participant and condition, and then compare the aggregate MAD values between conditions using a paired *t*-test.[9] These operations can be performed using *mt_aggregate_per_subject* and R's standard *t.test* function:

```
agg_mad <- mt_aggregate_per_subject(mt_data,
  use_variables="MAD", use2_variables="Condition",
  subject_id="subject_nr")

t.test(MAD~Condition, data=agg_mad, paired=TRUE)
```

In line with the hypothesis by Dale et al. (2007), the MAD values indicate larger curvature in the atypical ($M$ = 343.8 px, $SD$ = 218.6 px) than in the typical condition ($M$ = 172.2 px, $SD$ = 110.8 px), $t(59) = 6.73$, $p < .001$. A replication of the original analyses by Dale et al. using the current dataset can be found online at https://github.com/pascalkieslich/mousetrap-resources.

**Trajectory types.**

While aggregate response trajectories (cf., Figure 4) and curvature indices provide a first indication of the average curvature of the trajectories in each condition, they do not necessarily represent the shape of the individual trajectories well. Specifically, an aggregate curved trajectory might result from different types of trajectories, for example, a mixture of straight lines and triangular "change of mind" trajectories which first head directly to the non-chosen and then to the chosen option (cf., Chapter 9). If this is the case, the average trajectory might not be representative

[9] Analyses can also be performed on the trial level using mixed-effects models that can account for individual differences between participants as well as trial-level predictors (see https://github.com/pascalkieslich/mousetrap-resources).

of the movement patterns observed in the study, but purely an artefact of aggregation. Under these circumstances, the shape of the aggregate trajectory would provide only limited (and potentially misleading) information about the underlying cognitive processes.

Several methods have been suggested to assess the degree of heterogeneity of the individual trajectories on the trial level. Previous approaches have focused on the distribution of trial-level curvature indices (such as area under curve or MAD, cf., Table 1) and tested them for indications of bimodality. The assumption behind these approaches is that gradually curved trajectories on the trial level should result in a unimodal distribution, while a combination of straight and extremely curved trajectories should result in a bimodal distribution (Hehman et al., 2015). The bimodality of the distribution is frequently assessed by computing the bimodality coefficient (BC; Pfister, Schwarz, Janczyk, Dale, & Freeman, 2013) which is interpreted as bimodal for values > .555 (Freeman & Ambady, 2010). Alternative methods for identifying bimodality have been discussed, especially the Hartigan's dip statistic (Freeman & Dale, 2013). Both methods are implemented in the *mt_check_bimodality* function.

Instead of attempting to detect mixtures of distinct trajectory types based on the distribution of curvature indices (which condense each trajectory to a single numeric value), more recent analysis methods take into account the complete shape of each trajectory by using every point of the trajectory. The shape of individual trajectories can be assessed visually by plotting raw or smoothed heatmaps with the *mt_heatmap* function and by comparing heatmaps between conditions using the *mt_diffmap* function (code examples are provided at https://github.com/pascalkieslich/mousetrap-resources).

As can be seen in Figure 5 (middle panel), there appear to be different types of trajectories on the trial level in the current study, with a large proportion of straight and mildly curved trajectories and a small proportion of extremely curved, "change of mind" trajectories. More importantly, a difference heatmap reveals that the relative occurrence of these types differs between conditions, with a higher proportion of extremely curved trajectories in the atypical condition (orange areas in Figure 5, bottom panel). Analyses that go beyond a visual inspection to identify trajectory types and instead use a clustering approach based on spatial similarity (or the assignment of trajectories to different prototypes) are also implemented in the mousetrap R package and described in Chapter 9 (see also Wulff, Haslbeck, & Schulte-Mecklenbeck, 2018).

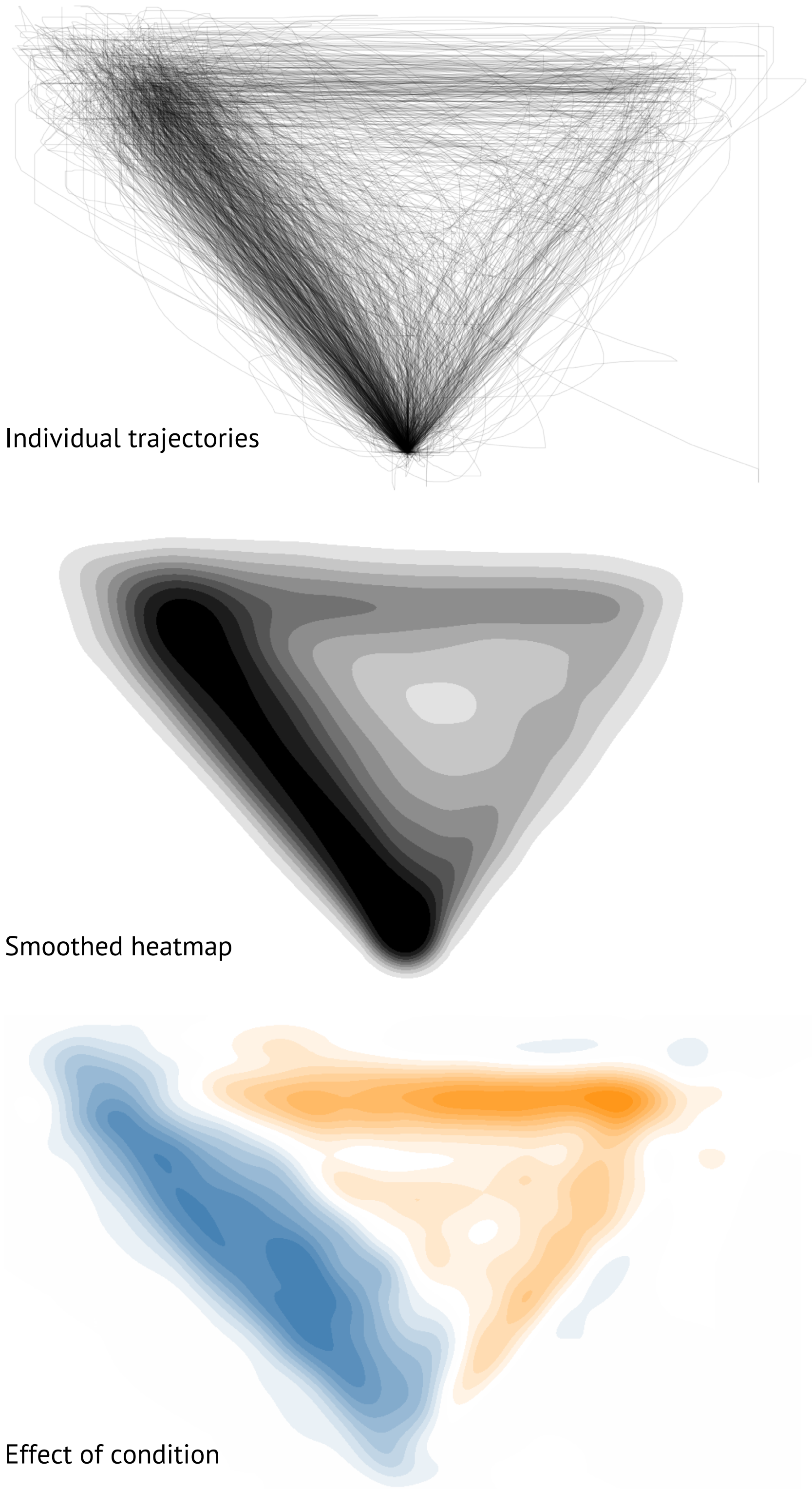


*Figure 5*. Heatmap of the (remapped) individual trajectories (top panel), smoothed heatmap (middle panel) and difference of smoothed heatmaps between conditions (bottom panel), where blue indicates higher density in the typical and orange higher density in the atypical condition (white indicates comparable density).

The different trajectory types and their frequency have also been used to distinguish between different theoretical models (see Chapters 9-10 for more information). When used to this end, it is important to keep in mind that the setup of mouse-tracking studies can influence the shape of individual trajectories (see section *Design considerations*). In the current study, the occurrence of rather “extreme” trajectory types (straight and “change of mind”) may have been caused by the relatively simplistic setup of the study with a static starting procedure, default mouse sensitivity settings and the use of a click instead of a mouse-over response.

**Complexity.**

In addition to curvature, mouse-tracking studies have also used the complexity of the movement as an indicator of response competition. If multiple response options simultaneously attract the cursor, this should result in more complex, less smooth trajectories compared to cases where only one option exerts an influence (Dale et al., 2007).

In two-alternative tasks, complexity is typically assessed with regard to movements along the horizontal axis, since this is the dimension that separates the options. The most common measure of complexity is x-flips, the number of directional changes along the x-axis (Freeman & Ambady, 2010), which is calculated by the *mt_measures* function (Table 1). As response competition might not always lead to directional changes, other mouse-tracking studies have used sample entropy (Dale et al., 2007; McKinstry, Dale, & Spivey, 2008) which quantifies the degree of unpredictability of movement along the x-axis. Sample entropy can be computed using *mt_sample_entropy*, which uses time-normalized trajectories by default, following the recommendation that each trajectory be represented by the same number of positions (Hehman et al., 2015):

```
mt_data <- mt_sample_entropy(mt_data, use="tn_trajectories")
```

Koop and Johnson (2013) propose a substantive interpretation of complexity-related measures in preferential choice tasks, based on the assumption that the x-position at a specific point during the trial is a proxy for the current absolute preference. They suggest that x-flips reflect changes in the momentary valence whereas x-reversals (i.e., the number of times the cursor crosses the vertical axis at the midpoint between the two options) indicate changes of absolute preference. Recently, the sequence in which certain areas of interest (one for each choice option) are visited with the mouse cursor has also been used to identify how often participants changed their mind during the decision-making process (Szaszi, Palfi, Szollosi, Kieslich, & Aczel, 2018; see also Travers, Rolison, & Feeney, 2016).

As with curvature indices, complexity indices can be analyzed either by aggregating values per participant and condition (using *mt_aggregate_per_subject*) and comparing the result across conditions, or on the trial level using mixed-effects models.

**Temporal analyses.**

Although many studies use it in this sense, mouse-tracking is not limited to the analysis of aggregate indices that collapse each trajectory to a single value. Analyses of trajectories' temporal development can shed light on the time course of response option activations across the trial and, in particular, how and when different cognitive processes influence the trajectory (Hehman et al., 2015). In the following, we will briefly illustrate some simple use cases.

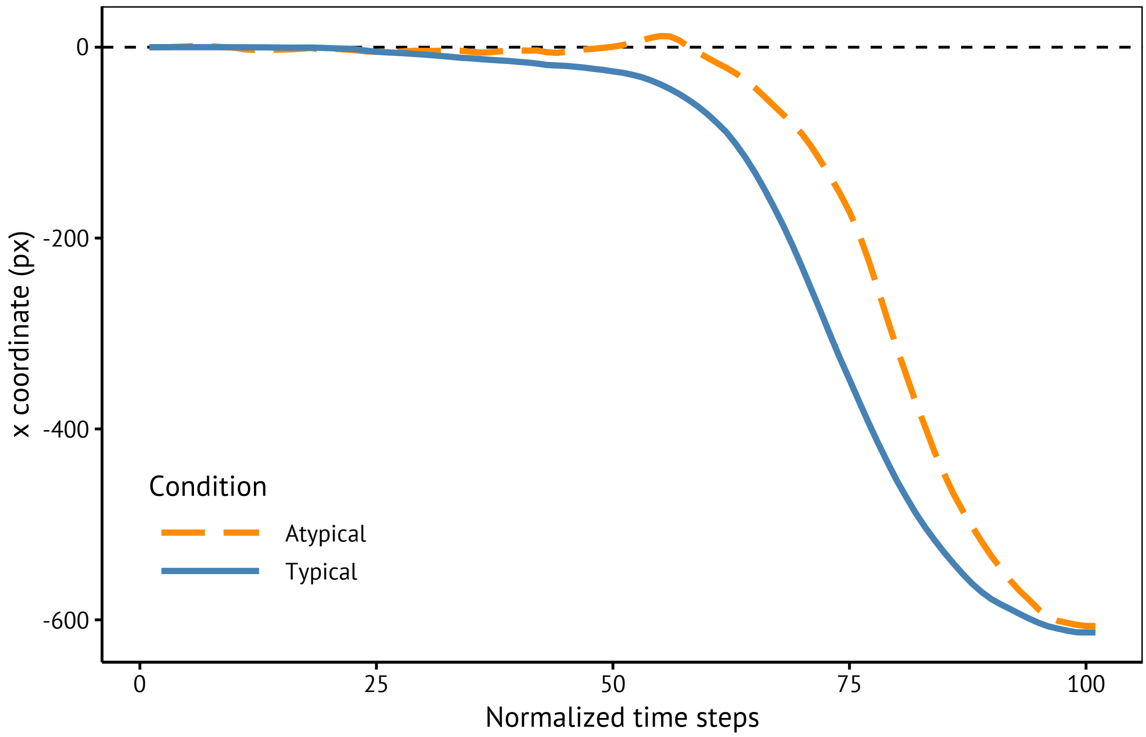


*Figure 6*. Plot of the average time-normalized x-position over time. For each time step, x-positions were first averaged within participants and condition.

One purpose of temporal analyses is to supplement aggregate analyses of trajectory curvature by showing at which point and for how long aggregate trajectories diverge between conditions. Previous studies (e.g., Dale et al., 2007) have examined this by comparing the horizontal positions of the time-normalized trajectories at each time step using a series of *t*-tests between conditions (code examples can be found at https://github.com/pascalkieslich/mousetrap-resources). Using this approach reveals that for time steps from 54 to 95 (of 101 steps) the average x- coordinates differed between conditions (Figure 6). If a theory provides specific predictions with regard to the temporal development, for example, whether the divergence between conditions should occur early or late in the decision-making process, this can be used to test them. Note that the comparison of trajectories between conditions can be problematic if response time differences between conditions are large, and that temporal analyses can also be conducted based on raw instead of time-normalized trajectories (see also Hehman et al., 2015).

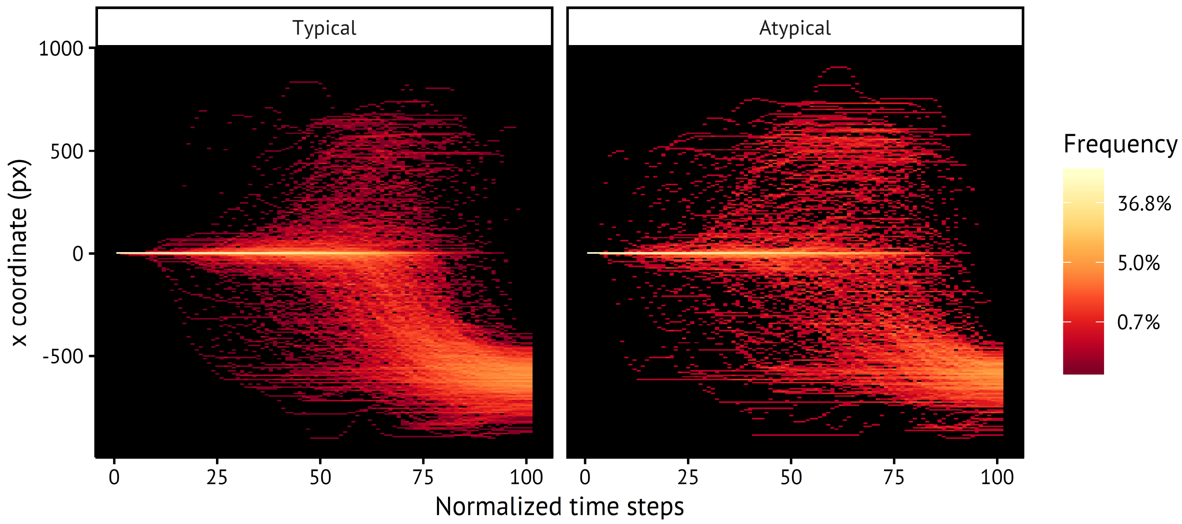


*Figure 7.* Riverbed plot of the distribution of x-positions across time for time-normalized trajectories separately for the two experimental conditions. For each time step, the colors indicate the relative frequency with which each bin of x-positions was observed.

As with aggregated trial-level indices, aggregated x-positions may not necessarily represent the underlying trial level trajectories well. To inspect whether this is the case it is useful to illustrate the full distribution of trial-level x-positions across normalized time using the *mt_plot_riverbed* function (following an approach by Scherbaum et al., 2010). As can be seen in Figure 7, the aggregate x-positions displayed in Figure 6 are a rather poor representation of the individual trajectories, which vary greatly. Specifically, while the majority of trajectories go directly to the eventually chosen option, a substantial number of trajectories first moves to the non-chosen option (crossing the midline). This means that the data may be better analyzed on the trial level using, for instance, mixed-effects models or type-based analyses (Wulff et al., 2018; see also Chapter 9). Moreover, Figure 7 reveals that in most trials of both conditions the cursor remained in a neutral position (in many cases it stayed on the start button) for more than half of the trial, a behavior that is probably related to the use of a static starting condition that did not enforce early movement initiation (cf., Scherbaum & Kieslich, 2018).

In addition to analyzing the temporal development of the cursor position, previous mouse-tracking studies have also focused on other variables derived from it, especially velocity, acceleration, and movement angle. The analysis of velocity and acceleration has been used to investigate response activation and competition (Hehman et al., 2015). In mousetrap, velocity and acceleration can be computed using *mt_calculate_derivatives* which attaches velocity and acceleration values to each of the recorded cursor positions.[10] Subsequent analyses can be performed as sketched above, using the velocity values instead of x-positions as the dimension of interest.

