



**HAL**  
open science

## Measuring sequences of keystrokes with jsPsych: Reliability of response times and interkeystroke intervals

S Pinet, C Zielinski, S M Mathôt, Stéphane Dufau, Alario F.-X., M Longcamp

### ► To cite this version:

S Pinet, C Zielinski, S M Mathôt, Stéphane Dufau, Alario F.-X., et al.. Measuring sequences of keystrokes with jsPsych: Reliability of response times and interkeystroke intervals. *Behavior Research Methods*, 2017, 49 (3), pp.1163-1176. 10.3758/s13428-016-0776-3 . hal-01472765

**HAL Id: hal-01472765**

**<https://amu.hal.science/hal-01472765v1>**

Submitted on 6 Mar 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Measuring sequences of keystrokes with jsPsych: Reliability of response times and interkeystroke intervals

S. Pinet<sup>1</sup> · C. Zielinski<sup>2</sup> · S. Mathôt<sup>1</sup> · S. Dufau<sup>1</sup> · F.-X. Alario<sup>1</sup> · M. Longcamp<sup>3</sup>

© Psychonomic Society, Inc. 2016

**Abstract** Although the precision and reliability of response time (RT) measurements performed via Web-based interfaces have been evaluated, sequences of keystrokes have never been investigated in detail. Keystrokes often occur in much more rapid succession than RTs, and operating systems may interpret successive or concomitant keyboard events according to both automatic and user-based settings. Sequence keystroke timing could thus be more sensitive than single RTs to noise in online measurements. Here, we quantified the precision and reliability of timing measures performed during sequences of keystrokes. We used the JavaScript jsPsych library to create an experiment involving finger-movement sequences, and ran it online with 633 participants. We manipulated the structure of three keystroke motor sequences, targeting a replication of previous findings regarding both RTs and interkeystroke intervals (IKIs). Our online data accurately reproduced the original results and allowed for a novel assessment of demographic variables such as age and gender. In parallel, we also measured the objective timing accuracy of the jsPsych interface by using specialized hardware and software, showing a constant 60-ms delay for RTs and a 0-ms delay for IKIs across the sequences. The distribution of IKIs revealed quantizing for a majority of participants, most likely due to the sampling

frequency of their USB keyboards. Overall, these findings indicate that JsPsych provides good reliability and accuracy in sequence keystroke timings for mental chronometry purposes, through online recordings.

**Keywords** Interkeystroke intervals · Online experiment · Motor sequence

Online research via Web interfaces is becoming increasingly important in the field of cognitive psychology (Gosling & Mason, 2015). Collecting large amounts of data over hundreds of participants in a short amount of time holds the promise of overcoming the statistical power limitations of typical laboratory samples (Reips, 2002). However, online experiments imply a trade-off between what is gained by dramatic increases in sample size and better sampling of the whole population, and what is lost to uncontrolled factors such as distracting environments and diversity in equipment configurations. The latter is acutely relevant when the experiments rely on mental chronometry (Posner, 1978).

A number of studies have evaluated the accuracy and reliability of response time measurements performed through various Web-based interfaces. Such empirical evaluations (reviewed in Reimers & Stewart, 2015) include direct comparisons of experimental results from Web and lab implementations (Reimers & Stewart, 2007; Schubert, Murteira, Collins, & Lopes, 2013), attempts to replicate classical experimental effects on response times (RTs) with online measures (Crump, McDonnell, & Gureckis, 2013; Enochson & Culbertson, 2015; Reimers & Maylor, 2005), and measures of the timing performance of Web-based testing setups using specialist software or hardware (Keller, Gunasekharan, Mayo, & Corley, 2009; Reimers & Stewart, 2015; Simcox & Fiez, 2014). In general, these studies have agreed that online RTs

---

**Electronic supplementary material** The online version of this article (doi:10.3758/s13428-016-0776-3) contains supplementary material, which is available to authorized users.

---

✉ M. Longcamp  
marieke.longcamp@univ-amu.fr

<sup>1</sup> Aix Marseille Univ, CNRS, LPC, Marseille, France

<sup>2</sup> Aix Marseille Univ, CNRS, BLRI, Aix-en-Provence, France

<sup>3</sup> Aix Marseille Univ, CNRS, LNC, 3 Place Victor Hugo, 13331 Marseille Cedex 3, France

are reliable, if slightly overestimated (in the range of tens of milliseconds; de Leeuw & Motz, 2016; Reimers & Stewart, 2007; Schubert et al., 2013).

The studies above have concerned the chronometry of single keypresses, with RTs being defined as the time elapsed between the onset of a stimulus and the unitary response. In contrast, tasks involving rapid sequences of keystrokes have never been investigated with online setups. Sequences of keystrokes are important for researchers interested in behaviors such as typing, musical performance, motor-sequence learning, serial RT tasks, rhythm production, and so forth. This line of research is especially interested in the structure of sequence programming and how temporal and ordinal forms of information are acquired during learning, and it relates to the general problem of serial order in behavior (for a review, see Rhodes, Bullock, Verwey, Averbeck, & Page, 2004). Two dependent variables can be derived from sequences of keystrokes: RTs, defined above, and interkeystroke intervals (IKIs), the time elapsing between two successive keystrokes. There are two broad reasons why IKIs might be more sensitive than RTs to the noise induced by online measurements.

First, IKIs are typically much shorter than RTs. For example, they can last a few tens of milliseconds for expert participants in typing studies (Rumelhart & Norman, 1982), and less than 200 ms for well-trained participants in serial-RT tasks (Nissen & Bullemer, 1987). A given level of (unavoidable) chronometric imprecision could have a larger impact on the accuracy of such shorter durations than on longer RTs. Standard keyboards are connected through USB ports sampled at a rate of 125 Hz (i.e., every 8 ms). Such quantization could distort small differences that happen to be near that range. Crump et al. (2013) and Reimers and Stewart (2015) previously highlighted the difficulties inherent to shorter timings.

Second, we were also concerned that operating system (OS) settings, such as accessibility features or keyboard hotkeys, might potentially interfere in a detectable way with the recordings and the expected pattern of results. OSs interpret successive or concomitant keyboard events according to both automatic and user-based settings (multiple key presses interpreted as one, combinations of key presses triggering a particular event, etc.). The impact of this intermediate layer of software on keyboard chronometry is not known.

There are preliminary indications of how reliably the timing of successive responses can be recorded. Simcox and Fiez (2014) and Keller et al. (2009) used specialized equipment to generate a stream of keystrokes with a fixed known interval. They measured the recovery of such interval through a Web-based software, showing good timing accuracy. A limitation of this approach is that their manipulation involved fixed intervals and a single button. A crucial and original feature of the present study is the use of three distinct keys from the keyboard, and the generation of variable delays

between keys, just as it happens in actual experiments involving sequence production.

To assess the accuracy of IKIs measured online, we adopted a two-fold strategy. First, we measured the timing accuracy of the jsPsych interface using a specialized hardware (Black Box Toolkit) that was modified such that three response switches from the keyboard could be alternatively and variably triggered without human intervention. Second, we ran an actual experiment online and performed a complete quality check of the data using descriptive and inferential statistics.

The experiment we designed involved finger movement sequences for which effects on the recorded RT and the IKIs are well-established (conditions 4 vs. 6 of Rosenbaum, Inhoff, & Gordon, 1984, Exp. 3). In a given block, participants produced two pretrained sequences of three consecutive finger responses. Across the two sequences, two of the responses were identical, whereas the third was different, which generates an uncertainty. In the original study, RTs decreased when the uncertain response occurred later in the sequence. In addition, the IKI preceding the uncertain response was lengthened. These effects of uncertainty on sequence programming and execution have been replicated (Rosenbaum, Hindorff, & Munro, 1987).

