



Full length article

Introducing otree_tools: A powerful package to provide process data for attention, multitasking behavior and effort through tracking focus[☆]



Philipp Chapkovski^{a,*}, Christian Zihlmann^b

^a Higher School of Economics, Moscow, Russia

^b University of Fribourg, Switzerland

ARTICLE INFO

Article history:

Received 28 November 2018

Received in revised form 29 March 2019

Accepted 29 April 2019

Available online 7 May 2019

Keywords:

Experimental methodology

Process data

oTree

Time tracking

Attention

Real effort

Effort

Cost of effort

Honesty

Multitasking

ABSTRACT

With the advent of online subject pools, conducting experiments outside the laboratory has become more popular among the scientific community. Unlike the lab, online and field environments tend to be accompanied by a loss of control. In this article we introduce *otree_tools*, a concise yet powerful add-on for oTree (Chen et al., 2016). *otree_tools* provides novel ways by which to measure behaviors that are potentially important in the social sciences, such as attention, multitasking and effort. The software also features a novel method of tracking time through the identification of noise. We demonstrate the utility of *otree_tools* with the help of experimental evidence. The software substantially increases control of the environment. Moreover, the original metrics can be employed for innovative outcome variables, opening the avenue for new research opportunities.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

One of the major features of running experiments outside of a standard laboratory is that the environment is generally more natural but less controllable by the researcher (Harrison and List, 2004; Mason and Suri, 2012; Palan and Schitter, 2018). Participants are likely preoccupied with other activities during the experiment, such as browsing the web, writing emails, checking social media or simultaneously engaging with other paid tasks on platforms like Amazon Mechanical Turk¹ (AMT) or prolific.ac². Research has investigated this conjecture and discovered multitasking behavior to be typical not only on AMT but also in campus and community samples, decreasing participants involvement with the experiment (Chandler et al., 2014; Necka et al., 2016).

Such naturally occurring activities might yet be of great interest to a researcher – if one is able to capture it.

otree_tools addresses these multitasking behavior practices with its key feature, a focus tracking mechanism. It is nearly impossible to examine what participants of online or field studies are actually doing during the experiment with any certainty. Fortunately however, modern browsers provide reliable tools to capture the moment in which a participant switches to another browser tab or any other computer application. *otree_tools* monitors and records these events and provides a concise package to embed this tracking mechanism with just a single line of code.

As a result, *otree_tools* is not only able to provide new measures for attention, but also a novel way to track time. The software corrects total reported time of events in which subjects switch away from the experimental window. *otree_tools focus time* filters out these noise-producing events from the data through identification and correction for participants who spend a certain amount of time not engaging with the experiment or survey. In short, *otree_tools* is able to differentiate between *unfocus* and *focus time* elapsed, while traditional time trackers fail to make this discrimination and only report the total time.

Potential use cases for implementation of the software are manifold. To give a short preview, a researcher is able to control

[☆] The software application described here is free of charge and licensed under the MIT open source license with a citation requirement of this article. The application is available for download at https://github.com/otree-tools/otree_tools. Any use of the software - whether as a whole or in parts - implies the acceptance of the license agreement available in the folder for download.

* Corresponding author.

E-mail addresses: fchapkovskiy@hse.ru (P. Chapkovski), christian.zihlmann@unifr.ch (C. Zihlmann).

¹ <https://www.mturk.com/>.

² <https://prolific.ac/>.

for subjects' inattention and multitasking behavior. Furthermore, *otree_tools* is able to identify sources for data contamination, such as unsophisticated bots. Importantly, the software might provide proxies for honesty, impatience and self-control. In addition, *focus time* measures unobserved effort: In Section 3, we elaborate why standard tracked time is subject to a systematic measurement error in measuring effort. With help of a recently conducted real effort experiment, we investigate and find evidence in favor of this hypothesis. We show that *otree_tools* *focus time* effectively filters out noise and produces lower standard deviations relative to the point estimate, lower inter-quartile ranges as well as better goodness-of-fit. Often, in contrast to total elapsed time, *otree_tools* *focus time* has significant predictive power for participants' performance in the real effort task. In short, in measuring effort, *focus time* outperforms classical time trackers.

We are confident that the process data collected with *otree_tools* effectively supports researchers' ability to control for subjects inattention and the resulting low involvement with ongoing studies, especially in non-laboratory environments. Beyond eliminating noise and increasing the power of study, we believe the software provides access to innovative explanatory and outcome variables, opening new opportunities in many fields, for example honesty, attention and multitasking research.

In contrast to many other methods of collecting process data, such as eye-tracking or neural and physiological recordings, *otree_tools* is easy to implement and does not require specific hardware or programming knowledge, can be applied to large-scale samples and is a cost-free, open-source software.

The article proceeds as follows. First, we briefly outline the main features included in *otree_tools*. In Section 3, we elaborate on important use cases. Before we conclude, we present initial experimental evidence in Section 4 for the functionality of *otree_tools*. Readers interested in using the software are encouraged to consult the Online Appendix C: Technical specifics of the focus tracking mechanism as well as the Online Appendix D: Users guide³.

2. Features of *otree_tools*

The focus tracker implemented in *otree_tools* identifies and records all events in which a user loses the technical focus⁴ with the experimental screen, for example by switching to another tab, browser window or any other computer application⁵. In other words, *focus* is only given if the current *oTree* window, on which the experiment is occurring, is technically the active window. Generally, an element is active if the keys typed on the keyboard are sent to it. Refer to Appendix A: Screen shots for visual illustrations of different scenarios. Briefly, every time a participant of an ongoing experiment switches to another tab or application, a loss of focus event is registered. Therefore, activities such as browsing the web, chatting on Skype, writing emails, scrolling through the lecture slides and so forth generate a loss of focus event, which allows for a reliable identification of such behavior by *otree_tools*.

These focus events are recorded in a database with an associated timestamp. The database is accessible through a GUI. *otree_tools* also provides functions that create metrics based on the collected data. Table 1 provides a brief overview over these functions. Importantly, the software is able to discriminate

between unfocus and *focus time*. Traditional time trackers are unable to make this distinction and simply report total time elapsed. *Focus time* thus identifies participants who spend more time on the task than actually being focused on it.

