Video Article
Methods to Test Visual Attention Online

Amanda Yung[1], Pedro Cardoso-Leite[2], Gillian Dale[3], Daphne Bavelier[2,4], C. Shawn Green[3]

[1]Center for Visual Science, University of Rochester

[2]Faculty of Psychology and Educational Sciences, University of Geneva

[3]Department of Psychology, University of Wisconsin-Madison

[4]Department of Brain and Cognitive Sciences, University of Rochester

Correspondence to: Amanda Yung at ayung@cvs.rochester.edu

Abstract

Online data collection methods have particular appeal to behavioral scientists because they offer the promise of much larger and much more representative data samples than can typically be collected on college campuses. However, before such methods can be widely adopted, a number of technological challenges must be overcome – in particular in experiments where tight control over stimulus properties is necessary. Here we present methods for collecting performance data on two tests of visual attention. Both tests require control over the visual angle of the stimuli (which in turn requires knowledge of the viewing distance, monitor size, screen resolution, *etc.*) and the timing of the stimuli (as the tests involve either briefly flashed stimuli or stimuli that move at specific rates). Data collected on these tests from over 1,700 online participants were consistent with data collected in laboratory-based versions of the exact same tests. These results suggest that with proper care, timing/stimulus size dependent tasks can be deployed in web-based settings.

Video Link

The video component of this article can be found at http://www.jove.com/video/52470/

Introduction

Over the past five years there has been a surge of interest in the use of online behavioral data collection methods. While the vast majority of publications in the domain of psychology have utilized potentially non-representative subject populations[1] (*i.e.,* primarily college undergraduates) and often reasonably small sample sizes as well (*i.e.,* typically in the range of tens of subjects), online methods offer the promise of far more diverse and larger samples. For instance, Amazon's Mechanical Turk service has been the subject of a number of recent studies, both describing the characteristics of the "worker" population and the use of this population in behavioral research[2-6].

However, one significant concern related to such methods is the relative lack of control over critical stimulus variables. For example, in most visual psychophysics tasks, stimuli are described in terms of visual angle. The calculation of visual angles requires precise measurements of viewing distance, screen size, and screen resolution. While these parameters are trivial to measure and control in a lab setting (where there is a known monitor and participants view stimuli while in a chin rest placed a known distance from the monitor), the same is not true of online data collection. In an online environment, not only will participants inevitably use a wide variety of monitors of different sizes with different software settings, they also may not have easy access to rulers/tape measures that would allow them to determine their monitor size or have the knowledge necessary to determine their software and hardware settings (*e.g.,* refresh rate, resolution).

Here we describe a set of methods to collect data on two well-known tests of visual attention – the Useful Field of View (UFOV) paradigm[7] and the multiple object tracking (MOT) task[8] – while avoiding as much as possible the sources of variability that are inherent in online measurements. These tasks can be run by any participant with an internet connection and an HTML5 compatible browser. Participants who do not know their screen size are walked through a measurement process utilizing commonly available items of standard size (*i.e.,* credit card/CD – see **Figure 1**).

Data on these two tasks were collected from over 1,700 participants in a Massive Online Open Course. Average performance of this online sample was highly consistent with results obtained in tightly controlled laboratory-based measures of the exact same tasks[9,10]. Our results are thus consistent with the growing body of literature demonstrating the efficacy of online data collection methods, even in tasks that require specific control over viewing conditions.

Protocol

The protocol was approved by the institutional review board at the University of Wisconsin-Madison. The following steps have been written as a guide for programmers to replicate the automated process of the web application described.

1. Login Participant

1. Instruct the participant to use an internet-enabled computer and navigate to the web application using an HTML5 compatible browser: http://brainandlearning.org/jove. Have the participant sit in a quiet room free of distractions, with the computer at a comfortable height.
   NOTE: Since the entire experiment is hosted online, the tasks can also be performed remotely without the presence of a research assistant. All instructions for the participant are included in the web application.
2. Have the participant input a unique ID that will be associated with the data collected and stored in a MySQL database. Have the participant reuse this ID if the online tasks are not completed within the same session. Before logging in, obtain consent from the participant via a consent form linked on the page.
   NOTE: A participant's progress is saved after each task in order to allow for completion of the 2 tasks at separate times if needed. Instruct the participant to always use the same ID in order to start where one left off.

2. Screen Calibration

NOTE: The web application guides the participant through the three steps outlined in the calibration page at: http://brainandlearning.org/jove/Calibration.