Stimulus display and response recording were controlled by the JavaScript library jsPsych, combined with HTML and CSS (de Leeuw, 2015). JavaScript offers technical advantages over its alternatives, such as Java or Adobe Flash (Reimers & Stewart, 2015): It is natively supported by all modern browsers, such as Firefox or Chrome, and it does not require any installation or updating of browser plugins. This ensures a responsive experimental design and an accurate measurement of RTs, because it is not affected by remote server and network latencies. This configuration has been successfully used to record single responses (e.g., de Leeuw & Motz, 2016, and the other references above).

## Hardware assessment of timing accuracy via the Black Box Toolkit

### Method

**Materials and procedure** To assess the reliability of the RTs generated by multiple consecutive keystrokes, we resorted to specialized hardware, the Black Box Toolkit (BBTK; Black Box Toolkit Ltd, Sheffield, UK), which can automatically generate and record triggers with submillisecond precision. Three mechanical keys (corresponding to the letters S, D, and F) of a Dell standard USB keyboard were wired to the Black Box Toolkit. With this wiring, the BBTK was able to close the three key switches on demand and generate keyboard response sequences. In our tests, the BBTK was

programmed to detect a visual stimulus through its opto-detector and then automatically generate a sequence of three responses. To handle the display of visual stimuli and the keyboard response collection, we used a jsPsych procedure similar to that used in the actual experimental task (described in the next section and available online: [https://github.com/blri/Online\\_experiments\\_jsPsych](https://github.com/blri/Online_experiments_jsPsych)). This procedure was run on an iMac 27 using the Safari Web browser (version 9.0). Three tests were run, all of which consisted in displaying a white @ character on a black background 40 times, responded to with three keystrokes, thus generating 120 automated keyboard responses. The key identities activated by the BBTK were randomized across and within trials. The programmed RTs in the first test were randomly chosen within the interval [100, 250] ms. In the second test, they were fixed at 150 ms. In the third test, the range was [350, 500] ms. The data are available at the following repository: <https://osf.io/r5dfg/>.

## Results

The programmed keystroke times were compared to those recorded through the jsPsych procedure. In each of the three tests, the recorded keystroke times (from stimulus to actual keystroke) for each of the three keystroke conditions (first, second, and third keystrokes) were all overestimated by about 60 ms ( $SD = 8$  ms; Fig. 1, upper panel). This observation was substantiated by an analysis of variance (ANOVA) with the crossed factors Keystroke Position and Test Block, in which none of the tests was significant (all  $ps > .1$ ). When IKIs were computed (second minus first and third minus second keystroke times), the difference between the programmed and recorded values had a mean value of 0 ms ( $SD = 10$  ms; Fig. 1, lower panel). Again, an ANOVA showed no significant effect of position nor of test block (all  $ps > .1$ ). In short, the first keystroke was recorded with a delay of ~60 ms (probably due to the computation of the stimuli and the physics of LCD screens),<sup>1</sup> and this delay was passed on reliably to subsequent keystrokes. Our measure of primary interest, IKIs, was unbiased, showing a mean error of 0 ms.

The variability in the difference between the programmed and recorded values,  $SD = 8$  ms, was in the range of the sampling frequency of the USB keyboard (frequency of 125 Hz, thus a period of 8 ms). This prompted a final analysis to assess the likely quantization of the measure. To show this, we divided each IKI by 8 and took the remainder of the

division. If sampling was somewhat biased toward 8, the distribution of the remainder values would not be homogeneous. On the contrary, if sampling happened every millisecond, the distribution should be homogeneous. The vast majority of IKIs appeared to be multiples of 8, which indeed corresponding to the sampling of the USB connection at 125 Hz (see Fig. 2). This quantizing cannot be related to the display of the visual stimulus, since only the onset of the stimulus, and therefore the activation of the opto-detector, depended on the screen parameters.

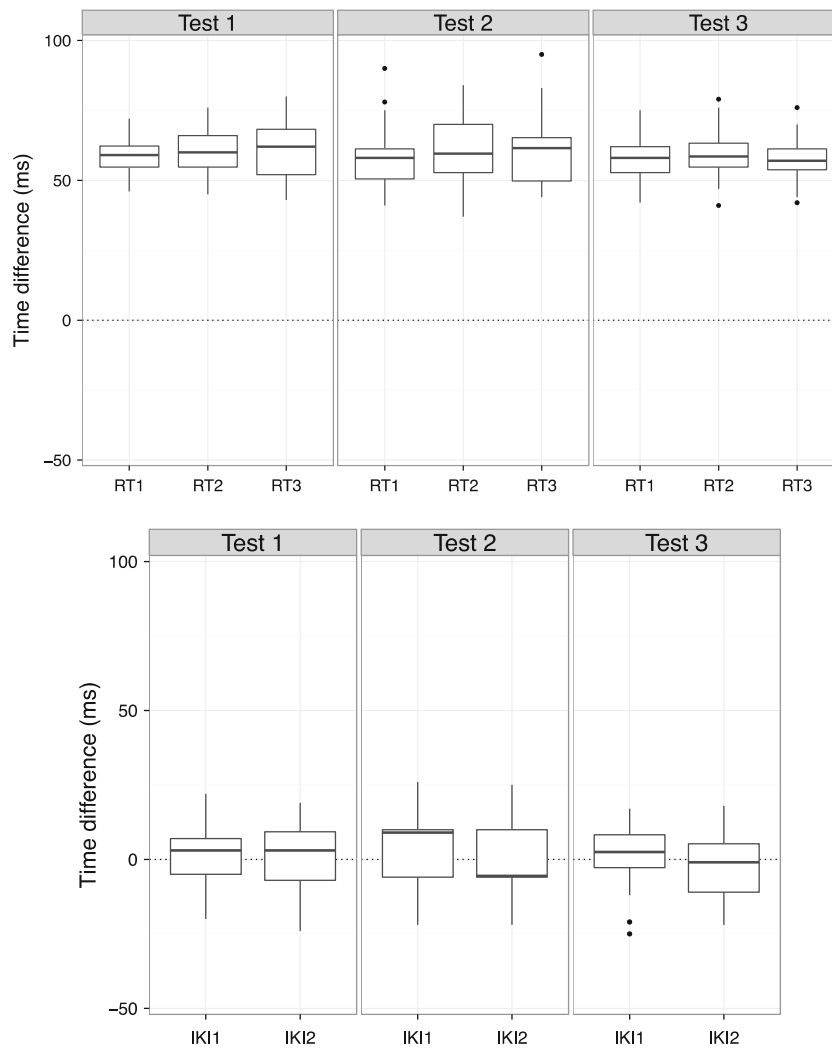
## Discussion

The automatic assessment of sequence keystroke timings generated by the BBTK and recorded by a jsPsych procedure revealed that IKIs are unbiased (mean deviation 0 ms) but largely quantized (at 125 Hz, the USB port sampling rate). We now examined how these objective recording conditions fared when attempting to replicate a classic effect on human sequence production.

## Online experiment

The jsPsych platform was used to collect data from a large sample of participants. To evaluate the reliability of the jsPsych platform to accurately measure the timing of sequences of keystrokes, we designed a task involving rapid sequences of keystrokes. The overall duration of the task had to be short (less than 20 min), paired with easy-to-understand instructions to be adapted to an online format. Therefore, we aimed to reproduce two classic effects on motor sequence performance with a design adapted from Experiment 3, conditions 4 versus 6, of Rosenbaum, Inhoff, and Gordon (1984). From the original design, we selected only two conditions, to show differences between RTs and IKIs related to the structure of the sequence performed while keeping the duration of the experiment as short as possible. The conditions were the most different in terms of the sequences used (different hands and fingers; see below) and yielded the largest difference in the RT and IKI measures. The original study explored the motor representations used to perform sequences in a choice-reaction-time task design, in which participants had to select a sequence of motor responses to a visual stimulus (X or O). In the original experiment, the sequences had three finger responses that differed by one element placed at either Position 2 or Position 3. The variable element involved both a change of hand and the choice of a nonhomologous finger (e.g., R-Ring to L-Index). Rosenbaum et al. (1984) found that the position of the uncertain response had an effect on RTs, with longer RTs for uncertain responses at Position 2 than at

<sup>1</sup> The delay could occur between the jsPsych pseudocommand “display stim” (with which the jsPsych initial clock read would occur) and the actual physical display on the screen (detected by the opto-detector; the BBTK initial clock read). This delay is probably due to the computation of the display screen, the communication between graphic electronic components and the physics of LCD screens, and the timing inaccuracies of jsPsych.