We consider the aforementioned features to be the key innovations and main contributions of our software, as they provide novel ways to measure attention and to track time. We believe these features may prove helpful for researchers in many fields, which is why we dedicate the remainder of this article solely to them. Remaining auxiliary features of *otree_tools* such as customized 'Next' buttons and Multiple-Choice fields are covered in the Online Appendix D: Users guide.

3. Use cases

3.1. A tool to measure inattention

Subject inattention is a key hazard in many surveys and experiments and is frequently exhibited through the skipping of instructions or responding randomly to surveys (Palan and Schitter, 2018). In an informal survey among social scientists, more than two thirds listed data quality and worker's attention as the greatest concern with AMT (Chandler et al., 2014). Yet, behavior that adversely affects data quality (e.g. completing a study in a sleepy state) is found to be common not only in online but also in community and campus samples. Similarly, multitasking behavior identified through self-reports is found to be common among different sample populations (Necka et al., 2016; Paolacci et al., 2010). In a classical lab setting, Oppenheimer et al. (2009) find that a substantial number of participants do not follow instructions and propose the implementation of instructional manipulation checks. The literature further suggests screening participants and adding comprehension questions and attention checks (Necka et al., 2016). Examining the total time spent to complete a task is proposed to be a fast and easy-to-implement means of detecting poor quality responses (Mason and Suri, 2012).

otree_tools collects several metrics to measure subjects inattention⁶. First, by analyzing *focus time*, a researcher is more accurately able to identify subjects who respond to a question randomly or without properly reading and reflecting, enabling an assessment of self-reported survey responses. *Focus time* can also be implemented for tracking the total duration that subjects are focused on instruction pages. The software also collects data on how many times a page has been submitted, which allows a researcher to identify whether a participant answered a comprehension question or attention check correctly for the first time or only after several tries. Importantly, the function `num_focusoff_events` collects the count of events in which focus has been lost, for each participant by each single page. Also, the duration of a loss of focus event is accessible. Because a loss of focus is likely accompanied by a loss of attention, these metrics likely represent participants' inattention. Moreover, it captures multitasking behavior, allowing for its identification through observable data instead of self-reports. While informally testing a beta version of *otree_tools* on AMT, we collected data about how often participants lose focus with the task window. Based on observed behavior, we have documented that multitasking is indeed common on AMT: In this short task, which lasted roughly only six minutes (median 351 s), more than half of the participants (57%) lost focus with the experimental window at least once⁷.

Beyond providing control variables to mitigate noise, the data collected with *otree_tools* may also make innovative explanatory and outcome variables in multitasking and attention span

³ The purpose of the body of this article is to introduce *otree_tools*, but not to provide practical guidance for implementation.

⁴ Note that we adopt the terminology used by the browser developers: Focus is defined in a technical sense, but does not necessarily correlate fully with one's cognitive state.

⁵ Refer to Online Appendix C: Technical specifics of the focus tracking mechanism for a discussion about the technical background of the tracking mechanism.

⁶ Contrary to many common attention checks, the software operates in the background and is thus non-invasive.

⁷ See Figure 6 in Online Appendix B: Additional experimental results.

Table 1
Features included in the *focus tracker* of *otree_tools*.

Function	Description
<i>Focus tracker Can</i>	
<code>get_time_per_page</code>	Total time a participant spends on the oTree page.
<code>get_focused_time</code>	Duration the oTree page has received technical focus.
<code>get_unfocused_time</code>	Duration the oTree page has lost technical focus.
<code>num_focusoff_events</code>	Number of loss of focus events for each participant by page.

research available. Human attention is limited and we are rarely able to process all of the information that is freely available. Indeed, many key findings in behavioral economics may reflect a form of inattention (Gabaix, 2019), and attention will legitimately become more important in the social sciences in the future. To make progress in this direction, “We need more measures of inattention.” (Gabaix, 2019, p.327). We are confident that *otree_tools* does so.

3.2. A tool to measure effort

Effort can be exerted in various ways, for example physically or mentally. The determination of the cost of expending effort may depend on the specific setting. Very likely, opportunity cost of time enters one’s cost-of-effort assessment. Other factors likely also determine cost-of-effort, for instance motivation, skills and cognitive ability. That is why effort is generally hard to measure. Process-tracking methods that attempt this include pupil dilation, eye-tracking and neural metrics (Schulte-Mecklenbeck et al., 2017).

In experimental economics, two major paradigms are widely used to measure effort: Stated-effort tasks induce a monetary cost function for effort choices, while real effort tasks are employed for participants to exert real effort with the goal of increasing external validity. In the latter case, exerted effort is not observable. Researchers thus commonly assume that a subject’s observed performance represents unobserved effort. This assumption implies a loss of control over the cost-of-effort function⁸. Crucially, unobserved effort exerted should be sufficiently strongly correlated with observed performance⁹. Researchers may reasonably judge how strong this connection is. Because exerted effort remains unobservable, empirical evidence is difficult to elicit. As a consequence, (comparative) evidence about the appropriateness of real effort tasks is rare. A few exceptions provide valuable insights (Brueggen and Strobel, 2007; Carpenter and Huet-Vaughn, 2017; Gill and Prowse, 2019; Lezzi et al., 2015). We recommend in particular a recent helpful overview from Charness et al. (2018).

Since the opportunity cost is most probably one determinant of effort, time elapsed measurement might be another way to estimate effort¹⁰. The importance of time also becomes obvious in real world observations, like contractual weekly hours or minimum wage laws. In these examples, time seems to be at least equally important as the outcomes. Hence, time on task may be a potential proxy for effort based on procedural rather than outcome data, which is work input rather than output.

Indeed, Gabaix (2019) advocates that collecting time on task is valuable and constitutes a valid proxy for unobserved effort.

