Video Article

Testing Ideas About Visual Perception from Real-world Observations Using Behavioral and Eye Tracking Measurements

Jennifer E. Corbett[1,2], Jaap Munneke[1,2]

[1]College of Health and Life Sciences, Division of Psychology, Brunel University London

[2]Centre for Cognitive Neuroscience, Brunel University London

Correspondence to: Jennifer E. Corbett at jennifer.e.corbett@gmail.com

URL: https://www.jove.com/video/59004
DOI: doi:10.3791/59004

Keywords: Visual perception, Abductive designs, Behavior, Eye tracking, Multiple object tracking, Gestalt grouping

Date Published: 10/2/2018

Citation: Corbett, J.E., Munneke, J. Testing Ideas About Visual Perception from Real-world Observations Using Behavioral and Eye Tracking Measurements. *J. Vis. Exp.* (), e59004, doi:10.3791/59004 (2018).

Abstract

The majority of scientific inquiry relies on deducing specific conclusions from general rules. However, when existing theories fail to fully account for a given observation, an abductive approach can allow for the formulation of new theories that can then be tested and gradually expanded to account for a wider range of similar observations. Here, we present an abductive approach using convergent behavioral and eye tracking measures to explain a real-world observation that grouping by color similarity facilitates observers in tracking the interactions between players on different sports teams.

Introduction

For over a century, investigations of visual perception have relied predominantly on deductive methods. Predictions of an existing theory are tested and the resultant data call for the theory to be supported, modified, or abandoned in favor of an alternative or newly developed theory. However, it is not always possible to deduce specific ideas from general principles when the goal is to understand these general principles. For example, our current understanding of how the limited capacity visual system is able to construct the illusion of stable and complete perception stems largely from abductive work pioneered by Gestalt Psychologists, who developed a set of heuristics for perceptual organization based on observations of inherent regularities in the surrounding environment[1,2,3]. Yet, this sort of abductive research that forms the bedrock of current theories has fallen out of fashion in the scientific community. Despite the inherently cyclical nature of inductive and deductive reasoning, our modern discussion of abductive research has become restricted to illusion contests and demo nights. In an attempt to promote the reintegration of this fundamental scientific link between observation and explanation, we provide an example of a method for observing a real-world phenomenon, recreating it in the lab, and generating testable hypotheses based on these observations. Given the long-standing scientific trend to move away from abductive reasoning, we promote using a range of converging behavioral and eye tracking methods to assess real-world phenomenon aimed to help gather as much information as possible when testing potential explanations.

Our particular investigation began with the observation that the different-colored jerseys worn by teams of players in a soccer match somehow seemed to help keep track of how the players were interacting, both in a collective sense of monitoring the movements of all the players on the field, and in terms of the relationships between the players on a given team. Importantly, studies of multiple object tracking suggest the opposite effects, such that observers are less able to track objects which are part of the same Gestalt group compared to tracking performance for ungrouped objects (*e.g.,* see references[4,5]). In an effort to explain our observations, we hypothesized that grouping information by similarity of color (the Gestalt principle of grouping by similarity) might allow observers to track more information than the typical "four-object limit" that has been widely accepted in the visual perception literature[6,7]. Along these lines, organizing players into fewer teams with more players on each team compared to more teams with fewer players on each team may facilitate observers in tracking players' interactions, as indexed by their performance in counting the number of times a given player gained possession of the ball. Furthermore, the effect of this grouping by color similarity might affect observers' tracking performance differently depending on whether they were attempting to monitor the interactions of all the players or only the interactions of a specific team. To explore these proposals, we began by abstracting the soccer match scenario to a set-up that could be experimentally manipulated in the laboratory.