Finally, an emerging class of analyses has focused on the movement angle, which quantifies the direction of movement over time indicating, for example, whether participants move towards or away from a specific response alternative. Previous studies have used movement angles to disentangle when and to which extent different factors influence the movement direction (e.g., Dshemuchadse, Scherbaum, & Goschke, 2013; Scherbaum et al., 2010; Sullivan, Hutcherson, Harris, & Rangel, 2015). For details on these approaches, see Scherbaum and Dshemuchadse (2018).

## Summary and conclusion

In mouse-tracking studies, participants' cursor movements are recorded as they choose between different options represented as buttons on a computer screen. Thereby, mouse-tracking aims to measure the degree of conflict between the alternatives and the temporal development of its resolution. While Chapter 9 provides a detailed look at the types of trajectories revealed in this paradigm, and Chapter 10 provides an introduction to this method and its use in the literature, this chapter has shown how to construct a mouse-tracking study using the mousetrap plugin for the graphical experiment builder OpenSesame, and how to analyze the resulting data in the

[10] Velocity and acceleration can be calculated for raw trajectories (by default) as well as for time-normalized trajectories. In addition, both can be computed based on the Euclidean distance traveled along the x- and y-dimension (by default) or for a single dimension only.

mousetrap R package. We have covered technical issues surrounding the application of this method, and highlighted design considerations and their influence on the collected data.

The strength of mouse-tracking lies in the ease with which it can be applied. Using only standard laboratory hardware, cognitive processes can be tracked at high temporal resolution. It is also a flexible tool that can be adapted to many different tasks, and which is even more powerful in combination with other process tracing methods (e.g., eye-tracking, cf., Koop & Johnson, 2013; Quétard et al., 2016). Data collection and processing as described in this chapter are handled entirely by free, open-source software (Kieslich & Henninger, 2017; Kieslich et al., 2018), making mouse-tracking easily accessible to interested researchers and transparent to those looking to replicate findings or adapt and extend the methods described herein.

As a fairly recent addition to the family of process tracing methods, many aspects of the method are not yet fully standardized. Therefore, the degrees of freedom with regard to data collection, processing, and analysis are substantial. Where available, we have pointed to the current state of knowledge regarding best practices, which is bound to grow over time. We advise users of mouse-tracking to seek convergence between analyses and indices where no standard has been established so far. In doing so, they should also consider the effects of aggregation by inspecting the distribution of trajectories and indices on the trial level (cf., Chapter 9). While researchers may often explore different experimental setups and analyses if they apply mouse-tracking in a new domain, (additional) pre-registered studies should be conducted to perform strictly confirmatory hypothesis testing (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012).

In sum, we have demonstrated the potential mouse-tracking has as a process tracing method for various areas of decision research. Given the limits of an introductory tutorial, we have only covered the most frequently used analyses. Similarly, the current chapter has limited itself to the frequently investigated two-option design, but mouse-tracking can easily be extended to situations with more than two alternatives (e.g., Koop & Johnson, 2011). Lastly, more sophisticated analysis methods are being developed to more fully harvest the rich potential of mouse-tracking data, such as time continuous multiple regression (Scherbaum & Dshemuchadse, 2018), entropy approaches (Calcagnì, Lombardi, & Sulpizio, 2017), generalized processing tree models (Heck, Erdfelder, & Kieslich, in press), and decision landscapes (Zgonnikov, Aleni, Piiroinen, O'Hora, & di Bernardo, 2017). Thus, we are confident that mouse-tracking will continue to offer researchers novel insights into how decision processes unfold over time.

## Recommended reading list

- https://github.com/pascalkieslich/mousetrap-resources: resources for creating mouse-tracking experiments and analyzing mouse-tracking data (including the examples from the current chapter).
- Kieslich and Henninger (2017): an introduction into and validation of the mousetrap plugin for OpenSesame, which also provides detailed information about the example study used in the current chapter.
- Kieslich, Wulff, Henninger, Haslbeck, and Schulte-Mecklenbeck (2018): a detailed tutorial on analyzing hand- and mouse-tracking data using the mousetrap R package.
- Hehman, Stolier, and Freeman (2015): a description of several analytic approaches for mouse-tracking data.

## References

Calcagnì, A., Lombardi, L., & Sulpizio, S. (2017). Analyzing spatial data from mouse tracker methodology: An entropic approach. *Behavior Research Methods*, *49*(6), 2012–2030.

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15–28.

Dshemuchadse, M., Scherbaum, S., & Goschke, T. (2013). How decisions emerge: Action dynamics in intertemporal decision making. *Journal of Experimental Psychology: General*, *142*(1), 93–100.

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

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241.

Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods*, *45*(1), 83–97.

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*, 59.

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*(4), 377–388.

Heck, D. W., Erdfelder, E., & Kieslich, P. J. (in press). Generalized processing tree models: Jointly modeling discrete and continuous variables. *Psychometrika*.

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, *18*(3), 384–401.

Kieslich, P. J., & Henninger, F. (2016). Readbulk: An R package for reading and combining multiple data files. Retrieved from https://doi.org/10.5281/zenodo.596649

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods*, *49*(5), 1652–1667.

Kieslich, P. J., Wulff, D. U., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2018). *Mouse- and hand-tracking as a window to cognition: A tutorial on implementation, analysis, and visualization*. Manuscript in preparation.

Koop, G. J., & Johnson, J. G. (2011). Response dynamics: A new window on the decision process. *Judgment and Decision Making*, *6*(8), 750–758.

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, *67*(4), 151–185.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324.

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*(1), 22–24.

Pfister, R., Schwarz, K. A., Janczyk, M., Dale, R., & Freeman, J. B. (2013). Good things peak in pairs: A note on the bimodality coefficient. *Frontiers in Psychology*, *4*, 700.

Quétard, B., Quinton, J. C., Mermillod, M., Barca, L., Pezzulo, G., Colomb, M., & Izaute, M. (2016). Differential effects of visual uncertainty and contextual guidance on perceptual decisions: Evidence from eye and mouse tracking in visual search. *Journal of Vision*, *16*(11), 28.

R Core Team. (2016). *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. (2018). Psychometrics based on continuous measures: Exploiting the dynamics of computer mouse movements with time continuous multiple regression. Manuscript submitted for publication.

Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: The temporal dynamics of action selection. *Cognition*, *115*(3), 407–416.

Scherbaum, S., & Kieslich, P. J. (2018). Stuck at the starting line: How the starting procedure influences mouse-tracking data. *Behavior Research Methods, 50*(5), 2097–2110.

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, *15*(5), 207–211.

Spivey, M. J., Dale, R., Knoblich, G., & Grosjean, M. (2010). Do curved reaching movements emerge from competing perceptions? A reply to van der Wel et al. (2009). *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 251–254.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10393–10398.

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences*, *22*(6), 531–543.

Sullivan, N., Hutcherson, C., Harris, A., & Rangel, A. (2015). Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. *Psychological Science*, *26*(2), 122–134.

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*(1), 23–32.

Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the Cognitive Reflection Test. *Cognition*, *150*, 109–118.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638.

Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2018). *Measuring the (dis-) continuous mind*. Manuscript in preparation.

Zgonnikov, A., Aleni, A., Piiroinen, P. T., O'Hora, D., & di Bernardo, M. (2017). Decision landscapes: Visualizing mouse-tracking data. *Royal Society Open Science*, *4*(11), 170482.

# Design Factors in Mouse-Tracking:
# What Makes a Difference?

Pascal J. Kieslich[a], Martin Schoemann[b], Tobias Grage[b],
Johanna Hepp[c] & Stefan Scherbaum[b]

University of Mannheim, Germany[a],
Technische Universität Dresden, Germany[b],
Central Institute of Mental Health, Mannheim, Germany[c]

Correspondence concerning this article should be addressed to Pascal J. Kieslich, Department of Psychology, School of Social Sciences, University of Mannheim, L13, 17, 68161 Mannheim, Germany, phone: +49 621 181 3645, e-mail: kieslich@psychologie.uni-mannheim.de.

**Abstract**

The investigation of cognitive processes by tracking and analyzing mouse movements has become a popular method in many psychological disciplines, including language, social cognition, perception, decision-making, and memory. When creating mouse-tracking experiments, researchers face a number of design choices, for example, whether participants indicate responses by clicking on the corresponding button or by just entering the button area. Hitherto, many different settings have been employed, but little is known about how these methodological differences affect mouse-tracking data and their theoretical interpretation. We conducted a series of experiments to systematically investigate the influence of three central design factors, using a classic mouse-tracking paradigm, in which participants classify typical and atypical exemplars into one of two categories. In separate experiments, we manipulated the design factors response indication, mouse sensitivity, and starting procedure. The core finding that mouse movements deviate more towards the non-chosen option for atypical exemplars was replicated in all conditions. However, the size of this effect varied considerably depending on the response indication and starting procedure. Besides, trajectory shapes were influenced by all design factors. In bimodality analyses, some setups led to a bimodal distribution of curvature indices and some to a unimodal distribution. 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 from mouse-tracking data. All methodological designs and analyses were implemented using open-source software and are freely available from https://osf.io/xdp7a/.

Key words: mouse-tracking; cognitive processes; experimental design; decision-making; response dynamics

## Design Factors in Mouse-Tracking: What Makes a Difference?

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 being a relatively new method, mouse-tracking has quickly spread across a broad range of psychological fields, as recent reviews demonstrate (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, 2017). As a result, the methodological setup has 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. Previous evidence potentially affected by methodological choices spans inferences about the influence of various psychological factors on decision conflict (e.g., semantic 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, it is a pressing issue to understand how design factors in mouse-tracking affect trajectories und 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 assess the impact of the most central design choices and discuss how their 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 the different design factors have on mouse-tracking data. Finally, we discuss implications for the connection of mouse-tracking data and theorizing, and will 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 Figure 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 towards 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.

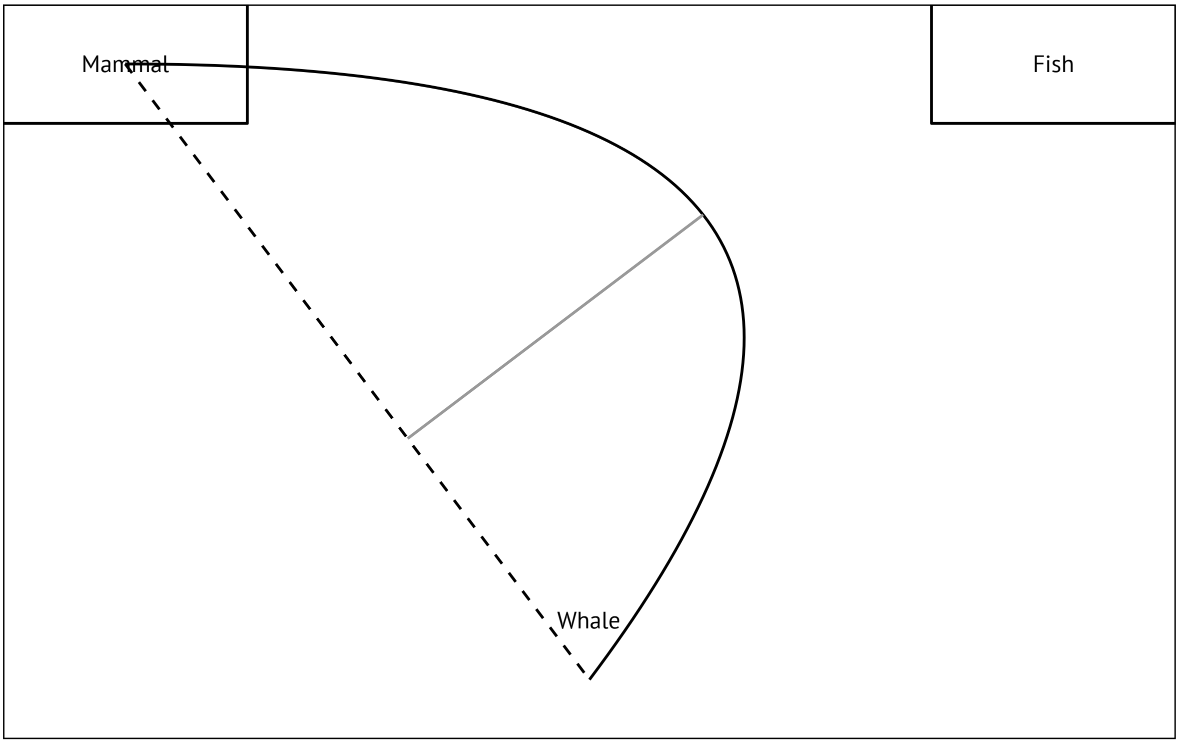


*Figure 1.* Setup of the mouse-tracking experiment including an exemplary cursor trajectory. The trial is initiated by clicking on a start button in the bottom center of the screen (not displayed) after which the name of the to-be-classified animal is presented. Participants indicate their classification decision by clicking on one of two response buttons. For the example trajectory, the maximum absolute deviation (MAD) is depicted (in gray) as the maximum perpendicular deviation of the trajectory from a straight line connecting the start and end point of the trajectory (dashed line).

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 to test predictions about individual differences and contextual factors that influence the amount of conflict in a specific decision. Second, mouse-tracking allows 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 gets gradually 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 where, 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 (Spivey, Grosjean, & Knoblich, 2005; Dale, Kehoe, & Spivey, 2007; Dale & Duran, 2011) 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, Dshemuchadse, Fischer, & Goschke, 2010), numerical cognition (see review by Faulkenberry, Witte, & Hartmann, in press), 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 atypical faces, mouse trajectories were more curved towards the opposite-gender stereotype. For instance, for an atypical female face (compared to a typical female face), the mouse trajectory was more curved towards 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 account best for the data, assuming a simultaneous co-activation of competing stereotypes that gradually gets resolved over time.

The second exemplar study used mouse-tracking to examine participants' preferential choices (Koop & Johnson, 2013, Experiment 1), asking them to indicate which of two pictures they preferred by clicking on it. 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 towards the non-chosen option, the smaller the pleasantness difference between the two pictures became. This study was used as a first step towards 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 example for 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 which was presented on the left versus right side. Participants' mouse trajectories reflected the typical Simon effect in that they were more curved towards the non-chosen 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 to disentangle 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 upwards at the beginning of the trial for the stimulus to be displayed and could indicate their response by moving the cursor onto the corresponding button (no click was required). Evidently, these three studies vary considerably with regard to their methodological setup, having implemented three different starting procedures and two different response indication procedures – further hard- 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 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) provide some recommendations for the basic set-up of mouse-tracking studies. Both suggest researchers should install measures that increase the likelihood of participants initializing their mouse movement early in the trial, to ensure cognitive processing takes place while participants are moving the mouse and not beforehand. Fischer and Hartmann additionally recommend reducing the cursor speed to better capture cognitive effects in the trajectory measures. Importantly, Hehman et al. (2015) note 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 on the effects of a specific design factor on mouse-tracking data, Scherbaum and Kieslich (2017) examined two different starting procedures. They compared data from the previously described experiment by Scherbaum et al. (2010) that used a dynamic starting procedure (i.e., a procedure in which participants have to move the mouse upwards for the stimulus to be displayed) to 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 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 provides a first indication that design factors may play an important role in mouse-tracking studies, it has only focused 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 is therefore to systematically investigate the influence of a set of design factors that commonly vary between mouse-tracking studies. For this purpose, we use different variations of a classic and simple mouse-tracking paradigm that we describe in the following. In addition, we provide researchers with open-source implementations for all 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 focus 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 analyses codes so that researchers can both replicate our analyses as well as apply them to their own data.

## Overview of 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 both features with the correct and the competing category (e.g., a whale sharing both features with 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 towards the non-chosen 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.

*Table 1.* Overview of experiments including basic design choices and their manipulation.

| 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 |

*Note.* default = 50% cursor speed (of maximum speed) with acceleration enabled, slow = 30% (Experiment 2) / 40% (Experiment 3) cursor speed with acceleration disabled, rtmax = static start with restricted total response time, initmax = static start with restricted initiation time.

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 tended towards the non-chosen option in a trial, the more the mouse trajectory deviated toward it. To quantify curvature, different indices have been proposed which 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 Figure 1) connecting the start and end position of the trajectory (with maximum deviations above the direct line, i.e., in the direction of the non-chosen 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.

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 (where both systems agree) and so-called change of mind trajectories where 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 while 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, would expect that both response options are 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). Based on a bimodality analysis, the original study by Dale et al. (2007) found support for dynamic models in their mouse-tracking data as trajectory curvature was 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 affected bimodality analyses, this would pose a general challenge for 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 was observed in one setup whereas a unimodal distribution was observed in another, it could imply that the setup directly influences the underlying cognitive process. However, an alternative (and in our view more plausible) interpretation could 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. 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 that are basically designed to answer the question whether there are one or two different types of trajectories, it has recently been argued that more fine-grained analyses are needed which allow for inferences about a variety of different trajectory shapes (Wulff et al., in press; Wulff, Haslbeck, & Schulte-Mecklenbeck, 2018). In this regard, one proposed procedure is the mapping of trajectories onto trajectory prototypes. Based on their meta-analysis of mouse- and hand-tracking studies, Wulff et al. (2018) suggest that a set of 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 Experiment 1).

## Experiment 1

In the first experiment, we examined the effect the response indication procedure has on mouse-tracking data. For this, 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 and no click was 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 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 for changes of mind (Freeman, 2014; Szaszi, Palfi, Szollosi, Kieslich, & Aczel, 2018; Wulff et al., 2018). 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 where a 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 like 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.

### Methods

**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 which was followed by the experiment currently under investigation. Participants received partial course credit for completing the studies.