1. Ask the participant to input the diagonal size of the screen in inches in the labeled textbox.
   1. However, if the participant does not know this information, have the participant find a CD or credit card as a calibration object (**Figure 1**). When one is selected, prompt the participant to place the object against the screen and align it with a representative image of the object displayed on the screen.
   2. Prompt the participant to adjust the size of the screen image to match the size of the physical object. Based on the measurements of a physical CD (diameter of 4.7") or a credit card (width of 3.2") in addition to the pixel size of the representative image, determine the ratio of pixels to inches for the screen.
   3. Retrieve the pixel resolution of the monitor via JavaScript's *screen.width* and *screen.height* properties to then calculate the diagonal size of the screen in pixels. Knowing this value and the previously estimated pixel-to-inch ratio (see step **2.1.2**), convert the diagonal size to inches. Have the participant confirm this value through a dialog box.
2. Prompt the participant to adjust the screen brightness settings until all 12 bands in a black-to-white gradient displayed on the screen are clearly distinguishable. Brightness setting controls vary by computer.
3. Ask the participant to sit an arm's length away from the monitor in a comfortable position and then set the browser window to full screen mode. The browser window must be in full screen mode to maximize the visual space used by the tasks and to remove any visual distractions, such as the browser toolbar and desktop taskbars.
4. Knowing the resolution of the participant's screen and the diagonal size of the monitor, use the web application to automatically calculate the pixels/degree conversion value, based on a 50 cm viewing distance. Resize the dimensions of the stimuli in the tasks using this value. All visual angle dimensions reported below are based on this assumed mean distance value from the monitor.
5. Once calibration is complete, ask the participant to complete the two tasks described below. Choose the order of the tasks or randomly assign the order via the web application.

3. Multiple Object Tracking Task (MOT) – Figure 2

1. Introduce and familiarize the participant with the MOT stimuli through a self-guided tutorial, seen at: http://brainandlearning.org/jove/MOT/practice.php. Ask the participant to read step-by-step instructions that demonstrate how the trials will work. Once the participant finishes reading the instructions, prompt the participant to go through the practice trials.
   1. Setup the practice stimuli to consist of 8 dots at 0.8° with a movement speed of 2°/sec. Use the HTML5 *requestAnimationFrame* API to optimize browser animation at a frame rate of 60 Hz in order to control this stimulus motion.
   2. Ensure the dots move within the boundaries of a circle of 2° eccentricity and a circle no larger than the height of the participant's screen, without the instructions obscured.
   3. Set the dots to move in a random trajectory, where at each frame a dot has a 60% chance of changing direction by a maximum angle of 0.2°. If a dot collides with another dot or the inner or outer radial limits, move the dot in the opposite direction.
   4. Prompt the participant to track the blue dots (varying between 1 and 2 dots per practice trial), with the yellow dots acting as distractors.
   5. After 2 sec, change the blue dots to yellow dots and continue to move them amongst the original yellow dots for another 4 sec. At the end of each trial, stop the dots and highlight one.
   6. Prompt the participant to respond via key press whether the highlighted dot was a tracked dot or a distractor dot. Next, prompt the participant to press the space bar to continue onto the next trial.
   7. After 3 consecutive correct trials, or a maximum of 6 trials, move the participant onto the full task.
2. Start the full MOT task for the participant. An example of the task can be found at: http://brainandlearning.org/jove/MOT.
   1. Setup the full task with 16 dots that move at 5°/sec within the space between 2° eccentricity and 10° eccentricity. If the participant's screen cannot fit a circle of 10° eccentricity, use the maximum size the screen can contain instead.
   2. Have the participant complete a total of 45 trials: a mixture of 5 trials consisting of 1 tracked dot and 10 trials each consisting of 2 - 5 tracked dots. Match all other parameters to the practice trials (see **steps 3.1.3 - 3.1.6**).

3. Record the participant's response and response time once the dot is highlighted.
4. For every 15 trials, suggest a break to the participant. At these breaks, display the participant's performance (percent of correct trials) within the block on the screen.

4. Moving from One Task to Another (Optional Step)

1. Allow the participant to take a break between the two tasks. However, repeat **steps 1 and 2** if the tasks are not completed during the same login session.