**Fig. 1** Boxplots (median, 1st and 3rd quartiles, lowest and highest values within 1.5 times the interquartile range) showing the differences between the response times (RTs) recorded by jsPsych and those programmed into the Black Box Toolkit. (Top) For each of the three tests, the three consecutive RTs are RT1, RT2, and RT3. All of the conditions had similar

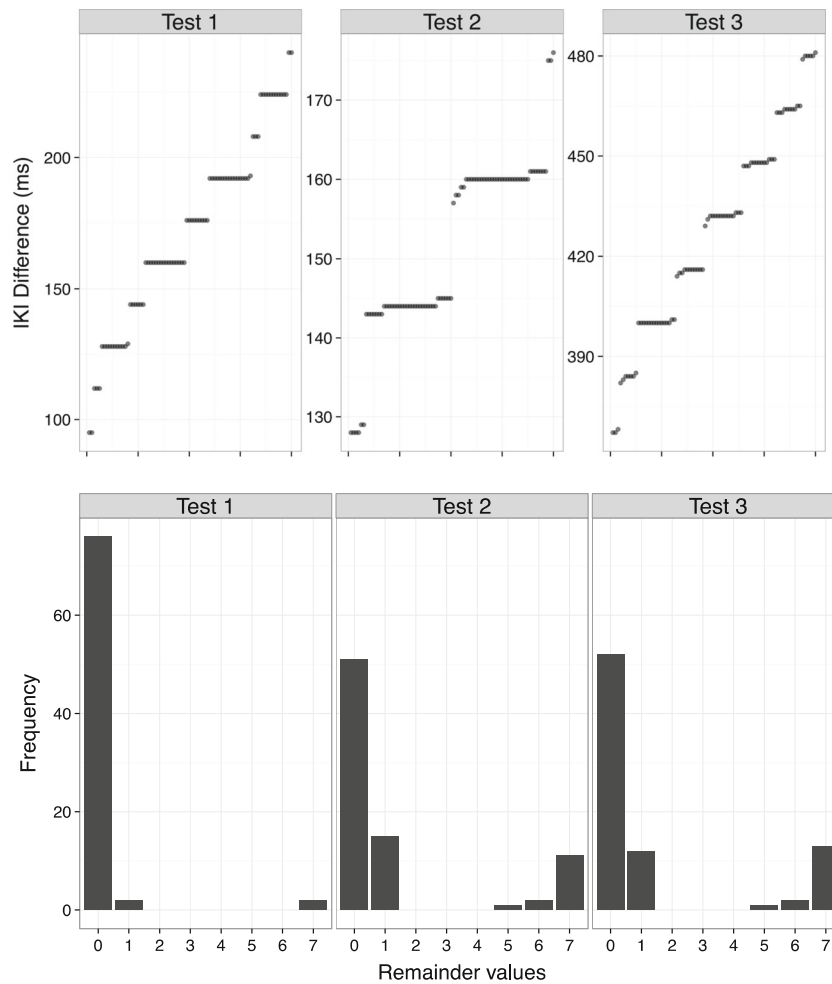
distributions centered around 60 ms. (Bottom) Interkeystroke interval (IKI) analysis, similar to that for RTs. The distributions are centered around 0 ms. Negative values of the difference occur for IKIs when, between the two responses used to calculate the actual IKI, the recorded first response is more delayed than the recorded second response

Position 3 (with means of about 460 and 380 ms, respectively; data from Rosenbaum et al., 1984). In addition, the IKI preceding the uncertain response was found to be longer, as was attested by a significant interaction between the position of the uncertainty and the position of the required response (with mean IKI1 uncertain = 193 ms; mean IKI1 certain = 177 ms; mean IKI2 uncertain = 233 ms; mean IKI2 certain = 163 ms; data from Rosenbaum et al., 1984). The original experiment involved six participants.

## Method

**Participants** Members of our university staff were invited by e-mail to participate in the online experiment. After 31 days,

we had collected data from over 600 participants (100 times more than in the original study), and data collection was discontinued. Ninety-two participants were excluded, mainly on the basis of self-reported technical or concentration issues, leaving a usable sample of 541 participants. Reported issues ranged from interruption during the task (17), difficulty understanding the instructions or staying focused (20), discomfort during the task (13), troubles with their specific devices (e.g., keyboard; 11), and various other issues (26). Any participant who reported an issue (whether specific or not) or was below 18 years of age was excluded from the final participant pool. Table 1 outlines the demographics of the final sample. We also collected information about the OSs and Web browsers used (Table 2). The most frequent combination was Windows and Firefox.



**Fig. 2** Interkeystroke intervals (IKIs) for the three tests. (Upper panels) For each test, IKIs sorted in order of increasing length are displayed. (Lower panels) IKI values were divided by 8 (in relation to the 125-Hz

keyboard sampling rate; see the main text for details). The histogram of the remainder values from this division shows that the vast majority of IKIs are multiples of 8

**Stimuli and design** The design is summarized in Table 3. The two constant sequences were unimanual and consisted of the sequence [Index, Ring, Middle fingers], performed with either the left or the right hand (factor Hand). Four varying sequences comprised a variation from the constant sequences on either the second or the third position (factor Uncertainty). A block contained two interleaved sequences, one constant and one varying. Each participant was assigned randomly to one constant sequence (thus, to one level of the factor Hand, either IRM or irm; see Table 3) and performed the two uncertainty conditions (2nd and 3rd) for his or her assigned hand in consecutive blocks. The association of a

given sequence with a visual stimulus (X or O) and the order of presentation of the uncertainty conditions were randomized across participants.

**Procedure** The online experiment consisted of a set of HTML, JavaScript, CSS, and PHP files. These files were stored on a server that was already set up and managed by S.M. (co-author of this study), and associated with the cogsci.nl domain name. The collected data were saved in the MySQL database that was configured on this server machine.

The experiment itself was mainly developed using the open-source jsPsych library (version 4.3, 2015; de Leeuw,

**Table 1** Sample characteristics

Handedness	471 right 69 left
Gender	324 women 214 male
Age	mean = 39.64, median = 39, Q1–Q3 = 30–47, range = 21–69
Affiliation	228 work at AMU, 21 study at AMU, 19 unrelated, 273 not specified

**Table 2** Web browsers and OS sample characteristics

	Linux	OS X	Windows
Firefox	27	54	254
Chrome	4	18	80
Explorer	0	0	51
Safari	0	45	0
Other	6	0	1

2015). This library contains predefined methods to manage experimental timelines, collect RTs and user actions (mouse or keyboard events), randomize stimuli, store the data, and prepare data for the backup. A plugin was devoted to the main task of the study, involving the typing of a three-key sequence following a visual stimulus (i.e., a character appearing on the screen). This plugin, called *jspsych-key-sequence*, allowed for recording the typing times of each pressed key through an adaptation of an available jsPsych plugin (*jspsych-multi-stim-multi-response*). Collection of each keystroke timing relied on the jsPsych method `getKeyboardResponse()`, which uses the general JavaScript object `Date()`. The visual stimulus, the corresponding expected typing sequences, the total number of stimulus, and the interstimulus duration were configurable as input parameters and were adapted for the blackbox testing. When given several types of stimuli and the associated sequences, the stimuli were first randomly shuffled to be displayed successively. The code for this experiment is available in the following GitHub repository: [https://github.com/blri/Online\\_experiments\\_jsPsych](https://github.com/blri/Online_experiments_jsPsych). The online experiment started with a welcome screen containing concise explanations. Within this Web page, a push button only visible to a computer-based browser allowed the participant to go further; in this way, participants using tablets or smartphones were automatically excluded from the experiment. A second Web page was then loaded in a full-size window, which embedded the JavaScripts (jsPsych core script and plugins, jQuery) necessary to launch the experiment.