The basic setup and procedures followed 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 material was 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. Afterwards, 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 Figure 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 the recording of 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 x 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 studies, 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, aged between 18 and 38 years, $M = 22.0$ years, $SD = 3.7$). The majority of 81 participants indicated a preference for the right hand (EHI score > 60), while another six participants indicated a preference for the left hand (EHI score < -60), the remaining 21 participants indicating no strong preference.

## Results

We focused on a set of typical analyses that are commonly performed in mouse-tracking studies. As such, we compared the trajectory curvature and trajectory shapes, both using 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 (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 percentage of correctly answered trials between the two design conditions. The number of correctly answered trials did not differ significantly between experimental conditions (93.5% correct answers in click condition, 93.1% in 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 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 trajectories per typicality condition first within and then across participants and separately for the click and the touch condition. The resulting aggregate trajectories are displayed in Figure 2. As expected, the aggregate trajectories deviated more towards the non-chosen option for atypical than for typical exemplars in both experimental conditions. However, this difference was considerably larger in the click condition.

[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 complete analyses online).

*Table 2.* Descriptive statistics of maximum absolute deviation (MAD) values (in pixel) per typicality condition and paired t-test results (for the comparison of the atypical and typical condition).

| | | | Typical | | | Atypical | | | *t*-test | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Experiment | Condition | *N* | *M* | *SD* | *BC* | *M* | *SD* | *BC* | *t* | *p* | $d_z$ |
| 1 | click | 53 | 142.7 | 111.5 | 0.632 | 287.5 | 237.0 | 0.641 | 4.43 | <.001 | 0.61 |
| 1 | touch | 55 | 52.2 | 78.3 | 0.442 | 79.0 | 91.6 | 0.500 | 2.69 | .009 | 0.36 |
| 2 | default | 59 | 157.8 | 158.8 | 0.558 | 283.4 | 225.7 | 0.576 | 5.49 | <.001 | 0.71 |
| 2 | slow | 59 | 73.4 | 113.1 | 0.573 | 150.9 | 141.2 | 0.593 | 4.52 | <.001 | 0.59 |
| 3 | static | 59 | 185.2 | 134.4 | 0.520 | 269.7 | 172.7 | 0.548 | 4.18 | <.001 | 0.54 |
| 3 | rtmax | 60 | 189.8 | 150.8 | 0.536 | 301.5 | 197.8 | 0.501 | 4.32 | <.001 | 0.56 |
| 3 | initmax | 66 | 304.8 | 140.7 | 0.510 | 470.9 | 203.2 | 0.473 | 7.39 | <.001 | 0.91 |
| 3 | dynamic | 60 | 297.0 | 111.6 | 0.560 | 364.1 | 154.0 | 0.508 | 3.95 | <.001 | 0.51 |

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

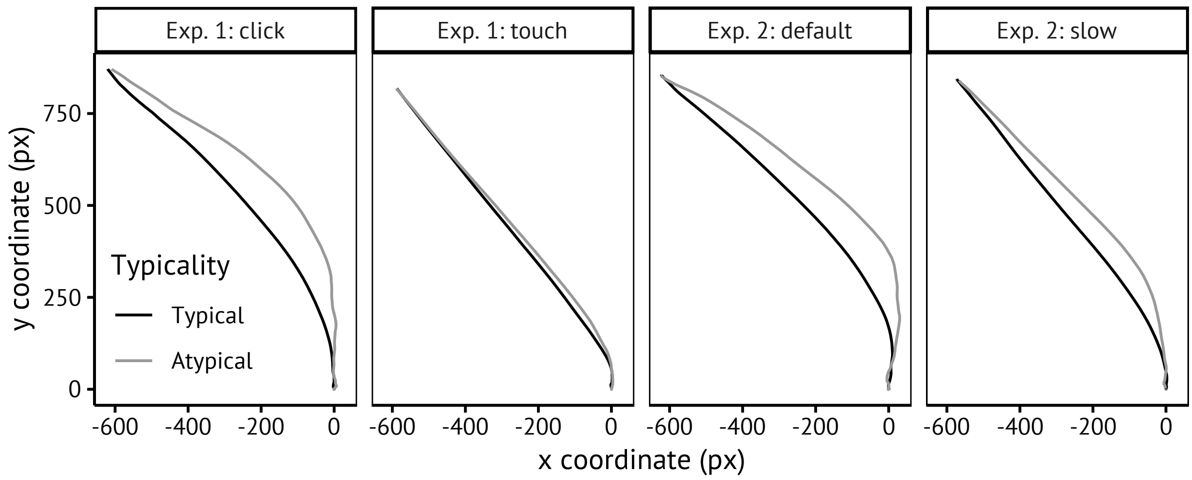


*Figure 2.* Aggregate mouse trajectories for Experiment 1 and 2. All individual trajectories were flipped to the left, time-normalized and aggregated separately per typicality and experimental condition.

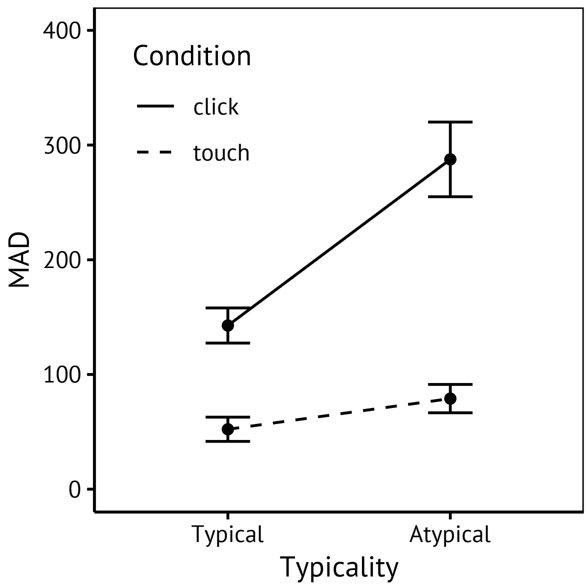


*Figure 3.* Mean of maximum absolute deviation values (MAD, in pixel) for Experiment 1 separately per typicality and experimental condition. Error bars indicate 1 *SEM*.

To statistically test these differences in aggregate trajectory curvature, we computed the MAD for each trajectory (see Figure 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 studies.

A repeated measures ANOVA using the aggregated MAD values per participant with the within factor typicality (atypical vs. typical) and the between 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 the touch condition, MAD values were significantly higher for atypical than for typical exemplars (Table 2). Besides, there was 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 (Figure 3).

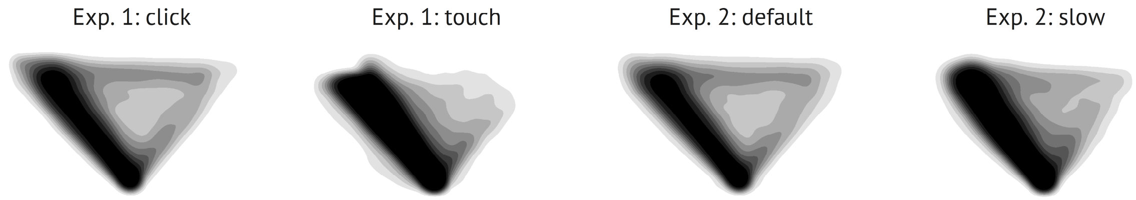


*Figure 4.* Smoothed heatmaps of the individual trajectories in Experiment 1 and 2 separately per experimental condition. Darker colors indicate higher density.

**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 analyses procedures reported in previous mouse-tracking studies (Freeman & Ambady, 2009, 2010; Freeman & Dale, 2013; Spivey et al., 2005). That is, we standardized MAD values per participant and then computed the bimodality coefficient separately for atypical and typical trials for each experimental condition. As can be seen in Table 2, both bimodality coefficients in the click condition were larger than .555, which – based on simulation studies – is used as a cut-off 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 whether there are one or two different types of trajectories, it has recently been argued that mouse-tracking researchers should perform more fine-grained analyses to make inferences about the presence or absence of different types of trajectory shapes (Wulff et al., in press, 2018). One such analysis is a graphical approach that plots a (smoothed) heatmap of all trajectories separately for each experimental condition (Kieslich et al., in press). The resulting plots for each experimental condition are displayed in Figure 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 non-chosen 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.

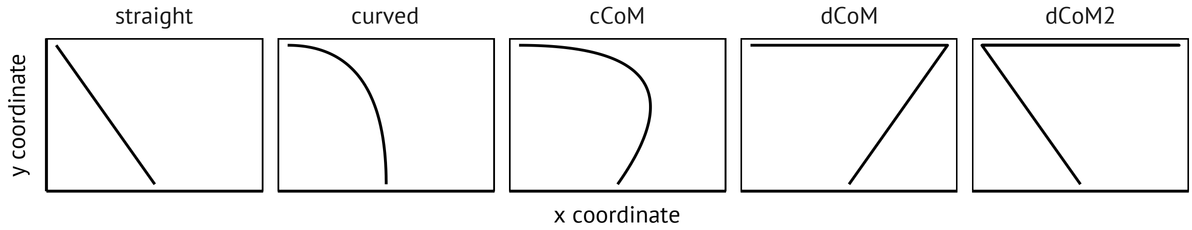


*Figure 5.* Set of prototype trajectories used in the current analyses (cCoM = continuous change of mind, dCoM = discrete change of mind, dCoM2 = double change of mind).

To quantify and statistically test for differences in the frequency of trajectory types between conditions, a recently proposed procedure is the mapping of trajectories onto trajectory prototypes (Wulff et al., in press). The prototypes used in the current study are depicted in Figure 5. They are based on prototype trajectories proposed in the meta-analysis of Wulff et al. (2018). 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 non-chosen option, discrete change of mind (*dCoM*) trajectories that first move straight to the non-chosen option and from there 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 non-chosen and the chosen option.[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.

[2] There are, of course, also other possible types of trajectories. However, we would argue that, based on the plots of individual trajectories per assigned prototype (Figures 6 and 11), this set of prototypes seems to well describe the vast majority of trajectories in the current experiments. See Footnote 14 for a reanalysis of Experiment 3, where a subset of trajectories seemed to be better accounted for by additional prototypes.

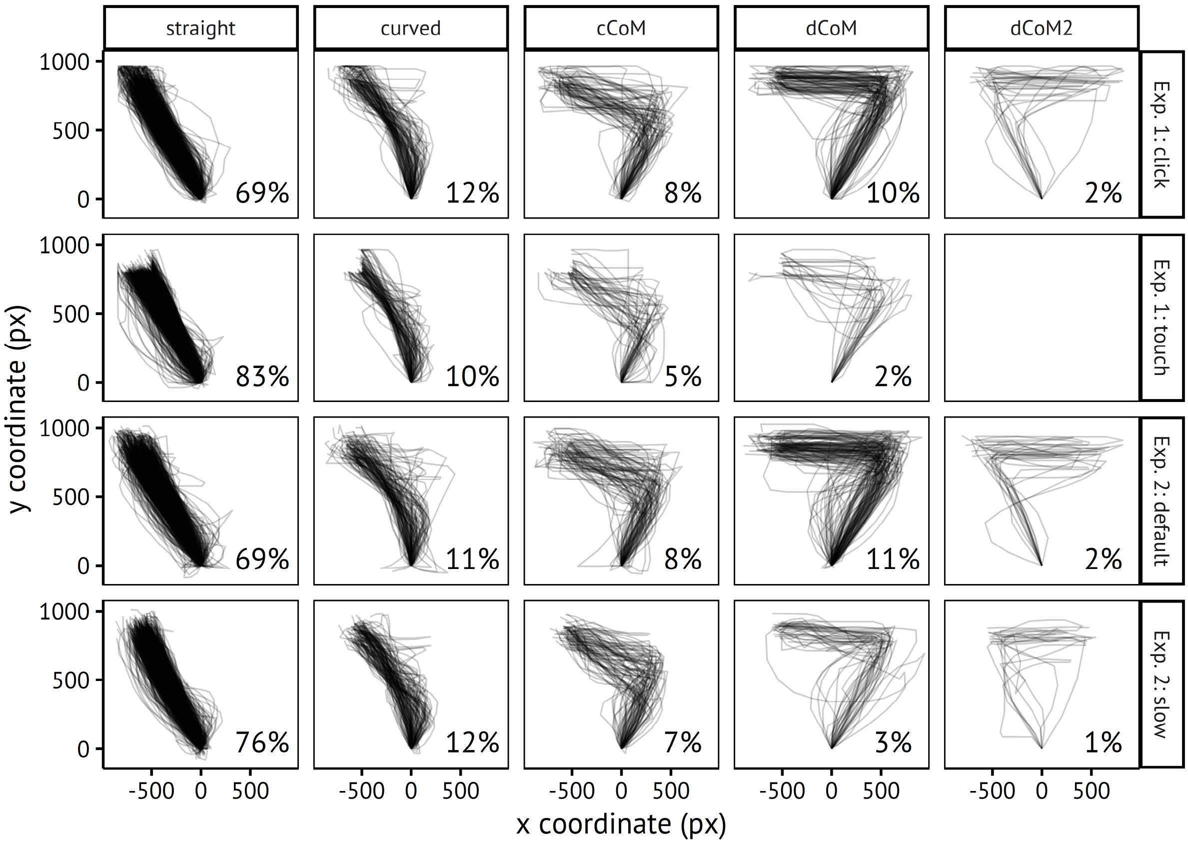


*Figure 6.* Individual trajectories per assigned prototype separately for the different experimental conditions from Experiment 1 and 2. For each prototype, the relative frequency of classifications per experimental condition is displayed.

The results of the prototype assignment and the relative frequency of each prototype classification are displayed in Figure 6. The majority of individual trajectories seems to be explained well by the current set of prototypes. The relative frequency of prototypes differed significantly between experimental conditions, $\chi^2(4) = 83.79$, $p < .001$. While most trajectories were classified as straight in the touch condition, there were relatively fewer straight classifications in the click condition. Besides, a considerably greater percentage of trajectories was classified as discrete changes of mind (both dCoM and dCoM2) in the click condition.

*Table 3.* Percentage of trajectories per assigned prototype in each experimental condition separately for typical / atypical exemplars.

| | | Assigned prototype | | | | |
|---|---|---|---|---|---|---|
| Experiment | Condition | 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 |

*Note.* cCoM = continuous change of mind, dCoM = discrete change of mind, dCoM2 = double change of mind.

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 non-chosen option indicated greater amounts of response conflict, we treated the assigned prototype as ordinal variable (straight < curved < cCoM < dCoM < dCoM2). We included a random intercept per participant as well as the effect-coded predictors typicality (atypical = 0.5, typical = -0.5), experimental condition (click = 0.5, touch = -0.5) 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 participants 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 that first move straight to the non-chosen 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 evidence 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 reviews by Freeman, 2018; Stillman et al., 2018). This experiment demonstrates 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 that could be interpreted as evidence for a dynamic model. While it is possible that changing a design aspect affects the underlying decision process, we deem it more plausible that the decision process remains unaffected and instead the design factor influences the mapping of the decision process onto the mouse movement. These findings 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 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 is under Windows 7/8 medium speed with acceleration enabled, e.g., Kieslich & Hilbig, 2014; Szaszi et al., 2018) and other studies that deliberately reduced the cursor speed and disabled 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) have 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 default settings, small movements of the wrist might already be enough to move the cursor to indicate a response due to enabled acceleration. While 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 (2014) 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 demonstrated. Alternatively, it could also be the case that the default condition leads to larger average effects by exaggerating small hand movements – although Freeman and Ambady (2010) note that extremely high speeds might lead to problems as the cursor movements may 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.

### Methods

**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 with the click condition of Experiment 1.[4] For the slow condition, the only change was that the cursor acceleration was disabled and the cursor speed was reduced from 50% to 30% (of the maximum speed). As the mouse

[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.

[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 position of all trajectories was aligned statistically during preprocessing.

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. Based on the power analysis reported in Experiment 1, we intended to ensure a minimum number of 54 participants per experimental condition. A total number of 118 participants completed the experiment (88 female, aged between 18 and 35 years, $M$ = 22.7, $SD$ = 3.3). The majority of 91 participants indicated a preference for the right hand, while seven participants indicated a preference for the left hand, the remaining 20 participants indicating 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 number of correctly answered trials did not differ significantly between experimental conditions (93.5% correct answers in default condition, 94.5% in 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].

[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”, the settings for the slow condition were “accel=off speed=6”.

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

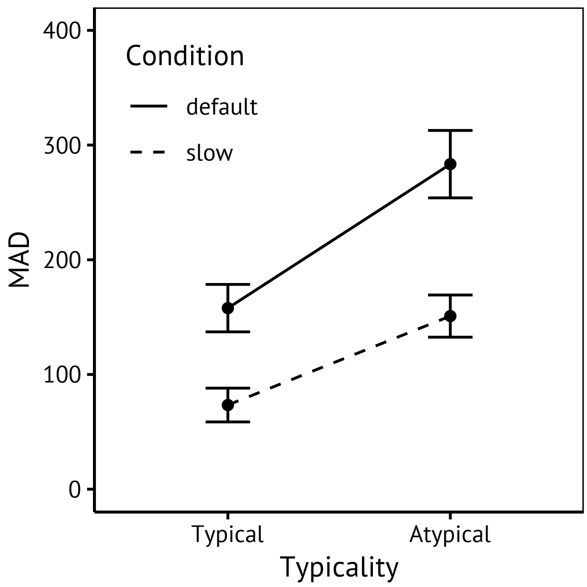


*Figure 7.* Mean of maximum absolute deviation values (MAD, in pixel) for Experiment 2 separately per typicality and experimental condition. Error bars indicate 1 *SEM*.