Given that existing theories could not explain our observations, we set-out to gather as much data as possible to provide the most complete assessment of the phenomenon in question. Specifically, we had access to psychophysical and eye tracking testing facilities that allowed us to concurrently measure observers' behavioral accuracy and pupil size while they tracked the interactions of numbered, colored-circle "players" passing a soccer ball. This convergent approach allowed us to measure not only the effects of grouping that could be directly observed from the participants' response behavior, but also cognitive factors that could be derived from eye movements not contaminated with the extra factors required to plan and execute a manual button press response. Pupil diameter has been demonstrated to index the amount of mental effort observers invest, such that pupil size increases with difficulty and set size in visual search and counting tasks[8]. Along these lines, more effortful top-down task demands tend to elicit larger pupil diameters, and performance in similar tasks; for example, visual search is typically facilitated by top-down attention[9]. Therefore, pupil diameter can index potential differences in the amount of mental effort participants invested to track the ball target as a function of the top-down set elicited by instructing them to count the possessions by players on a specific color team when there were fewer teams of more players each and when there were more teams of fewer players each, as compared to when they were instructed to simply count the possessions of all the players.

Protocol

All aspects of the protocol were approved by the local research ethics committee.

1. Participant Recruitment

1. Recruit participants aged 18-35 years to ensure that their eyes are fully developed but likely not subject to the effects of aging (*e.g.*, presbyopia).
2. Stress during recruitment that participants must have normal or corrected-to-normal vision.

2. Experimental Design

NOTE: Using a stimulus presentation software, construct stimuli that can be manipulated along the dimensions that are key to answering the experimental question while holding all other parameters constant or randomized. This protocol focuses on the effects of grouping by color similarity on observers' abilities to track the interactions between multiple objects. Therefore, we used numbered, colored circles and a JPEG image of a soccer ball as visual stimuli.

1. **Conditions: Construction of a design matrix of the experimental factors**
   NOTE: See **Figure 1**.
   1. To examine the effects of the number of different colored teams (NTeams) and the particular task the participant was attempting to perform (CountTask), design a scenario where the same number of circle stimuli (*e.g.*, 20) can be divided equally into a small number and a large number of individual groups (*e.g.*, two groups of 10 circles each, and five groups of four circles each).
   2. Importantly, include a baseline condition to obtain a measure of the observers' performance when there is no grouping (*e.g.*, 20 circles that are all the same color).

2. **Randomization in space and time**
   1. Randomize an equal number of trials for each of the (five) combinations of the experimental factors (CountAll2Teams, CountAll5Teams, CountTeam2Teams, CountTeam5Teams, and 1Team/Baseline) separately for each participant.
      NOTE: Ensure that there are enough trials for proper statistical analysis given potential data loss due to human and mechanical error. Typically, 25-50 trials per condition are sufficient in behavioral and eye tracking experiments.
   2. Randomly select the colors of the different (two or five) teams on each trial from a base set of RGB values to ensure that any differences in luminance akin to differences in real-world team jerseys will average out over trials.
      NOTE: We used Red = (255 0 0); Green = (0 255 0); Blue = (0 0 255); Yellow = (255 255 0); Magenta = (255 0 255).
   3. Make a base set of screen locations with more locations than the total number of stimuli that cover the central 3/4 of the visible screen. Randomly select where each circle will start for each trial, and then add a small x- and y-jitter to each location so that the stimuli do not appear to be grouped into any geometric formations that might otherwise confound the results.
   4. Randomly select the number of each circle on each team from one to the total number of stimuli so that there is an equal probability of the same number (*e.g.*, 15) appearing in any of the teams presented on the trial.
   5. Randomize the starting location/player who has possession of the soccer ball.
   6. Randomly select the length of a given trial from an interval that allows time for a sufficient number of possessions for participants to count (*e.g.*, 12-14 s allowing for an average of 11 possessions per team).
   7. Cut the trial into shorter intervals (*e.g.*, randomly selected intervals from 400-600 ms in 50 ms steps) and assign each player a 50% chance of changing location at the start of each interval.
   8. If a player is assigned to change location, immediately represent this stimulus at a new location randomly selected from any of the locations that do not currently have a player assigned (offset the new location by small, randomly selected amounts of jitter in the x- and y-directions to avoid any grid configurations).
   9. Ensure that the ball has a probability of changing players on a large enough proportion of the intervals in a given trial (*e.g.*, 90%).
   10. Randomly select one of the team colors at the start of the trial, and give players on this team approximately half of the possessions to ensure conditions are approximately equal in difficulty in terms of the number of possessions observers must count.
   11. To ensure that individual elements are not truncated, construct an imaginary boundary from the top, bottom, and sides of the screen and restrict the players and the ball from moving within this region.

