Journal of Emerging Computer Technologies
Review Article
Received: *2025-02-23* | ***Reviewing:*** *2025-02-24 & 2025-07-20* | ***Accepted:*** *2025-07-20* | ***Online:*** *2025-08-21* | ***Issue Date:*** *2025-12-31*
Doi: 10.57020/ject.1645604

Software Solutions For Web-Based Experiments: A Comprehensive Review For Cognitive and Performance Research

Beyzanur KARAGÜZEL
Psychology
TED University
Ankara, Türkiye
0009-0003-5690-4818
beyzanur.karaguzel@tedu.edu.tr

Merve ERDOĞDU
Forensic Psychology
Hacettepe University
Ankara, Türkiye
0000-0002-3745-2639
merve.erdogdu@hacettepe.edu.tr

***Abstract*— The rapid growth of experimental software packages presents both opportunities and challenges for psychology and cognitive science researchers. These tools offer flexible study designs, but the increasing variety complicates software selection. This review provides a practical guide by comparing major experiment-building platforms across key criteria: timing precision, stimulus presentation, usability, and implementation demands. We systematically evaluate both lab-based and web-based software, highlighting their strengths and limitations. Lab-based systems generally deliver superior experimental control, while modern web platforms have significantly improved reliability and flexibility. We recommend PsychToolbox and PsychoPy for studies requiring maximum timing precision, E-Prime for clinical and applied settings, and Gorilla or jsPsych for online data collection. Our analysis emphasizes aligning software choice with research context, considering technical expertise, participant accessibility, and experiment complexity. PsychToolbox and PsychoPy are suited for advanced programmers, whereas E-Prime and Gorilla offer accessible solutions for researchers with limited coding skills. For complex behavioral paradigms, PsychoPy and jsPsych provide versatile options. This review functions as both a comparative analysis and practical handbook, enabling researchers to select appropriate software tailored to their experimental needs. By synthesizing performance benchmarks and implementation considerations, we deliver actionable recommendations to optimize study design across diverse research scenarios.**

***Keywords*— Psychology, cognition, attention, web-based experiments, software packages**

I. Introduction

Psychology encompasses a broad spectrum of research domains aimed at understanding and explaining human behavior by examining its underlying causes and consequences. As a cornerstone of scientific inquiry, experimental methods have long been employed to investigate behavioral, cognitive, and neurological processes. With advancements in computer technology, web-based experimentation has emerged as one of the most widely adopted methodologies, experiencing rapid growth in recent years. For instance, the number of articles indexed in Web of Science containing the keywords 'MTurk' or 'Mechanical Turk' increased fivefold between 2013 (121 articles) and 2018 (642 articles) [1]. Similarly, MTurk-based studies in the social sciences surged nearly 20-fold, from 61 in 2011 to 1,200 in 2015 [2].

Although the terms online and web-based experiments are frequently used interchangeably, key distinctions exist. Online experiments broadly include any internet-mediated research, such as mobile applications, email surveys, or social media platforms. In contrast, web-based experiments specifically denote studies conducted via web browsers, typically hosted on websites or web applications. Thus, while all web-based experiments fall under the umbrella of online experiments, the reverse is not necessarily true. For clarity, this paper adopts the term web-based to align with its focus on browser-delivered experimental paradigms.

Traditional lab-based experiments prioritize stringent control over confounding variables, striving to approximate real-world conditions as closely as possible. However, this approach has faced criticism for its ecological validity the degree to which findings generalize to natural settings [3]. To address this, researchers proposed Cognitive Ethology, a paradigm advocating for the direct observation of behavior in natural environments before transitioning to controlled lab studies. This shift mitigates discrepancies between artificial and real-world contexts. Lab-based cognitive procedures further demand rigorous methodologies, including focus groups, cognitive interviews (employing direct and indirect questioning), and assessments of comprehension, recall, and decision-making processes [4, 5]. Notably, many of these challenges align with the capabilities offered by web-based experimentation.

Comparative studies have extensively evaluated the reliability of traditional lab-based versus web-based methods, scrutinizing potential limitations of the latter [6, 7, 8, 9]. A large-scale analysis, demonstrated that web-based methods retain validity despite concerns about non-serious or repetitive respondents, yielding results consistent with traditional approaches. Such comparisons underscore the importance of data quality, prompting ongoing investigations into the reliability and validity of web-based cognitive and perceptual measures [10, 11, 12].

Web-based cognitive tests such as the Cambridge Face Memory Test (CFMT), Reading the Mind in the Eyes (RMIE), Verbal Paired Associates Memory (VPAM), and Forward Digit Span (FDS) have been developed to assess memory, attention, and emotional perception [13, 14, 15]. Researchers conducted a landmark study evaluating data quality in web-based cognitive and perceptual testing, focusing on three metrics: mean performance, performance variance, and

Cite (APA): Karaguzel, B., Erdogdu, M. (2025). Software Solutions For Web-Based Experiments: A Comprehensive Review For Cognitive and Performance Research. *Journal of Emerging Computer Technologies*, 5(1), 70-83. Doi: 10.57020/ject.1645604

internal reliability [16]. Their findings revealed no systematic differences between web-collected and lab-collected data, challenging the assumption that uncompensated, unsupervised participants compromise data quality, even in cognitively demanding tasks.

The evolution of web-based experimentation has introduced increasingly sophisticated design practices, enhancing the precision and scope of measurable outcomes. These advancements are reshaping computational neuroscience, elucidating links between cognition, perception, and behavior [17]. For example, the Stroop test, a classic measure of cognitive interference, has been adapted into modern paradigms like the visual probe task, which successfully quantified attention biases toward emotional faces in socially anxious individuals [18]. While debates persist about the limitations of self-report diagnostics in clinical psychology, computational methods offer robust alternatives. Studies demonstrated that large-scale, web-based cognitive testing improves diagnostic accuracy by transcending the constraints of traditional symptom categories [19]. Similarly, gamified smartphone apps and platforms like Amazon Mechanical Turk have enabled large-sample data collection, refining the neurocognitive understanding of psychiatric symptoms [20, 21, 22].

In sports psychology, web-based experiments have proven invaluable for investigating attention, perception, and motor performance. Studies employing tools like the Vienna Determination Test and Visual Pursuit Test have highlighted differences in reaction speeds between athletes and non-athletes [23]. Research on volleyball players further revealed superior motor command skills in athletes when countering distracting stimuli [24]. Such findings underscore the potential of web-based methods to advance both theoretical knowledge and training methodologies [25, 26, 27].

The primary advantage of web-based experiments lies in their scalability, enabling researchers to recruit large, diverse participant pools without extensive logistical demands [28]. It is argued that studies based on self-report give inaccurate results due to biases and pose serious reliability problems [29]. Studies shows that there is no found statistical difference between online and in-person testing [30, 31].

While web-based testing offers significant advantages in scalability and accessibility, researchers must carefully address data quality challenges to achieve laboratory-grade reliability. A recent large-scale comparison evaluated data quality across 196 MTurk participants, 300 Prolific participants, and 255 university students, revealing that participant pool characteristics significantly influenced outcomes more than testing modality itself [32]. Critically, the study demonstrated only marginal reductions in data quality for web-based testing when using rigorously vetted platforms, with Prolific participants performing comparably to laboratory-tested students, while MTurk samples showed greater variability. These findings underscore the importance of strategic participant recruitment in online research, suggesting that platform selection and screening protocols can mitigate traditional concerns about web-based data quality. This evidence reinforces our comparative analysis of experiment software by highlighting how methodological rigor, in both tool selection and participant management, can bridge the gap between online and laboratory research standards.)

This review synthesizes current knowledge on the most widely used software packages for web-based experimentation, comparing their precision, timing accuracy, and usability. By providing a comparative analysis and practical guidance for novice users, this paper aims to serve as a foundational resource for researchers designing experiments in cognitive and perceptual domains.

II. Methodology

A. Common Variables Used in Web-Based Experimentation

A critical consideration in the design of web-based experiments is the interplay of precision, accuracy, and timing. These metrics collectively determine the quality of data generated by experimental software and the validity of subsequent findings [33, 34].

In Table I, comparative table of common variables in web-based experimentation is shown.