Yet, time might partially be spent without effort. Concretely, in computerized experimental studies conducted online or in the field, participants are likely engaged with other activities such as checking social media or simultaneously working on other paid tasks¹¹. These activities do not represent effort provision in relation to the task under investigation. As a consequence, total time elapsed is a noisy measure of effort. In contrast to traditional time trackers, such effortless behavior is identified by *otree_tools* and classified as unfocus time. Remember that *otree_tools* provides three measurements for time elapsed: Total time is simply the total time elapsed on a task as observed by standard time trackers,¹² *focus time* is the time the participant had technical focus on the experiment while unfocus time is the contrary. Consequently, total time is composed of unfocus and *focus time*. Hence, *focus time* is endogenous to total time. It becomes evident that *focus time* produces, by definition, lower (or at most equal) means and medians than total time. Now assume that *focus time* is indeed a determinant of cost-of-effort, while unfocus time represents inactivity and is simply noise. Since unobserved effort is only a function of invested *focus time*, total time measures true effort within a systematic positive measurement error because it also includes unfocus time. *Focus time* separates this bias and is thus superior to total time¹³. Consequently, in measuring effort, total time is subject to an attenuation bias: the estimated $\hat{\beta}$ will always underestimate the true β . This downward bias becomes larger with an increase in the share of noise represented through unfocus time. For further elaboration on the topic of measurement error, refer for example to Antonakis et al. (2010) and Wooldridge (2016).

In other words, *focus time* filters out noise from the data and is consequently likelier to capture the signal, which is the true relationship between time invested into the task and effort. The utility of total time, and with it the strength of the dominance of *focus time*, depends on the magnitude of unfocus time. Certainly, both *focus time* and total time should yield statistical significance in predicting effort in large sample sizes. However, and importantly, the downward attenuation bias does not vanish. To sum up, compared to total time, *focus time* should always (i) result in lower standard deviations relative to $\hat{\beta}$ (ii) exhibit a stronger relationship with effort, i.e. yield a higher β -coefficient (iii) explain more of the statistical relationship, i.e. result in a higher goodness-of-fit (R^2). In short, *focus time* increases the power of the statistical test.

⁸ Ideally, the researcher is able to control for heterogeneity and hence to control for the cost-of-effort function.

⁹ In fact, real effort tasks are all likely to suffer from an omitted variable bias, resulting in greater sample sizes needed to capture treatment effects. As an anonymous referee correctly points out, a larger sample size does not necessarily provide relief since the bias may persist even in the asymptotic limit. We are thankful for this comment.

¹⁰ Some researchers use this approach to determine exerted effort in valuing goods or to predict choice outcomes, to name a few examples (Krajbich and Rangel, 2011; Ofek et al., 2007).

¹¹ In addition, be it in the lab or field, a participant may also engage in offline activities such as daydreaming or restroom breaks. *otree_tools* is able to identify such events under the assumption that the computer screen is locked or in sleep mode. If the screen still receives technical focus, *otree_tools*, as well as traditional time trackers, are not able to capture such offline activities.

¹² Refer to Online Appendix C: Technical specifics of the focus tracking mechanism for further technical background.

¹³ Recall that measurement error in the dependent variable generally does not jeopardize the validity of OLS. Yet, systematic measurement error in the independent variable turns OLS inconsistent. The larger the variance in the measurement error, relative to the variance of the true values, the higher the OLS inconsistency (Wooldridge, 2016).

Since true exerted effort is unobserved, real effort tasks are ideal for studying the appropriateness of time elapsed as a proxy for effort, thanks to the availability of objective outcome data, i.e. participants performance, which can be used as the dependent variable. We have done just that and have provided the initial evidence supporting our claim in Section 4 Demonstration of usefulness.

To summarize, in real effort experiments, *focus time* complements performance as a procedural proxy for effort. In other experimental settings, in which an objective outcome-based proxy, such as performance, is absent, *focus time* may well serve the purpose of measuring effort and thus providing valuable insights.

3.3. A tool to identify honesty

otree_tools may also prove helpful in honesty research. We illustrate its usefulness with a simple example. [Hugh-Jones \(2016\)](#) asks subjects six quiz questions. Subjects receive a payoff for correct answers. The experimenter requests that the subjects not look on the internet for the answers. Because the experiment is conducted online, it is naturally easy to do so. To alleviate this concern, the questions chosen were too difficult, if not impossible, for almost anyone to answer without cheating. Consequently, a correct response inspired an assumption of cheating. To test the validity of this assumption, the authors ran a control experiment in a standard laboratory with 144 subjects, where cheating was difficult (but not impossible). With *otree_tools*, a researcher can actually check if and when participants switch windows and lose focus. The result is a more accurate identification of honesty on an individual level and thus more precise outcome variables, while simultaneously employing more natural questions. We believe that the data collected with *otree_tools* complements many measures in honesty research and allows for new research opportunities.

3.4. Other applications

Data contamination. In August 2018, news that AMT might be contaminated by so-called bots¹⁴ spread rapidly among the scientific community ([Dreyfuss, 2018](#); [Stokel-Walker, 2018](#)). Since bot behavior likely undermines the validity of academic studies, it is crucial to identify (or optimally prevent) it. Noticeably, more recent evidence demonstrates that the drop in data quality on AMT may not only be connected to bots, but also to low-quality workers,¹⁵ who are employing Virtual Private Servers (VPS) farms to conceal their physical location ([Dennis et al., 2018](#); [Moss and Litman, 2018](#)). Common checks such as CAPTCHAs or test questions are easily passed by these so-called farmers and are thus useless ([Moss and Litman, 2018](#)). Also, standard screening methods, such as only allowing U.S. workers, do not provide relief¹⁶ ([Dennis et al., 2018](#)). *otree_tools* attempts to combat both potential sources of data contamination. First, the software is able to identify unsophisticated bots and automated scripts because the focus tracker would register zero *focus time* as

these robots work in the background. We successfully tested this approach with the two native bot types implemented in oTree, command-line and browser bots¹⁷. Second, farmers are, in our opinion, harder to identify, as they are human workers using a commercial VPS. Often, subjects who do not engage with a study sincerely are identifiable by extremely fast or super slow responses ([Spiliopoulos and Ortmann, 2018](#)). By analyzing *focus time* instead of total time, identification of such subjects becomes more accurate. While it is not perfect, we hope *otree_tools* provides complementary ways to detect data contamination.