The experiment was divided into two parts, one for each uncertainty condition; each part comprised a training phase and a test phase. Written instructions described the task and introduced the pairing between visual stimuli (X and O) and

**Table 3** Design table

Sequence	Uncertainty	Hand	
		Left	Right
Constant	–	irm	IRM
Varying	2nd	ilm	IiM
	3rd	irI	IRi

Lowercase letters correspond to left-hand fingers, uppercase letters to right-hand fingers. I, index; R, ring; M, middle

sequences for each part. During the test phase, one of the two visual stimuli was displayed and stayed on until the participant hit three keys, followed by an interval of 500 ms. The test phase comprised two blocks of 20 trials each, in which ten trials of each sequence were intermixed. Participants were familiarized with the sequences during a training phase, which ended when each sequence was correctly performed four times. Before the test, the plugin was set in a training mode that could be switched on from the input settings, and allowed to provide feedback by changing the color of the stimulus before it disappeared (the initially black symbol became green if the key-sequence was correct, red otherwise). After the experiment, participants were asked to answer a few questions (handedness, gender, age, employed or not employed by the university) and had the opportunity to report whether any problem occurred during the experiment. No monetary compensation was offered.

Before launching the large-scale study, the online experiment was pretested on a small number of participants ( $N = 54$ ), to assess the clarity of the instructions and the functioning of the program in various material configurations. A spy library was included in the experiment file, sending us the JavaScript error message encountered by any user's browser (<https://github.com/posabsolute/jQuery-Error-Handler-Plugin>). Five errors were reported due to four JavaScript methods that were not interpreted by old browser versions. To avoid subsequent errors, a polyfill was added to the experiment file (see the code in the dedicated GitHub repository).

The data related to the experimental task comprised strings of sequences and the associated typing times and survey answers. They were recorded as a JavaScript object during the experiment, which was subsequently transformed to plain text before being sent to the server. At the end of each experiment, the data were transferred to the server as character arrays, including the participant's anonymous identifier, OS/browser information, and the jsPsych data. The transferred dataset was finally stored in the MySQL database with the help of the PHP files. Access to the database and the writing of the data inside the data table were done using the PHP Data Object extension, which ensured database portability and security.

All data are available at the following repository: <https://osf.io/r5dfg/>.

**Statistical analyses** To replicate the original study, the data were first assessed via ANOVAs performed on the RTs and IKIs averaged per participant and cell design. Then, mixed linear regressions were used to estimate the actual effect sizes, and also the effects of additional variables: trial number, gender, age, handedness, and OS and Web browser. These variables were tested as linear predictors, except for age. The relation between age and performance has been reported as nonlinear (Baltes & Lindenberger, 1997), and spline

interpolation has been successfully used in cognitive-aging research to approximate the age effect trajectory (Fozard, Vercruyssen, Reynolds, Hancock, & Quilter, 1994). Visual inspection of our data suggested that the effects of age on performance (RTs) could be nonlinear. Therefore, we used restricted cubic splines with three knots, which allowed the effect to be modeled separately in two intervals, without a priori knowledge of the point of separation between these intervals. In the model, we included random intercepts for participants and items (i.e., the different finger sequences). Since how to compute  $p$  values in this kind of analysis has been debated (Bates, Mächler, Bolker, & Walker, 2015), we took  $t$  values to approximate  $z$  values, and considered any value above 1.96 significant.

To characterize the data reliability and the added value of increasing the number of observations, we calculated the means and confidence intervals of RTs and IKIs over random samples of increasing size, then ran the same regression models on samples of increasing size.

Finally, on the basis of the results of the BBTK assessment, we searched for quantization in our data. We focused mainly on the sampling bias that could result from USB keyboards' sampling rate (125 Hz; i.e., a value being sampled every 8 ms), and used the same methodology as reported above.

## Results

### *Replication of original study*

**RT and IKI distributions** Twelve out of the sample of 541 participants were excluded because they did not reach 85% accuracy on the task, leaving a final sample of 529 participants. Only correct trials were included in the following analysis. We also excluded trials in which any IKI was equal to zero, since the order of key pressing then could not be determined.

Figure 3 presents the RT and IKI distributions of correct responses, which were right-skewed as expected (RT skewness = 117.1, IKI skewness = 5.46). The shape of the IKI distribution was also consistent with the typically described shapes of IKI distributions from laboratory studies of typing, which are less right-skewed than typical RT distributions (e.g., Gentner, 1983). No deadline was imposed during the course of a trial, and a few very extreme values were recorded. To study the processes of interest, we choose a 3,000-ms high cutoff for RTs, to identify trials on which participants were actively engaged in the task. On the basis of rough approximations of neural conduction time, we considered RTs below 200 ms to be anticipations that were not directly triggered by to stimulus presentation, and excluded those trials. These cut-offs for RTs had very little effect on the IKI distribution (0.57% of the data were removed). Similarly, to study trials in which participants were engaged in the task for the whole

course of the trial, we kept those with IKIs lower than 1,000 ms. This whole procedure left 92.5% of the data.

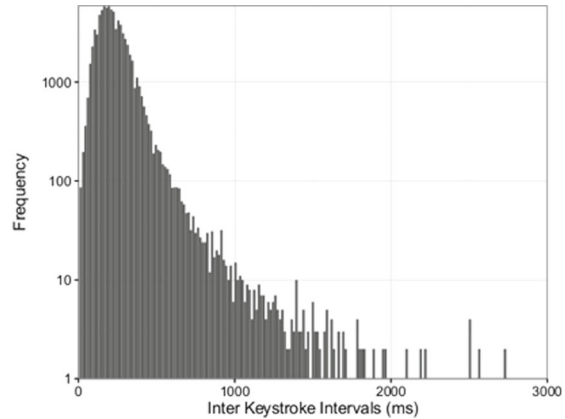
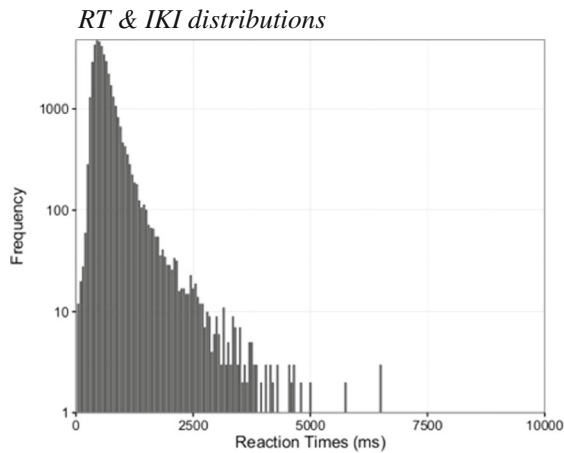
**Effects of uncertainty on RT and IKI** For RTs, the design included the following factors: Position of Uncertainty (on 2nd or 3rd keystroke), Type of Sequence (constant or varying), and Hand of the constant sequence (left or right). We did not include any interaction between Hand and the other factors in the design. An ANOVA revealed a main effect of uncertainty [ $F(1, 528) = 373, p < .001$ ]; RTs were shorter for sequences varying on the 3rd rather than on the 2nd key ( $M_{2nd} = 677$  ms,  $M_{3rd} = 560$  ms). This is in good agreement with the original results of Rosenbaum et al. (1984). The main effect of sequence also reached significance [ $F(1, 528) = 187, p < .001$ ]: RTs were shorter for constant sequences ( $M_c = 597$  ms,  $M_v = 640$  ms). Finally, the Sequence  $\times$  Uncertainty interaction was also significant [ $F(1, 528) = 7.3, p < .01$ ]. The main effect of hand was not significant ( $F < 1$ ).