3. **Ecological validity: Keep it as real world as possible**
   1. Use numbers to individuate different players.
   2. Present the numbers centered on the top half of each circle like numbers on actual soccer jerseys.
   3. To ensure that the ball never obscures the possessing player's number, center the ball horizontally with the player, and position it vertically from the center to the bottom of the player (at the player's "feet").
   4. Allow players to overlap as they would in a real-world match.

4. **Other design considerations**
   1. Use a gray background to minimize the likelihood that visual afterimages or transients confound any observed effects.
   2. Divide the trials into several blocks to ensure that participants take regular breaks every 3-5 min and rest their eyes.

3. Instructing the Participants

NOTE: See **Figure 2**.

1. Ensure that the participant fully understands the task and how it relates to the real-world scenario before beginning the experimental blocks.
2. Present each participant with written, illustrated instructions and read these instructions aloud with the participant at the start of the experimental session.
3. Instruct participants to estimate the number of possessions on each trial as accurately as possible.
4. Ensure that the participant understands that the players might overlap at times like in a real soccer match and to make their best guess if this happens.
5. Ensure that the participant understands what does count as a possession.
   NOTE: Possessions are defined as receiving the ball from a player on the same team, or from a player on a different team.
6. Ensure that the participant understands what does not count as a possession.
   NOTE: The starting position/player of the ball does not count as a possession and the same player moving with the ball does not count as a possession.
7. Before setting up the eye tracker, give participants a practice block with a few repetitions of each condition (in randomized order) and monitor their performance during this practice block to ensure that they are performing as instructed.
    1. Impose (recommended) some performance criteria during the practice block to ensure participants are performing as instructed, such as requiring them to complete the practice block with an average error below a certain threshold.
    2. Make sure to ask whether participants have any questions and ensure that they understand the answers before beginning the experimental blocks.

4. Eye Tracking Set Up

1. Determine the sampling rate of the eye tracking system.
   NOTE: The measures available from the eye data are largely dependent on the sampling rate of the eye tracking hardware. For example, if the eye tracking system can only sample up 60 times per second (60 Hz), then a single sample is taken approximately every 16.67 ms. This translates to the potential for up to 16.67 ms of noise to be introduced into individual samples, which usually allows for the measurement of pupil size, but often precludes measuring eye movements like saccades and fixations that depend on more precise sampling rates. If the tracking system can sample at a higher rate (*e.g.*, 500 Hz), then such additional measures can also be included to the extent that they can provide information relevant to answering the experimental question. The eye tracker used in the original study by Akyuz and colleagues[10] tracked participants' right eye positions at 60 Hz. For best results, always follow the instructions of the specific eye tracker used.
2. Synchronize the eye tracker with the stimulus presentation software and eye tracking recording software according to manufacturer's instructions.
3. Position the participant at a constant distance from the center of the monitor.
4. Ask the participant to sit with their chin in a chinrest and adjust the chair height and chinrest position as needed to ensure that the participant's head is centered horizontally and vertically with the monitor.
5. Before beginning each experimental block, calibrate the eye tracker to ensure it is accurately tracking participants' gaze within 1° of accuracy by having each participant complete a calibration task.
    1. Ask participants to follow the white center of a black dot with their eyes as it moves in random order through nine different points arranged in an imaginary 3 x 3 square grid covering the central 90% of the screen.
    2. Remind the participant to move as little as possible for the duration of each block.
6. After successfully calibrating the eye tracker, ensure the participant is ready to begin the task, remind them to sit as still as possible, and begin recording the eye data.
7. If the eye tracking software indicates that the eye position has been lost, stop recording, recalibrate, and redo the block.