TABLE I. COMPARATIVE TABLE OF COMMON VARIABLES IN WEB-BASED EXPERIMENTATION

Metric	Definition	Measurement Methods	Example Benchmarks	Notes
Precision	Reproducibility of measurements	Test-Retest Variability, Inter-Trial Consistency, Hardware & Software Diagnostics	PsychoPy ±0.5 ms visual stimuli; OpenSesame ±2.1 ms (Bridgers et al., 2020)	High ICC (>0.9) indicates reliability
Accuracy	Deviation from expected values	Calibration protocols, Benchmark comparisons, Error detection	Keyboard response within 1-2 ms; Display lag ≤ 16.7 ms per frame	Automatic trial exclusion if limits exceeded
Timing	Precision of stimulus delivery intervals	Stimulus delivery tests, Jitter, Drift	<1 ms variation in flashes; ≤ 2 ms jitter in RSVP; ≤ 50 ms drift/hour	Timing errors affect reaction times

** This table summarizes critical metrics used to evaluate web-based experimental software: precision, accuracy, and timing. It defines each metric, outlines measurement approaches, and provides benchmark examples from leading platforms. Highlighting these variables helps researchers assess software performance, ensuring data quality and experimental validity in web-based studies.*

1. Precision

Precision reflects the reproducibility of measurements under identical conditions, critical for establishing reliability. In experiment software, precision is operationally defined and measured through:

- Test-Retest Variability: Calculated as the standard deviation of timing differences across repeated stimulus presentations (e.g., ±0.5 ms in PsychoPy vs. ±2.1 ms in OpenSesame for visual stimuli) [35].

- Inter-Trial Consistency: Assessed via intraclass correlation coefficients (ICCs) for response times in repetitive tasks.
- Measurement Tools:
  - Hardware synchronization
  - Software diagnostics

Practical Example: A high-precision auditory task would show <1% variance in tone-onset times across 100 trials [34].

2. Accuracy

Accuracy measures deviations from expected values, validated through:

- Calibration Protocols:
  - Visual: Display calibration ensures a 100 ms stimulus appears for exactly 100 ms, verified with high-speed cameras.
  - Auditory: A 1000 Hz tone played at 60 dB is confirmed using professional sound level meters.
- Benchmark Comparisons:
  - A keyboard response to a visual cue should register within 1-2 ms of the actual press time when tested with mechanical input simulators.
  - Display lag should not exceed 1 frame refresh cycle
- Error Detection:
  - Software logs should flag stimuli that render 5+ ms late as warnings.
  - Automatic exclusion of trials where response times exceed physiological limits.

Practical Example: If a participant reacts to a stimulus in 250 ms, accurate software will record this as 250±2 ms, whereas inaccurate systems might report 230 ms or 270 ms due to improper latency compensation [34].

3. Timing

Timing precision is quantified through:

- Stimulus Delivery Tests:
  - A sequence of flashes should appear at exact intervals.
  - Audio-visual sync should maintain alignment (e.g., a beep and flash appear simultaneously).
- Critical Metrics:
  - Jitter: In a rapid serial visual presentation task, image transitions should vary across trials.
  - Drift: Over a 1-hour session, cumulative timing errors should not exceed 50 ms total.
- Real-World Impact:
  - A delay in Stroop task stimuli can inflate measured reaction times.
  - In attentional blink paradigms, jitter reduces detectable effects.

Practical Example: Software with poor timing might display a 50 ms stimulus for 45-55 ms, while precise systems guarantee 50±0.5 ms durations [35].

In Table II, comparisons for software packages were shown.

TABLE II. COMPARISON BY KEY FEATURES

Software	Timing Precision (ms)	Auditory Stimulus Accuracy	Visual Stimulus Accuracy	Ease of Use	Platform Compatibility
PsychoPy	< 1	High	High	Medium	Windows, macOS, Ubuntu
Psychtoolbox	< 1	High	High	Low (Matlab-based)	Windows, macOS, Linux
E-Prime	< 1	High	High	High	Windows only
NBS Presentation	< 1	High	High	Medium	Windows only
OpenSesame	~10	Low	Medium	High	Windows, macOS, Ubuntu
Expyriment	~5	Medium	Medium	Medium	Windows, macOS, Linux
Inquisit	~10	Medium	Medium	High	Windows, macOS
SuperLab	~10	Medium	Medium	High	Windows, macOS
jsPsych	~10	Medium	Medium	Medium	Web-based (All platforms)
Lab.js	~10	Medium	Medium	High	Web-based (All platforms)
Gorilla	~10	Medium	Medium	High	Web-based (All platforms)
Vision Egg	~5	N/A	High	Medium	Windows, macOS, Linux (Python required)

**This table provides a side-by-side comparison of major experimental software packages based on critical technical and practical factors. These include timing precision (in milliseconds), accuracy in presenting auditory and visual stimuli, ease of use, and platform compatibility. The table highlights that software such as PsychoPy, Psychtoolbox, E-Prime, and Presentation achieve sub-millisecond accuracy, making them suitable for high-precision lab experiments. Meanwhile, web-based tools like jsPsych, Lab.js, and Gorilla trade off a slight reduction in timing precision for increased accessibility and platform flexibility.*

B. Common Software Packages

A critical challenge in online experimentation is ensuring participant compliance and data integrity. Unlike lab-based studies where conditions are controlled and monitored by researchers, online experiments rely on participants completing tasks independently using their own hardware and software systems [36, 37, 38]. To address this, experimental software packages incorporate rigorous precision, accuracy, and timing controls to minimize variability and maintain experimental validity. In Table III, a summary for software packages were shown.

TABLE III. SUMMARY OF STRENGTHS AND USE CASES

Software	**Strengths**	**Best For**
PsychoPy	Open-source, high timing accuracy, flexible coding	Lab and online experiments requiring precision
Psychtoolbox	Most precise timing, widely validated	Expert users with Matlab experience
E-Prime	Widely used in cognitive research, user-friendly	Psych researchers needing precision without coding
NBS Presentation	Excellent for auditory experiments	Complex multimodal experiments
OpenSesame	Easy to use, open-source	Students, educators, small-scale research
Expyriment	Python-based, good for reaction time studies	Academic researchers familiar with Python
Inquisit	Good range of templates, fast prototyping	Implicit testing and bias research
SuperLab	Educational use, simple interface	Teaching classic paradigms
jsPsych	Customizable, works in browser	Web-based experiments with visual/audio stimuli
Lab.js	Drag-and-drop UI, open-source	Web studies with complex branching
Gorilla	Participant recruitment, payment management	Scalable online studies
Vision Egg	Visual psychophysics, OpenGL-based	High-end visual stimulus control in Python

**This summary table presents a qualitative overview of the strengths and ideal use cases for each software package. It offers readers quick guidance on which software best fits their research needs whether it's precise lab-based studies, flexible web-based data collection, or beginner-friendly interfaces. For example, Gorilla and Lab.js are well-suited for large-scale online studies, while Psychtoolbox is ideal for expert users conducting high-precision psychophysics experiments.*

4. PsychoPy

PsychoPy is an open-source software package designed for creating experiments in psychology, cognitive science, and neuroscience research. Built on Python, it offers both programming flexibility for complex designs and a user-friendly graphical interface for simpler implementations [39].

Key features include:

- Stimulus Presentation: Precise millisecond-level timing for visual (images, text) and auditory stimuli, critical for reaction-time and perceptual studies [40].
- Accessibility: A drag-and-drop interface enables researchers without programming expertise to design experiments (e.g., surveys, cognitive tasks).
- Data Handling: Automated data collection and export in CSV or other formats for analysis [41].

As a free, open-source tool with an active user community, PsychoPy provides extensive documentation, sample experiments, and support forums. Further details are available at: https://www.psychopy.org/. An image of the software package taken from its website is shown in Figure I.

2. PsychToolbox

PsychToolbox is an open-source MATLAB/Octave-based package for psychophysics, neuroscience, and cognitive psychology research. It specializes in high-precision stimulus control and data acquisition [42].