**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 (Figure 2). In both experimental conditions, the aggregate trajectories deviated more towards the non-chosen 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 factor typicality (atypical vs. typical) and the between 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, there was 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 (Figure 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. Both in the default and in the slow condition the coefficients were larger than .555 indicating a bimodal distribution (Table 2). The smoothed heatmaps (Figure 4) indicated that in both conditions there were a considerable number of straight trajectories but also a number of change of mind trajectories where the cursor was moved all the way to the non-chosen 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 well explained by the prototypes (Figure 6). The relative frequency 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 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 = 0.5, typical = -0.5), experimental condition (default = 0.5, slow = -0.5), 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 studies.** 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 to replicate the typicality effect and examine its stability.

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

The relative frequency of the classified prototypes did not differ significantly between studies, $\chi^2(4) = 1.01$, $p = .908$. In an ordinal mixed regression predicting the assigned prototype with typicality (atypical = 0.5, typical = -0.5), study (study 1 = 0.5, study 2 = -0.5), and their interaction, there was a significant effect of the typicality predictor ($z = 6.85$, $p < .001$). However, there was no significant effect of study ($z = 0.10$, $p = .923$) nor was there a significant interaction between typicality and study ($z = 0.37$, $p = .713$).

### 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 size 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 settings, which amplify even small movements towards one of the options. Interestingly, while 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 also was an increase of more extreme movement types for atypical exemplars in the slow condition. At the same time, there was also no evidence that a slow condition would increase the 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 studies, as the default condition was virtually identical with the click condition of Experiment 1. Across all analyses, there were no significant differences between studies, 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. While 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 produces 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 impact on the observed effects is 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 that we will return to in the next experiment that compares 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 movement early in the trial, hoping to ensure that the complete decision process is reflected in the movement (Fischer & Hartmann, 2014; Hehman et al., 2015; Scherbaum & Kieslich, 2017). We implemented and tested three of these 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*) 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 each trial if participants initiated their movement too slowly (see Hehman et al., 2015, for a discussion; see Freeman & Ambady, 2009, 2011; Stolier & Freeman, 2016; Papesh & Goldinger, 2012; Yu, Wang, Wang, & Bastin, 2012, for exemplary applications). The fourth starting procedure that we include methodologically ensures a movement initiation even before stimulus presentation. In this *dynamic* starting procedure, participants have to initiate an upwards movement for the stimulus to be displayed. This procedure has been employed by a number of mouse-tracking studies (see Scherbaum & Kieslich, 2017, for a discussion; see Scherbaum et al., 2010; Huette & McMurray, 2010; Dshemuchadse et al., 2013; Frisch et al., 2015, for exemplary applications).

Previous recommendations in the literature have stated that starting procedures which 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 compared to a static starting procedure. This should hold in particular for the dynamic and the initmax starting procedure, 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, 2017). This study found that a dynamic starting procedure did not lead to significantly larger cognitive effects found in aggregate curvature measures than a static starting procedure in which the stimulus was presented after a fixed, short delay. However, a dynamic starting procedure led to larger cognitive effects in temporal analyses that assessed how the cursor movement direction was affected by different factors at a specific time point. While 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 studies (and, as a consequence, without random assignment). Besides, it only considered two possible starting procedures. Therefore, a study that experimentally compares a larger set of commonly used starting procedures is needed.

## Methods

**Procedure and materials.** The 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 of the study, participants provided demographic information and answered the EHI. After completing the study, participants had the chance to win one of several vouchers for local coffee shops (and other businesses, including a voucher for a German soccer league game) or sweets.

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

Apart from the starting procedure that was manipulated between participants (static vs. rtmax vs. initmax vs. dynamic), all experimental conditions were identical. 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

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

to initiate an upwards 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 upwards 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 upwards for the stimulus to be presented.[9]

**Participants.** Based on the power analysis reported in Experiment 1, we intended to ensure a minimum number of 54 participants per experimental condition. A total number of 245 participants completed the experiment and was included in the analysis. The sample comprised 162 women and participants were between 18 and 50 years old ($M$ = 21.9, $SD$ = 3.3). The majority of 172 participants indicated a preference for the right hand, 16 participants indicated a preference for the left hand, and 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 were conducted as manipulation checks. Besides, as 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 and the static starting procedure served as baseline condition.

**Correctness.** Across all trials, the number 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) while the initmax condition led to a significantly lower performance ($z$ = -3.12, $p$ = .002). The rtmax condition also led to a significantly lower

[8] To calculate the initiation time, we used the same upwards movement criterion as in the dynamic condition (an upwards 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, 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 for performing this upwards movement (in order to keep the setup as simple and understandable as possible).

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 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.

*Table 4.* Mean (SD) of the per participant aggregated timing variables in Experiment 3 presented separately for each condition (in ms).

| Condition | $RT_{initial}$ | Initiation time | RT |
|---|---|---|---|
| static | 808.5 (324.1) | 508.7 (215.6) | 2110.4 (654.1) |
| rtmax | 650.1 (176.6) | 437.3 (160.0) | 1521.6 (183.4) |
| initmax | 377.4 (159.5) | 243.1 (142.8) | 1471.7 (248.6) |
| dynamic | 773.4 (752.2) | 348.7 (233.2) | 2805.4 (1199.8) |

*Note*. $RT_{initial}$ = Time until cursor was moved 50 px upwards.

As a first variable, we computed the time it took participants to move the mouse upwards for 50 px ($RT_{initial}$). As participants in the initmax condition were specifically instructed to initiate their movement within 0.6 s, we expected that the $RT_{initial}$ should be lower in the initmax compared to the static condition. The average $RT_{initial}$ in the initmax condition was considerably smaller than the instructed time limit (although participants still exceeded the time limit in 12.4% of trials).

[10] Initially, we planned to exclude trials exceeding the movement initialization time limit in the initmax condition. We noticed during pilot trials 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 that would have to be excluded because they did not initiate their movement in time for 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 this procedure. All main results can be replicated when excluding trials in the initmax condition where the time limit was not met and excluding, as a consequence, five participants for whom no correctly answered trials in either typicality condition remained (see complete analyses online).

The $RT_{initial}$ 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 the $RT_{initial}$ 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 is initiated. The initiation time 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 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 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 is only displayed after the initiation of the upwards movement and, consequently, the processing may start later. Interestingly, if we calculate 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 is on average ($M$ = 2021.1 ms, $SD$ = 671.1 ms) quite comparable to the total RT in the static condition.

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

[12] When repeating the analyses using median instead of mean values per participant, the general pattern could be 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.

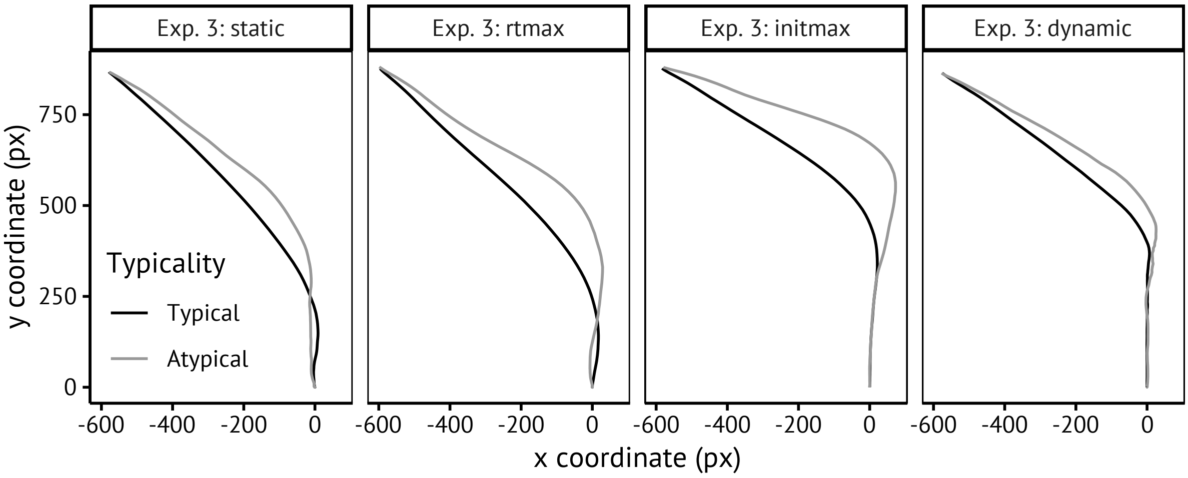


*Figure 8.* Aggregate mouse trajectories for Experiment 3. All individual trajectories were flipped to the left, time-normalized and aggregated separately per typicality and experimental condition.

**Aggregate trajectory curvature.** To get a general impression of the effect of the starting procedure on trajectory curvature, we inspected the aggregate time-normalized trajectories (Figure 8). In all experimental conditions, the aggregate trajectories deviated more towards the non-chosen option for atypical than for typical exemplars. The dynamic and initmax conditions generally led to prolonged vertical upwards movements compared to the static and rtmax 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 factor typicality (atypical vs. typical) and the between 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, there was 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 compared to the static condition, $t(241) = 4.10$, $p < .001$. The MAD values in the rtmax condition did not differ significantly from the static condition, $t(241) = 0.72$, $p = .470$.

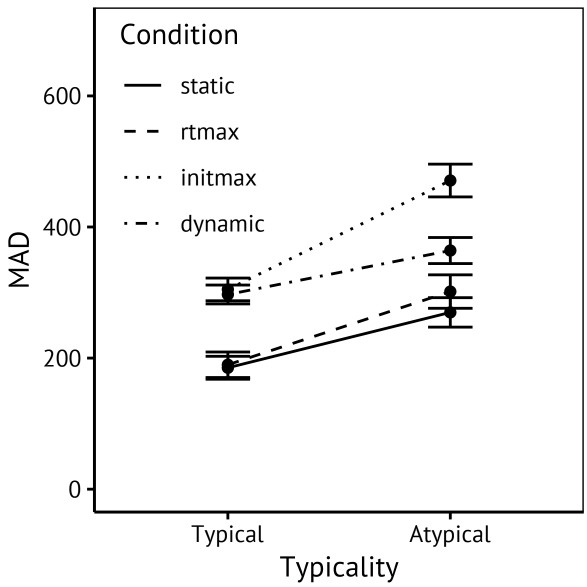


*Figure 9.* Mean of maximum absolute deviation values (MAD, in pixel) for Experiment 3 separately per typicality and experimental condition. Error bars indicate 1 *SEM*.

There also was 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 Figure 9, the typicality effect was significantly larger in the initmax than in the static condition, $t(241) = 2.68$, $p = .008$. There was no significant difference in the size of the typicality effect between the dynamic and the static condition, $t(241) = -0.56$, $p = .576$, nor between the rtmax and the static condition, $t(241) = 0.87$, $p = .383$.

**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, both for typical and for atypical trials, with the exception of the typical trials in the dynamic condition where the value of .560 was slightly larger than the cut-off.

The smoothed heatmaps (Figure 10) indicated that in the static and rtmax conditions there were a considerable number of straight trajectories. In the dynamic and the initmax condition there were fewer straight trajectories but instead many trajectories that moved upwards for a distance (longer than the required movement criterion). In all conditions, there also seemed to be a number of change of mind trajectories where the cursor was moved all the way to non-chosen option before moving to the chosen option.

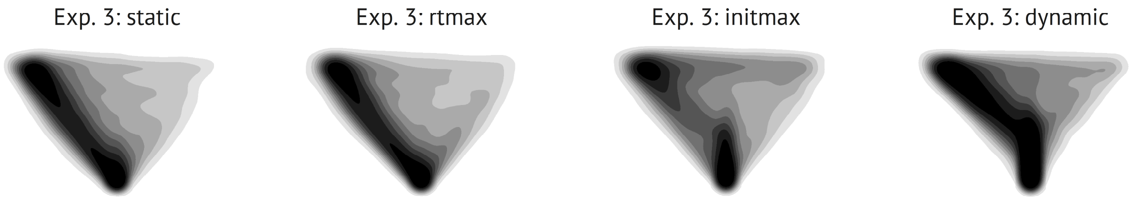


*Figure 10.* Smoothed heatmaps of the individual trajectories in Experiment 3 separately per experimental condition. Darker colors indicate higher density.

To quantify and statistically test for differences in the frequency of trajectory types between conditions, we mapped 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 (Figure 11).[14] The relative frequency 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 was classified as straight. Similar results were also obtained in the rtmax condition. In the dynamic condition, the majority of trials were classified as curved, while the initmax condition led to a roughly even split of straight and curved classifications and a considerable increase of dCoM classifications compared with 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 = 0.5, typical = -0.5) and experimental condition (dummy coded, static serving as 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 condition led to significantly more extreme trajectories compared with the static condition, $z = 5.99$, $p < .001$, and $z = 4.37$, $p < .001$. For the initmax condition both more curved and change of mind trajectory types occurred while for the dynamic condition, there seemed to be especially more curved trajectories (see Figure 11). With regard to the interaction of typicality and condition, there were no significant interactions for the rtmax and initmax

[14] Upon closer inspection of the data, we discovered a small number of trajectories in the initmax and dynamic condition that were not captured well by the existing prototypes as they went upwards all the way to the top of the screen and then either moved left to the chosen option or first right to the non-chosen 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 complete analyses online).

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 compared to the static condition, $z$ = -2.19, $p$ = .029.

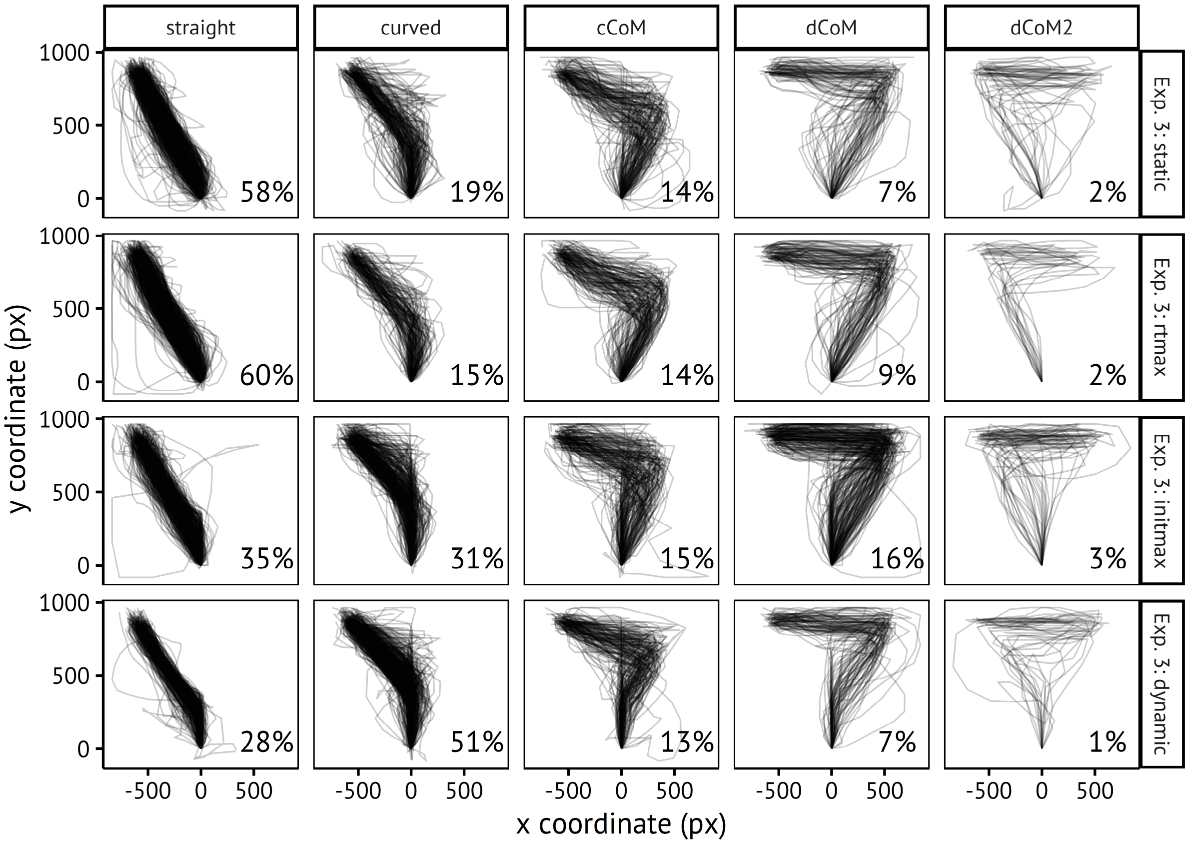


*Figure 11.* Individual trajectories per assigned prototype separately for the different experimental conditions from Experiment 3. For each prototype, the relative frequency of classifications per experimental condition is displayed.

### Discussion

In this experiment, we examined the influence of four different starting procedures on mouse-tracking data. Several previous mouse-tracking studies have used a static starting procedure in which the stimulus is immediately presented after participants have clicked on the start button and participants do not receive any instruction regarding movement initiation. However, this poses the risk that participants may 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 is 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 upwards for the stimulus to be displayed (dynamic).