5. Useful Field of View Task (UFOV) – Figure 3

1. Introduce and familiarize the participant with the UFOV stimuli through a self-guided tutorial, seen at: http://brainandlearning.org/jove/UFOV/practice.php. Ask the participant to go through 4 stages of step-by-step instructions that demonstrate the two target stimuli that must be attended to during the task.
   1. Set the central target stimulus as a 1° smiley that flashes at the center of the screen with either long or short hair. Randomize the smiley's hair length across trials.
   2. Set the peripheral target stimulus as a 1° star that flashes at 4° eccentricity at one of 8 locations around the circle (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°). Randomize the location of the star across trials.
   3. Control stimulus duration via number of frames used for presentation time. Optimize frame refresh at about 17 msec per frame by using the HTML5 *requestAnimationFrame* API.
   4. To check if the expected presentation time was achieved, use JavaScript's *getTime()* method to obtain the stimulus duration's start time and end time based on the participant's system clock. Calculate the measured presentation time from these two values and use this value for data analysis.
   5. For each practice trial, wait 500 msec before displaying the stimuli for approximately 200 msec (about 12 frames).
   6. Follow stimulus presentation with a noise mask composed of a randomly generated grayscale dot array for 320 msec (about 19 frames).
   7. For stage 1, only display the central target and then prompt the participant to respond via key press which hair length was displayed.
   8. For stage 2, only display the peripheral target and then prompt the participant to click on one of 8 radial lines, representing the 8 possible target locations, to indicate where the star appeared.
   9. For stage 3, display both central and peripheral target stimuli and then prompt the participant to provide responses to both the type of smiley and the location of the star.
      NOTE: Participants can freely choose the order of these two responses.
   10. For stage 4, display both target stimuli in addition to peripheral distractors, and then prompt the participant to respond to both target stimuli. For the distractors, display 1° squares presented at the remaining 7 locations at 4° eccentricity, in addition to 8 more squares at 2° eccentricity.
   11. After the participant's response, show the participant feedback (a green checkmark for a correct answer or a red cross for an incorrect answer) for each target response after each trial.
   12. Move the participant onto the next practice stage after getting 3 consecutive correct trials. After stage 4, move the participant onto the full task.
2. Prompt the participant to start the full UFOV task. An example of the task can be found at: http://brainandlearning.org/jove/UFOV.
   1. Present the same central stimulus as in the practice session (see **step 5.1.1**). Display the peripheral target at 7° eccentricity at one of the previously mentioned 8 locations (see **step 5.1.2**). 24 distractor squares are also displayed at 3° eccentricity, 5° eccentricity, and the remaining 7° eccentricity locations.
   2. Use a 3-down, 1-up staircase procedure to determine the presentation time of the stimuli: decrease the duration of stimuli after 3 consecutive correct trials and increase after each error trial.
   3. Before the first 3 reversals in the staircase, use a step size of 2 frames (approximately every 33 msec). After 3 reversals, use a step size of 1 frame. Vary the delay before the stimulus onset between 1 frame and 99 frames per trial, and keep the noise mask duration at 320 msec (about 19 frames).
      NOTE: Reversals are the points at which the duration changes either from increasing to decreasing, or decreasing to increasing.
   4. End the task when one of three conditions is met: the staircase procedure reaches 8 reversals; the participant completes 10 consecutive trials at either the ceiling duration (99 frames) or the floor duration (1 frame); or the participant reaches a maximum of 72 trials.
   5. Record the participant's response and response time for both the central stimulus and the peripheral stimulus.

Representative Results

Outlier Removal

A total of 1,779 participants completed the UFOV task. Of those, 32 participants had UFOV thresholds that were greater than 3 standard deviations from the mean, suggesting that they were unable to perform the task as instructed. As such, the UFOV data from these participants were removed from the final analysis, leaving a total of 1,747 participants.

Data were obtained from 1,746 participants for the MOT task. Two participants had mean accuracy scores that were more than 3 standard deviations below the mean, thus the data from these participants were removed from the final MOT analysis, leaving a total of 1,744 participants.

UFOV

For the UFOV task, performance was calculated by averaging the presentation time over the final 5 trials in order to obtain a detection threshold. The presentation time reflected the measured stimulus presentation duration on each participant's screen: the time from the start of the first stimulus frame until the end the last stimulus frame was recorded in milliseconds using the participant's system clock. The detection threshold reflects the minimum presentation duration at which the participants can detect the peripheral target with approximately 79% accuracy, given our use of a 3-down, 1-up staircase procedure. The mean UFOV threshold was 64.7 msec (SD = 53.5, 95% CI [62.17, 67.19]) and scores ranged from 17 msec to 315 msec with a median threshold of 45 msec (see **Figure 4**). The threshold distribution was positively skewed, with skewness of 1.92 (SE = 0.06) and kurtosis of 3.93 (SE = 0.12).