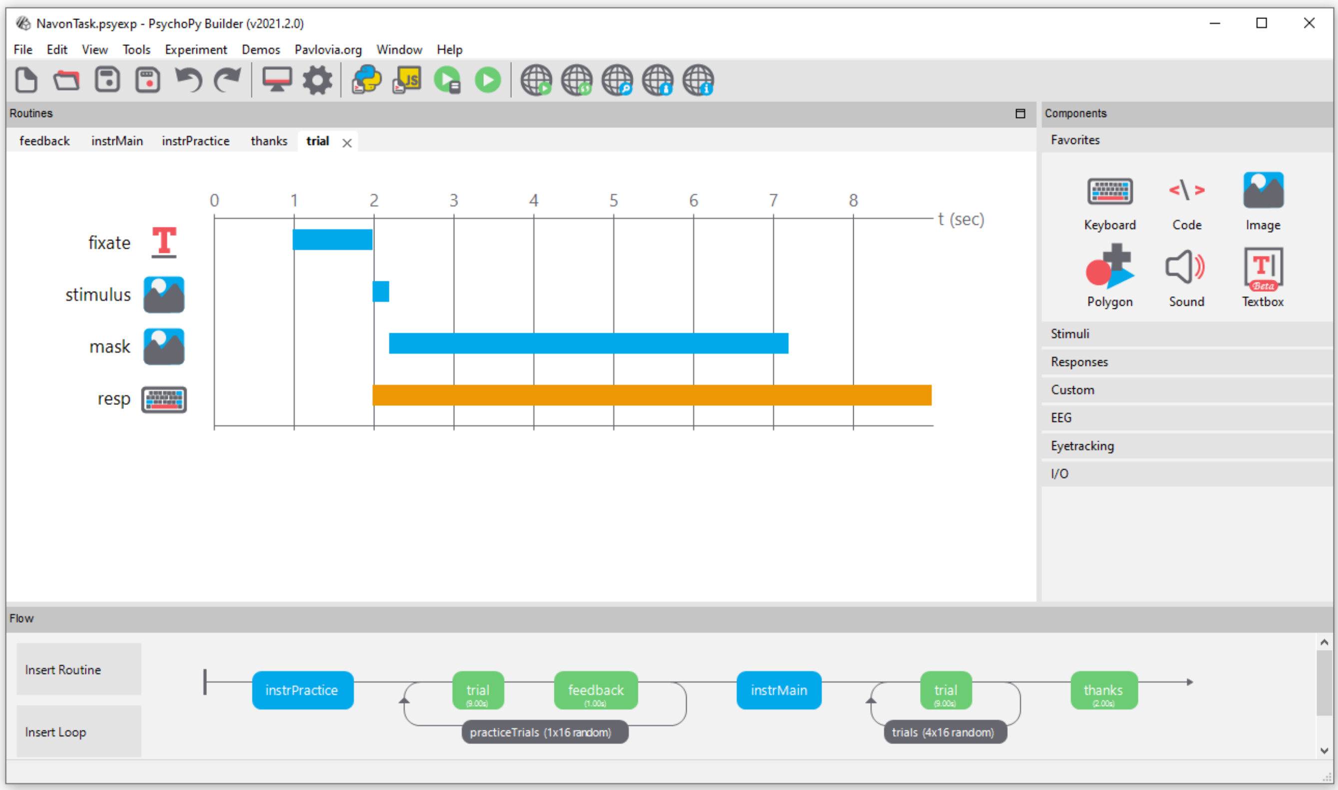

FIGURE I. PSYCHOPY BUILDER SCREEN

Notable capabilities:

- Precision Timing: Millisecond-accurate synchronization of visual/auditory stimuli.
- Versatility: Supports diverse paradigms, from image processing to neural coding.
- Integration: Compatible with MATLAB's analytical tools for data processing.

PsychToolbox is ideal for researchers proficient in MATLAB. Documentation and resources can be accessed at: http://psychtoolbox.org/. A visualization of the software package from a work [43] is shown in Figure II.

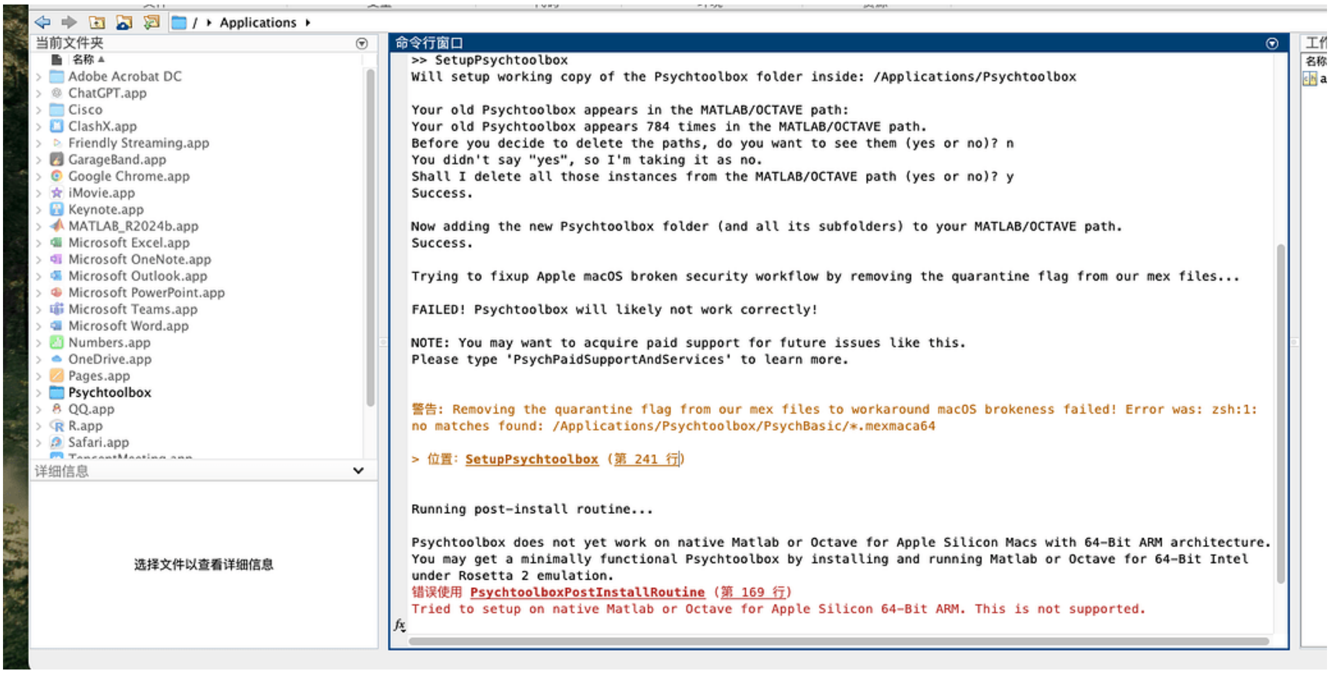

FIGURE II. PSYCHTOOLBOX SETUP SCREEN

3. *NBS Presentation*

NBS Presentation is a specialized platform for neuroscience and cognitive psychology experiments, offering robust tools for stimulus delivery, response recording, and data analysis.

Advantages include:

- User-Friendly Design: Drag-and-drop interface for creating experiments without programming.
- Multimodal Stimuli: Support for text, images, videos, and audio with precise timing.
- Data Management: Automated data export to statistical software for analysis.

Suitable for complex experimental designs, NBS Presentation balances accessibility with advanced customization. Learn more at: https://www.neurobs.com/. An image of the software package taken from its website is shown in Figure III.

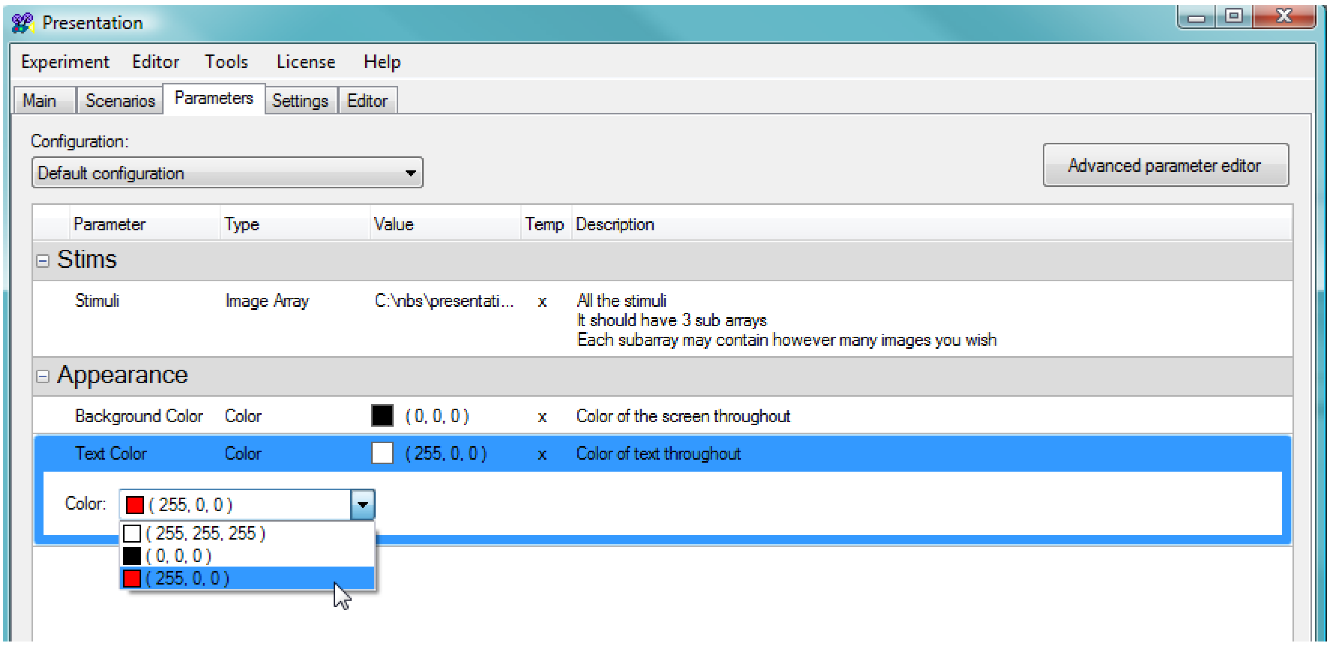

FIGURE III. NBS PRESENTATION SCREEN

4. E- Prime

E-Prime is a widely used software package for designing, implementing, and analyzing psychological experiments, particularly in experimental psychology, cognitive science, neuroscience, and education research. It enables precise control over visual, auditory, and tactile stimulus presentation, as well as response timing [44].