For IKIs, the design was similar, with the additional factor Position of the interval within the sequence (1st or 2nd IKI). An ANOVA revealed main effects of sequence [ $F(1, 528) = 37, p < .001$ ] and position [ $F(1, 528) = 7.4, p < .01$ ]; IKIs were overall shorter in varying than in constant sequences ( $M_v = 221$ ,  $M_c = 228$  ms), and were also shorter at the second than at the first position ( $M_{2nd} = 222$  ms,  $M_{1st} = 227$  ms). The interactions Sequence  $\times$  Position [ $F(1, 528) = 68.8, p < .001$ ], Sequence  $\times$  Uncertainty [ $F(1, 528) = 5.0, p < .05$ ], and Sequence  $\times$  Uncertainty  $\times$  Position [ $F(1, 528) = 27.6, p < .001$ ] were also significant. Crucially, the Uncertainty  $\times$  Position interaction was highly significant [ $F(1, 528) = 191, p < .001$ ], indicating that IKIs were longer at the position of uncertainty (Fig. 4). The Uncertainty  $\times$  Position interaction was robust and was also observed on subsets of the data (see the [supplementary material](#) for the OSs Windows and OS X and the Web browsers Chrome and Firefox). In short, the ANOVA revealed that the main results from the original study were replicated.

**Effects of participant characteristics on RTs and IKIs** We also used regression to better estimate the effect sizes over and above the individual characteristics (participants and computer configurations; see Tables 4 and 5). The contrasts in categorical variables were assessed against the most frequent category (Windows operating system and Firefox Web browser).

Summarizing the results on RTs (see Table 4), we obtained the same significant effects as with the original ANOVA, although the Sequence  $\times$  Uncertainty interaction did not reach significance and presented a small estimate. Importantly, the estimates for the main effects of sequence and uncertainty were the largest. Regarding IKIs (see Table 5), all main effects and interactions reached significance, except the main effect of hand. Our interaction of interest, Position  $\times$  Uncertainty, presented one of the largest estimates.





**Fig. 3** Distributions of response times and interkeystroke intervals for correct trials. The  $y$ -axis in each panel is log-transformed. Three very

extreme values (46.4, 153, and 174 s) were removed from the response time distribution for display purposes

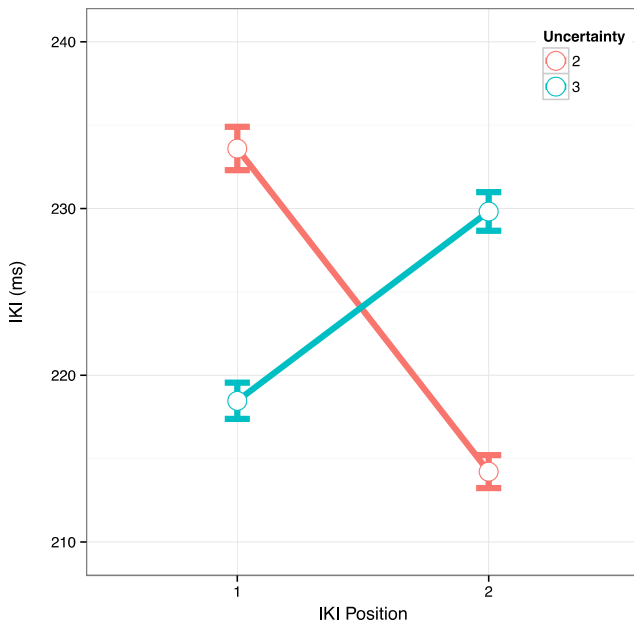
Introducing personal characteristics in the model yielded a significant main effect of gender, with male participants being faster than female participants, in terms of both RTs and IKIs. We also observed a slight slowing down of both RTs and IKIs with age, but no significant effect of handedness (Tables 4 and 5). Regarding computer configurations, only the contrast of Chrome against Firefox on RTs approached significance, with responses collected from Chrome being faster (see the [supplementary material](#)). This contrast did not reach significance for IKIs.

### Additional analyses

Having reproduced the result from Rosenbaum et al. (1984), we aimed to further characterize the data collected via the online platform.

**Estimation of data quantization** Similar to the analyses performed on the hardware test data above, we evaluated the extent of data quantization in the online experiment. Toward this end, the IKIs were sorted in increasing order. Figure 5 (upper panel) displays two representative participants: The IKIs for the participant in the left panel take almost all possible values between 130 and 190 ms; in contrast, the IKIs for the participant in the right panel are concentrated around multiples of 8.

This sampling bias was assessed on the transformed IKI distributions, generated by taking the remainder values from the division by 8. The homogeneity of these distributions was quantified over the whole sample by means of independent chi-square tests for each participant (see the Method section), which revealed that 449 participants (83% of the sample) presented a sampling bias (as indexed by significant chi-square tests, FDR-corrected). This confirms that we should not expect a precision higher than 8 ms on actual IKI measurements. The reason why some of the data from some participants did not show this quantization could not be meaningfully traced to the specific features of their computer configurations that were available to us.



**Fig. 4** Mean interkeystroke intervals (IKIs) as a function of uncertainty (on 2nd or 3rd element of the sequence) and position within the sequence (first or second). Error bars represent mean standard errors estimated by bootstrap

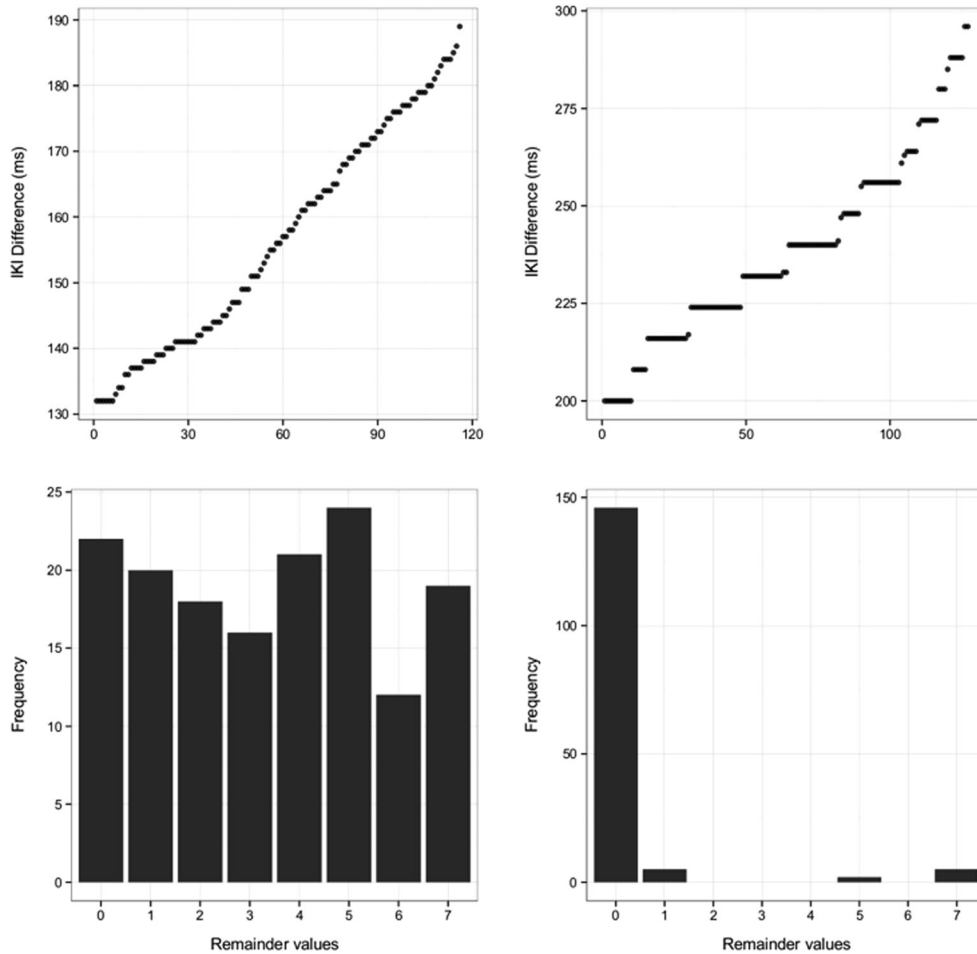
**Data reliability** Rosenbaum's original study had included only six participants. Our final sample size ( $N = 541$ ) allowed us to assess the relationship between sample size and effect reliability. From the data, we randomly selected samples that ranged from six to 100 participants (with replacement). For each sample size, 20

