HHS Public Access

Author manuscript

Biol Psychol. Author manuscript; available in PMC 2020 July 03.

Published in final edited form as:

Biol Psychol. 2019 September ; 146: 107712. doi:10.1016/j.biopsycho.2019.05.012.

Methodological considerations in the use of Noldus EthoVision XT video tracking of children with autism in multi-site studies

Maura Sabatos-DeVito[a,*,1], **Michael Murias**[b], **Geraldine Dawson**[a], **Toni Howell**[a], **Andrew Yuan**[a], **Samuel Marsan**[a], **Raphael A. Bernier**[c], **Cynthia A. Brandt**[d], **Katarzyna Chawarska**[e], **James D. Dzuira**[f], **Susan Faja**[g], **Shafali S. Jeste**[i], **Adam Naples**[e], **Charles A. Nelson**[g,h], **Frederick Shic**[j], **Catherine A. Sugar**[k], **Sara J. Webb**[j], **James C. McPartland**[e,**], **Autism Biomarkers Consortium for Clinical Trials**

[a]Department of Psychiatry and Behavioral Sciences, Duke Center for Autism and Brain Development, Duke University Medical Center, Durham, NC, USA

[b]Duke Institute for Brain Sciences, Duke University Medical Center, Durham, NC, USA

[c]Department of Psychiatry, University of Washington, Seattle, WA, USA

[d]Yale School of Medicine, New Haven, CT, USA

[e]Yale University School of Medicine, Child Study Center, New Haven, CT, USA

[f]Department of Emergency Medicine, Yale Center for Analytical Sciences, Yale University, CT, USA

[g]Boston Children's Hospital and Harvard Medical School, Boston, MA, USA

[h]Harvard Graduate School of Education, Boston, MA, USA

[i]Department of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles, Los Angeles, CA, USA

[j]University of Washington and Seattle Children's Research Institute Center for Child Behavior, Health, and Development, Seattle, WA, USA

[k]Departments of Biostatistics, Statistics and Psychiatry, University of California, Los Angeles, CA, USA

Abstract

Animal models of autism spectrum disorders (ASD) contribute to understanding of the role of genetics and the biological mechanisms underlying behavioral phenotypes and inform the development of potential treatments. Translational biomarkers are needed that can both validate these models and facilitate behavioral testing paradigms for ASD in humans. Automated video tracking of movement patterns and positions recorded from overhead cameras is routinely applied in behavioral paradigms designed to elicit core behavioral manifestations of ASD in rodent

[*]Corresponding author at: Duke Center for Autism and Brain Development, Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, 2424 Erwin Road, Suite 501, Durham, NC, 27705, USA., maura.devito@dm.duke.edu (M. Sabatos-DeVito). [**]Corresponding author at: Yale Child Study Center, 230 S Frontage Rd, New Haven, CT, 06519, USA. james.mcpartland@yale.edu (J.C. McPartland).

[1]Mailing Address: DUMC 2737, Durham, NC 27710.

models. In humans, laboratory-based observations are a common semi-naturalistic context for assessing a variety of behaviors relevant to ASD such as social engagement, play, and attention. We present information learned and suggest guidelines for designing, recording, acquiring, and evaluating video tracking data of human movement patterns based on our experience in a multi-site video tracking study of children with ASD in the context of a parent-child, laboratory-based play interaction.

Keywords

Autism spectrum disorder; Parent-child interaction; Play-based assessment; Automated behavioral assessment; Video tracking