Key Features:

- User-Friendly Interface: Drag-and-drop functionality allows experiment design without programming expertise.
- Precision Timing: Ensures accurate stimulus presentation and reaction time measurement.
- Customization: Supports text, images, videos, and audio, with pre-made templates for common paradigms.
- Data Analysis: Exports response data for statistical processing.

E-Prime's robust synchronization capabilities and large user community make it a versatile tool for experiments of varying complexity. For details, visit: https://pstnet.com/products/e-prime/. An image of the software package taken from its website is shown in Figure IV.

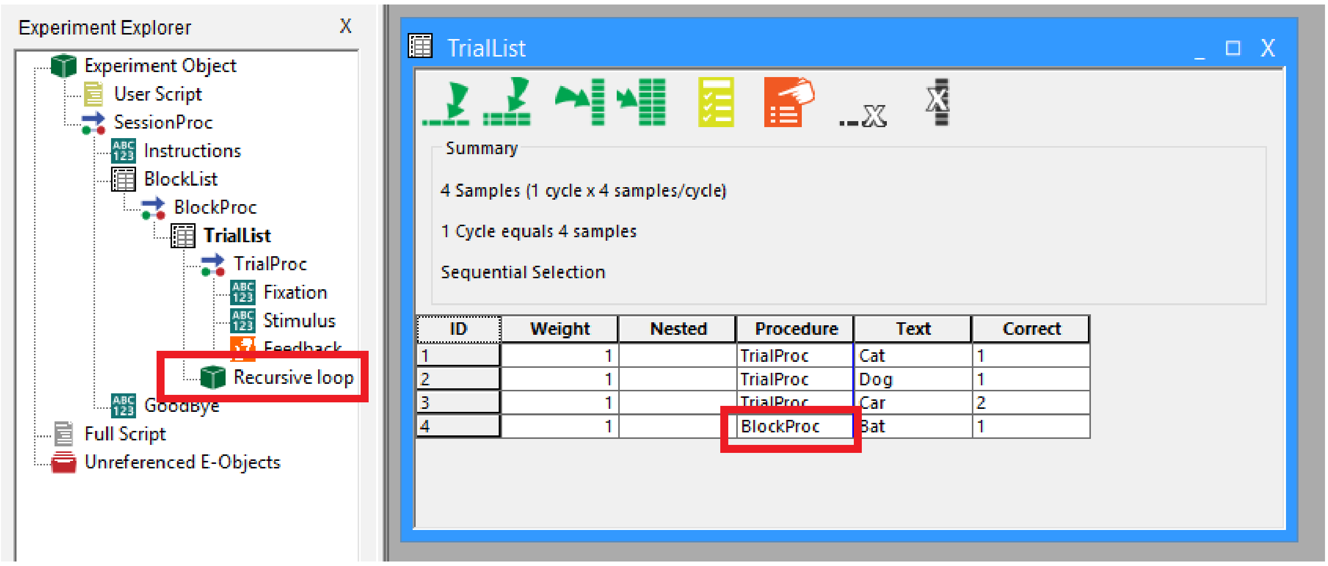

FIGURE IV. E-PRIME DESIGN SCREEN

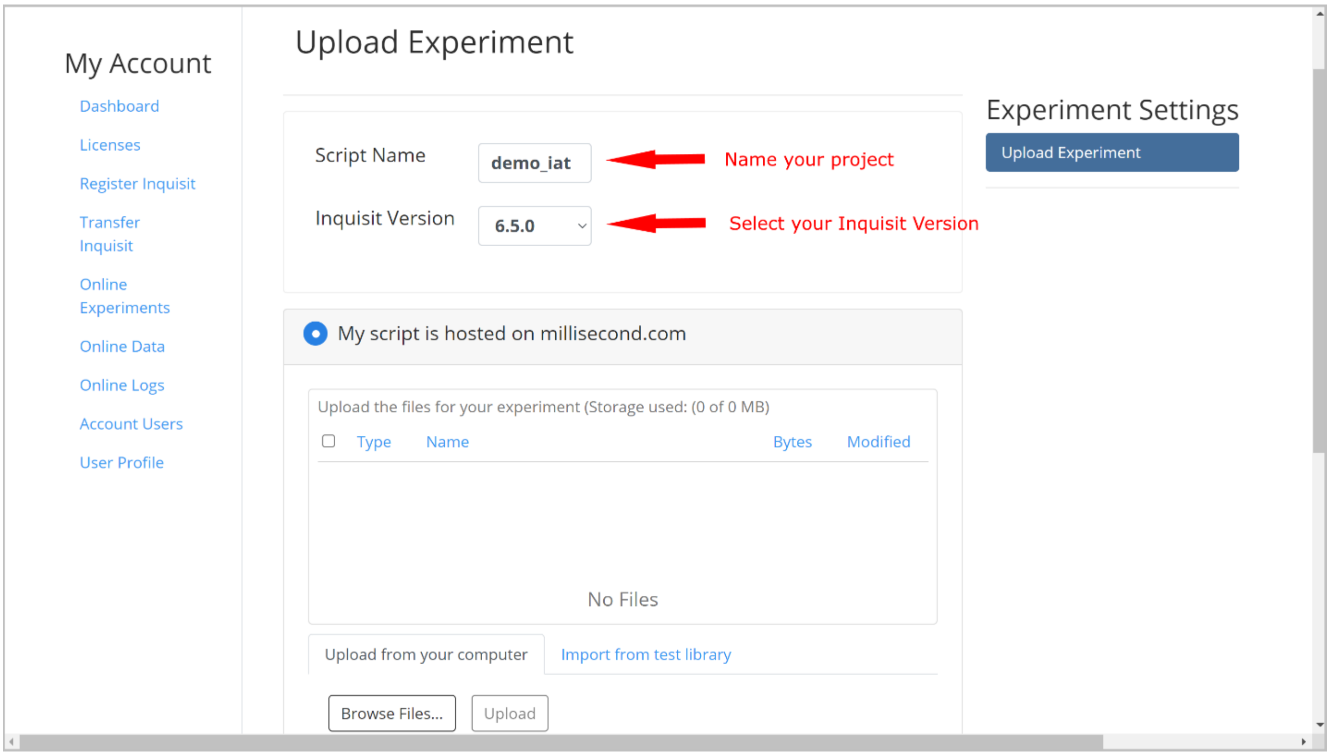

FIGURE V. INQUISIT SETUP SCREEN

5. INQUISIT

Developed by Millisecond Software, INQUISIT specializes in implicit attitude tests, priming experiments, and attention bias studies. It is favored by cognitive scientists and neuroscientists for its ease of use [45].

Key Features:

- Accessibility: Intuitive drag-and-drop interface requires no coding.
- Stimulus Control: Precise timing for visual, auditory, and tactile stimuli.
- Security: Prioritizes data privacy and offers pre-built templates.

Ideal for rapid experiment deployment, INQUISIT streamlines data collection and analysis. Learn more: https://www.millisecond.com/. An image of the software package taken from its website is shown in Figure V.

6. SuperLab

SuperLab is a cross-platform software package for designing and analyzing psychological experiments, widely used in education and research [46, 47].

Key Features:

- Multimodal Stimuli: Supports text, images, audio, and video with millisecond timing.
- Educational Utility: Facilitates replication of classic cognitive psychology studies.
- Data Export: Enables statistical analysis in external tools [48].

SuperLab's synchronization of audiovisual stimuli makes it suitable for diverse experimental paradigms. Details: https://cedrus.com/superlab/index.htm.

7. Expyriment

Expyriment is a Python-based open-source tool for reaction time experiments and stimulus presentation, catering to psychologists and neuroscientists [49].

Key Features:

- Python Integration: Leverages Python's flexibility for custom designs.
- Cross-Platform: Runs on Windows, macOS, and Linux.
- Precision: Controls timing for visual/auditory stimuli and response collection.

Expyriment balances accessibility for beginners with advanced features for programmers. Visit: https://expyriment.org/. An image of the software package taken from its website is shown in Figure VI.

FIGURE VI. EXPYRIMENT PROGRAMMING SAMPLE

8. OpenSesame

OpenSesame is an open-source platform for designing behavioral experiments, emphasizing transparency and community collaboration [50].