**Table 4** Mixed model regression coefficient for RTs

	$\beta$	SE	t	95% CI	(sig.)
(Intercept)	592.55	57.09	10.38		
Experimental variables					
Hand	-5.87	13.66	-0.43	[-32.69, 20.96]	
Sequence	46.71	5.43	8.61	[36.05, 57.37]	*
Uncertainty	-116.48	3.34	-34.90	[-123.04, -109.92]	*
Sequence $\times$ Uncertainty	-11.75	6.36	-1.85	[-24.24, 0.74]	
Trial number	-1.05	0.03	-31.65	[-1.11, -0.98]	*
Participants' variables					
Gender	-54.32	13.78	-3.94	[-81.38, -27.26]	*
Age (spline 1)	3.50	1.64	2.13	[0.27, 6.73]	*
Age (spline 2)	3.94	2.07	1.90	[-0.13, 8.01]	
Handedness	25.51	19.53	1.31	[-12.85, 63.87]	
Computers' variables					
OS: OS X/Windows	4.19	19.95	0.21	[-35, 43.37]	
OS: Linux/Windows	-45.02	28.72	-1.57	[-101.43, 11.39]	
Web Browser: Chrome/Firefox	-55.77	17.51	-3.19	[-90.16, -21.38]	*
Web Browser: IE/Firefox	27.45	23.23	1.18	[-18.19, 73.09]	
Web Browser: Safari/Firefox	-50.50	29.71	-1.70	[-108.86, 7.86]	
Web Browser: Autres/Firefox	36.51	62.11	0.59	[-85.5, 158.53]	

**Table 5** Mixed model regression coefficients for IKIs

	$\beta$	SE	t	95% CI	(sig.)
(Intercept)	205.14	26.11	7.86		
Experimental variables					
Hand	-8.16	6.13	-1.33	[-20.21, 3.89]	
Sequence	-21.18	1.65	-12.84	[-24.42, -17.94]	*
Uncertainty	-23.13	1.06	-21.91	[-25.2, -21.05]	*
Position	-30.81	1.06	-29.18	[-32.89, -28.74]	*
Sequence $\times$ Uncertainty	12.93	1.95	6.62	[9.09, 16.77]	*
Sequence $\times$ Position	22.71	1.49	15.22	[19.78, 25.64]	*
Uncertainty $\times$ Position	39.30	1.49	26.37	[36.38, 42.23]	*
Sequence $\times$ Uncertainty $\times$ Position	-17.94	2.10	-8.53	[-22.07, -13.81]	*
Trial number	-0.28	0.01	-37.89	[-0.3, -0.27]	*
Participants' variables					
Gender	-35.56	6.32	-5.63	[-47.97, -23.16]	*
Age (spline 1)	1.46	0.75	1.94	[-0.02, 2.94]	
Age (spline 2)	2.44	0.95	2.57	[0.58, 4.31]	*
Handedness	9.41	8.95	1.05	[-8.18, 26.99]	
Computers' variables					
OS: OS X/Windows	-1.50	9.14	-0.16	[-19.46, 16.47]	
OS: Linux/Windows	-5.40	13.16	-0.41	[-31.26, 20.46]	
Web browser: Chrome/Firefox	-0.68	8.02	-0.08	[-16.44, 15.09]	
Web browser: IE/Firefox	1.78	10.65	0.17	[-19.14, 22.7]	
Web browser: Safari/Firefox	0.18	13.62	0.01	[-26.58, 26.93]	
Web browser: Autres/Firefox	1.63	28.48	0.06	[-54.32, 57.57]	



**Fig. 5** Data sampling of two representative participants (left and right columns). The upper row represents interkeystroke intervals (IKIs), ranked in increasing order. The lower row represents the distribution of

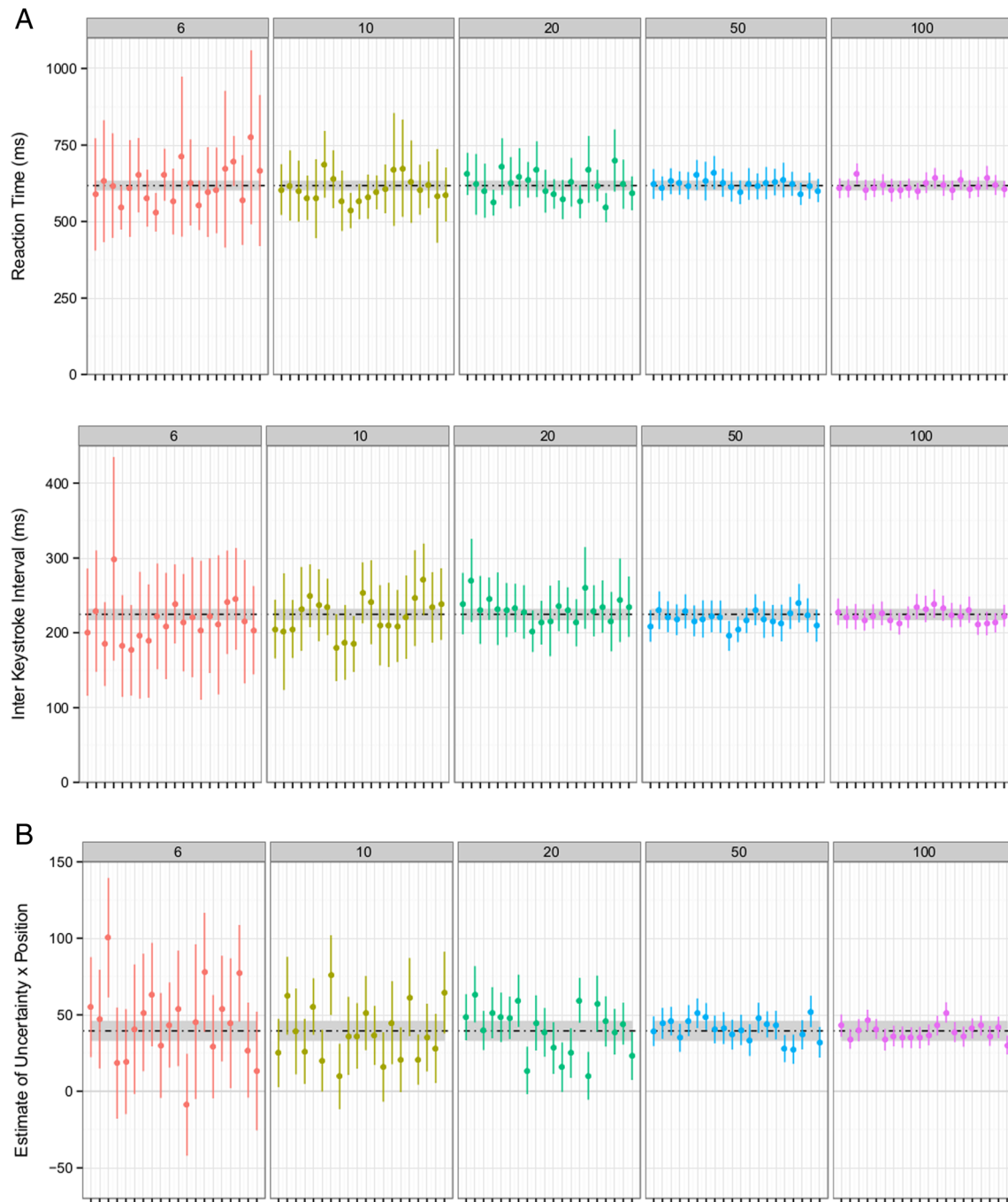
the remainders from division of the IKIs by 8. The data on the right show quantization around multiples of 8 (see the main text for details)