Estimating rewards. Often, participants of online studies are rewarded based on the estimated duration of completing the task or survey. It may be worthwhile to compare summary statistics of *focus time* and total time in order to assess and calculate the appropriate monetary reward.

Activity-based costing. Research in Managerial Accounting documents a bias in the estimation of one's own time spent, captured through benchmarking the self-reported estimated total time against time tracked by the computer ([Schuhmacher and Burkert, 2014](#)). With its *focus time* feature, *otree_tools* may facilitate field experiments and improve the accuracy of the measurement in this strand of literature.

Response time. In contrast to psychology, the analysis of response time (RT) data is scarce in experimental economics. Generally, the literature associates longer RT with choice options that are similar in utility, leading to more conflicted decision-making ([Frydman and Krajbich, 2018](#)). RT is process-based data, which has been proven valuable in various economic settings, for example allowing for the inference of social, risk and time preferences ([Chabris et al., 2008](#); [Konovalov and Krajbich, 2017b](#)). Furthermore, RT can reveal an agents' private information, demonstrated by the buyers private value of the good in a bargaining experiment ([Konovalov and Krajbich, 2017a](#)). In their recent overview, [Spiliopoulos and Ortmann \(2018\)](#) conclude that RT analysis holds considerable potential for experimental economics. *Focus time* also provides a better metric for RT because it eliminates measurement error of time tracked, which standard trackers are likely subject to. Also, the focus tracker may support the execution of field experiments in RT research. To sum up, *otree_tools* facilitates further research in this field at no cost.

Inter-temporal choices. Research in inter-temporal choices relies largely on experiments of monetary allocations, an experimental design often criticized ([Ericson et al., 2015](#); [Sprenger, 2015](#)). That is why experimental economics started to employ real effort tasks in lieu of monetary choices to study inter-temporal consumption choices ([Augenblick and Rabin, 2018](#); [Augenblick et al., 2015](#); [Carvalho et al., 2016](#)). *otree_tools* might be helpful for the study of time preferences over direct consumption: After all, *focus time* may well serve the purpose of investigating, namely the allocation of effort provision over time. *otree_tools* is also better able to control for learning effects occurring over time compared to classical time trackers.

Impatience and self-control. Closely related, time-preferences have primarily been investigated in the laboratory thus far. But "individuals pursue many activities that indirectly reveal a preference for early gratification" ([DellaVigna and Paserman, 2005](#), p.545). The author's proxy exhibited impatience through several types of behavior. One of them is an interviewer's subjective assessment of the respondent's impatience and restlessness while taking the survey. The number of loss of focus events might represent an accurate measurement of impatience, since participants who switch away from the experiment likely trade for an immediate reward (e.g. checking social media) for delayed costs (finalizing the task). Moreover, switching tasks can be seen as a direct measure of a lack of self-control, opening still more research opportunities in this field.

¹⁷ Because browser bots mimic real browsers, the approach does only work if the oTree window is not active.

¹⁴ A bot is essentially a robot that runs automated tasks.

¹⁵ Low quality is defined by the authors as random answers, poor grammar and inconsistent behavior.

¹⁶ Another commonly suggested method to identify bots and farmers is to collect IPs and the related geolocation coordinates, scan the data for duplicates and identify if these locations are associated with a commercial data center ([Dennis et al., 2018](#); [Gautam et al., 2018](#)). Nonetheless, duplicate IPs may well originate from human workers sitting for example in the same university building. Also, the true location can be concealed through the use of a Virtual Private Network (VPN). Furthermore, collection of such sensitive data is often not allowed by the IRB or by legislation.

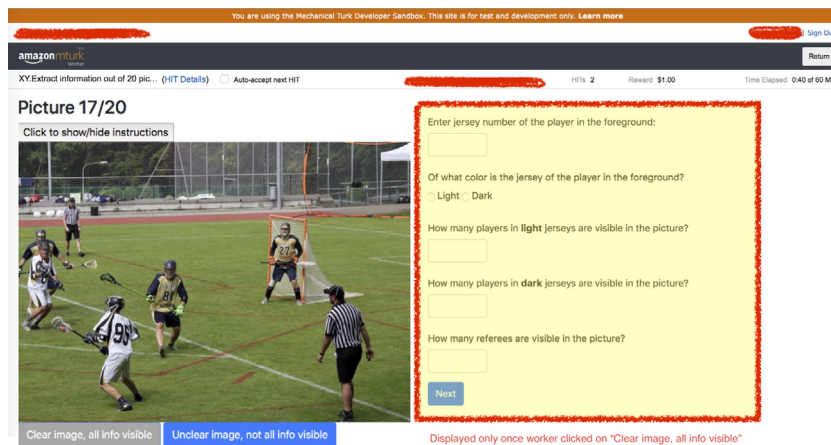


Fig. 1. Real effort task.

4. Demonstration of usefulness

This section demonstrates the usefulness of the *otree_tools* focus tracker on the basis of recent experimental data, obtained during the performance of a pilot test in November 2018¹⁸.

4.1. Experimental design: Real effort task

The natural field experiment was conducted on AMT. All parts of the experiment are coded with the software *oTree* (Chen et al., 2016). Furthermore, version 0.3.8 of *otree_tools* was implemented. Workers were not aware that they participating in an experiment and engaged in a visual search task: extracting and categorizing information from a picture. A real effort activity like this is common and natural on AMT. We presented workers with pictures from game-play situations of a lacrosse game and asked them to extract five pieces of information from each picture. Consult the screen shot in Fig. 1 for a visual illustration¹⁹. There were two stages, a pre-treatment stage and a treatment stage²⁰. In each stage, workers were tasked with categorizing a set of 20 pictures. The relevant variables for the purpose of this article are summarized in Table 2. The following analysis proceeds with a sample size of 61 observations²¹.

4.2. Relationship between time and focus time

Fig. 2 displays a scatter plot, displaying the relationship between the two variables *TIME* and *FOCUS*. If both time tracking methods lead to equal results, all observations would be found on the red plotted 45-degree line. One can clearly see that the focus tracker produces different data points than a traditional time tracker. For many observations, the deviation is impressive. Consider the observation plotted with a green filled dot: In stage

1, this worker spends approximately 23 min (~ 1400 s) on the task but is only actually focused on it for roughly 11 min (~ 660 s). Box plots, summary statistics and histograms are enclosed in the Online Appendix B: Additional experimental results and reveal that *FOCUS* produces lower means, medians, inter-quartile ranges and relative standard deviations than *TIME*, as predicted in Section 3.2.