5. Trial Sequence

1. Begin each trial by instructing each participant to press the space bar on the keyboard when ready.
2. Present the specific task instructions on a gray background (*e.g.*, "ALL" in the CountAll conditions, "GREEN" in the count team conditions, and "YELLOW ALL" in the 1Team/Baseline conditions). Ask the participant to press the space bar again after reading the trial instructions to initiate the trial.
3. Present the players passing the ball as described in the "Experimental Design" section (see section 2). Keep the task instructions at the corner of the screen throughout the entire duration of the trial in case the participant forgets what task to perform.
4. At the end of the trial, present a gray screen prompting the participant to respond "How many possessions?" using the number keys on the computer's keyboard.

6. Data Analysis

1. Co-register the eye tracking and behavioral data.
2. For each participant, exclude from analysis any trials with absolute errors greater than a pre-determined criterion, trials without responses, and trials where the eye tracker lost the pupil for more than a given number of samples.
3. Calculate the behavioral percent correct as follows.
    1. Calculate the percent correct on each trial using the following formula:

$$\%Correct = \frac{NPos - AbsError}{NPos} \times 100$$

where *NPos* = the actual number of possessions, and *AbsError* = the absolute error on each trial.

2. Subtract the resultant average value in the baseline condition from the average of each of the four main conditions, such that negative values indicate lower accuracy than baseline.

4. Measure the pupil diameter.
NOTE: Typically, pupil diameter is measured in arbitrary units specific to the particular eye tracking software used.
   1. Divide each participant's average pupil diameter in a given condition by their pupil diameter averaged over all trials to calculate the given participant's average relative pupil diameter for that condition.
   2. Subtract the resultant average value in the baseline condition from the average of each of the four main conditions, such that negative values indicate smaller pupil diameters than baseline.

Representative Results

The current protocol was previously used to test the effects of grouping by color similarity on participants' abilities to track the interactions between individual objects[10]. In that study, we tested 29 students from Bilkent University (mean age = 19.8 years, 20 females), who voluntarily participated in exchange for course credit or monetary compensation. All aspects of the protocol were approved by Bilkent University's Ethics Committee. We present statistical analyses based on these previously obtained results, and illustrations of representative results based on hypothetical data to best illustrate the protocol's possible outcomes.

The experiment consisted of a 2 x 2 design with an additional baseline condition. The experimental factors consisted of "CountTask" (count either the ball possessions within one team or within all teams) and "Nteams" (either two teams of 10 players, or five teams of four players were present in the display on a given trial). The baseline condition consisted of one group of 20 homogenously colored circles. The dependent measures were accuracy in counting the number of possessions and the average pupil diameter throughout a given trial. Crucially, all statistical analyses were performed on baseline corrected data: For each participant, the average performance obtained in the baseline condition was subtracted from the participant's average performance on each of the four experimental conditions. Using baseline-corrected data provides a better measurement of the influence of the experimental factors on the participant's tracking performance as it negates the large amount of noise that is the result of using dynamic displays rendered over multiple seconds.

Throughout the experiment, participants were tasked with counting the number of ball possessions of players of either the same team or of all players present in the display, independent of team membership. **Figure 2a** provides an overview of what the dynamic displays in this experiment looked like, and **Figure 2b** instructs what does and does not constitute a possession.

The applied methodology provided a number of quantitative results. First, using the difference in accuracy relative to the baseline as a dependent measure, a repeated measures ANOVA with CountTask (CountAll vs CountTeam) and Nteams (2Teams vs 5Teams) as factors indicated that participants were significantly more accurate when counting the possessions of all players present in the display, as compared to counting the possessions of players within one specified team ($F(1, 28) = 26.759$, $p < 0.001$, $\eta_p^2 = 0.489$). Similarly, participants were more accurate when trials consisted of two teams of 10 players as compared to trials on which five teams of four players were presented ($F(1,28) = 6.735$, $p = 0.015$, $\eta_p^2 = 0.194$; see **Figure 3a**). Subsequent analyses suggested that participants were more accurate when counting the number of possessions of a specific team when there were two teams of 10 players, as compared to five teams of four players ($t(28) = 2.732$, $p = 0.011$, Cohen's $d = 0.41$. This difference between two and five teams was absent when participants were counting possessions of all players regardless of team ($p = 0.66$). In addition, when tracking the number of possessions of a specified team, participants performed significantly better than in the baseline condition regardless of whether two or five teams were presented in the display. This was confirmed by a one sampled t-test against 0 (baseline; both ts(28) > 5.981, both ps < 0.001, both Cohen's ds ≥ 1.1). When players were counting possessions of players in all teams, performance was marginally better than in the baseline condition (both ts(28) > 1.91, both ps < 0.066, both Cohen's ds ≥ 0.355.