Results showed that the initmax condition, in which participants were instructed to initialize an upwards mouse movement within 0.6 s, led to a significantly larger typicality effect than 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 non-chosen 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 the initiation of the mouse movement (see Hehman et al., 2015, for a discussion). With the current setting (move upwards 50 px within 0.6 s), we found that participants could not always meet the time limit for initiating their mouse movement. While 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 of 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 upwards 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 might come closest to the idea of continuously curved trajectories – potentially even more so if used in combination 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). These studies also often employed a dynamic starting procedure in combination with restrictions regarding the time for initiating the upwards 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, 2017).[15]

[15] Older studies used a different movement initiation criterion (moving upwards 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, 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 (that could alter basic decision processes, e.g., through an 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 in the static condition and participants initiated their upwards mouse movements earlier in the trial. Still, future studies could explore the use of a stricter total time limit, at the risk of altering 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 (where the bimodality coefficient had indicated bimodality). The main methodological difference for the static condition in the current study 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 compared to the slow condition from Experiment 2 (see Figures 4 and 8). This, in turn, led to higher average MAD values (Table 2) and deviations from this higher baseline have a smaller effect on the bimodality coefficient. Still, the inspection of the heatmaps 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 as well as their shape. As 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 while a static start leads to a majority of straight trajectories. While the former type of trajectories is usually associated with dynamic process models, the latter type is rather interpreted 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 the initmax starting procedures increasing the likelihood that the decision process (and especially its early stages) are more continuously mapped onto the mouse movement.

## General Discussion

Over the past decade, mouse-tracking has spread to a multitude of psychological areas and was 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 entailed considerable variation in the methodological setup of previous mouse-tracking studies. To improve understanding of the empirical and ultimately also theoretical consequences of methodological differences for mouse-tracking data, the present study reported a systematic investigation of three central design factors. In a series of experiments, the design factors response indication, starting procedure, and mouse sensitivity were varied while the basic experimental setup remained constant and followed a classic mouse-tracking experiment by Dale and colleagues (2007). In all methodological setups, the postulated typicality effect was replicated in that mouse trajectories deviated more towards the non-chosen 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. Besides, trajectory shapes were influenced by all design factors: In traditional bimodality analyses, some setups led to the distribution of curvature indices being classified as unimodal, while other setups were 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, could explain differences in the size of the typicality effect between conditions.

### Implications for Interpreting Mouse-Tracking Results

The findings of the current study have general implications for interpreting 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. However, the size of this effect varied considerably between setups. Therefore, it is well possible that smaller and less robust effects than the typicality effect we investigated would not be detected in certain methodological setups. As a consequence, before concluding that a certain manipulation does not influence the 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 current experiments anything from a small (touch condition from Experiment 1) to a large effect (initmax condition from Experiment 3) 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 aimed to differentiate between dynamic and dual-system models that should lead to a unimodal versus a bimodal distribution, respectively. The current study demonstrates that, depending on the methodological setup, both unimodal and bimodal distributions can be obtained in the very same psychological task. Assuming that cognitive processing is not influenced by the setup (which we deem unlikely but cannot completely rule out based on the current data), this implies that different shapes can occur for the same cognitive process. In our view, this indicates that the mapping of cognitive processes onto mouse movements can vary depending on the methodological setup and an interpretation of trajectory shapes needs to consider the methodological conditions under which they were obtained.

More specifically, the starting procedure influences the degree to which early aspects of the decision process are reflected in the mouse movement, with an initmax and dynamic starting procedure increasing the likelihood that early aspects are captured, while 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 it 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 towards that option. The click condition allowed participants to move all the way to an option (if this is 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 click response mode is more likely to produce a mix of straight and change of mind trajectories than a study with a dynamic starting procedure and touch response mode.

Third, the results demonstrate the usefulness of the newly proposed analysis method for identifying different types of movement trajectories (Wulff et al., in press, 2018). On the one hand, it allows unpacking the effect 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 current 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 different types of trajectories are present in the data or not. In Experiment 1 and Experiment 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 compared to 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 upwards for the stimulus to be displayed. This, in turn, is in line with previous findings by Scherbaum and Kieslich (2017) 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 current 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 upwards 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 turn off acceleration when comparing the different starting procedures – a setup which we would 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 a response by touching the button with the cursor. However, this was related to an occurrence of more extreme trajectories. Besides, 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 which 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 which setup is suited for which psychological tasks. While encouraging early mouse movements through the respective starting procedure should work well for studies with stimuli that can be processed very quickly and where 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 may be better suited so participants can initiate the movement after acquiring the stimulus information. This is at the risk that they may 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 current study in conditions that used a static starting procedure).

### Limitations

The current study investigated the influence of design factors on mouse-tracking data by replicating a classic mouse-tracking experiment (Dale et al., 2007, Experiment 1) with different methodological setups. To investigate the impact of these different setups on mouse-tracking data, we had to make a number of choices, each of which entail 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 carry-over effects. Therefore, we could only implement a limited amount 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. Lastly, 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 to 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 it is combined with a starting procedure that encourages early movement initiation. Besides, for starting procedures that encourage an early movement initiation, a fast cursor speed might lead to problems (as previously discussed). Consequently, the study of design factors in mouse-tracking warrants further investigation and extension of the present results to investigate the effects of each design factor with different combinations of the other factors.

Besides, while 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). There are, of course, many more speed settings that could be used – both faster and slower speed – 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 which was not included in the current study. Besides, 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 (2017) 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, there are likely additional design factors relevant in mouse-tracking studies. One potentially relevant factor is the stimulus position. This is already suggested by the fact that the shape 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 studies. Another potentially relevant factor concerns the response button position. Research on these and additional factors is currently under way (Grage, Schoemann, Kieslich, & Scherbaum, 2018; Schoemann, Lüken, Grage, Kieslich, & Scherbaum, 2018). In addition, there are also 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. Compared to the paradigm by Dale and colleagues 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 would argue that the basic setup of the current study is representative of many mouse-tracking studies and that the effects observed in each study were reliable (as indicated by the cross-study comparison of two experimental conditions with identical setups that did not differ significantly in any analyses). Still, all of this highlights the need for future studies that examine the effects of the individual design factors in other mouse-tracking paradigms.

Finally, in our analyses we focused on the influence design factors have on trajectory curvature and the 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 may also have an influence in each of these analyses (see Scherbaum & Kieslich, 2017, who perform a number of additional analyses in their comparison of two starting procedures). As the data for all experiments is freely provided in an open format (along with open-source software for their analysis), interested researchers are invited to use it to explore how the design factors in the current studies influence data in their particular analysis of interest.

### Conclusion

We presented one of the first comprehensive investigations of three central design factors in mouse-tracking – response indication, mouse sensitivity, and starting procedure – and 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. Lastly, an extension of the present investigation to further setups and tasks seems imperative – an endeavor that we would like to encourage further researchers to pursue with help of the experiments, analysis code and data that we have made available.

## Acknowledgments

This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences funded by the German Research Foundation. We thank Arndt Bröder for providing helpful comments on an earlier version of this manuscript and Rebecca Heck, Daniel Megally, Anke Bosse, Rika Groß, Johanna Kaiser, Eva Kickhöfen, Silvia Martin, and Anna Lena Metzger for their assistance in data collection.

## References

Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science*, *35*(5), 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*(1), 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*(1), 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*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Faulkenberry, T. J., Witte, M., & Hartmann, M. (in press). Tracking the continuous dynamics of numerical processing: A brief review and editorial. *Journal of Numerical Cognition*. Retrieved from https://psyarxiv.com/pruz7/

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*(1), 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*. Advance online publication. 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*(10), 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*(1), 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*(4), 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*(4), 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*(1), 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*(4), 377–388. https://doi.org/10.1007/s10339-015-0666-0

Grage, T., Schoemann, M., Kieslich, P. J., & Scherbaum, S. (2018). *Validate mouse-tracking: How design factors influence action dynamics in the Simon task*. Manuscript in preparation.

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*(7), 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*(3), 384–401. https://doi.org/10.1177/1368430214538325

Huette, S., & McMurray, B. (2010). Continuous dynamics of color categorization. *Psychonomic Bulletin & Review*, *17*(3), 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*(1), 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*(5), 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*(6), 510–522.

Koop, G. J. (2013). An assessment of the temporal dynamics of moral decisions. *Judgment and Decision Making*, *8*(5), 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*(5), 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*(8), 750–758.

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, *67*(4), 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*(4), 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*(2), 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*(1), 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*(5), 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*(3), 407–416. https://doi.org/10.1016/j.cognition.2010.02.004

Scherbaum, S., & Kieslich, P. J. (2017). Stuck at the starting line: How the starting procedure influences mouse-tracking data. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-017-0977-4

Schoemann, M., Lüken, M., Grage, T., Kieslich, P. J., & Scherbaum, S. (2018). *Validate mouse-tracking: How design factors influence action dynamics in intertemporal decision making*. Manuscript submitted for publication.

Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., & Willemsen, M. C. (2017). Process-tracing methods in decision making: On growing up in the 70s. *Current Directions in Psychological Science*, *26*(5), 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*(5), 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 of the United States of America*, *102*(29), 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*(9), 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*(6), 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*(6), 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*(2), 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*(1), 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. (2018). *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*(3), 289–306. https://doi.org/10.1521/soco.2012.30.3.289

Behav Res (2018) 50:2097–2110
DOI 10.3758/s13428-017-0977-4

# Stuck at the starting line: How the starting procedure influences mouse-tracking data

**Stefan Scherbaum**[1] · **Pascal J. Kieslich**[2]

Published online: 3 November 2017


**Abstract** Mouse-tracking is an increasingly popular method to trace cognitive processes. As is common for a novel method, the exact methodological procedures employed in an individual study are still relatively idiosyncratic and the effects of different methodological setups on mouse-tracking measures have not been explored so far. Here, we study the impact of one commonly occurring methodological variation, namely whether participants have to initiate their mouse movements to trigger stimulus presentation (dynamic starting condition) or whether the stimulus is presented automatically after a fixed delay and participants can freely decide when to initiate their movements (static starting condition). We compared data from a previous study in which participants performed a mouse-tracking version of a Simon task with a dynamic starting condition to data from a new study that employed a static starting condition in an otherwise identical setup. Results showed reliable Simon effects and Congruency Sequence effects on response time (RT) and discrete trial-level mouse-tracking measures (i.e., average deviation) in both starting conditions. In contrast, within-trial continuous measures (i.e., extracted temporal segments) were weaker and occurred in a more temporally compressed way in the static compared to the dynamic starting condition. This was in line with generally less consistent movements within and across participants in the static compared to the dynamic condition. Our results suggest that studies that use within-trial continuous measures to assess dynamic aspects of mouse movements should apply dynamic starting procedures to enhance the leakage of cognitive processing into the mouse movements.

**Keywords** Mouse-tracking · Methodology · Boundary conditions · Simon task

**Electronic supplementary material** The online version of this article (https://doi.org/10.3758/s13428-017-0977-4) contains supplementary material, which is available to authorized users.

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

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

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

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

## Stuck at the starting line: How the starting procedure influences mouse-tracking data

To understand how the cognitive system brings forth an astonishing spectrum of behavior, the study of cognitive processes is a central endeavor. The tracing of cognitive processes has been an important tool, starting with process tracing methods such as verbal protocol analyses (Ericsson & Simon, 1984; Newell & Simon, 1972), complemented later by objective measures, such as eye-tracking. The latter method allowed researchers to trace cognitive processes from behavior instead of relying on introspective self-reports. In recent years, a further method extended the arsenal of process tracing methods: Mouse movement tracking offers a simple way to trace participants' cognitive processing while they make choices and execute response movements. The central assumption behind mouse-tracking is that cognitive processing is continuously revealed in hand (and mouse) movements (Spivey, 2007; Spivey & Dale, 2006; Spivey, Grosjean, & Knoblich, 2005). In return, the analyses of mouse movement data can be used to make inferences about the development of the cognitive processes leading up to the final decision. The advantages of mouse-tracking are manifold: the hardware is cheap, mouse movement measuring can be implemented in most experimental software, and most participants are highly familiar with moving a computer mouse. Hence, mouse-tracking flourished in recent years (for a review, see Freeman, Dale, & Farmer, 2011), finding application in studies of language and semantic processing (Dale, Kehoe, & Spivey, 2007; Dshemuchadse, Grage, & Scherbaum, 2015; Spivey et al., 2005), conflict resolution (Scherbaum, Dshemuchadse, Fischer, & Goschke, 2010), and value-based decision making (Dshemuchadse, Scherbaum, & Goschke, 2013; Kieslich & Hilbig, 2014; Koop & Johnson, 2013; Scherbaum, Dshemuchadse, Leiberg, & Goschke, 2013; van Rooij, Favela, Malone, & Richardson, 2013).

As it is typical for uprising new methods, a variety of methodological approaches can be found that vary between application domains and even between research groups within the same domain. For example, in some mouse-tracking studies, participants need to start a trial actively by initiating a mouse movement that, in turn, starts the presentation of the imperative stimulus (e.g., Dshemuchadse et al., 2013; Scherbaum et al., 2010), while in other studies, participants respond by starting their movement *after* the imperative stimulus has already been presented (e.g., Dale et al., 2007; Kieslich & Hilbig, 2014; Koop & Johnson, 2011). Inevitably, such methodological differences pose two challenges to the community of mouse-tracking users: first, implementing a study becomes a relatively idiosyncratic process in which a researcher has to weigh different methodological options to the best of her knowledge. Second, without systematic investigation, it remains unclear in how far methodological differences influence the results with respect to the posed research question (Fischer & Hartmann, 2014). Therefore, it seems of urgent importance to start investigating how far methodological differences influence results of mouse-tracking studies, first to allow for a consistent interpretation and comparability of studies employing different methodological setups, and, second, as a basis for developing methodological standards as they are common for other process tracing techniques, for example, electroencephalography or eye-tracking.

Here, we present a first humble step into this direction by comparing how differences in the way participants start a trial influence mouse-tracking data and results. In this regard, two common approaches are compared: (1) a "dynamic starting procedure" in which participants have to initiate a movement first to trigger stimulus presentation, and (2) a "static starting procedure" in which a stimulus is presented after a fixed time interval and participants can freely initiate their movement. The discussion in how far these (and other) differences influence data quality – especially consistency and reliability – is an ongoing debate in the community, though still mainly at conferences and meetings.

To study differences in the starting procedure, we used the Simon task, a paradigm that is well established in cognitive psychology. In this task, participants have to select one of two possible response options depending on the magnitude of a number shown on the screen (e.g., the left option if the number is smaller than 5, otherwise right), but have to ignore the location of the number on the screen (e.g., left or right). This arrangement can lead to two types of trials. In conflict trials, the direction indicated by number magnitude (e.g., left) differs from the location of the number on the screen (e.g., right). In non-conflict trials, the direction indicated by number magnitude corresponds to the location of the number on the screen. Two reliable effects are commonly observed in the Simon task. The Simon effect refers to slower response times in conflict trials compared to non-conflict trials. The congruency-sequence effect refers to a decrease in the Simon effect if the current trial was preceded by a conflict trial compared to a preceding non-conflict trial. In an original study, we have investigated these effects and their dynamics via mouse-tracking using a dynamic starting procedure (Scherbaum et al., 2010). Participants clicked on a box at the bottom-center of the screen and then started to move the mouse cursor upwards. After meeting a movement threshold, the number was presented so that participants had to select their left-right response while already moving. We chose this procedure, first, to ensure that the cognitive processes influencing response selection leaked as strongly as possible into the mouse movements and, second, to establish a high level of consistency within and across participants regarding the movements at the start of the trial. In our study, we found that the Simon effect and the congruency sequence effect affected mouse movements, as mouse movements were more curved toward the incorrect response option in conflict trials and this conflict effect was reduced if the previous trial also was a conflict trial. We further analyzed the timing profile of these influences, determining when and how strongly congruency and congruency sequence influenced mouse movement direction. The pattern of results could be replicated across two consecutive studies speaking for the robustness of both the effects per se and their dynamics.

Hence, these effects and the results of the original study offer an ideal platform to investigate in how far differences in the starting condition, that is, a static versus a dynamic starting procedure, influence the consistency of movements within and across participants, and in how far these properties of the data influence different mouse-tracking measures. Currently, many mouse-tracking studies rely on discrete measures of effects on the trial level, calculating initiation times, movement times, movement deviation, or the number of changes in movement direction for statistical analysis (Freeman & Ambady, 2010). As many discrete measures integrate information over the course of the whole trial they should be relatively robust against changes in design.

However, other groups have gone further and analyzed the movements as time-series within the trial (Dshemuchadse et al., 2013; Scherbaum et al., 2010, 2013; Sullivan, Hutcherson, Harris, & Rangel, 2015), similar to the analysis of EEG. Due to their higher temporal resolution, within-trial continuous measures should be more prone to changes in the setup or procedure of mouse-tracking.

Taken together, in a mouse movement version of the Simon task, we will investigate to what extent a specific change in the methodological setup influences the consistency of mouse-tracking data. Specifically, we will examine to what degree a static starting condition, in which the stimulus is presented automatically after a fixed delay and participants can freely decide when to initiate their movements, might decrease data quality compared to a dynamic starting condition that requires participants to initiate mouse movements in order to trigger stimulus presentation. We combined data from our original study (Scherbaum et al., 2010, Experiment 2) in which we used the dynamic starting condition, with data from a new sample of participants who performed the identical task, except that we used a static starting condition. We expected (1) that cognitive effects on discrete movement measures would only slightly be influenced by differences in the starting condition whereas (2) cognitive effects on within-trial continuous movement measures would be larger and more reliable in the dynamic starting condition compared to the static starting condition. Underlying this latter phenomenon, we expected (3) that the consistency of movements within trials, across trials, and across participants would be higher in the dynamic starting condition than in the static starting condition.

## Method

### Participants

Twenty right-handed students (17 female, mean age = 20.5 years) of the Technische Universität Dresden, Germany, participated in the experiment. In the original study, 20 right-handed students (17 female, mean age = 21.1 years) of the Technische Universität Dresden had participated. All participants had normal or corrected-to-normal vision. They gave informed consent to the study and received either class credit or 5€ payment.