As one can see, *otree_tools* identifies and corrects for observations of who spent a large amount of time connected to, but not focused on, the experimental window. *Focus time* therefore differs from time, as it identifies noise and thus provides a more accurate metric for time elapsed.

4.3. Predicting performance: Focus time as a proxy for effort

Is *focus time* a powerful predictor of effort, as represented through observed performance? In Section 3.2, we elaborate on why time tracked by standard time trackers is subject to a measurement error, resulting in a downward attenuation bias in predicting performance. Consequently, we predict that *FOCUS* should, compared to *TIME*, always result in higher point estimates, lower standard errors relative to the point estimate and a higher R^2 . As a consequence, *focus time* should be a better predictor of effort and yield higher statistical power.

Table 3 displays the correlation matrix for stage 1. Performance measured through errors, our proxy for effort, does not correlate with *TIME*, but with *FOCUS*. The direction of the effect is as expected negative, with a correlation coefficient of -0.401 . Table 4 displays the results of OLS estimations with raw and log-transformed data. As predicted, *TIME* exhibits higher standard errors, lower point estimates and a lower goodness-of-fit, compared with *FOCUS*. The expected relationship between performance and work input is identified at a statistically significant level by *FOCUS* in both models (2) and (4), and by *TIME* in the log-log model (3). The magnitude of the relationship is very relevant: Spending one percent of *focus time* more on the task reduces errors by 0.737 percent (model (4)), while this elasticity is only half of the magnitude with *TIME* (model (3)), a clear indication on how serious the downward attenuation bias can be. As outlined previously, statistical significance is a matter of sample size and amount of noise. We hypothesized that *FOCUS* must always yield higher (or at least equal) significance compared to *TIME* in finite samples, as it filters out noise that is not representative of effort, captured by unfocus time. This is exactly what the data demonstrates. The results for stage 2 are enclosed in the Online Appendix B: Additional experimental results and corroborate the findings obtained from stage 1 in a surprisingly

¹⁸ The primary purpose of this pilot was to test the code and to evaluate experimental parameters. Despite being limited in scale, the data powerfully reveals the usefulness of *otree_tools*. The main study is preregistered, see Herz and Zihlmann (2018).

¹⁹ The task has many advantages: First, it is a natural task with a potentially productive output. Second, counting tasks are assessed to be relatively independent of individual abilities and learning opportunities (Charness et al., 2018). Lastly, the task is tedious and requires real human effort.

²⁰ Treatment conditions induced in stage 2 are neglected in this methodological article.

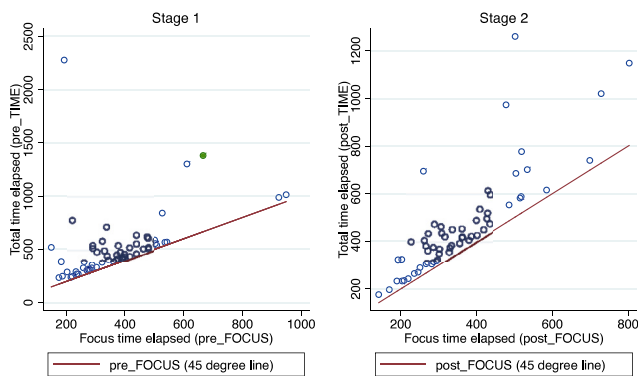
²¹ Due to technical reasons outlined in the Online Appendix C: Technical specifics of the focus tracking mechanism, the focus tracker returned zero focus time for two workers. For consolidation purposes, we excluded one observation who used a smart phone to complete the task. The qualitative results and conclusions are robust towards the exclusion.

Table 2
Summary of key variables.

Variable	Type	Description	Properties
pre_ERRORS	Endogenous	Proxy for unobserved effort based on work output. Performance of a worker in terms of delivered quality (measured through its counterpart, number of errors). In each picture, there are five pieces of information to extract. Hence, in each task with a set of 20 pictures, a worker can maximally score 100 errors.	Discrete: min 0 max 100
pre_TIME	Exogenous	Proxy for unobserved effort based on work input, measured through total time elapsed to complete the task.	Continuous
pre_FOCUS	Exogenous	Proxy for unobserved effort based on work input, measured through <i>focus time</i> spent to complete the task, i.e. the duration worker's oTree window received technical focus.	Continuous

Note: The prefix *post* in place of *pre* indicates stage 2, i.e. the treatment stage.

Scatter plots
TIME vs. FOCUS



Note: All units in seconds. Graphs with different scale on y and x axis.

Fig. 2. Scatter plot: Time vs. focus time.

Table 3
Correlation matrix stage 1.

	(1)		
	pre_ERRORS	pre_TIME	pre_FOCUS
pre_ERRORS	1		
pre_TIME	−0.0310 (0.813)	1	
pre_FOCUS	−0.401 (0.001)	0.424 (0.001)	1

p-values in parentheses.

unequivocal manner. For both stages, we also employed various robustness checks (robust variances, censored Tobit, and a set of robust regressions to account for influential observations). The results remain robust: In general, *focus time* outperforms total time as regards the magnitude of the point estimates, standard errors, statistical significance and goodness-of-fit.

We demonstrated that *FOCUS* dominates *TIME* in predicting observed performance due to its higher statistical power and the elimination of the measurement error. As a result, in our limited-in-scale sample, *focus time* has predictive power for observed performance while this necessary power is often absent with traditionally tracked time. Therefore, we put forward that *focus time* represents a novel alternative proxy for effort, based on work input (the process) rather than work output (the outcome).

5. Conclusion

This article introduces *otree_tools*, a powerful extension for the oTree platform. The software allows for uncomplicated collection of process data at virtually no cost. It integrates novel

Table 4
Regressing time and focus on performance, stage 1.