Second, a similar ANOVA on the average relative pupil diameter (calculated as the mean conditional diameter divided by the overall pupil diameter), showed that participants had a smaller relative pupil size when counting possessions of all players independent of team, compared to counting possessions by players from one team (independent of number of teams; $F(1, 28) = 15.14$, $p = 0.001$, $\eta p^2 = 0.351$; see **Figure 3b**). In addition, participants displayed a smaller relative pupil size when two teams of 10 players were presented as compared to five teams of four players ($F(1, 28) = 11.223$, $MSE = 0.001$, $p = 0.002$, $\eta_p^2 = 0.286$. Post-hoc comparisons showed a smaller relative pupil size when participants were counting possessions by players of a specific team, when only two, as compared to five teams were presented ($t(28) = 4.429$, $SEM = 0.005$, $p < 0.001$, Cohen's $d = 0.521$ See **Figure 3b**). Finally, one-sampled t-tests showed that in all conditions, pupil size was smaller than in the baseline condition (all ts(28) ≥ 2.54, all ps ≤ 0.018, all Cohen's ds ≥ 0.474).

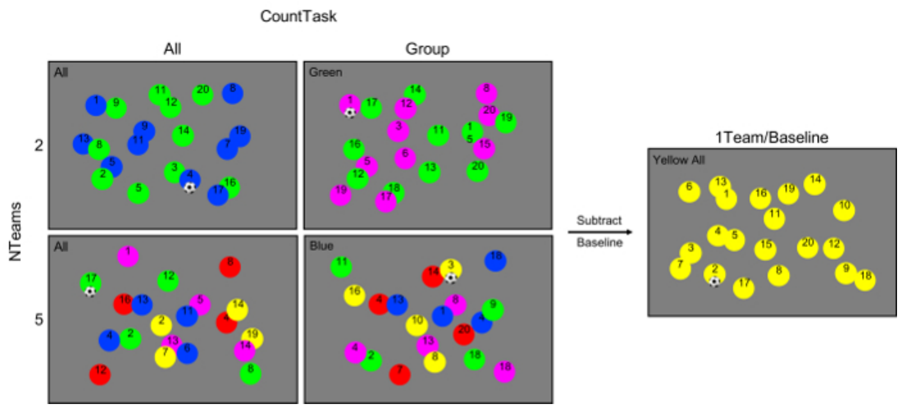

Figure 1: Design matrix. A 2 x 2 design was used to investigate the role of grouping by color similarity on observers' ability to track a soccer ball in a dynamic display. Critical factors were the number of teams in the display (*i.e.,* the number of different colors; NTeams) and whether the participants were instructed to track possessions within one specified team or to track the possessions of all players in the display (CountTask). A baseline condition was introduced consisting of 20 players from the same team. To ensure measurements with the least amount of noise, we subtracted performance in the baseline condition (the condition with one team/color) from the average performance in each of the four experimental conditions, separately for each participant. Please click here to view a larger version of this figure.