MOT

MOT performance was measured by calculating the mean accuracy (percent correct) for each set size (1 - 5). Accuracy ranged from 0.4 - 1.0 for set size 1 to 0.1 - 1.0 for set size 5, and mean accuracy ranged from 0.99 (SD = 0.06, 95% CI [0.983, 0.989]) for set size 1 to 0.71 (SD = 0.17, 95% CI [0.700, 0.716]) for set size 5. The median accuracy scores ranged from 1.0 to 0.70 for set size 1 and 5 respectively (see **Figure 5**).

A repeated-measures ANOVA was conducted to examine whether accuracy differed as a function of set size. There was a significant main effect of set size ($F(4, 6968) = 1574.70$, $p < 0.001$, $\eta_p^2 = 0.475$) such that accuracy decreased as set size increased, demonstrating a typical MOT effect.
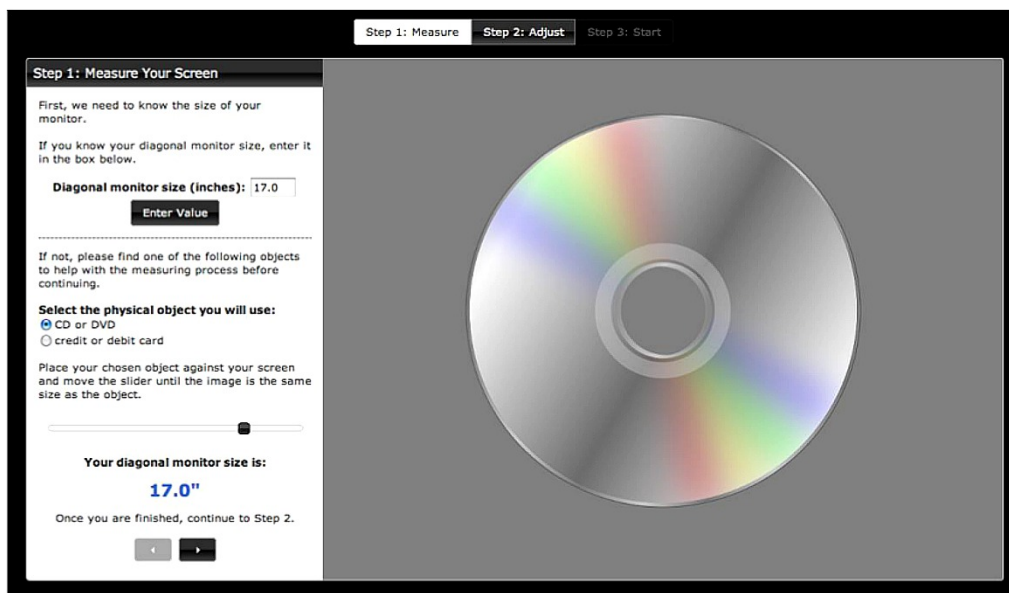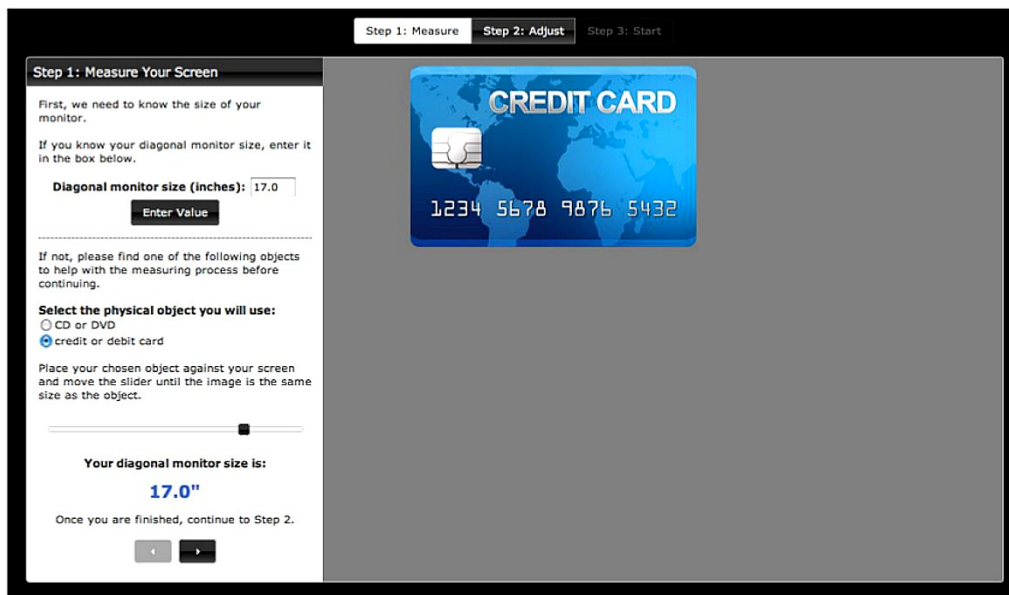
Figure 1. Screen measurement. Because not all online participants know their screen size – or have easy access to a ruler/tape measure to assess their screen size – the calibration process asked subjects to utilize commonly available items of standard size (credit card – above; CD – below). Please click here to view a larger version of this figure.