### Apparatus and stimuli

The apparatus and stimuli in the new experiment were identical to the apparatus and stimuli in the original experiment. Target stimuli (numbers 1–4 and 6–9) were presented in white on a black background on a 17-in. screen running at a resolution of 1,280 × 1 024 pixels (75 Hz refresh rate). They had a width of 6.44° and a horizontal distance to the screen center of 20.10°. Except for one procedural difference (see below), the setup of the current study was the same as in the original study. Response boxes (11.55° in width) were presented at the top left and top right of the screen. As presentation software, we used Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997) in Matlab 2006b (the Mathworks Inc., Natick, MA, USA), running on a Windows XP SP2 personal computer. Responses were carried out by moving a standard computer mouse (Logitech Wheel Mouse USB). In the driver settings, non-linear acceleration ("optimize movements" option) was switched off to enable a linear ballistic arm movement and to ensure that the upwards movement (within the trial) and the downwards movement (in the inter-trial interval) cancelled out each other. Furthermore, the mouse speed was set to one-quarter of maximum speed, a setting that ensured that participants could reach the target box with one continuous upwards movement while at the same time ensuring that the movement range was as large as possible. Mouse trajectories were sampled with a frequency of 92 Hz and recorded from stimulus presentation until response in each trial.

### Procedure

The procedure in the new experiment was identical to the procedure of the original experiment, with the exception of the starting condition. Participants were asked to move the cursor into the upper left response box for digits smaller than five and into the upper right response box for digits larger than five. Each trial consisted of three stages: the alignment stage, the start stage, and the response stage. In the alignment stage, participants had to click on a red box (11.55° in width) at the bottom of the screen within a deadline of 1.5 s. This served to align the starting area for each trial. After clicking on this box, the start stage began and two response boxes in the right and left upper corner of the screen were presented. The procedure of the start stage differed between the new experiment, in which we implemented a static starting condition, and the original experiment, in which we had implemented a dynamic starting condition. In the static starting condition, the start stage simply lasted 200 ms (this was the average duration of the start stage in the original experiment that used the dynamic starting condition) and participants simply had to wait for the start of the response stage. In contrast, in the dynamic starting condition, participants were required to start the mouse movement upwards within a deadline of 1.5 s. Specifically, the response stage only started after participants moved the mouse upwards for at least 4 pixels in each of two consecutive time steps. Usually, this procedure is applied to force participants to be already moving when entering the decision process to assure that they do not decide first and then only execute the final movement (Scherbaum et al., 2010). In the response stage, the imperative stimulus (the number) was presented. For this stage, participants in both starting conditions were instructed to respond as quickly and accurately as possible and to move the mouse continuously upwards once they had initialized their movement.

The trial ended after moving the cursor into one of the response boxes within a deadline of 2 s (see Fig. 1). If subjects missed the respective deadline in one of the three stages, the next trial started with the presentation of the red start box. Response times (RTs) were measured as the duration of the response stage, reflecting the interval between the onset of the target stimulus and the arrival of the mouse cursor in the response box area.

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

### Design

The Simon task used here is based on the conflict between the direction indicated by the number (left vs. right) and the position on screen where the number was presented (left vs. right). Hence, we varied these properties orthogonally for the current trial and for the preceding trial resulting in the following independent variables: the direction and location of the number in the current trial (direction$_N$ [left vs. right] and location$_N$ [left vs. right]), and the direction and location of the number in the previous trial (direction$_{N\text{-}1}$ [left vs. right] and location$_{N\text{-}1}$ [left vs. right]). This resulted in four combinations for the current trial and four combinations for the previous trial. The sequence of trials was balanced within each block by pseudo randomization resulting in a balanced Trial$_N$ (4) × Trial$_{N\text{-}1}$ (4) × trial repetition (16) transition matrix. This way, we obtained a balanced sequence of 256 trials with systematically manipulated congruency of direction/ location within the current trial (congruency$_N$), congruency of direction/location within the previous trial (congruency$_{N\text{-}1}$), and sequences of designated responses. Three such sequences were generated, resulting overall in three blocks and 256 trials per block.

### Data preprocessing and statistical analyses

We excluded erroneous trials and trials following an error (4.2 %). To avoid any bias in data analysis of the two methodologically different sets, we refrained from outlier analysis as performed in the original study. Mouse trajectories were remapped so that all trajectories would end in the left response box and horizontally aligned for common starting position (horizontal middle position of the screen corresponds to 0 pixels, and values increase towards the right, i.e., the non-chosen option). Each trajectory was normalized into 100 equal time steps (following Spivey et al., 2005).

Data preprocessing and aggregation was performed in Matlab 2010a (the Mathworks Inc.) and in R (R Core Team, 2016) using the mousetrap R package (Kieslich, Wulff, Henninger, Haslbeck, & Schulte-Mecklenbeck, 2016). Statistical analyses were performed in Matlab, R, and JASP 0.7.5.6 (JASP Team, 2016).

## Results

### Comparison of groups

Since our analysis builds on two independent groups of participants from different studies, we first checked for differences between these groups other than the start condition. All tested individuals were right-handed, and groups showed no significant differences in age, $t(38) = 0.881$, $p = .384$ or in sex (both groups contained 17 female and 3 male participants). All other descriptive variables also showed no significant differences (all $p > 0.125$, see Supplementary Material). To check for general differences in speed in the task, we analyzed the inter-trial interval (ITI), that is, the time between reaching the response box in the previous trial and clicking into the start box to begin the next trial. As we do not see a methodological reason why a difference in the setup of the starting condition should affect the ITI, we used it as a general indicator of speed differences in the task that are related to differences between participant groups. We found no significant differences between groups for the ITI, $t(38) = 1.51$, $p = .140$. Taken together, we found no significant differences between the two groups on indicator variables that should (or could) not be affected by the starting condition.

### Cognitive effects

Next, we were interested in how far the study of cognitive processes via mouse movements would be influenced by the starting condition. We expected that discrete measures – which describe the whole movement in a trial by one value – would be relatively robust against differences in the starting condition (hypothesis 1), whereas effects for continuous measures – which capture the variation of the movement at each time point – would be weaker in the static starting condition compared to the dynamic starting condition (hypothesis 2).

#### *Discrete effects*

We first inspected discrete measures for the Simon effect (congruent vs. incongruent trials reflected in the factor congruency$_N$) and congruency sequence effects (the modulation of the Simon effect by the previous trial's congruency reflected in the interaction congruency$_N$ × congruency$_{N\text{-}1}$). As dependent variables, we computed the response time and

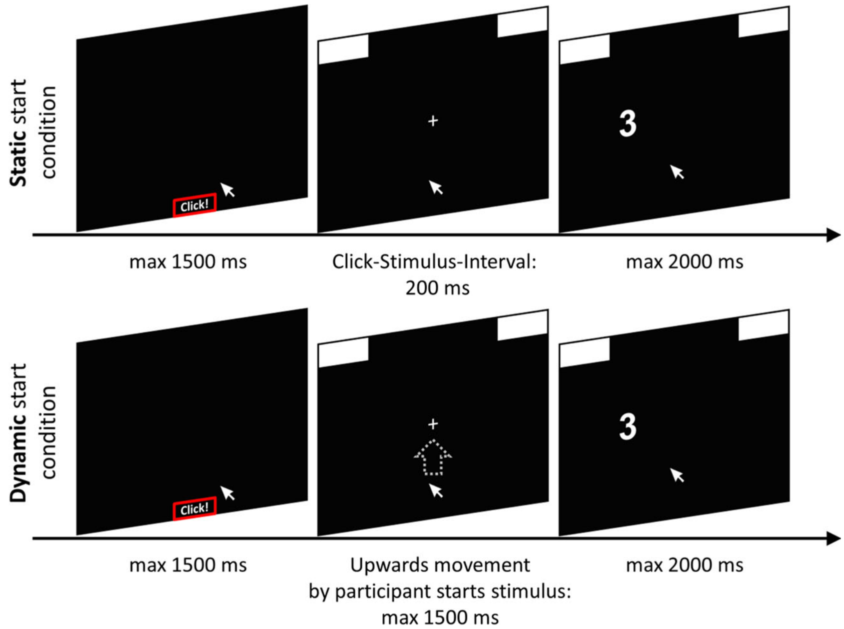


**Fig. 1** Setup of the experiment: Participants had to click with the mouse cursor on a red box at the bottom of the screen. After clicking, response boxes appeared at the upper edge of the screen. In the static starting condition, the stimulus was presented 200 ms afterwards. In the dynamic starting condition, participants had to move the cursor upwards, in order to start stimulus presentation – only after reaching a movement threshold, the stimulus was presented. To respond, participants had to move the mouse cursor to the left or the right response box depending on the magnitude of the number (left response if < 5, right response if > 5)

the average deviation (AD) per condition and participant. AD is the average perpendicular deviation between the actual movement and a hypothetical straight line from the start to the end point of the movement. For both measures, we conducted repeated measures analyses of variance (ANOVA) with the within subject factors congruency$_N$ and congruency$_{N\text{-}1}$ and the between-subject factor starting condition (dynamic vs. static).

ANOVA on AD revealed a significant main effect of the starting condition, $F(1,38) = 59.08$, $p < .001$, $\eta^2_p = 0.61$, with higher AD values in the dynamic than in the static condition. In addition, the main effects for congruency$_N$, $F(1,38) = 77.84$, $p < .001$, $\eta^2_p = 0.67$, and congruency$_{N\text{-}1}$, $F(1,38) = 23.35$, $p < .001$, $\eta^2_p = 0.38$, were significant as well as the interactions congruency$_N$ × congruency$_{N\text{-}1}$, $F(1,38) = 94.05$, $p < .001$, $\eta^2_p = 0.71$ and congruency$_{N\text{-}1}$ × starting condition, $F(1,38) = 8.17$, $p = .007$, $\eta^2_p = 0.18$. With regard to the variability of the theoretically important effects, both the Simon effect (congruency$_N$) and the congruency sequence effect (congruency$_N$ × congruency$_{N\text{-}1}$) did not significantly interact with the starting condition, congruency$_N$ × starting condition, $F(1,38) = 3.07$, $p = .09$, $\eta^2_p = 0.07$, and congruency$_N$ × congruency$_{N\text{-}1}$ × starting condition, $F(1,38) = 0.65$, $\eta^2_p = 0.02$, $p = .43$. However, both effects were descriptively larger in the dynamic starting condition (congruency$_N$: $\eta^2_p = 0.73$; congruency$_N$ × congruency$_{N\text{-}1}$: $\eta^2_p = 0.77$) than in the static starting condition (congruency$_N$: $\eta^2_p = 0.59$; congruency$_N$ × congruency$_{N\text{-}1}$: $\eta^2_p = 0.68$). Hence, in both conditions, mean AD showed the expected Simon effects and congruency sequence effects, though with descriptively lower effect sizes in the static starting condition than in the dynamic starting condition (Fig. 2, top panels).

Response time (RT) was calculated as the time difference between the moment of stimulus onset and the moment when the mouse cursor reached the response box (note that elsewhere this measure is also called movement time. Since in the dynamic starting condition participants are already moving, we chose referring to this measure as RT as it includes the whole process of response selection).

ANOVA revealed significant main effects for congruency$_N$, $F(1,38) = 108.03$, $p < .001$, $\eta^2_p = 0.74$, congruency$_{N\text{-}1}$, $F(1,38) = 4.93$, $p = .03$, $\eta^2_p = 0.11$, and starting condition, $F(1,38) = 8.30$, $p = .006$, $\eta^2_p = 0.18$, and a significant interaction congruency$_N$ × congruency$_{N\text{-}1}$, $F(1,38) = 156.95$, $p < .001$, $\eta^2_p = 0.81$. The other interactions were not significant, congruency$_N$ × starting condition, $F(1,38) = 0.07$, $p = .79$, congruency$_{N\text{-}1}$ × starting condition, $F(1,38) = 3.12$, $p = 0.09$, and congruency$_N$ × congruency$_{N\text{-}1}$ × starting condition, $F(1,38) = 1.36$, $p = .25$. Looking at the effect sizes in each starting condition for the Simon effect (congruency$_N$) and congruency sequence effects (congruency$_N$ × congruency$_{N\text{-}1}$) indicates similar effect sizes for the dynamic starting condition (congruency$_N$: $\eta^2_p = 0.71$; congruency$_N$ × congruency$_{N\text{-}1}$: $\eta^2_p = 0.80$) and the static starting condition (congruency$_N$: $\eta^2_p = 0.77$; congruency$_N$ × congruency$_{N\text{-}1}$: $\eta^2_p = 0.81$). Hence, in both conditions, RT showed the expected Simon effects and congruency sequence effects with similar effect sizes (Fig. 2, bottom panels).

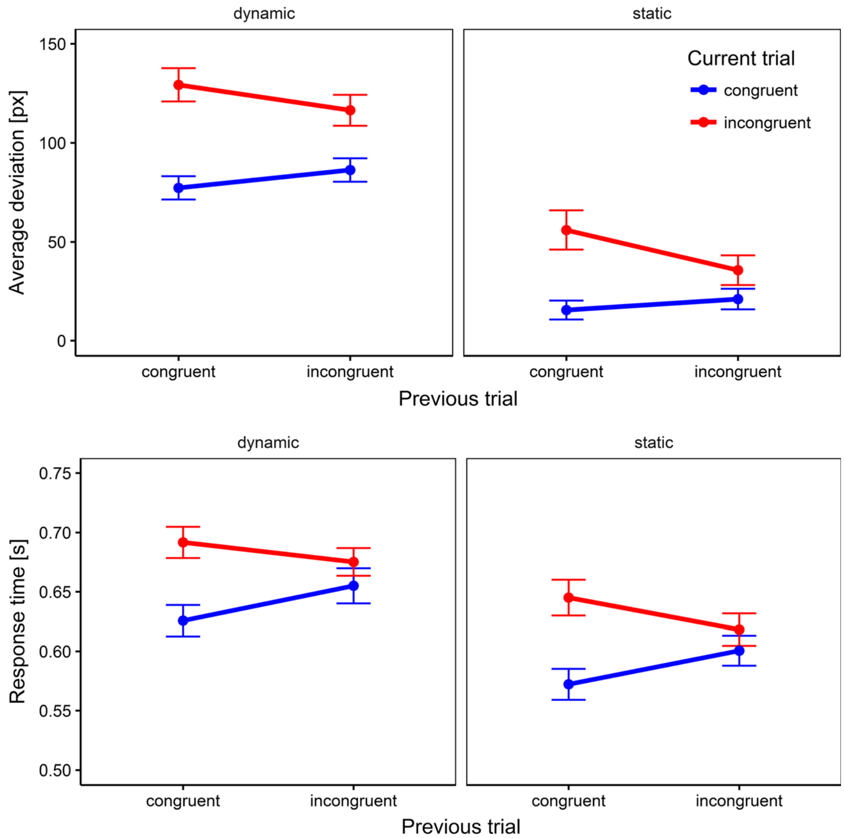


**Fig. 2.** Average deviation of mouse movements (top panels) and response times (bottom panels) as a function of previous trial congruency and current trial congruency separately for the dynamic and static starting condition. Error bars indicate 1 *SE*

Taken together, the discrete measures show the expected robustness against differences in the starting condition, though for AD descriptively weaker effect sizes were found in the static starting condition.

*Continuous effects*

In the next step, we inspected continuous mouse-tracking measures. We expected these measures to be more strongly influenced by differences in the starting condition compared to discrete measures, since they do not integrate information across the whole trial but are based on the instantaneous information in each time step. Hence, if one starting condition leads to lower consistency of movements, this should increase the noise in the data and particularly influence measures with a higher temporal resolution. Hypothesis 2 stated that effects on continuous measures will be weaker and less reliable in the static starting condition compared to the dynamic starting condition. Several mechanisms are assumed to contribute to this effect in the static starting condition: First, influences on mouse movements should show a time-lag and be compressed at the end of the trial due to the prolonged start of the main movement. Second, effects of cognitive processes should be weaker, because these processes can take place before the movement is initiated and hence only partly leak into the movement. Third, the lower data quality further decreases reliability by inducing noise into the strength and the timing of processes.

Visual inspection of heatmaps (Fig. 3) of mouse movements along the X-axis over time reveals a smoother, though wider spread distribution of movements in the dynamic starting condition (Fig. 3, left) for both, congruent and incongruent trials, compared to the trials in the static starting condition (Fig. 3, right). Averaged mouse movements for congruency$_N$ and congruency$_{N\text{-}1}$ indicate a similar pattern of effects in the dynamic and the static starting condition, though time-lagged and less pronounced in the static starting condition compared to the dynamic starting condition (Fig. 4).