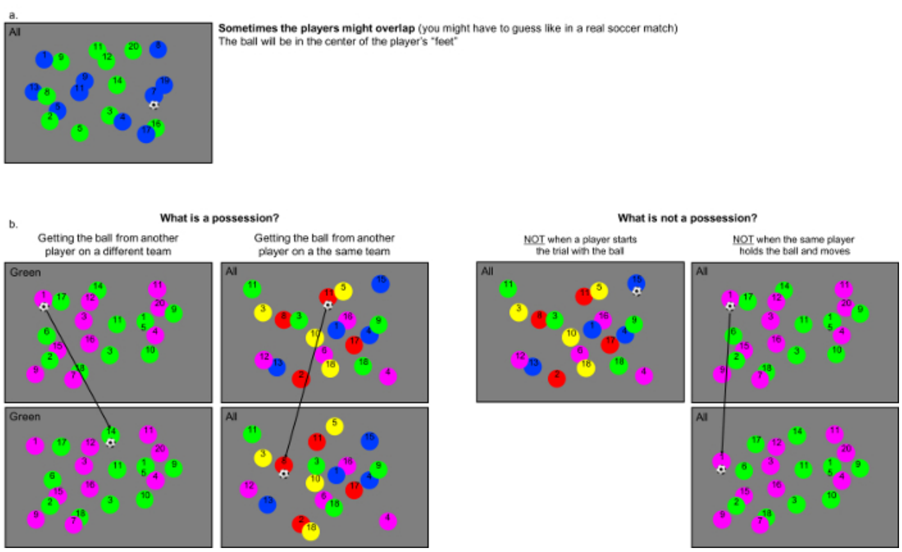

Figure 2: Written, illustrated instructions for participants. (**a**) Players of teams with differently colored "jerseys" moved in small steps through the display. Individual players could overlap, reflecting real-world soccer conditions (where one player can occlude another player). However, to ensure accurate task performance both the soccer ball and the jersey-number were always clearly visible on top of each player. (**b**) Example displays of what does and does not constitute a possession. A possession is defined as a player receiving the ball from another player, regardless of whether the player that "played" the ball was from the same or a different team. Crucially, the first possession when the trial starts does not count as a possession, nor is it a possession when the player moves with the ball to a new location. Please click here to view a larger version of this figure.

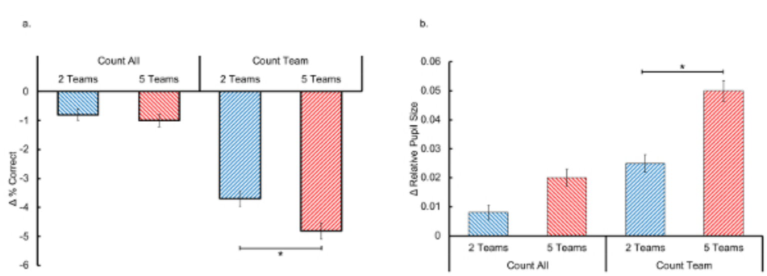

Figure 3: Measurements. (**a**) Relative accuracy in counting the number of possessions in the experimental conditions minus the performance in the baseline condition. Negative values signify an increase in accuracy as compared to baseline. (**b**) Relative pupil size during the experimental conditions, divided by the overall pupil size over all trials. Negative values indicate that pupil size was smaller compared to baseline conditions. All data is hypothetical data that follow numerical and statistical patters presented in Akyuz *et al.*[10]. Error bars reflect the standard error of the mean. Please click here to view a larger version of this figure.

Discussion

The protocol presented here promotes a form of abductive research, beginning with a real-world observation and then using behavioral and eye tracking measures to seek out the most parsimonious explanation. Although the current research climate almost exclusively promotes deductive research, there are numerous circumstances in which an observation does not fit within current theories and an alternative explanation must be formed. The present protocol focusses on a situation where a benefit of being able to pay attention to more objects organized into fewer groups compared to fewer objects arranged in more groups was abstracted after observing a soccer match with players in different colored jerseys. As this increase in information processing could not be explained by current theories which posit that the human visual system can keep track of about four individual objects at one time[4,5], it was necessary to develop a new account for these observations. Importantly, the present protocol is aimed to provide researchers an avenue for testing hypotheses when observations cannot be interpreted within an already established theoretical framework.

The most critical steps in this protocol were using convergent measures obtained from available behavioral and eye tracking recording equipment, keeping the experimental stimuli as close to the real-world situation as possible while holding all other parameters constant or randomizing parameters so that only the key variables of interest (in this case, color similarity, grouping, and the observer's task) could account for any differences in the observed results.