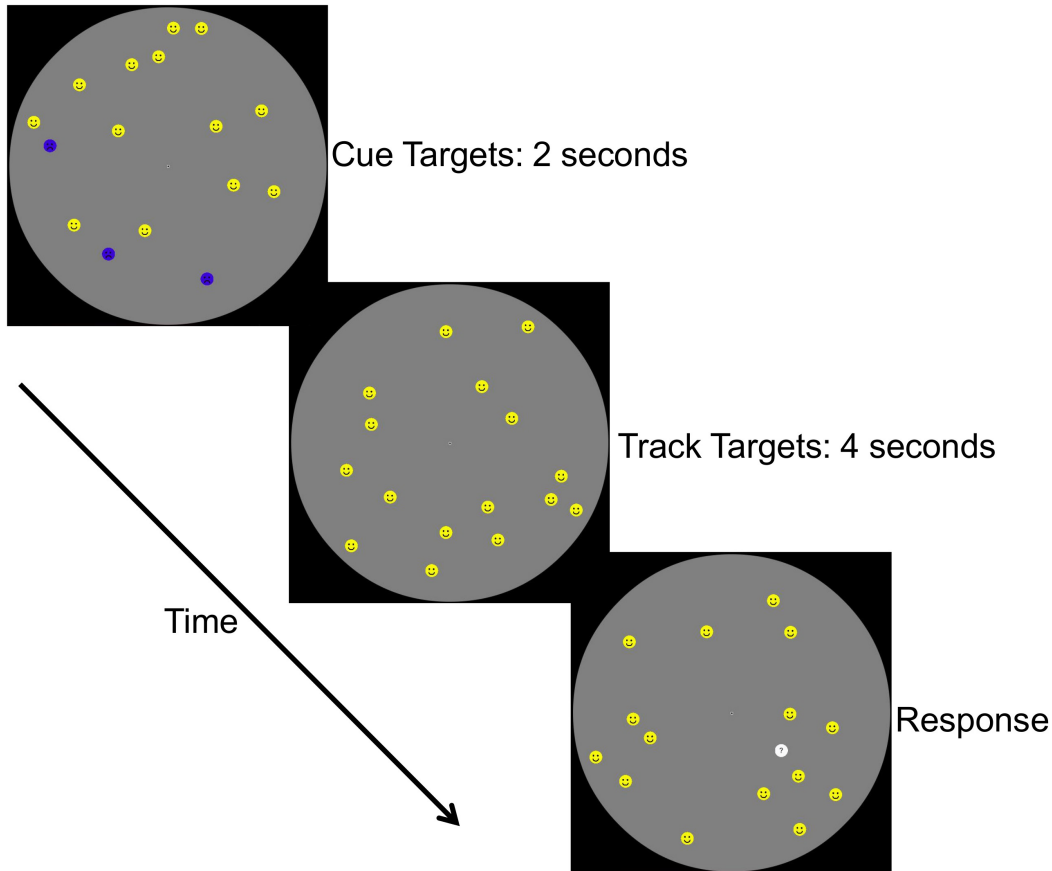
Figure 2. MOT Task. Participants viewed a set of randomly moving dots. At trial onset, a subset of these dots was blue (targets), while the remainder were yellow (distractors). After 2 sec the blue target dots changed to yellow, making them visually indistinguishable from the distractors. Participants had to mentally track the formerly blue target dots for 4 sec until a response screen appeared. On this screen one of the dots was white and the subject made a "yes (this was one of the original targets)" or "no (this was not one of the original targets)" decision (with a key press). Please click here to view a larger version of this figure.
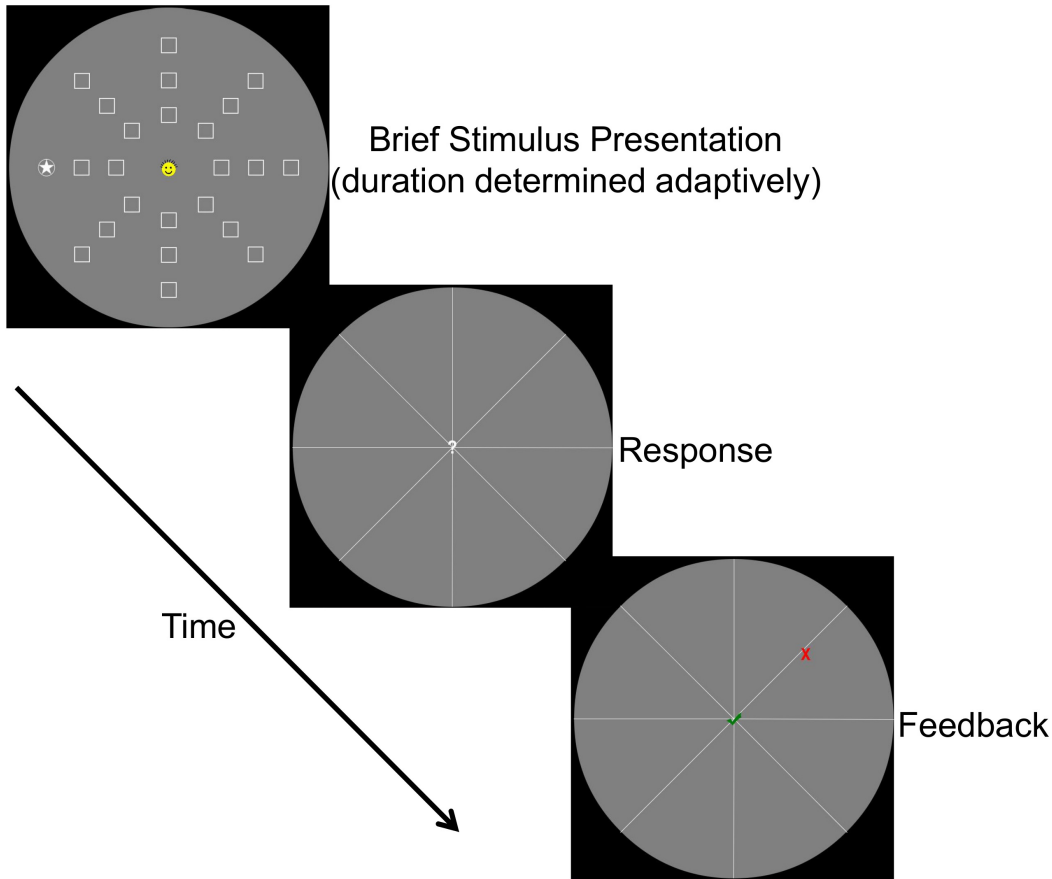
Figure 3. UFOV Task. The main screen consisted of a central stimulus (a yellow smiley that could have either short or long hair), a peripheral stimulus (a filled white star inside a