For the statistical analysis of movement dynamics, we performed time continuous multiple linear regression on mouse movement angles on the X/Y plane as done in the original study (for an analysis with linear-mixed models leading to comparable results, see Supplementary Material). Based on the remapped, time-normalized trajectory data, movement angle was calculated as the angle relative to the Y-axis for each difference vector between two time steps. This measure has two advantages over the raw trajectory data. First, it better reflects the instantaneous tendency of the mouse movement since it is based on a differential measure compared to the cumulative effects in raw trajectory data. Second, it integrates

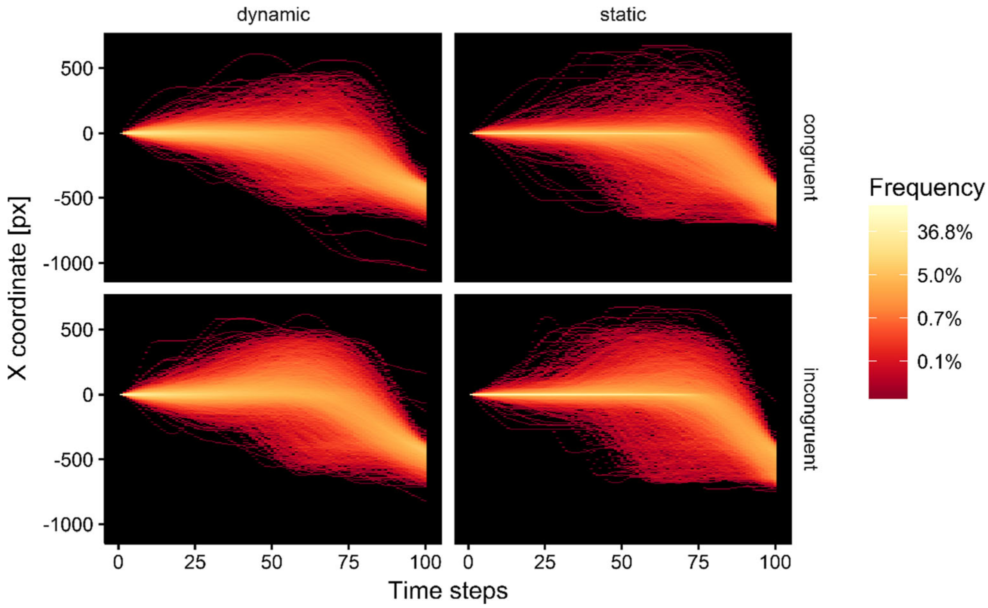


**Fig. 3** Heatmaps of pooled mouse movements along the X-axis as a function of time and current trial congruency separately for each starting condition

the movement on the X/Y-plane into a single measure. Based on this angle, we then dissected the influences of the independent variables on mouse movements within a trial. We applied a three step procedure. In the first step, we coded for each participant three predictors for all trials: location$_N$ (congruent/incongruent), congruency sequence (same/different), and previous response (same/different). Location$_N$ reflects the influence of the current stimulus location – the information that should be ignored and that induces the Simon effect. Congruency sequence reflects the expected influence of the previous trial's congruency on the strength of the potentially conflict inducing location$_N$ influence of the current trial. Hence, it reflects the interaction of congruency$_N$ × congruency$_{N\text{-}1}$, predicting how strong the mouse trajectory is deflected into the direction of the current stimulus location depending on the previously induced conflict. Previous response reflects a potential bias by the previously performed response. To provide comparable beta weights in the next step, we coded the predictors with values of -1 and 1. In the second step, we calculated multiple regressions with these predictors (angles were available for 99 time steps leading to 99 multiple regressions) on the trajectory angle that had also been standardized for each participant from -1 to 1 to provide comparable results. This yielded three time-varying beta weights (3 weights × 99 time steps) for each participant. Finally, in the third step, we computed grand averages of these three time-varying beta weights yielding a time-varying strength of influence curve for each predictor (Fig. 5).

We analyzed the dynamics of these three influences in two ways. First, we performed peak analysis, extracting strength and timing of peaks of the three influences via a jack-knifing procedure as has been used previously, for example, for peak

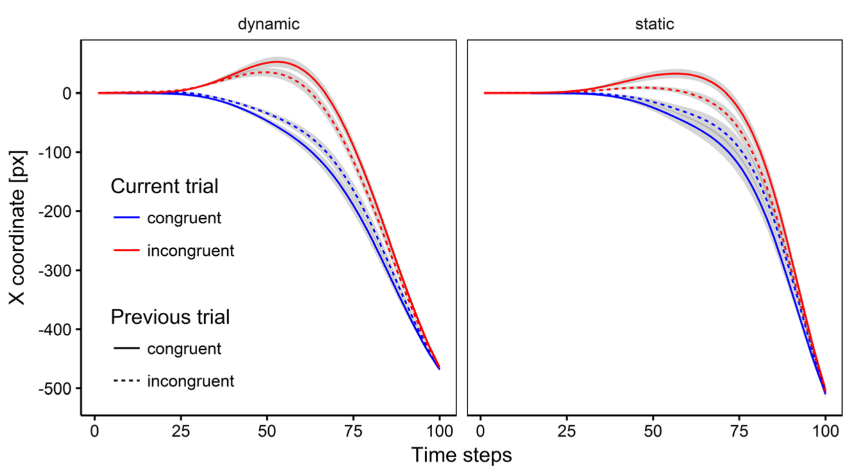


**Fig. 4** Average X coordinate per time step depending on congruency and starting condition. Coordinates were first averaged within and then across participants. Confidence bands indicate 1 *SE*

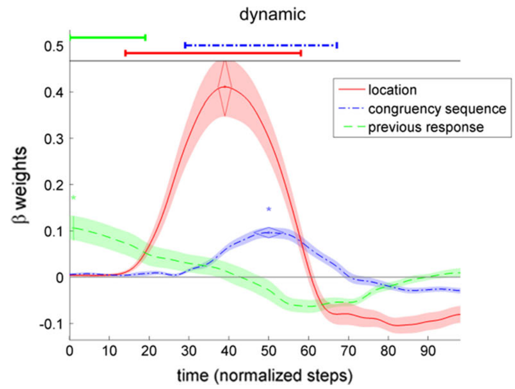


**Fig. 5** Results of time continuous regression analysis. Beta weights indicate the strength of influence of each regressor on the mouse movement angle in the dynamic condition (**left**) and the static condition (**right**). Peaks are marked by diamonds indicating jack-knifed standard errors. Lines above the graphs indicated segments of beta-weights that were significantly greater than zero ($t$-test, minimum of ten consecutive significant time steps)

detection in lateralized readiness potentials (Miller, Patterson, & Ulrich, 2001). We tested peak values and timing statistically with one-sided $t$-tests corrected for jack-knifing. Second, we detected significant temporal segments of influence by calculating $t$-tests against zero for each time step of the three time-varying beta-weights (Dshemuchadse et al., 2013; Scherbaum et al., 2010). We compensated for multiple comparisons of temporally dependent data by only accepting segments of more than ten consecutive significant $t$-tests (see Dale et al., 2007; Scherbaum, Gottschalk, Dshemuchadse, & Fischer, 2015, for a Monte Carlo analysis on this issue).

Results of peak analysis are shown in Table 1. With regard to peak timing, the static and the dynamic starting condition show the same order of peaks of influences. However, the peaks in the static starting condition show a significant lag compared to the dynamic starting condition, for location$_N$, $t_j(38) = 3.61$, $p_j < .001$, $d = 0.57$, and the previous response, $t_j(38) = 2.19$, $p_j = .02$, $d = 0.35$, but not for congruency sequence, $t_j(38) = 0.71$, $p_j = .15$. Furthermore, the static starting condition shows a higher amount of noise. Hence, a repeatedly found effect (Scherbaum et al., 2010; Scherbaum, Frisch, Dshemuchadse, Rudolf, & Fischer, in press), the timing difference between the peaks for location$_N$ and congruency sequence, cannot be replicated in the static starting condition, $t_j(19) = 0.96$, $p_j = .12$, while it is significant for the dynamic starting condition, $t_j(19) = 3.57$, $p_j < 0.01$.

Concerning peak strength, the static starting condition showed a lower beta weight than the dynamic starting condition for location$_N$, $t_j(38) = 2.06$, $p_j = .025$, $d = 0.33$, but not for congruency sequence, $t_j(38) = 1.09$, $p_j = .11$ and the previous response, $t_j(38) = 1.36$, $p_j = .16$.

Results of time segment analysis are shown in Table 2. In concordance with peak analysis, the dynamic starting condition yields more distinct time windows for all influences, especially showing less overlap between location$_N$ and congruency sequence (29 time steps, 186 ms) than the static starting condition (36 time steps, 231 ms). The larger overlap in the static starting condition is mainly caused by the pronounced time-lag of location$_N$ in the static starting condition compared to the dynamic starting condition. A similar lag is present for the influence of the previous response.

## Movement consistency

The measures of process dynamics show a pronounced influence of the starting condition on all three cognitive effects, the Simon effect (as reflected in the influence of the current stimulus' location), the congruency sequence effects, and biases

**Table 1** Results from peak analysis on beta weights from continuous regression analysis separately for each starting condition. Segment times represent the projection of time steps to each participant's mean RT. SE represent jack-knifed standard errors of the mean (see main text)

| | Dynamic start | | | | | | Static start | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Location$_N$ | | Congruency sequence | | Previous response | | Location$_N$ | | Congruency sequence | | Previous response | |
| | *M* | *SE* | *M* | *SE* | *M* | *SE* | *M* | *SE* | *M* | *SE* | *M* | *SE* |
| Strength | 0.41 | 0.06 | 0.10 | 0.01 | 0.11 | 0.03 | 0.26 | 0.04 | 0.08 | 0.01 | 0.06 | 0.02 |
| Time step | 38.95 | 1.67 | 49.95 | 3.77 | 1.00 | 0.00 | 48.10 | 1.90 | 57.10 | 9.34 | 23.85 | 10.43 |
| Time (ms) | 250 | 11 | 320 | 24 | 6 | 0 | 309 | 12 | 366 | 60 | 153 | 67 |

**Table 2** Significant segments of beta weights from continuous regression analysis separately for each starting condition

| | Dynamic start | | | Static start | | |
|---|---|---|---|---|---|---|
| | $\text{Location}_N$ | Congruency sequence | Previous response | $\text{Location}_N$ | Congruency sequence | Previous response |
| Time step | [14, 58] | [29, 67] | [1, 19] | [29, 65] | [26, 71] | [1, 41] |
| Time (ms) | [89, 372] | [186, 429] | [0, 121] | [186, 417] | [166, 455] | [0, 263] |

due to the previously performed response. This indicates that within-trial continuous measures are less robust against changes in the starting condition than discrete measures. As stated in our third hypothesis, we expected mouse movements to be less consistent within trials, across trials, and across participants in the static starting condition compared to the dynamic starting condition.

To check whether the manipulation of the starting condition led to different starts of participants' movements and less consistent mouse movements within trials, we visually inspected heatmaps of mouse movements along the Y-axis and velocity profiles – the speed of movement at each time step (measured as the Euclidean distance traveled [in px] divided through the time passed [in ms]) – both pooled across all participants (Fig. 6). Heatmaps of movements along the Y-axis indicated that participants in the dynamic starting condition moved smoothly and consistently upwards, whereas participants in the static starting condition often stayed at the bottom of the screen for more than half of the trial before moving upwards quickly in the second half of the trial. Velocity profiles corroborated this interpretation, with low, but consistent movement speed in the dynamic starting condition, and, in contrast, a strongly increasing and inconsistent movement speed in the static starting condition.

To quantify the consistency of movements within trials, we created a continuous movement index. This index is calculated for each trial as the correlation of the actual Y-axis position at each time step and a projected Y-axis position assuming a constant straight upwards movement from the first to the last point of the trial. An index of 1 hence indicates a smooth and constant upwards movement. In concordance with the visual impression from the heatmaps, the movement index was significantly higher in the dynamic starting condition ($M = 0.94$, $SE = 0.01$) than in the static starting condition ($M = 0.80$, $SE = 0.02$), $t(38) = 5.21$, $p < .001$, $d = 1.65$. This indicates that the different start instructions indeed influenced how consistently participants moved at the start and across the whole trial.

So far, all mouse movement analyses were based on the mouse movements recorded from stimulus presentation until response. However, given that in the static starting condition participants could freely decide when to initialize their movement, the movement could also have started after the stimulus was already presented (and, consequently, after tracking onset). Therefore, an alternative analysis approach in the static starting condition could only focus on the part of each trial after movement has already been initiated (Buetti & Kerzel, 2008). This could, in principle, increase the similarity of the analyzed parts between the dynamic and the static starting condition for the movement index and potentially also for the continuous measures. While restricting each trial only on the part after movement initiation indeed improved the movement index in the static condition ($M = 0.89$, $SE = 0.01$), it still was significantly smaller than the movement index in the dynamic starting condition, $t(38) = 1.97$, $p = 0.028$, $d = 0.62$. Analyzing the restricted movements in the time-continuous regression analysis yielded worse results than the analysis of unrestricted movements reported above, with wider spread peaks for the Simon effect and a loss of the influence of the previous response (see Supplementary Material).

To check the consistency of data across participants, we calculated the movement initiation time, a frequently used measure in mouse-tracking studies. Specifically, in the dynamic starting condition, movements were initiated before stimulus presentation and triggered stimulus presentation when the movement criterion was fulfilled (4 pixels in two consecutive time steps). We hence took the time difference between the click on the start box and the triggering of stimulus presentation by the mouse movement as initiation time. In the static starting condition, the stimulus was presented 200 ms after the click on the start box and participants could freely decide when to initiate their movement. We hence determined initiation time for each trial as the first time step after stimulus presentation in which participants had moved the mouse by more than 8 pixels (matching the criterion of the dynamic starting condition). We averaged the initiation times of each trial per participant and compared them between conditions. Initiation times in the dynamic starting condition ($M = 0.19$, $SE = 0.01$) were comparable to the static starting condition ($M = 0.21$, $SE = 0.02$), as also indicated by a $t$-test, $t(38) = 0.51$, $p = .61$[1]. However, the static starting condition showed a significantly higher variance ($SD = 0.09$) than the dynamic starting condition ($SD = 0.04$), as indicated by Levene's test, $F(1, 38) = 8.64$, $p = .01$. This indicates a lower consistency in

[1] Given the significant difference in variances between groups, we repeated the statistical analysis for different means of initiation times based on Welch's $t$. Again, we found no statistical differences, $t(26.1) = 0.51$, $p = 0.61$.

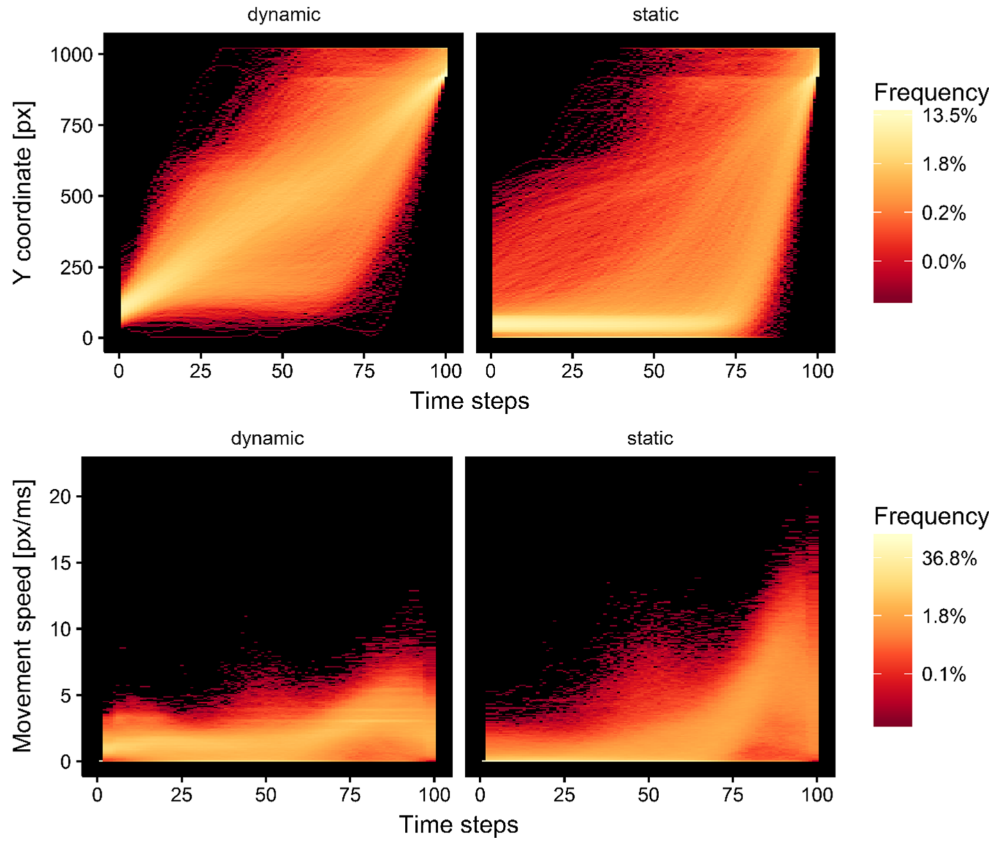


**Fig. 6** Heatmaps of movements along the Y-Axis (upper figures) and of movement speed (lower figures) as a function of time separately for each starting condition. Colors show the log-scaled probability of movements to cross the respective bin at a specific time step

movement initiation strategies across participants in the static compared to the dynamic starting condition.

To assess the consistency of movements across trials, we calculated the bimodality coefficient of the distributions of AD and checked in how far this index differed between groups. Specifically, we calculated the bimodality coefficient of each participant's distribution of AD which indicates how broadly and potentially bimodally distributed AD is across trials of a participant.