There are several parts of the current protocol that may potentially require some troubleshooting. First, depending on the eye tracking equipment and setup, and the participants, there may be a need to recalibrate frequently. For example, it is crucial that the experimenter monitors the eye tracking software during recording to ensure that the pupil is not lost. If the eye tracker loses the pupil, the block should be stopped and re-done to ensure proper data collection. For this reason, it is essential both to include enough trials per condition and to organize trials into short blocks that allow participants a break every 3 to 5 min. However, there are some situations where the particular participant has great difficulty in performing the eye tracking task over the entire duration of the experiment, due to fatigue, dry eyes, contact lens problems, *etc.* In such cases, it may be necessary to dismiss the participant from the experiment. In addition, participants will occasionally make mistakes in their behavioral responses, for example, by accidentally typing "99" instead of "09." Depending on the task and the experimental question, it will be necessary to determine whether to allow participants the opportunity to correct their responses on each trial, or whether extreme responses are trimmed at the data analysis stage. If the experimental question is more concerned with perceptual processes, as in the present scenario, filtering out extreme responses at the analysis stage can help to avoid situations in which participants spend more time thinking about their responses than they normally would observing the parallel real-world event. However, if the experiment is designed to investigation longer-duration processes such as memory or information retention, allowing participants the opportunity to correct their responses during the experiment may be more beneficial.

The present method is limited in two main ways. First, as with all abductive designs, the goal is to find the simplest explanation of the observed phenomenon. Obviously, there are plenty of opportunities for the resultant explanation to fail to extend to similar circumstances. Therefore, it is absolutely essential that any findings from abductive studies be used to form further, more specific hypotheses that can then be tested using more traditional deductive methods. In addition, our study was limited by the relatively low sampling rate of the available eye tracking equipment. As we could sample every 16.67 ms, we were able to record a relative measure of observers' pupil size. Using equipment with a sampling rate of 500 Hz or greater would allow future investigations to measure more temporally fine-grained aspects such as ballistic eye movements (saccades), which can provide further insight into how the experimental parameters affect perceptual fluency and speed.

Abductive research is meant to find the simplest explanation when none currently exist. Then, this explanation can be posed as a hypothesis with predictions that can be tested in a more traditional, systematic, deductive framework. For example, building on our initial findings that observers are more accurate and exert less effort when tracking the interactions of more objects organized into fewer groups *versus* fewer objects organized into more groups, future studies can extend the external validity of this paradigm by investigating the contributions of social cues such as body language to tracking the interactions of players in real-world sports matches, perhaps by superimposing simple colored-circles on players from real-world footage to retain their actual movement patterns.

Disclosures

The authors have nothing to disclose.

Acknowledgements

The authors have no acknowledgements.

References

1. Koffka, K. *Principles of Gestalt psychology.* Harcourt, Brace. Oxford, England. (1935).
2. Kanizsa, G. Subjective Contours. *Scientific American*. **234** (4), 48-53 (1976).
3. Wertheimer, M., Riezler, K. GESTALT THEORY. *Social Research.* **11**(1), 78-99 (1944).
4. Erlikhman, G., Keane, B. P., Mettler, E., Horowitz, T. S., Kellman, P. J. Automatic feature-based grouping during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*. **39**(6), 1625 (2013).
5. Scholl, B. J., Pylyshyn, Z. W., Feldman, J. What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*. **80**(1-2), 159-177.4 (2001).
6. Luck, S. J., Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature*. **390**, 279-281 (1997).

7. Pylyshyn, Z.W., Storm, R.W. Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*. **3**, 179-197 (1988).
8. Porter, G., Troscianko, T., Gilchrist, I. D. Effort during visual search and counting: Insights from pupillometry. *The Quarterly Journal of Experimental Psychology*. **60**(2), 211-229 (2007).
9. Chun, M. M., Jiang, Y. Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*. **10**, 360-365 (1999).
10. **10.** Akyuz, S., Munneke, J., Corbett, J. E. Set similarity modulates object tracking in dynamic environments. *Attention, Perception & Psychophysics*. **80**(7), 1744-1751 (2018).