Key Features:

- Open-Source: Customizable code for tailored experiments.
- Drag-and-Drop Interface: Simplifies design for non-programmers.
- Multi-Platform Support: Compatible with major operating systems.

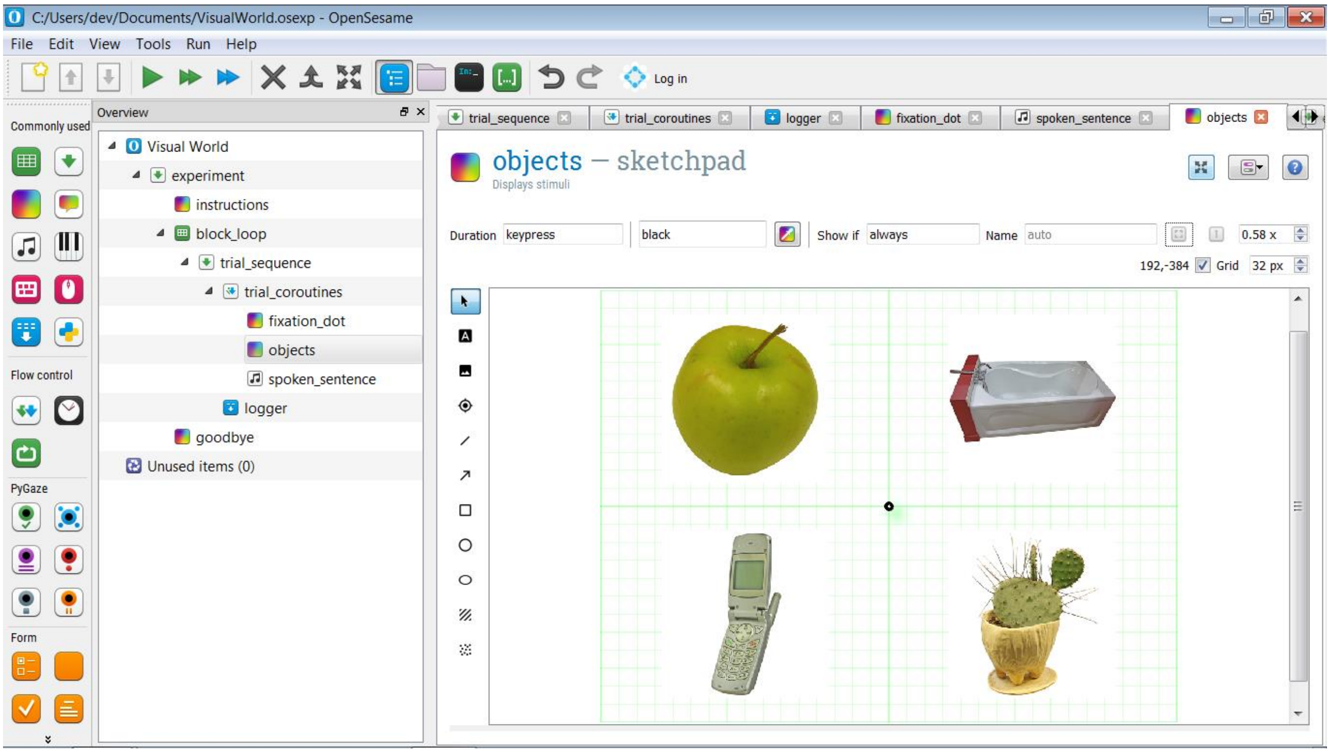

FIGURE VII. OPENSESAME VISUAL WORLD

Its active user community and adaptability make it a robust choice for open science initiatives. Explore: http://osdoc.cogsci.nl/. An image of the software package taken from its website is shown in Figure VII.

9. Gorilla

Gorilla is a web-based platform for creating and deploying online experiments, designed to modernize data collection in psychology and social sciences [51].

Key Features:

- Device-Agnostic: Runs on computers, tablets, and smartphones.
- Participant Management: Simplifies recruitment and compensation.
- Data Security: Encrypted storage and export to analysis tools [52].

Gorilla's scalability and interactive task support address the growing demand for remote research. More info: https://gorilla.sc/. An image of the software package taken from its website is shown in Figure VIII.

10. jsPsych

jsPsych is a JavaScript library designed for creating and deploying online behavioral experiments, particularly in psychology, cognitive science, and neuroscience research [53]. Its browser-based framework enables cross-platform compatibility, allowing participants to complete experiments on any device.

Key Features:

- Modular Design: Supports flexible experiment construction through reusable plugins (e.g., for visual/auditory stimuli, surveys, and reaction time tasks).
- Precision Timing: Ensures accurate stimulus presentation and response collection [54].
- Data Handling: Exports results in multiple formats (CSV, JSON) for analysis.
- Community Support: Active user forums and extensive documentation facilitate troubleshooting.

jsPsych is ideal for researchers seeking a balance between customization (via JavaScript) and accessibility. For details, visit: https://www.jspsych.org/7.3/.

11. Lab.js

Lab.js is an open-source tool for building online experiments in cognitive psychology, behavioral economics, and neuroscience. It combines a drag-and-drop interface with advanced scripting capabilities [55].

Key Features:

- Flexible Design: Customizable modules for stimuli, surveys, and complex experimental flows.
- Programming Integration: JavaScript-based for sophisticated experimental logic (e.g., adaptive designs).
- Data Export: Supports integration with statistical analysis tools.

Lab.js is particularly suited for researchers requiring both ease of use and computational power. Learn more: https://lab.js.org/. An image of the software package taken from its website is shown in Figure IX.