Distributions of AD differed between conditions, as the bimodality index was higher in the static starting condition ($M = 0.57$, $SE = 0.03$) than in the dynamic starting condition ($M = 0.41$, $SE = 0.02$), $t(38) = 4.92$, $p < .001$, $d = 1.56$ (see Supplementary Material for histograms of AD per condition). This is also reflected in the plot of individual response trajectories in which the dynamic starting condition shows a more coherent distribution of movements than the static condition, where many single movements leave the main area of movement (Fig. 7). More

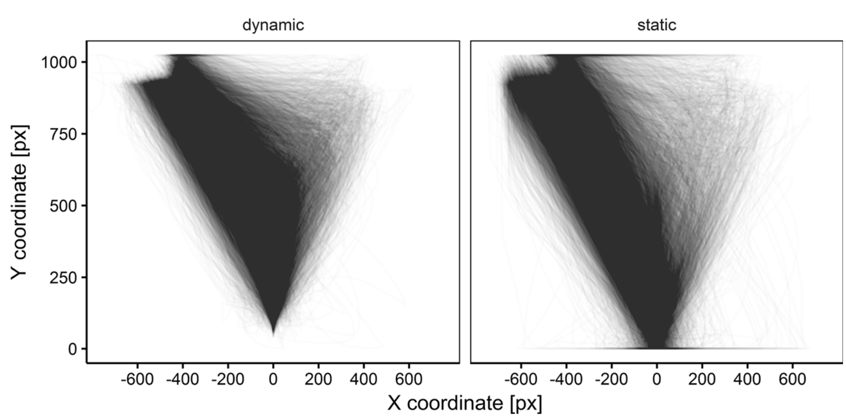


**Fig. 7** Plot of individual time-normalized trajectories per starting condition. Movements start in the bottom center of the screen and end in the upper left target box (as all trajectories were remapped to the left and their starting position was horizontally aligned)

precisely, the dynamic condition shows a smooth spread of movements while the static condition shows a combination of mainly straight movements and a few strongly curved trajectories.

Taken together, our analysis of consistency of movements within trials, across trials, and across participants indicates that the static starting condition yielded less consistent movements within a trial, across trials, and across participants. This is in line with, and may indeed be the underlying cause of, the weaker cognitive effects on the continuous measures observed in the static starting condition.

## Discussion

The present study investigated in how far methodological differences in the setup of mouse-tracking studies influence the consistency of mouse movements as well as the theoretically expected effects on mouse-tracking measures. In a mouse-tracking version of the Simon task, participants indicated their response by moving a computer mouse to a response box on the screen. The current study compared data from two experiments that varied in the way participants initialized their mouse movement. In a previously published experiment, a dynamic starting procedure was employed in which participants had to initialize their mouse movement in order to trigger the stimulus presentation. In a new experiment, a static starting procedure was used in which the stimulus was presented after a fixed delay and participants could freely decide when to initialize their mouse movement. As expected, we found that the static starting procedure yielded less consistent movements than the dynamic starting procedure. Concerning the influence of the starting procedure on theoretically predicted effects, we split our analysis in discrete measures that summarize the mouse movement in each trial in a single value, and continuous measures that examine the development of a specific movement characteristic within a trial over time. Effects on discrete measures (i.e., average deviation) were relatively robust against influences of the starting condition. In contrast, effects on continuous measures examined in the time continuous multiple regression analysis were weaker and more temporally compressed in the static condition compared to the dynamic condition.

Our results indicate that differences in the setup of mouse-tracking studies – here, specifically, in the starting condition – can indeed influence mouse movements and to some degree also the theoretically important effects investigated in such studies. The Simon task produces relatively robust experimental effects and hence all dynamic effects were present in both starting conditions – though they were much smaller and temporally compressed in the static starting condition. However, more subtle effects, for example, in value-based decision making (Dshemuchadse et al., 2013; Scherbaum et al., 2016) or semantic judgments (Dshemuchadse et al., 2015) might be more strongly affected when one studies them using a static starting condition, especially with within-trial continuous measures.

Does this mean that a dynamic starting procedure should always been applied? Our results indicate that at least for strong behavioral effects, a static starting procedure could be used (even for continuous measures). Such a static start setup might even be indispensable, if the logic of the experiment dictates strict sequences of stimulus timing which do not allow for participants starting the stimulus presentation themselves, for example, in priming experiments. Besides, other methodological considerations (with currently largely unknown consequences) might play a role when deciding whether the implementation of a dynamic starting procedure is feasible: This includes the question whether explicit instructions about the mouse movement should be provided as they might increase participants' awareness of mouse-tracking, which might be especially relevant when studying behaviors where influences of social norms might be expected. Besides, the requirement of a continuous upwards movement might be challenging if the stimulus information is more complex and its acquisition more time-consuming (though stimuli with a high amount of decision critical information per trial represent a general challenge for mouse-tracking studies, cf. Kieslich & Hilbig, 2014); this challenge is also amplified as a dynamic starting procedure is typically implemented in combination with limited time for responding. Finally, depending on what constitutes the mouse-tracking variable of interest a different methodological setup might be desirable (see Fischer & Hartmann, 2014).

It is hence in the judgment of the experimenter, whether the implementation of a dynamic starting procedure is desired and feasible in a study and, if not, whether an additional explicit instruction to start moving as quickly as possible might be sufficient (Freeman & Ambady, 2010), especially when the pursued effects are robust enough for a static starting procedure. Besides, the implementation of a dynamic starting procedure typically is methodologically more demanding and requires extensive pretesting to determine the exact spatial and temporal setup – although mouse-tracking studies in general require careful design and pretesting. Regarding the methodological implementation, a recently presented open-source software (mousetrap) can be used that allows creating mouse-tracking experiments via a graphical user interface without programming (Kieslich & Henninger, 2017) and that also allows implementing a dynamic starting procedure by specifying tracking boundaries.

### Integration

The present work adds to the emerging discussion about boundary conditions and standards for mouse-tracking studies (e.g., Faulkenberry & Rey, 2014; Fischer & Hartmann, 2014), but could also be applied to hand movement tracking studies in general (e.g., Buetti & Kerzel, 2008; Song & Nakayama, 2008, 2009). The basic idea of these studies is that cognitive processing leaks into the execution of the movements and hence cognitive processes become accessible to investigation by studying the differences in movements between different conditions. Our study indicates that the setup of the study influences the link between cognitive processing and mouse movements: Participants' upwards movements were less consistent within trials, across trials, and across participants when participants were not forced by the setup to start moving. In this regard, another methodological precaution that has been helpful in our experience is watching participants during practice trials and – if necessary – reminding them to keep moving.

On a general level, our study has two implications for future mouse-tracking studies: first, researchers should provide a detailed description of the methodological setup of the mouse-tracking task to enable researchers to compare findings from different mouse-tracking studies (cf. Fischer & Hartmann, 2014). Second, if a study aims to interpret mouse movements as the *continuous* tracking of response selection (in cognitive tasks) or the preference development (in value-based decision tasks), it should strive to maximize the likelihood of "processing while moving" through the appropriate methodological setup, for example, by using a dynamic starting procedure (if feasible). If response selection or preference development is (partly) performed before the movement is started by the participants, this might weaken the direct link between cognitive processing and mouse movements, and the duration of the initial period without mouse movement (the initiation time) might also contain information about the competition between response alternatives that is not visible in the actual mouse movement (Fischer & Hartmann, 2014). More importantly, when participants act inconsistently within a study and sometimes think before moving while at other times move before thinking, the effect under study could be split up and found partially in initiation times and partially in movement measures. Such a split up might decrease the chances of studies to find the predicted effects in the movement measures. Furthermore, the split up might lead to more bimodally distributed movement measures. This bimodal distribution, in turn, could be interpreted as evidence for two distinct cognitive processes taking place in the psychological task that is studied. However, following the reasoning outlined previously, this might be (partially) methodologically confounded with the fact that in a static start condition people sometimes think before moving (leading to a straight line) while other times they think while moving. Both cases, thinking before moving, and inconsistent movements, might considerably complicate dynamic analyses of the ongoing processes.

Surprisingly, the application of dynamic analyses of mouse movements (regression analysis, e.g., Dshemuchadse et al., 2013; Scherbaum et al., 2010; Sullivan et al., 2015; decision spaces, e.g., O'Hora, Dale, Piiroinen, & Connolly, 2013) is still in its infancy and most published studies so far focus on discrete measures of movements, which might raise the question what exactly is gained when using AD or maximum deviation of movements (MAD) instead of RT of key presses. Some studies indicate that AD might be more sensitive to certain influences (e.g., Scherbaum et al., 2015) or might indeed reflect different processes, as indicated by dissociations of RT and MAD (compare Koop & Johnson, 2011). In addition, Koop and Johnson (2013) argue that discrete mouse-tracking measures can provide researchers with more specific indicators for aspects of the preference development, such as changes of the absolute preference (which – in a typical two-choice mouse-tracking task – may be captured through crossings of the Y-axis) and changes of the momentary valence (via directional changes along the X-axis). Our study indicates that to uncover the full potential of mouse-tracking studies and to fully harvest the dynamics of decision processes by using dynamic analyses of mouse movements, a thorough design of the starting condition, including a dynamic start, might be necessary. Otherwise, one risks losing potentially present effects in the noise of inconsistent movements.

## Limitations

Our study is a first attempt to assess the influence of the starting condition on movements and theoretically important effects in mouse-tracking studies. It faces several limitations that we discuss in the following.

First, as we compared data from a previous study with data from a new study, participants were not randomly assigned to the starting condition. By nature of such a design, we cannot fully exclude that participants in the first sample (dynamic starting condition from the original study) were simply different to participants in the second sample (static starting condition from the new experiment). However, we found no significant differences in any of the sample characteristics that were assessed for each participant in both studies. Besides, the theoretically expected effects could be replicated in both starting conditions, and the pattern of differences between conditions was specific and mostly as expected from a methodological point of view. This makes us confident that the found pattern in the data is not due to inherent group differences, but caused by the difference in the starting condition between groups. Still, a future study that randomly assigns participants to either starting condition could be used to experimentally ensure full comparability between groups. Of course, the ideal design to avoid any differences between the two groups would have been a within-subjects design. However, in such a design it cannot be excluded that participants carry over a certain mode of movement from one condition to the other one, so that a between-subjects design seems preferable.

Second, an unexpected finding was that RT in the static starting condition was shorter than in the dynamic starting condition. A look at the speed profiles of both conditions (Fig. 6) indicates that the dynamic condition shows a more consistent movement speed across time. Participants start their movements and do not accelerate much even in the end phase of their movement. In contrast, the static starting condition shows considerable variation in movement speed across time. Participants start slowly but sharply increase their speed in the end phase of the movement. In our view, this pattern indicates that in the dynamic condition, participants permanently coordinate the processing of information and the movement of the mouse cursor, while in the static condition, participants already start response selection before initiating their movement and hence after initiating their movement quickly finish response selection and execute their movement directly to the target box. The latter strategy could easily yield the found advantage of approximately 50 ms for the static condition compared to the dynamic one. However, it also underlines that in the static condition cognitive processing does not always continuously leak into the mouse movements. Instead, in several trials cognitive processing might take place before movement initiation reducing the effects on mouse movements. The initially slow upwards movement in the static starting condition presumably also contributes to an on average lower AD in the static than in the dynamic starting condition as does the requirement in the dynamic condition to initially move upwards even before stimulus onset. The differences in AD between the starting conditions certainly questions the validity of absolute comparisons of AD between the two conditions. However, here we performed relative comparisons of the Simon effect and the congruency sequence effect. Our approach is also bolstered by the fact that both effects did not interact with the starting condition: Hence, in principle, the effects were equal irrespective of the starting condition.

Third, a comparison of initiation times revealed no significant differences between the starting conditions. However, given the different procedures for movement initiation in the two conditions, it is difficult to create a measure for initiation time that is comparable across conditions. In the dynamic starting condition, the initiation time was defined as the time it took participants to fulfill the upwards movement criterion (to trigger stimulus presentation) after clicking on the start box. In the static starting condition, the stimulus was presented automatically 200 ms after the click on the start box; therefore, the initiation time was defined as the time it took participants to fulfill the upwards movement criterion (as in the dynamic starting condition) after stimulus presentation. Whether these two measures are indeed comparable is an open question (e.g., one could also argue that the initiation time is underestimated in the static starting condition as participants can also prepare and start moving the mouse before the stimulus is presented, and, as a consequence, the initiation time should also be computed starting with the click on the start box). Hence, the results from this comparison should be handled with care and not be over generalized.

Fourth, given that in the static starting condition participants could freely decide when to initialize their movement, an alternative analyses approach for the static condition could only focus on the part of each trial after movement has already been initiated (cf. Buetti & Kerzel, 2008). While restricting each trial only on the part after movement initiation indeed improved the movement index, time continuous regression analysis revealed less reliable results and hence a loss in data quality for continuous analyses. Hence, we conclude that even the exclusion of the initial non-movement period cannot fully compensate for the lack of leakage of cognitive processing into the mouse movements in the static starting condition.

Finally, it should be stressed that aside from the starting condition a lot of other methodological factors often vary between mouse-tracking studies, for example, whether participants have unlimited time (e.g., Kieslich & Hilbig, 2014) for responding or whether there is a time limit (e.g., Dshemuchadse et al., 2013), whether participants have to click on a button (e.g., Koop & Johnson, 2013) to indicate a response or simply “touching” the button without a click suffices (e.g., Scherbaum et al., 2010), or whether participants receive explicit instructions about continuously moving upwards (e.g., Scherbaum et al., 2015) or no instructions about mouse movements (e.g., Kieslich & Hilbig, 2014; Koop & Johnson, 2013) are given. So, in order to enable a comparison of findings across different mouse-tracking, the influence of these and other additional methodological factors needs to be investigated.

## Conclusion

The present study is a first step in assessing the impact of methodological differences between mouse-tracking studies. We found that a static starting condition that did not enforce participants to initiate mouse movements before stimulus presentation led to less consistent mouse movements. While this did not have significant consequences for the investigation of effects with discrete mouse-tracking measures, effects on within-trial continuous measures were reduced. Moving to a higher ground in the studies of cognitive processes hence requires that experimenters understand the consequences of the individual methodological setup of a study and that they ensure methodologically (if desired) that the processes continuously leak into the movements of their participants.

**Author note** This research was partly supported by the German Research Council (DFG grant SFB 940/2 and DFG grant SCH1827/1-2 to Stefan Scherbaum). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding was received for this study.

The data of the dynamic starting condition in this article has been used in a previous publication (Scherbaum et al., 2010, Cognition). The data were reanalyzed completely for the current article.

## References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Buetti, S., & Kerzel, D. (2008). Time course of the Simon effect in pointing movements for horizontal, vertical, and acoustic stimuli: Evidence for a common mechanism. *Acta Psychologica*, *129*(3), 420–428.

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15–28.

Dshemuchadse, M., Grage, T., & Scherbaum, S. (2015). Action dynamics reveal two components of cognitive flexibility in a homonym relatedness judgment task. *Frontiers in Cognition*, *6*, 1244.

Dshemuchadse, M., Scherbaum, S., & Goschke, T. (2013). How decisions emerge: Action dynamics in intertemporal decision making. *Journal of Experimental Psychology: General*, *142*(1), 93–100.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge: MIT Press.

Faulkenberry, T. J., & Rey, A. E. (2014). Extending the reach of mousetracking in numerical cognition: A comment on Fischer and Hartmann (2014). *Frontiers in Psychology*, *5*, 1436.

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

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241.

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in Motion Reveals Mind in Motion. *Frontiers in Psychology*, *2*, 59.

JASP Team. (2016). JASP (Version 0.7.5.6). Retrieved from https://jasp-stats.org/

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods*, *49*(5), 1652–1667.

Kieslich, P. J., & Hilbig, B. E. (2014). Cognitive conflict in social dilemmas: An analysis of response dynamics. *Judgment and Decision Making*, *9*(6), 510–522.

Kieslich, P. J., Wulff, D. U., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2016). Mousetrap: An R package for processing and analyzing mouse-tracking data. https://doi.org/10.5281/zenodo.596640

Koop, G. J., & Johnson, J. G. (2011). Response dynamics: A new window on the decision process. *Judgment and Decision Making*, *6*(8), 750–758.

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, *67*(4), 151–185.

Miller, J., Patterson, T., & Ulrich, R. (2001). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, *35*(1), 99–115.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Oxford: Prentice-Hall.

O'Hora, D., Dale, R., Piiroinen, P. T., & Connolly, F. (2013). Local dynamics in decision making: The evolution of preference within and across decisions. *Scientific Reports*, *3*, 2210.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: The temporal dynamics of action selection. *Cognition*, *115*(3), 407–416.

Scherbaum, S., Dshemuchadse, M., Leiberg, S., & Goschke, T. (2013). Harder than expected: Increased conflict in clearly disadvantageous intertemporal choices in a computer game. *PLOS ONE*, *8*(11), e79310.

Scherbaum, S., Frisch, S., Dshemuchadse, M., Rudolf, M., & Fischer, R. (in press). The test of both worlds: Identifying feature binding and control processes in congruency sequence tasks by means of action dynamics. *Psychological Research*.

Scherbaum, S., Frisch, S., Leiberg, S., Lade, S. J., Goschke, T., & Dshemuchadse, M. (2016). Process dynamics in delay discounting decisions: An attractor dynamics approach. *Judgment and Decision Making*, *11*(5), 472–495.

Scherbaum, S., Gottschalk, C., Dshemuchadse, M., & Fischer, R. (2015). Action dynamics in multitasking: The impact of additional task factors on the execution of the prioritized motor movement. *Frontiers in Cognition*, *6*, 934.

Song, J. H., & Nakayama, K. (2008). Numeric comparison in a visually-guided manual reaching task. *Cognition*, *106*(2), 994–1003.

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

Spivey, M. J. (2007). *The Continuity of Mind*. New York: Oxford University Press.

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, *15*(5), 207–211.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10393–10398.

Sullivan, N., Hutcherson, C., Harris, A., & Rangel, A. (2015). Dietary Self-Control Is Related to the Speed With Which Attributes of Healthfulness and Tastiness Are Processed. *Psychological Science*, *26*(2), 122–134.

van Rooij, M. M., Favela, L. H., Malone, M., & Richardson, M. J. (2013). Modeling the Dynamics of Risky Choice. *Ecological Psychology*, *25*(3), 293–303.