	Errors		Log of errors	
	(1)	(2)	(3)	(4)
pre_TIME	−0.00140 (0.00386)			
pre_FOCUS		−0.0257*** (0.00726)		
log of pre_TIME			−0.377* (0.204)	
log of pre_FOCUS				−0.737*** (0.222)
Constant	14.99*** (2.747)	23.60*** (3.159)	4.721*** (1.262)	6.688*** (1.298)
Controls	Yes	Yes	Yes	Yes
Adj.R2	0.025	0.198	0.078	0.181
N	61	61	61	61

Standard errors in parentheses

Note: OLS. The dependent variable is *pre_ERRORS* in models (1) and (2) and its log-transformation in models (3) and (4). All models include controls for worker's device in use and familiarity with lacrosse. Variables in model (3) and (4) are (natural) log-transformed to account for non-linearities.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

metrics for inattention, multitasking, effort and honesty. As a main functionality, the package features a new time tracking method, discriminating between unfocus and *focus time*. We provide evidence that *otree_tools* outperforms the native time tracker implemented in oTree and that *focus time* is a valid proxy for effort.

A major purpose of *otree_tools* is to increase researcher's control over the environment in non-standard settings, such as online and field environments. The software increases the signal-to-noise ratio and with it, statistical power of a study and eventually the efficiency of empirical research.

Moreover, the metrics provided by *otree_tools* may also be used for outcome variables. We are confident that the tool supports researchers in implementing unexplored experimental designs to conduct innovative research in various fields of the social sciences.

Acknowledgement

Chris Wickens provided valuable technical advice - thank you. We are very grateful to Holger Herz, Sven Grossrieder and Elisa Matthewes for the careful reading of our work which helped us to improve considerably the manuscript. We are also thankful to two anonymous referee's for helpful suggestions and feedback. The constructive comments were greatly appreciated.

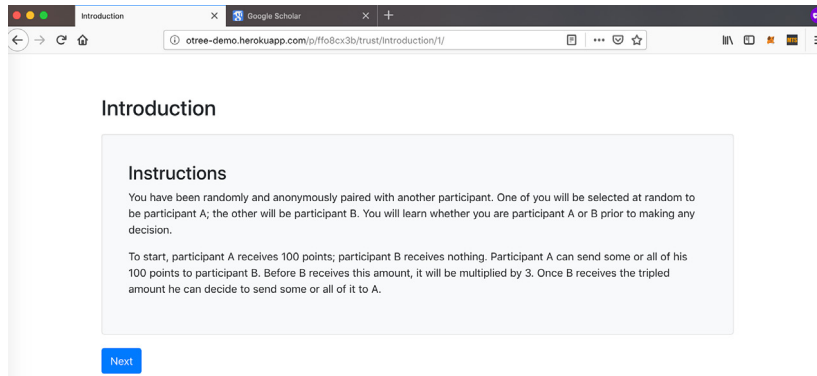
The article was prepared within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) and supported within the framework of a subsidy by the Russian Academic Excellence, Project '5-100'.

In future work, the implementation of an eye-tracker into otree_tools may be a promising endeavour. We gladly welcome thoughts, feedback and suggestions for the further development of otree_tools.

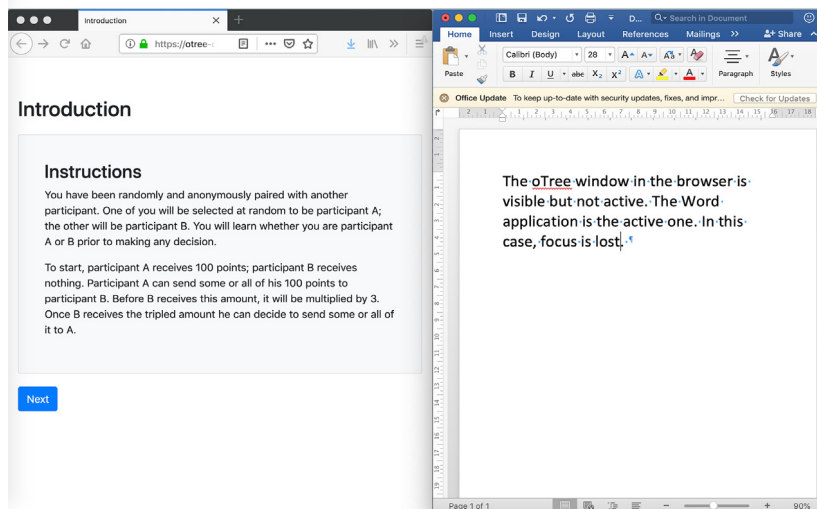
Appendix A. Screen shots

These screen shots visualize examples for focus being on and off.

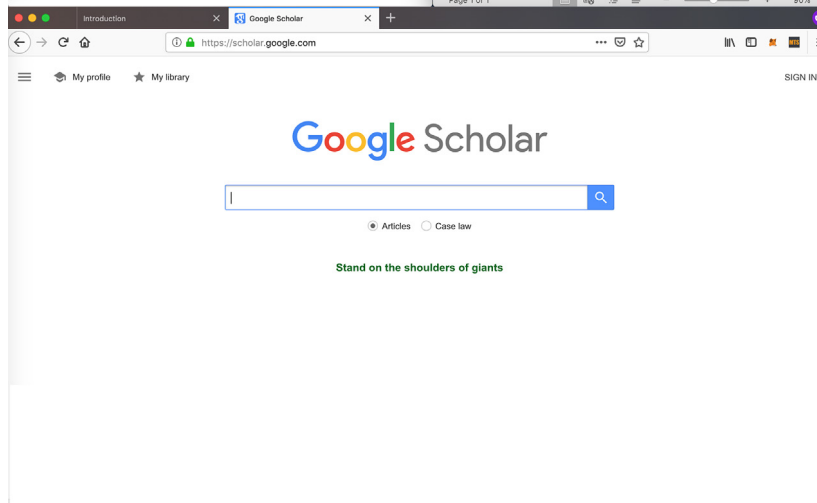
Focus on



Focus off



Focus off



Appendix B. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jbef.2019.04.010>.