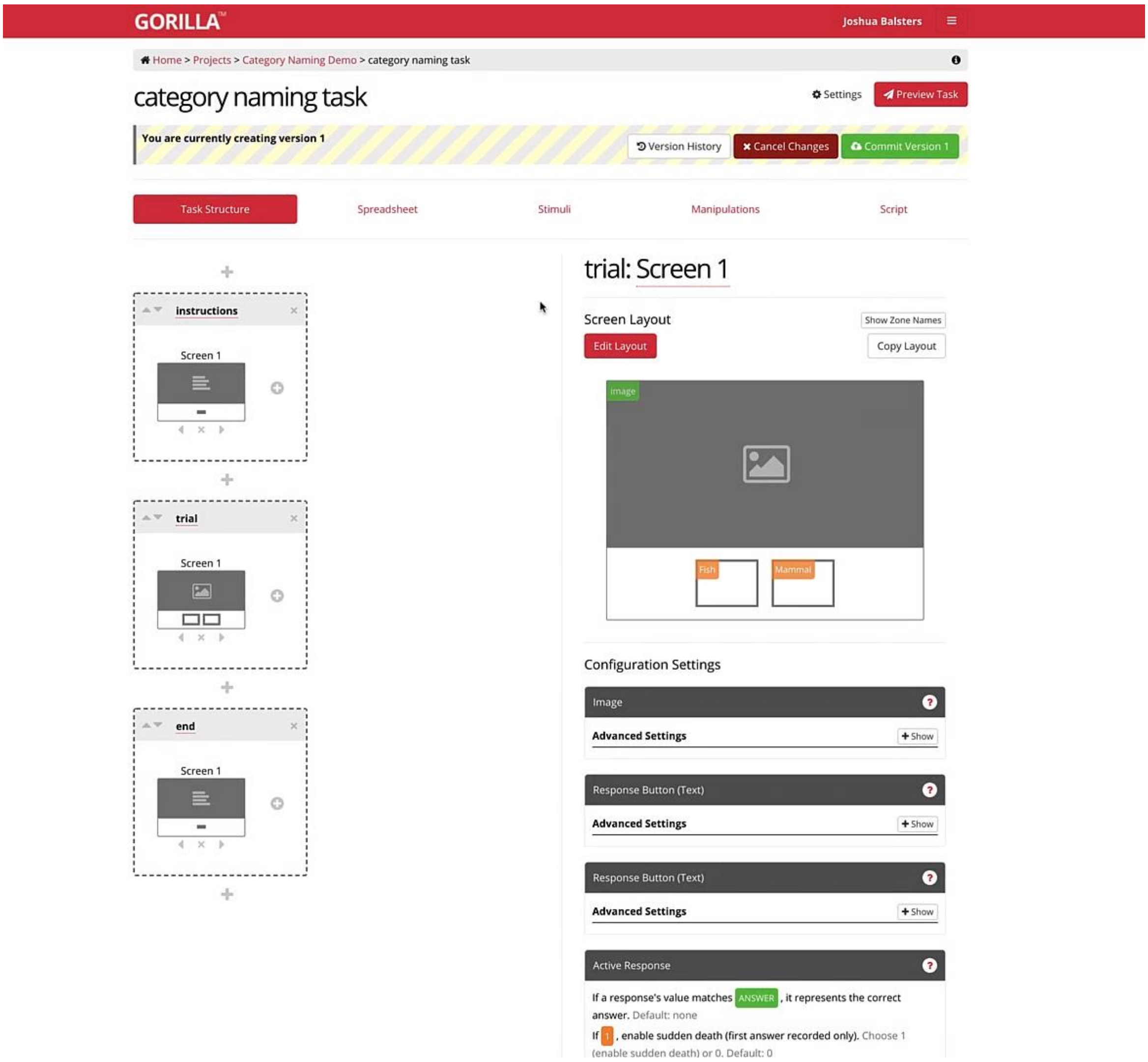

FIGURE VIII. GORILLA TRIAL SCREEN

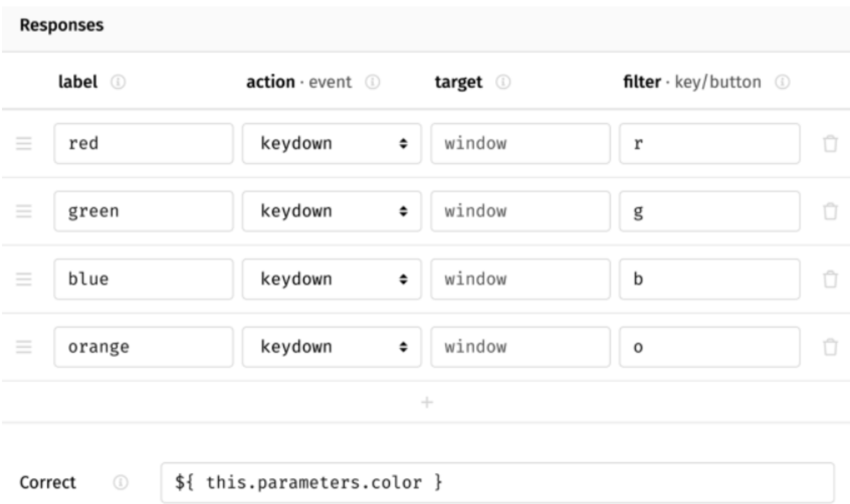

FIGURE IX. LAB.JS RESPONSES SCREEN

12. The Vision Egg

The Vision Egg is a Python-based platform for generating high-precision visual stimuli, widely used in psychophysics and neuroscience [56].

Key Features:

- Stimulus Control: OpenGL-powered rendering for dynamic visuals (e.g., moving shapes, textures).
- Timing Accuracy: Millisecond-level precision for reaction time studies.

- Customization: Modular Python scripts support tailored experimental designs.

This tool is invaluable for vision research, offering granular control over stimulus parameters. Documentation: https://visionegg.org/. An image of the software package taken from its website is shown in Figure X.

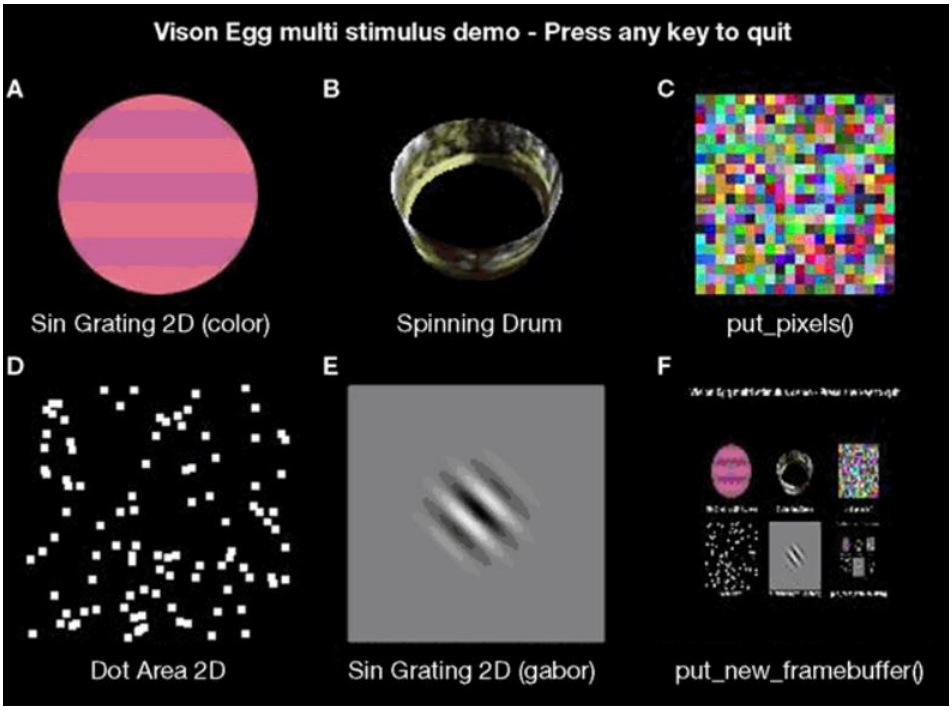

FIGURE X. VISION EGG DEMO SCREEN

In *Figure XI*, timing precisions for software packages were shown.

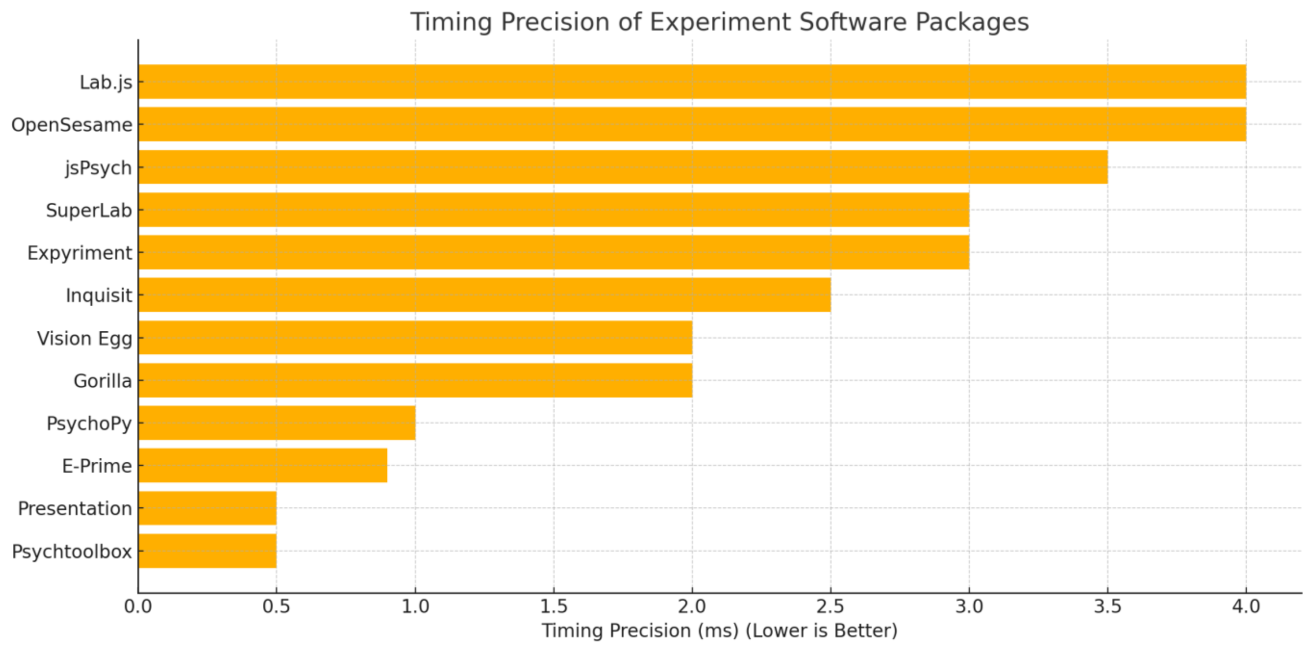

FIGURE XI. TIMING PRECISION OF EXPERIMENTAL SOFTWARE PACKAGES

** This figure provides a side-by-side comparison of major experimental software packages based on critical technical and practical factors. These include timing precision (in milliseconds), accuracy in presenting auditory and visual stimuli, ease of use, and platform compatibility. The table highlights that software such as PsychoPy, Psychtoolbox, E-Prime, and Presentation achieve sub-millisecond accuracy, making them suitable for high-precision lab experiments. Meanwhile, web-based tools like jsPsych, Lab.js, and Gorilla trade off a slight reduction in timing precision for increased accessibility and platform flexibility.*

III. CONCLUSION

This review has evaluated prominent software packages for designing web- and lab-based experiments, focusing on their precision, accuracy, and timing performance in presenting stimuli and recording responses. A critical benchmark study by Bridgers et al. [35] compared ten major platforms (PsychoPy, E-Prime, NBS Presentation, PsychToolbox, OpenSesame, Expyriment, Gorilla, jsPsych, Lab.js, and Testable) across Windows, macOS, and Ubuntu systems using the Black Box Toolkit (www.blackboxtoolkit.com) [57]. Key findings revealed that lab-based software particularly PsychToolbox, PsychoPy, NBS Presentation, and E-Prime achieved sub-millisecond precision in visual, auditory, and response timing. OpenSesame exhibited lower precision, especially for auditory stimuli. Across operating systems, Ubuntu marginally outperformed Windows, while macOS demonstrated the poorest performance for visual stimuli.