circle) and peripheral distractors (white outlined squares). This screen was briefly flashed (with the timing determined adaptively based on participant performance). When the response screen appeared the participant had to make two responses: they had to indicate (with a key press) whether the smiley had long or short hair and they had to indicate (by clicking) on which of the 8 radial spokes the target stimulus appeared. They then received feedback about both responses (here they chose the correct answer for the central task, but the incorrect answer for the peripheral task). Please click here to view a larger version of this figure.
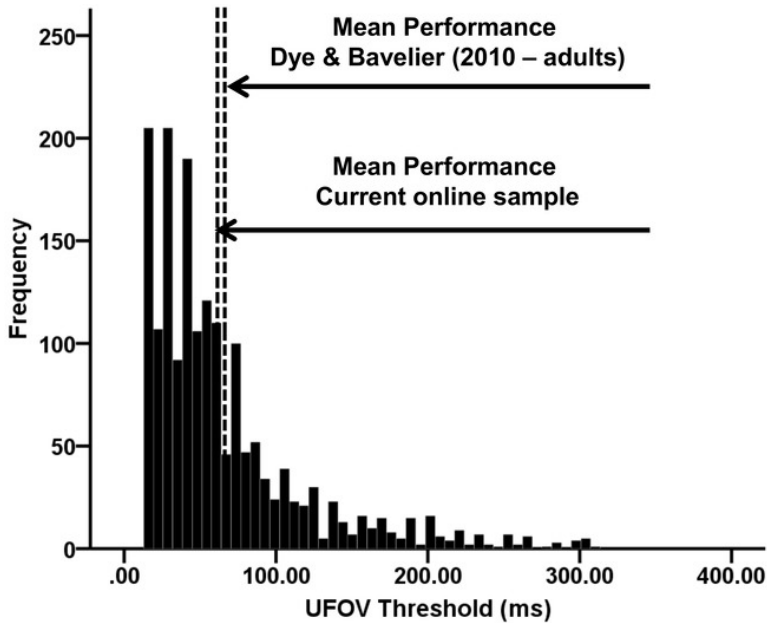
Figure 4. UFOV Results. As is clear from the histogram of subject performance, not only could the vast majority of the participants perform the task as instructed (~1% removed for poor/outlier performance), the mean performance was squarely in the range expected from lab based measures on the exact same task[9].
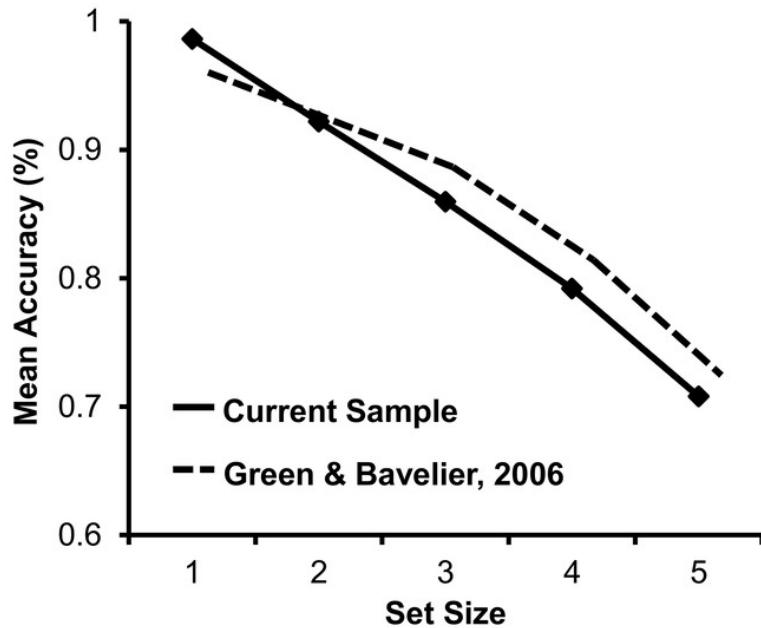