samples were drawn from the original distribution, and we calculated the mean (RT or IKI) and estimated the confidence interval of each sample. As sample size increased, the sample means became less dispersed around the mean of the whole distribution, and the range of the confidence intervals decreased: That is, the sample means became more stable estimates of the mean of the whole sample (Fig. 6A).

To evaluate the effect of sample size on the experimental effects, we followed the same procedure, taking random samples of increasing size from the whole distribution and running the mixed regression model described above on each sample. Figure 6B presents the evolution of the beta estimate for the Uncertainty  $\times$  Position interaction across samples. It shows that the effects are well estimated from a sample size of 50, because from this value and above, all of the confidence intervals include the mean of the whole distribution.

## General discussion

We aimed at quantifying the precision and reliability of timing measures performed during sequences of keystrokes. We used the JavaScript jsPsych library to create an experiment involving finger-movement sequences. Black Box Toolkit hardware tests showed a systematic delay centered around 60 ms for RTs, and an unbiased measure with an average 0-ms delay for IKIs. Thus, the delay in RT measurements did not increase when several keystrokes were collected in rapid succession. Then, online experimental data that we collected accurately reproduced the original results (Rosenbaum et al., 1984). Random subsampling of the data revealed that, in paradigms such as this, samples of at least 50 participants are necessary for accurate estimation of the data distributions and experimental effects. Finally, both the BBTK tests and online experiment revealed substantial quantizing of the IKI data, most likely due to the sampling frequency of the USB keyboards



**Fig. 6** (A) Means and confidence intervals of response times (RTs; top) and interkeystroke intervals (IKIs; bottom) over random samples of increasing size (see text for details). The horizontal line represents the mean of the whole sample with confidence intervals as a surrounding shaded area. Each point represents a random sample, sample size is indicated above each panel. (B) Beta estimates and confidence intervals

for the Uncertainty  $\times$  Position interaction, evaluated in mixed regression models on IKIs over random samples of increasing size. The horizontal line represents the beta estimate from the model on the whole sample (reported in Table 5), with confidence intervals as a surrounding shaded area

(125 Hz). This did not prevent the assessment of experimental effects with reasonable sample sizes.

The finding of a null difference between the programmed and recorded IKIs with the BBTK is a good indication that online testing can be used for assessing differences in timing

between the elements of motor sequences performed through the keyboard. It generalizes the findings of previous experiments that used a stream of identical keystrokes separated by fixed intervals in comparable conditions (Keller et al., 2009; Simcox & Fiez, 2014). Here, we measured a sequence of

responses to a visual signal, with variable interresponse intervals. In addition, we showed that even for short time intervals, the online system performed with very good timing precision. The standard deviations for the IKIs were very small (around 10 ms), a value that is considered accurate in RT measurements (see Reimers & Stewart, 2007, 2015). The BBTK tests therefore also indicate a good reliability of the measures.

The BBTK tests revealed a constant lag of about 60 ms on the RT measurements using our jsPsych configuration. The finding of a lag is consistent with another study that compared jsPsych with programs designed for in-lab recordings (de Leeuw & Motz, 2016). Comparable amounts of RT overestimation were also observed by Reimers and Stewart (2015) in a test with the BBTK and JavaScript in a design very similar to ours, except for the motor sequences feature and the use of predefined jsPsych functions. Those authors tested various browsers and OSs with the BBTK, using JavaScript and html5, and showed a general overestimation of RTs that varied from 30 to nearly 100 ms (on a Dell Optiplex machine). Some OSs seem to introduce longer lags; for instance, in Reimers and Stewart's (2015) experiment, the two Windows 7 machines measured RTs that were 30–40 ms longer than those on XP machines. In our case, the delay between programmed and recorded responses is probably a mixed result of the load of ongoing programs running in the OS, hardware constraints (of the Safari browser and Macintosh computer), and the jsPsych program itself. In addition, a standard deviation of 8 ms is a small value, similar to those found in Reimers and Stewart (2007, 2015) for measures of RTs. On the basis of those *SD* values, those authors, by comparing online and traditional measures, acknowledged that online measures overestimate RTs but do not add extra noise to the data. Relative to the recording of keystroke timings using a JavaScript implementation relying on the "Date" object, the present use of the `getKeyboardResponse()` jsPsych function could possibly add a small processing overhead.<sup>2</sup> If this overhead impacts the recording of keystroke timings, this impact will likely be very small, since our values are very close to those of Reimers and Stewart (2015). This supports the generalizability of the present findings to any JavaScript experiment, not just using jsPsych. In addition, this indicates that jsPsych can be used by researchers in experimental psychology who need to implement experiments with high timing precision.

The quantizing of IKIs occurred for all three of the BBTK tests and for the online measurements. A similar phenomenon has also been shown by Neath, Earle, Hallett, and Surprenant

(2011), when testing various configurations of program/computers and keyboards. In their study, quantizing occurred only occasionally and seemed to be related to one particular configuration of a given Macintosh computer and a particular type of keyboard, and it was not discussed further. Quantizing of responses was also observed by Reimers and Stewart (2015) on the basis of cumulative frequency distributions. Quantizing occurred more frequently under certain configurations, but the authors could not find any systematic predictor of quantizing in the data. Here, in addition to showing quantizing, as usual, with the cumulative frequency distributions, we used a procedure that allowed for the quantification of the intervals between which data were sampled. In the BBTK tests, data were "packed" by multiples of 8 ms, a value that corresponds well to the sampling rate of a USB port (125 Hz). This was also the case in data recorded online with actual participants and variable computer settings. A great majority of the sample displayed the same quantizing by steps in multiples of 8 ms. This is a strong suggestion that quantizing is due to the sampling of the keyboard by the system. Such noncontinuous sampling of the data should be acknowledged, and some researchers for whom quantizing matters should be aware of it and choose non-USB input devices. However, a specific study has shown that under typical RT measurement conditions, the variability in human performance outweighs the imprecision in response devices (Damian, 2010).