Web-based platforms, though inherently more variable, approached lab-level precision in some cases. PsychoPy and Gorilla delivered near-millisecond accuracy across multiple browser/OS combinations, with PsychoPy achieving exceptional response-time precision (3.5 ms). These results underscore that while web-based methods introduce minor compromises in timing, select platforms can yield data comparable to controlled lab environments.

Further analyses highlight practical considerations for researchers:

- PsychToolbox remains the most widely cited and accessible option for novice users [58].
- PsychoPy balances advanced functionality with Python's flexibility, ideal for researchers avoiding MATLAB.
- E-Prime and DirectRT excel in usability for standard paradigms, with E-Prime's self-contained design reducing reliance on external programming [28].

For experiments demanding maximal precision, lab-based tools (PsychToolbox, PsychoPy, NBS Presentation, E-Prime) are optimal. However, web-based packages like Gorilla and jsPsych provide scalable alternatives without sacrificing significant accuracy. Researchers should select software based on their specific needs:

- Stimulus complexity (e.g., Vision Egg for 3D visuals),
- Technical expertise (e.g., Lab.js for JavaScript proficiency),
- Experimental setting (lab vs. online).

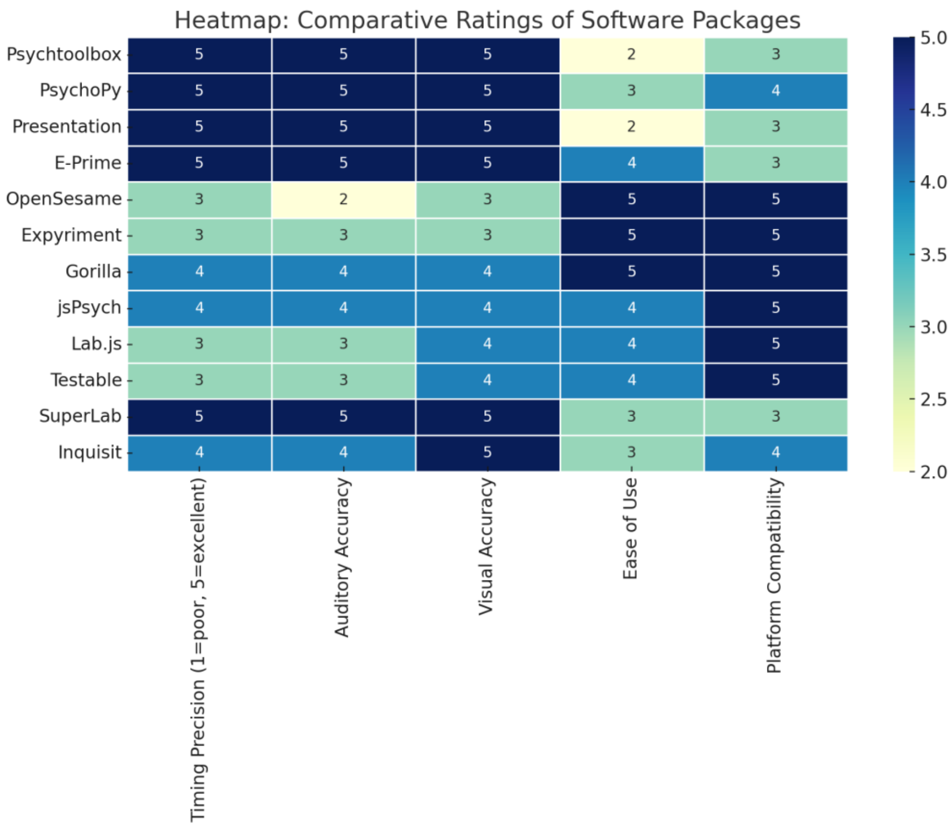

FIGURE XII. HEATMAP OF COMPARATIVE RATINGS

**This heatmap provides a visual overview of how each software package performs across several critical dimensions: timing precision, stimulus presentation accuracy, ease of use, flexibility, and platform compatibility. Darker shades indicate higher performance or better support in the respective category. This visualization helps readers quickly identify which tools offer the best combination of precision and usability. For instance, Psychtoolbox and PsychoPy rank highest in timing and accuracy, while Gorilla and jsPsych offer the best cross-platform and browser-based flexibility.*

As the field evolves, ongoing benchmarking will be essential to validate emerging tools. This synthesis serves as a guide for researchers navigating the trade-offs between precision, accessibility, and methodological flexibility in experimental design.

This review systematically evaluated leading software packages for designing web- and lab-based experiments, with a focus on their precision (variability in measurements), accuracy (deviation from true values), and timing performance in stimulus presentation and response recording. Empirical data from benchmark studies reveal critical differences across platforms, enabling evidence-based recommendations for researchers. In *Figure XII*, heatmap for comparative rating were shown and *Table IV* shows the key findings from benchmark studies.

TABLE IV. KEY FINDINGS FROM BENCHMARK STUDIES

Aspect	Lab-Based Software	Web-Based Software	Notes
Timing Precision	≤ 0.5 ms mean deviation (PsychToolbox, PsychoPy, NBS Presentation, E-Prime)	1.2–3 ms (PsychoPy, Gorilla); 3–10 ms (jsPsych/Lab.js)	Ubuntu > Windows > macOS for precision
Auditory Stimuli	High precision, low variability	Higher variability; Gorilla ~2–3 ms latency	OpenSesame shows 2–5 ms delays
Response Time Accuracy	≤ 1 ms resolution (keyboard/button responses)	3–15 ms; minimized to 3–5 ms with optimizations	Browser event-loop limits accuracy
Operating Systems	Ubuntu marginally best; macOS poorest	Browser/OS variability impacts timing	Hardware and software configurations vary
Best Use Cases	High-precision psychophysics, clinical studies	Large-scale behavioral and online studies	Selection based on complexity and scale

** The table summarizes benchmark results comparing lab-based and web-based experimental software packages across key performance metrics. It highlights differences in timing precision, auditory stimulus handling, response time accuracy, and operating system effects. These findings provide practical guidance for selecting software based on experimental requirements, balancing precision, accessibility, and platform variability.*

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

None.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

AKNOWLEDGEMENNT

Not applicable.

DATA AVAILABILITY

The data supporting the findings of this study are available upon request from the authors.

FUNDING

No funding was received for conducting this study.

AUTHORS' CONTRIBUTIONS

All authors contributed equally to this work.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

ETHICAL STATEMENT

In this article, the principles of scientific research and publication ethics were followed. This study did not involve human or animal subjects and did not require additional ethics committee approval.

REFERENCES

[1] Anwyl-Irvine, A. L., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2020a). Online Timing Accuracy and Precision: A comparison of platforms, browsers, and participant's devices.

[2] Bohannon, J. (2016). Mechanical Turk upends social sciences. *Science, 352*(6291), 1263–1264. doi: 10.1126/science.352.6291.1263.

[3] Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology, 99*(3), 317-443. doi: 10.1348/000712607X251243

[4] Willis, G. B. (1997). The use of the psychological laboratory to study sensitive survey topics. In: L. Harrison, & A. Hughes (Eds), *The Validity of Self-reported Drug Use: Improving the Accuracy of Survey Estimates* (pp. 416–438). Rockville, MD: National Institute on Drug Abuse.

[5] Jobe, J. B. (2003). Cognitive psychology and self-reports: Models and methods. *Quality of life research, 12*(3), 219-227.doi: 10.1023/a:1023279029852

[6] Krantz, J. H., Ballard, J., & Scher, J. (1997). Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness. *Behavior Research Methods, Instruments, & Computers, 29*, 264–269.doi: 10.3758/BF03204824

[7] Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology, 90*, 125–144.doi: 10.1348/000712699161189

[8] Buchanan, T., Johnson, J. A., & Goldberg, L. (2005). Implementing a Five-Factor personality inventory for use on the Internet. European *Journal of Psychological Assessment, 21(*2), 115–127.doi: 10.1027/1015-5759.21.2.115

[9] Buchanan, T. (2007). Personality testing on the Internet: What we know, and what we do not. In A. N. Joinson, K. Y. McKenna, T. Postmes, & U.-D. Reips (Eds.), The Oxford handbook of Internet psychology.

[10] Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. *American Psychologist, 59(*2), 93–104.doi: 10.1037/0003-066X.59.2.93

[11] Birnbaum, M. H. (Ed.), (2000). Psychological experiments on the Internet. Academic Press.