Figure 5. MOT Results. Consistent with previous work[10], MOT accuracy fell off smoothly with increasing set size.

Discussion

Online data collection has a number of advantages over standard laboratory-based data collection. These include the potential to sample far more representative populations than the typical college undergraduate pool utilized in the field, and the ability to obtain far greater sample sizes in less time than it takes to obtain sample sizes that are an order of magnitude smaller in the lab[1-6] (*e.g.,* the data points collected from 1,700+ participants in the current paper were obtained in less than one week).

The described online methods were able to replicate results obtained from previously conducted lab-based studies: calculated means and ranges for UFOV thresholds and MOT accuracy in the online tasks were comparable to results reported by Dye and Bavelier[9] for the UFOV task and Green and Bavelier[10] for the MOT task. However, the large participant sample did have an impact on the distribution of the results, particularly in the UFOV task. The online UFOV threshold distribution was more right skewed than previous laboratory-based results[9]. This difference in skew may be attributed to the greater diversity of participants recruited online, particularly in regards to their wider variation in age: the online sample ranged from 18 - 70 years, while the laboratory-based sample ranged from 18 - 22 years[9].

Furthermore, collecting data via online methods does require solving several technical challenges – particularly when close stimulus control is necessary for the validity of the measures. The two tasks employed here required control over both the visual angle of the stimuli that were presented and the timing of the stimuli. Visual angle in particular can be difficult to control in online settings as its calculation requires knowing viewing distance, monitor size, and screen resolution. This is particularly problematic given that many online participants may not know their monitor size or have easy access to a tape measure in order to measure their monitor size.