References

- Antonakis, John, Bendahan, Samuel, Jacquart, Philippe, Lalive, Rafael, 2010. On making causal claims: A review and recommendations. *Leadersh. Quart.* (ISSN: 1048-9843) 21 (6), 1086–1120. <http://dx.doi.org/10.1016/j.leaqua.2010.10.010>, URL <http://www.sciencedirect.com/science/article/pii/S1048984310001475>, Leadership Quarterly Yearly Review.
- Augenblick, Ned, Niederle, Muriel, Sprenger, Charles, 2015. Working over time: Dynamic inconsistency in real effort tasks *. *Q. J. Econ.* 130 (3), 1067–1115. <http://dx.doi.org/10.1093/qje/qjv020>.
- Augenblick, Ned, Rabin, Matthew, 2018. An experiment on time preference and misprediction in unpleasant tasks. *Rev. Econom. Stud.* <http://dx.doi.org/10.1093/restud/rdy019>, rdy019.
- Brueggemann, Alexander, Strobel, Martin, 2007. Real effort versus chosen effort in experiments. *Econom. Lett.* (ISSN: 0165-1765) 96 (2), 232–236. <http://dx.doi.org/10.1016/j.econlet.2007.01.008>, URL <http://www.sciencedirect.com/science/article/pii/S0165176507000158>.
- Carpenter, Jeffrey, Huet-Vaughn, Emiliano, 2017. In: Schram, Arthur, Ule, Aljaz (Eds.), *Real Effort Tasks*. In: *Handbook of Research Methods and Applications in Experimental Economics*, Edward Elgar, Cheltenham, UK.
- Carvalho, Leandro S., Meier, Stephan, Wang, Stephanie W., 2016. Poverty and economic decision-making: Evidence from changes in financial resources at payday. *Amer. Econ. Rev.* 106 (2), 260–284. <http://dx.doi.org/10.1257/aer.20140481>, URL <http://www.aeaweb.org/articles?id=10.1257/aer.20140481>.
- Chabris, Christopher F., Laibson, David, Morris, Carrie L., Schuldt, Jonathon P., Taubinsky, Dmitry, 2008. Measuring intertemporal preferences using response times, Working Paper Series, 14353, National Bureau of Economic Research, <http://dx.doi.org/10.3386/w14353>, URL <http://www.nber.org/papers/w14353>.
- Chandler, Jesse, Mueller, Pam, Paolacci, Gabriele, 2014. Nonnaïveté Among amazon mechanical turk workers: Consequences and solutions for behavioral researchers. *Behav. Res. Methods* (ISSN: 1554-3528) 46 (1), 112–130. <http://dx.doi.org/10.3758/s13428-013-0365-7>.
- Charness, Gary, Gneezy, Uri, Henderson, Austin, 2018. Experimental methods: Measuring effort in economics experiments. *J. Econ. Behav. Organ.* (ISSN: 0167-2681) 149, 74–87. <http://dx.doi.org/10.1016/j.jebo.2018.02.024>, URL <http://www.sciencedirect.com/science/article/pii/S0167268118300556>.
- Chen, Daniel L., Schonger, Martin, Wickens, Chris, 2016. OTree – An open-source platform for laboratory, online, and field experiments. *J. Behav. Exp. Financ.* (ISSN: 2214-6350) 9 (Supplement C), 88–97. <http://dx.doi.org/10.1016/j.jbef.2015.12.001>, URL <http://www.sciencedirect.com/science/article/pii/S2214635016000101>.
- DellaVigna, Stefano, Paserman, M. Daniele, 2005. Job search and impatience. *J. Labor Econ.* 23 (3), 527–588. <http://dx.doi.org/10.1086/430286>.
- Dennis, Sean A., Goodson, Brian Matthew, Pearson, Chris, 2018. MTurk Workers' Use of Low-Cost 'Virtual Private Servers' to Circumvent Screening Methods: A Research Note, <http://dx.doi.org/10.2139/ssrn.3233954>, URL <https://ssrn.com/abstract=3233954>.
- Dreyfuss, Emily, 2018. A Bot Panic Hits Amazon's Mechanical Turk, URL <https://www.wired.com/story/amazon-mechanical-turk-bot-panic/> (Accessed: 26.11.18).
- Ericson, Keith M., Marzilli, White, John Myles, Laibson, David, Cohen, Jonathan D., 2015. Money earlier or later? simple heuristics explain intertemporal choices better than delay discounting does. *Psychol. Sci.* 26 (6), 826–833. <http://dx.doi.org/10.1177/0956797615572232>, PMID: 25911124.
- Frydman, Cary, Krajbich, Ian, 2018. Using response times to infer others' beliefs: An application to information Cascades, Working paper, <http://dx.doi.org/10.2139/ssrn.2817026>.
- Gabaix, Xavier, 2019. Chapter 4 – behavioral inattention. In: Bernheim, B. Douglas, DellaVigna, Stefano, Laibson, David (Eds.), *Handbook of Behavioral Economics – Foundations and Applications 2*. In: *Handbook of Behavioral Economics: Applications and Foundations 1*, vol. 2, North-Holland, pp. 261–343. <http://dx.doi.org/10.1016/bs.hesbe.2018.11.001>, URL <http://www.sciencedirect.com/science/article/pii/S2352239918300216>.
- Gautam, R., Kerstein, M., Moss, A. J., Litman, L., 2018. Understanding Geolocations and Their Connection to Data Quality [blog post], URL <https://blog.turkprime.com/understanding-geolocations-and-their-connection-to-data-quality> (Accessed: 24.01.19).
- Gill, David, Prowse, Victoria, 2019. Measuring costly effort using the slider task. *J. Behav. Exp. Financ.* (ISSN: 2214-6350) 21, 1–9. <http://dx.doi.org/10.1016/j.jbef.2018.11.003>, URL <http://www.sciencedirect.com/science/article/pii/S2214635018302193>.
- Harrison, Glenn W., List, John A., 2004. Field experiments. *J. Econ. Lit.* 42 (4), 1009–1055. <http://dx.doi.org/10.1257/0022051043004577>, URL <http://www.aeaweb.org/articles?id=10.1257/0022051043004577>.