1. Introduction

Rodent models of autism contribute to our understanding of the etiology of autism spectrum disorder (ASD); identify biological mechanisms underlying behavioral manifestations of ASD (Kas et al., 2014; Wöhr & Scattoni, 2013); and offer insights into potential treatments for ASD. Behavioral phenotyping assays for animal autism models (Kas et al., 2014; Silverman, Yang, Lord, & Crawley, 2010; Wöhr & Scattoni, 2013) can assess the presence of core autism symptoms including social interaction (e.g., three-chambered social approach task; Moy et al., 2004; Yang, Silverman, & Crawley, 2011) and restricted, repetitive, and stereotyped behavior patterns (e.g., reversal learning tasks like T-maze and Morris water maze; Moy et al., 2007). Such tasks typically employ manual coding quantify behaviors; however, advances in video tracking technology have made it possible to automatically quantify physical movement, interactions, and behaviors of animals (e.g., Noldus, Spink & Tegelenbosch, 2001). There is a need for translational behavioral biomarkers that can validate animal models for autism in humans. A recent study applied video tracking to children with ASD in the context of a parent-child interaction to explore this technology's ability to measure core autism features and symptom severity (Cohen, Gardner, Karmel, & Kim, 2014). Automated video tracking technology offers the potential to yield more precise, reliable, and objective quantifications of human behaviors than manual, observer-based coding. However, further work is warranted to determine methodological guidelines for the use of video tracking in laboratory-based human behavior research.

Automated tracking systems such as Noldus EthoVision XT were designed to record the environment from an overhead camera and automatically quantify the movement and behavioral interactions of animals (for an overview, see Noldus, Spink, & Tegelenbosch, 2001). Such automated technology can be applied to experimental methods that are designed to probe constructs that are relevant to human behavioral research such as anxiety and social approach. For example, Simon, Dupuis, & Costentin (1994) used a video tracking system in the context of an open field experiment to quantify thigmotaxis (a tendency to stay near walls when stressed) as a measure of anxiety in mice. Hong et al. (2015) used automated methods including video tracking and machine learning to quantify social behaviors during interactions between rodents.

Noldus EthoVision XT video tracking technology was recently adapted to measure movement patterns of children with ASD in the context of a parent-child interaction as an index of core autism symptoms of social-communication and repetitive behaviors (Cohen et al., 2014). Based on informal observations that children with ASD often remain at a distance from their parent and move about the periphery of a room, Cohen et al. (2014) set up a parent-child play task using EthoVision XT software to quantify these behaviors (e.g., proportions of time and latencies to first approach the parent region or the periphery/border region of the room). Young children (ages 2.5 to 8 years old) with ASD, developmental delays, anxiety, or attention deficits wore a red shirt and were recorded from an overhead camera during two 3-minute parent-child play sessions. EthoVision XT color detection methods tracked the child's red shirt, thereby recording the x, y coordinates of the child. Latency to first approach the parent and the proportion of time spent in the periphery were associated with parent-reported and clinician-rated measures of social-communication and repetitive movements. Cohen et al. (2014) suggested that automated video tracking of children's movement behavior during parent-child play may serve as an objective, quantitative measure of early behavioral signs of core ASD features. They further suggested that these measures could be used as indicators of change in response to treatment and could be applied consistently across research sites for large-scale trials.

Laboratory-based, naturalistic contexts such as parent-child play offer an opportunity to observe and quantify core social-communication and repetitive/restrictive autism symptoms and track behavioral change in response to pharmacological and behavioral treatments. Play-based parent-child interaction tasks are an especially fruitful method of behavioral measurement, as they have been used to elicit a wide array of complex psychological constructs, including attention (e.g., Ruff & Lawson, 1990); maternal sensitivity (e.g., Berry, Blair, Willoughby, Granger, & Mills-Koonce, 2017); joint engagement (e.g., Adamson, Bakeman, & Deckner, 2004); and self-regulation (e.g., Jahromi et al., 2009). Intervention studies for children with ASD have also used parent-child interaction tasks to assess change in targeted social-communication behaviors (e.g., Kasari, Gulsrud, Wong, Kwon, & Locke, 2010; Kasari, Gulsrud, Paparella, Hellemann, & Berry, 2015). Furthermore, such tasks are non-invasive, can be applied with minimal risk to human participants, and can be adapted for participants representing a wide range of ages and cognitive and language levels. For these reasons, naturalistic observations are an ideal method for studying typical and atypical development, but current measurement approaches are not scalable for large-scale research.