We devised a series of steps to overcome some of these issues. While we can perfectly resolve monitor size, we still cannot precisely control the actual viewing distance. We suggest to participants to sit an arm's length away from the monitor, although this distance may vary among participants. Arm length was chosen, as U.S. anthropometric data indicates that the difference in length of a forward arm reach (the position in which participants would use to judge their distance away from the screen) between male and female adults is small, such that the median male reach is 63.8 cm while the median female reach is 62.5 cm[11]. Although the experiment setup procedure attempts to avoid introducing sex biases by using this measurement, there may be potential height biases; future studies that collect participants' height information would need to be conducted to assess this possibility.

As for stimulus timing, we took into account the discrepancies between expected duration and recorded duration of stimulus presentation when calculating threshold values. Rather than relying on the expected presentation duration, we measured the duration of the stimulus frames using the participant's system clock with millisecond precision. However, inherent disparities between monitor displays still were present and cannot be controlled for without physical *in situ* measurements. It is well known that Liquid Crystal Displays (LCD)— the most likely monitors our participants have access to—have long response times that typically vary depending on the start and end values of the pixel luminance changes. The latter issue is not a concern in our study because we always switched from the same background level to stimulus level. A greater concern is that variability in displays across participants causes a large portion of the measured variance. We believe that this is not an issue as pixel response times are typically smaller than 1 frame rate (*i.e.*, 17 msec)[12,13], which seems acceptable in comparison to the large inter individual variability in UFOV thresholds.

The methods employed here overcome the aforementioned challenges and thus allowed us to measure performance on two tasks – the UFOV and the MOT - that both require control over visual angle and screen timing properties. The results obtained by these methods were consistent with those obtained in standard laboratory settings, thus demonstrating their validity. Additionally, because these tasks require only an internet connection and an HTML5 compatible browser, these tasks can be employed not only to easily gather a large sample from a generally representative population, but can also be used to reach specific sub-types of individuals that may be geographically separated and thus difficult to bring to a common lab setting (*e.g.,* patients with a certain type of disease or individuals with a certain ethnic background). Furthermore, with the rise of use of iPads and other tablets, the design of the web application could easily be adapted for better compatibility with touchscreen technology in order to reach an even greater number of participants. While the web application can currently run on tablets via an HTML5 browser, future iterations could remove the requirement of a keyboard and replace response keys with interface buttons or gestures.

Disclosures

Acknowledgements

The authors have nothing to disclose.

References

1. Henrich, J., Heine, S. J., Norenzayan, A. The weirdest people in the world. *The Behavioral And Brain Sciences.* **33**, 61-135 (2010).
2. Buhrmester, M., Kwang, T., Gosling, S. D. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data. *Perspectives on Psychological Science.* **6**, (1), 3-5 (2011).
3. Goodman, J. K., Cryder, C. E., Cheema, A. Data collection in a flat world: the strengths and weaknesses of mechanical turk samples. *Journal of Behavioral Decision Making.* **26**, (3), 213-224 (2013).
4. Mason, W., Suri, S. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods.* **44**, (1), 1-23 (2012).
5. Crump, M. J., McDonnell, J. V., Gureckis, T. M. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One.* **8**, e57410 (1371).
6. Lupyan, G. The difficulties of executing simple algorithms: why brains make mistakes computers don't. *Cognition.* **129**, 615-636 (2013).
7. Ball, K., Owsley, C. The useful field of view test: a new technique for evaluating age-related declines in visual function. *J Am Optom Assoc.* **64**, (1), 71-79 (1993).
8. Pylyshyn, Z. W., Storm, R. W. Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision.* **3**, (3), 179-197 (1988).
9. Dye, M. W. G., Bavelier, D. Differential development of visual attention skills in school-age children. *Vision Research.* **50**, (4), 452-459 (2010).
10. Green, C. S., Bavelier, D. Enumeration versus multiple object tracking: The case of action video game players. *Cognition.* **101**, (1), 217-245 (2006).
11. Chapter 1. Kodak Company. Ergonomics Design Philosophy. *Kodak's Ergonomic Design for People at Work.* Chengalur, S. N., Rodgers, S. H., Bernard, T. E. John Wiley & Sons, Inc Hoboken, NJ (2004).
12. Elze, T., Bex, P. P1-7: Modern display technology in vision science: Assessment of OLED and LCD monitors for visual experiments. *i-Perception.* **3**, (9), 621 (2012).
13. Elze, T., Tanner, T. G. Temporal Properties of Liquid Crystal Displays: Implications for Vision Science Experiments. *PLoS One.* **7**, (9), e44048 (2012).