[12] McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of Web-delivered experiments: Can you trust the data? *Psychological Science, 11*(6), 502–506.doi: 10.1111/1467-9280.00296

[13] Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry, 42*(2), 241–251.doi: 10.1017/S0021963001006643

[14] Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia, 44*(4), 576–585. doi: 10.1016/j.neuropsychologia.2005.07.001

[15] Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. Proceedings of the *National Academy of Sciences, 107*(11), 5238–5241. doi: 10.1073/pnas.0913053107

[16] Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review, 19*(5), 847–857. doi: 10.3758/s13423-012-0296-9

[17] Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience, 21*, 1148–1160. doi: 10.1038/s41593-018-0210-5

[18] Bantin, T., Stevens, S. & Gerlach, A. L., Hermann, C. (2016). What does the facial dot probe task tell us about attentional processes in social anxiety? A systematic review. Journal of Behavior Therapy and *Experimental Psychiatry, 50*, 40-51. doi: 10.1016/j.jbtep.2015.04.009

[19] Gillan, M. C., Kosinski, M., Whelan, R., Phelps, E. A. & Daw, N. D. (2016a). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. eLife.

[20] Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron, 69*(6), 1204-1215.doi: 10.1016/j.neuron.2011.02.027

[21] Gillan, M. C. & Daw, N. D. (2016b). Taking Psychiatry Research Online. Neuron, 91(1), 19-23.

[22] Gillan, M. C., Otto, A. R., Phelps, E. A. & Daw, N. D. (2015). Model-based learning protects against forming habits. Cognitive, Effective & *Behavioral Neuroscience, 15*, 523-536. doi: 10.3758/s13415-015-0347-6

[23] Doğan, B. (2009) Multiple-choice reaction and visual perception in female and male elite athletes. *Journal of Sports Medicine and Physical Fitness, 49*(1), 91-96.

[24] Kokubu, M., Ando, S., Kida, N., & Oda, S. (2006). Interference Effects between Saccadic and Key-Press Reaction Times of Volleyball Players and Nonathletes. *Perceptual and Motor Skills, 103*(3), 709–716. doi: 10.2466/pms.103.3.709-716

[25] Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience, 11*(9), 1109–1116. doi: 10.1038/nn.2182

[26] Fissler, P., Küster, O., Schlee, W., & Kolassa, I.-T. (2013). Novelty Interventions to Enhance Broad Cognitive Abilities and Prevent Dementia. *Progress in Brain Research, 207*, 403–434. doi: 10.1016/B978-0-444-63327-9.00017-5

[27] Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review, 120*(3), 439–471. doi: 10.1037/a0033138

[28] Stahl, C. (2006). Software for Generating Psychological Experiments. *Experimental Psychology, 53*(3), 218–232. doi: 10.1027/1618-3169.53.3.218

[29] Anvari, F., Efendić, E., Olsen, J., Arslan, R. C., Elson, M., & Schneider, I. K. (2022). Bias in Self-Reports: An Initial Elevation Phenomenon. *Social Psychological and Personality Science, 14*(6), 727-737. doi: 10.1177/19485506221129160

[30] Sauter, M., Stefani, M., & Mack, W. (2022). Equal Quality for Online and Lab Data: A Direct Comparison from Two Dual-Task Paradigms. *Open Psychology, 4(*1), 47–59.

[31] Gagné, N., & Franzen, L. (2023). How to Run Behavioural Experiments Online: Best Practice Suggestions for Cognitive Psychology and Neuroscience. *Swiss Psychology Open, 3*(1), 1–21. doi: 10.5334/spo.34

[32] Uittenhove, K., Jeanneret, S., & Vergauwe, E. (2023). From Lab-Testing to Web-Testing in Cognitive Research: Who You Test is More Important than how You Test. Journal of Cognition, 6(1): 13, 1–17. doi: 10.5334/joc.259

[33] Anwyl-Irvine, A. L., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. Behavior Research Methods, 53, 1407–1425. doi: 10.3758/s13428-020-01501-5

[34] Plant R. R. (2016). A reminder on millisecond timing accuracy and potential replication failure in computer-based psychology experiments: an open letter. *Behavior Research Methods 48*(1), 408–411.doi: 10.3758/s13428-015-0577-0

[35] Bridges, D., Pitiot, A., MacAskill, M. R. & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. PeerJ, 8.

[36] Dandurand, F., Shultz, T. R. & Onishi, K. H. (2008). Comparing online and lab methods in a problem solving experiment. *Behavior Research Methods, 40*, 428–434.doi: 10.3758/brm.40.2.428

[37] Erdoğdu, M., Artuner, H., Demirbaş, H., Aytaç, G. & Karasırt D. (2022). Sporcu Sağlığında Güncel Yaklaşım: Bilgisayar Tabanlı Psikolojik Ölçümler. *Türkiye Sağlık Enstitüleri Başkanlığı Dergisi, 5*(3), 43-55. doi: 10.54537/tusebdergisi.1173181

[38] Erdoğdu, M., Aytaç, G., Deliceoğlu, G., (2023). Bilgisayar Temelli Uygulamalar ile Sporcularda Dikkat ve Alt Bileşenlerinin Tespit Edilmesi: Bir Labaratuvar Çalışması. *Bilgi Teknolojileri ve İletişim Dergisi, 1*(1), 117-146.

[39] Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*(10), 1-8. doi: 10.3389/neuro.11.010.2008

[40] Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*, 8–13. doi: 10.1016/j.jneumeth.2006.11.017

[41] Garaizar, P. & Vadillo, M. A. (2014). Accuracy and Precision of Visual Stimulus Timing in PsychoPy: No Timing Errors in Standard Usage. PLoS ONE, 9(11), e112033.

[42] Lin, Z., Yang, Z., Feng, C. & Zhang, Y. (2022). PsyBuilder: An Open-Source, Cross-Platform Graphical Experiment Builder for Psychtoolbox With Built-In Performance Optimization. *Association for Psychological Science, 5*(1), 1-20. doi: 10.1177/25152459211070573

[43] Lorca, M. H. & Capilla, A. (2018). Psychtoolbox. A Brief Guide To Start Programming Experiments In Psychology. UAM Ediciones.

[44] Hairston, W. D. & Maldjian, J. A. (2009). An adaptive staircase procedure for the E-Prime programming environment. *Computer Methods and Programs In Biomedicine, 93*, 104-108. doi: 10.1016/j.cmpb.2008.08.003

[45] Clercq, A., Grombez, G., Buysse, A. & Roeyers, H. (2003). A simple and sensitive method to measure timing accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 109-115. doi: 10.3758/BF03195502

[46] Ragozzine, F. (2002). SuperLab LT: Evaluation and Uses in Teaching Experimental Psychology. *Teaching of Psychology, 29*(3), 251-254.

[47] Haxby, J. V., Parasuraman, R., Lalonde, F. & Abboud, H. (1993). SuperLab: General-purpose Macintosh software for human experimental psychology and psychological testing. *Behavior Research Methods, Instruments, & Computers, 25*(3), 400-405. doi: 10.3758/BF03204531

[48] Vaz-Rebelo, P., Otero, J., Costa, C., Morgado, J. & Ishiwa, K. (2014). Questioning About Science Texts and Reading Time Through The Software SUPERLAB. *Social and Behavioral Sciences, 159*, 620 – 624. doi: 10.1016/j.sbspro.2014.12.436

[49] Krause, F., & Lindemann, O. (2014). Expyriment: A Python library for cognitive and neuroscientific experiments. *Behavior Research Methods, 46*(2), 416–428.

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

[51] Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020b). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*, 388–407. doi: 10.3758/s13428-019-01237-x

[52] Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics, 10*(48), 1-7. doi: 10.1186/1471-2105-10-48

[53] 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

[54] Pinet, S., Zielinski, C., Mathôt, S., Dufau, S., Alario, F.-X., & Longcamp, M. (2017). Measuring sequences of keystrokes with jsPsych: Reliability of response times and interkeystroke intervals. *Behavior Research Methods, 49*(3), 1163–1176.doi: 10.3758/s13428-016-0776-3

[55] Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J. & Hilbig, B. J. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods, 54*, 556- 573. doi: 10.3758/s13428-019-01283-5

[56] Straw, A. D. (2008). Vision Egg: An open-source library for realtime visual stimulus generation. *Frontiers in Neuroinformatics, 2*(4), 1–10. doi: 10.3389/neuro.11.004.2008

[57] Pandian, V. P. S. & Suleri, S. (2020). BlackBox Toolkit: Intelligent Assistance to UI Design. Art, Computer Design ArXiv, 25–30. doi: 10.48550/arXiv.2004.01949

[58] Yoonessi, A. & Yoonessi, A. (2011). A Glance at Psychophysics Software Programs. *Basic and Clinical Neuroscience, 2*(3), 73-75.