- Herz, Holger, Zihlmann, Christian, Hidden costs of control: evidence from the field, AEA RCT Registry, October 29, 2018. URL <https://www.socialscienceregistry.org/trials/3475/history/36334>.
- Hugh-Jones, David, 2016. Honesty, beliefs about honesty, and economic growth in 15 countries. *J. Econ. Behav. Organ.* (ISSN: 0167-2681) 127, 99–114. <http://dx.doi.org/10.1016/j.jebo.2016.04.012>, URL <http://www.sciencedirect.com/science/article/pii/S016726811630052X>.
- Kononov, Arkady, Krajbich, Ian, 2017a. On the Strategic Use of Response Times, Working paper, <http://dx.doi.org/10.2139/ssrn.3023640> URL <http://dx.doi.org/10.2139/ssrn.3023640>.
- Kononov, Arkady, Krajbich, Ian, 2017b. Revealed indifference: Using response times to infer preferences, Working paper, <http://dx.doi.org/10.2139/ssrn.3024233>.
- Krajbich, Ian, Rangel, Antonio, 2011. Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proc. Natl. Acad. Sci.* (ISSN: 0027-8424) 108 (33), 13852–13857. <http://dx.doi.org/10.1073/pnas.1101328108>, URL <https://www.pnas.org/content/108/33/13852>.
- Lezzi, Emanuela, Fleming, Piers, Zizzo, Daniel John, 2015. Does It Matter Which Effort Task You Use? A Comparison of Four Effort Tasks When Agents Compete For A Prize, Working paper, <http://dx.doi.org/10.2139/ssrn.2594659> URL <https://ssrn.com/abstract=2594659>.
- Mason, Winter, Suri, Siddharth, 2012. Conducting behavioral research on amazon's mechanical turk. *Behav. Res. Methods* 44 (1), 1–23. <http://dx.doi.org/10.3758/s13428-011-0124-6>.
- Moss, A.J., Litman, L., 2018. After the bot scare: Understanding what's been happening with data collection on MTurk and how to stop it [blog post], URL <https://blog.turkprime.com/after-the-bot-scare-understanding-whats-been-happening-with-data-collection-on-mturk-and-how-to-stop-it> (Accessed: 26.11.18).
- Necka, Elizabeth A., Cacioppo, Stephanie, Norman, Greg J., Cacioppo, John T., 2016. Measuring the prevalence of problematic respondent behaviors among mturk, Campus, and community participants. *PLOS One* 11 (6), 1–19. <http://dx.doi.org/10.1371/journal.pone.0157732>.
- Ofek, Elie, Yildiz, Muhamet, Haruvy, Ernan, 2007. The impact of prior decisions on subsequent valuations in a costly contemplation model. *Manage. Sci.* 53 (8), 1217–1233. <http://dx.doi.org/10.1287/mnsc.1060.0689>.
- Oppenheimer, Daniel M., Meyvis, Tom, Davidenko, Nicolas, 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *J. Exp. Soc. Psychol.* (ISSN: 0022-1031) 45 (4), 867–872. <http://dx.doi.org/10.1016/j.jesp.2009.03.009>, URL <http://www.sciencedirect.com/science/article/pii/S0022103109000766>.
- Palan, Stefan, Schitter, Christian, 2018. Prolific.aca subject pool for on-line experiments. *J. Behav. Exp. Financ.* 17, 22–27. <http://dx.doi.org/10.1016/j.jbef.2017.12.004>, URL <http://www.sciencedirect.com/science/article/pii/S2214635017300989>.
- Paolacci, Gabriele, Chandler, Jesse, Ipeirotis, Panagiotis G., 2010. Running experiments on amazon mechanical turk. *Judgm. Decis. Mak.* 5, 411–419, URL <https://ssrn.com/abstract=1626226>.
- Schuhmacher, Karl, Burkert, Michael, 2014. Error in time-driven costing: An experimental comparison of different approaches to estimating time consumption. AAA 2014 Management Accounting Section (MAS) Meeting Paper, <http://dx.doi.org/10.2139/ssrn.2312085> URL <https://ssrn.com/abstract=2312085>.
- Schulte-Mecklenbeck, Michael, Johnson, Joseph G., Bckenholt, Ulf, Goldstein, Daniel G., Russo, J. Edward, Sullivan, Nicolette J., Willemssen, Martijn C., 2017. Process-tracing methods in decision making: On growing up in the 70s. *Curr. Dir. Psychol. Sci.* (ISSN: 1573-6938) 26 (5), 442–450. <http://dx.doi.org/10.1177/0963721417708229>.
- Spiliopoulos, Leonidas, Ortmann, Andreas, 2018. The BCD of response time analysis in experimental economics. *Exp. Econ.* (ISSN: 1573-6938) 21 (2), 383–433. <http://dx.doi.org/10.1007/s10683-017-9528-1>.
- Sprenger, Charles, 2015. Judging experimental evidence on dynamic inconsistency. *Amer. Econ. Rev.* 105 (5), 280–285. <http://dx.doi.org/10.1257/aer.p20151086>, URL <http://www.aeaweb.org/articles?id=10.1257/aer.p20151086>.
- Stokel-Walker, Chris, 2018. Bots on Amazon's Mechanical Turk are ruining psychology studies, URL <https://www.newscientist.com/article/2176436-bots-on-amazons-mechanical-turk-are-ruining-psychology-studies/> (Accessed: 26.11.18).
- Wooldridge, Jeffrey M., 2016. *Introductory Econometrics: a Modern Approach*, sixth ed. Cengage Learning.

Further reading

- Crockford, Douglas, 2006. The application/json media type for javascript object notation (json), Tech. Report.
- Goldberg, Lewis R., 1992. The development of markers for the big-five factor structure.. *Psychol. Assess.* 4 (1), 26.
- Lubbers, Peter, 2011. Html5 web sockets: A quantum leap in scalability for the web, <http://www.websocket.org/quantum.html>.
- Rammstedt, Beatrice, John, Oliver P., 2007. Measuring personality in one minute or less: A 10-item short version of the big five inventory in English and German. *J. Res. Personal.* 41 (1), 203–212.
- Spurlock, Jake, 2013. Bootstrap: Responsive Web Development. "O'Reilly Media, Inc."