Our online study on a large sample of participants also replicated the original results of Rosenbaum et al. (1984): The position of the uncertain response had an effect on RTs, with longer RTs for uncertain responses at Position 2 than to those at Position 3. In addition, we found the same interaction between the position of the uncertainty and the position of the required response on IKI measurements. The main effect of uncertainty reported in the original experiment did not reach significance in our sample, whereas the other effects we reported had not been significant in the original study. This discrepancy is probably related to the instability of the effects estimated by Rosenbaum et al. (1984) in the original study. The interaction of interest was replicated later on (Rosenbaum et al., 1987) and can be considered reliable. It should nonetheless be kept in mind that the original effects were reported with a sample of only six participants. With regard to our specific conditions (e.g., the same number of trials as in the original study, but only a subset of the original experimental conditions) and the experimental context (online measurements), we found that a minimum sample size of 50 participants was necessary to provide a relatively good and reliable estimation of the effect of interest. This threshold is nonetheless specific to the present experiment, and clearly is not a recommendation for a minimum sample size in any online experiment, since its value is likely to depend on a number of factors, such as

<sup>2</sup> Due to the addition of two *if* statements linked to the parameterization of the function. Time measurements themselves are done basically by instantiating JavaScript Date objects. In addition, profilers have indicated that the overall execution time of `getKeyboardResponse()`, including the definition of `start_time`, `key_time`, and the associated conditions, ranges between 3.4 and 5.5 ms (using Chrome DevTools and Firebug profiling, respectively).

the number of trials and the effect sizes. Our analysis nonetheless provides an illustration of the trade-off between sample size and the precision of estimates of the effects of interest, which depends on the constraints of a given experiment. Systematic tests of this type in methodological or experimental online studies would be useful to get a better overview of the minimally required sample sizes in various contexts.

Our measured RTs were longer than those from Rosenbaum et al.'s (1984) study: This could be due to the time lag (as measured with the BBTK) introduced by the operating system, computer, screen, and keyboard used (see Neath et al., 2011, for an example for the keyboard), as well as by the online configuration of the browser and jsPsych. This finding is in agreement with previous studies that had compared in-lab and Web-based experiments and typically found delays from 25 to 100 ms (Crump et al., 2013; de Leeuw & Motz, 2016; Reimers & Stewart, 2007; Schubert et al., 2013).

Regression analyses also indicated that the material configuration had no measurable effect on IKIs, although an effect of browser was evidenced for RTs. Moreover, our online interface could capture some variability linked to demographic variables: For instance, it allowed for extra assessments of the effects of variables such as age or gender. We found slight influences of age and gender on both RTs and IKIs. Previously, no effects of age have been reported on typing rates (Salthouse, 1984) or on indirect measures of motor-sequence learning (Howard & Howard, 1989). However, a male advantage on motor speeds has been reported (Nicholson & Kimura, 1996). Such effects should be taken cautiously, because predictors such as age, gender, and computer configurations might be correlated, as was suggested by Reimers and Stewart (2015), and might lead to spurious effects of the covariates. Mixed regression analyses then present the advantage of accounting for all predictors and their specific variabilities at the same time.

In conclusion, online measurement using jsPsych appears to be an accurate way to test for fine differences between the IKIs in various conditions. It offers a promising tool for researchers interested in motor-sequence learning and execution.

**Author note** This study was partly supported by the French Agence Nationale de la Recherche (MEGALEX Grant No. ANR-12-CORP-0001; AMIDEX Grant No. ANR-11-IDEX-0001-02; Brain and Language Research Institute Grant No. ANR-11-LABX-0036; MULTIMEX Grant No. ANR-11-BSH2-0010), the European Research Council (FP7/2007–2013, Grant No. 263575), the Ministère de l'Enseignement et de la Recherche (doctoral MNRT grant to S.P.), and the Fédération de Recherche 3C (Aix-Marseille Université).

## References

- Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging, 12*, 12–21. doi:10.1037/0882-7974.12.1.12
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. doi:10.18637/jss.v067.i01
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE, 8*, e57410. doi:10.1371/journal.pone.0057410
- Damian, M. F. (2010). Does variability in human performance outweigh imprecision in response devices such as computer keyboards? *Behavior Research Methods, 42*, 205–211. doi:10.3758/BRM.42.1.205
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*, 1–12. doi:10.3758/s13428-014-0458-y
- de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods, 48*, 1–12. doi:10.3758/s13428-015-0567-2
- Enochson, K., & Culbertson, J. (2015). Collecting psycholinguistic response time data using Amazon Mechanical Turk. *PLoS ONE, 10*, 1–17. doi:10.1371/journal.pone.0116946
- Fozard, J. L., Vercruyssen, M., Reynolds, S. L., Hancock, P. A., & Quilter, R. E. (1994). Age differences and changes in reaction time: The Baltimore Longitudinal Study of Aging. *Journals of Gerontology: Psychological Sciences, 49*, 179–189. doi:10.1093/geronj/49.4.P179
- Gentner, D. R. (1983). The acquisition of typewriting skill. *Acta Psychologica, 54*, 233–248.
- Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual Review of Psychology, 66*, 877–902. doi:10.1146/annurev-psych-010814-015321
- Howard, D. V., & Howard, J. H. (1989). Age differences in learning serial patterns: Direct versus indirect measures. *Psychology and Aging, 4*, 357–364. doi:10.1037/0882-7974.4.3.357
- Keller, F., Gunasekharan, S., Mayo, N., & Corley, M. (2009). Timing accuracy of Web experiments: A case study using the WebExp software package. *Behavior Research Methods, 41*, 1–12. doi:10.3758/BRM.41.1.12
- Neath, I., Earle, A., Hallett, D., & Surprenant, A. M. (2011). Response time accuracy in Apple Macintosh computers. *Behavior Research Methods, 43*, 353–362. doi:10.3758/s13428-011-0069-9
- Nicholson, K. G., & Kimura, D. (1996). Sex differences for speech and manual skill. *Perceptual and Motor Skills, 82*, 3–13. doi:10.2466/pms.1996.82.1.3
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*, 1–32. doi:10.1016/0010-028590002-8
- Posner, M. (1978). *Chronometric explorations of mind*. Hillsdale: Erlbaum.
- Reimers, S., & Maylor, E. A. (2005). Task switching across the life span: Effects of age on general and specific switch costs. *Developmental Psychology, 41*, 661–671. doi:10.1037/0012-1649.41.4.661
- Reimers, S., & Stewart, N. (2007). Adobe Flash as a medium for online experimentation: A test of reaction time measurement capabilities. *Behavior Research Methods, 39*, 365–370. doi:10.3758/BF03193004
- Reimers, S., & Stewart, N. (2015). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript Web experiments. *Behavior Research Methods, 47*, 309–327. doi:10.3758/s13428-014-0471-1
- Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology, 49*, 243–256. doi:10.1027/1618-3169.49.4.243

- Rhodes, B. J., Bullock, D., Verwey, W. B., Averbeck, B. B., & Page, M. P. A. (2004). Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives. *Human Movement Science, 23*, 699–746. doi:10.1016/j.humov.2004.10.008
- Rosenbaum, D. A., Hindorff, V., & Munro, E. M. (1987). Scheduling and programming of rapid finger sequences: Tests and elaborations of the hierarchical editor model. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 193–203. doi:10.1037/0096-1523.13.2.193
- Rosenbaum, D. A., Inhoff, A. W., & Gordon, A. M. (1984). Choosing between movement sequences: A hierarchical editor model. *Journal of Experimental Psychology: General, 113*, 372–393. doi:10.1037/0096-3445.113.3.372
- Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science, 6*, 1–36. doi:10.1207/s15516709cog0601\_1
- Salthouse, T. A. (1984). Effects of age and skill in typing. *Journal of Experimental Psychology: General, 113*, 345–371. doi:10.1037/0096-3445.113.3.345
- Schubert, T. W., Murteira, C., Collins, E. C., & Lopes, D. (2013). ScriptingRT: A software library for collecting response latencies in online studies of cognition. *PLoS ONE, 8*, e67769. doi:10.1371/journal.pone.0067769
- Simcox, T., & Fiez, J. A. (2014). Collecting response times using Amazon Mechanical Turk and Adobe Flash. *Behavior Research Methods, 46*, 95–111. doi:10.3758/s13428-013-0345-y