Naturalistic lab-based methods traditionally require researchers to quantify behavioral patterns through manual coding procedures, a process that is time- and labor-intensive. Accurate coding also relies on quality video recordings with good camera angles, which can be hard to obtain due to frequent movements and positioning of parent and child. Automated video tracking technology has potential to yield precise, quantitative, objective, and scalable measures of human movement patterns that index ASD core symptoms and related behaviors, and offer a translational behavioral method that might validate animal models for autism if appropriately adapted for human participants.

Recent behavioral phenotyping efforts for rodent autism models have focused on measuring multiple core autism symptom domains (social, communication, and RRB) at once, yielding

a composite autism severity score (El-Kordi et al., 2013). This composite score was highly accurate and was deemed a potentially useful measurement approach for treatment studies compared to behavioral paradigms that target only one symptom of autism (Wöhr & Scattoni, 2013). In human behavioral research, parent-child play-based tasks offer the potential of eliciting behaviors across multiple autism-relevant domains at once. Applying video tracking to this context may offer the opportunity to identify an autism severity composite for children with ASD.

The application of video tracking technology to human movement presents methodological challenges and considerations that need to be addressed to ensure that the method is reliable and valid. The goal of this paper is to detail methodological concerns and suggest guidelines for applying video tracking technology to child behaviors. These recommendations were developed based on our application of the Noldus EthoVision XT (Version 11.5; Noldus Information Technology, 2015a,2015b) video tracking system during a parent-child play interaction with school-aged children with ASD and typical development in the NIH Autism Biomarkers Consortium for Clinical Trials, a multi-site study to develop biomarkers and automated measures of social-communication. Although these guidelines were generated from that context, they have the potential to be flexibly applied to other clinical populations and other types of tasks and interactions that probe a range of behaviors, including but not limited to self-regulation, repetitive movement patterns, and hyperactivity.

We organized our guidelines into five sections: Developing a protocol; Choosing and setting up recording equipment; Setting up the video tracking arena; Identifying and troubleshooting factors that can influence data pre-acquisition; and Identifying and troubleshooting factors that can influence data post-acquisition.

2. Developing a protocol

Optimal contexts for video tracking have been developed in animal research, but are largely unexplored in human behavioral research. Here, we detail several considerations when developing a protocol to measure child behaviors in the laboratory. While our examples are specific to our experience of applying video tracking in the context of a parent-child, play-based task probing social approach and exploratory behaviors, video tracking has the potential to be applied to a wide range of tasks and interactions.

2.1. Task design and constructs

First and foremost, the behavioral interaction or paradigm must be developmentally appropriate for the age range of the children being assessed in order to successfully elicit the psychological constructs and observable behaviors of interest. For example, to capture children's joint engagement behaviors (i.e., shared attention to people and objects), the investigator's target population might be children who are developing joint attention, such as typically developing toddlers and preschoolers (Adamson et al., 2004) and children with ASD (Adamson, Bakeman, Deckner, & Romski, 2009; Adamson, Deckner, & Bakeman, 2010). Task length must provide enough opportunity for the subjects being tracked to demonstrate the behaviors of interest. We implemented a 6-minute parent-child play session with school-aged children both to allow enough time for the participants to acclimate to the

environment and demonstrate behaviors of interest (social approach, object exploration) and to remain consistent with Cohen et al.s' (2014) task, which involved two 3-minute play sessions.

Second, it is critical to consider how the psychological constructs and target behaviors elicited by the task might be meaningfully measured in terms of physical activity and location. The degree to which a child and parent need to be physically close to successfully interact depends on the nature of task activities and the child's social-communication behaviors. Some shared activities may not require proximity for high levels of interaction (e.g., racing cars or rolling a ball back and forth). If the child has sophisticated language, s/he could be highly engaged in conversation with the parent without being physically close. Thus, it may also be important to consider using additional data sources, such as measures of speech production. In sum, when developing a task to be measured with video tracking, consider the developmental level and age range of the sample, and the ways in which constructs and behaviors of interest are elicited through the tasks and stimuli and manifested in terms of physical location and movement.

Third, it is critical to consider all aspects of task administration, particularly task instructions. For parent-child interactions, parents are typically given specific instructions on how to carry out the task protocols, and parents' ability to carry out those instructions will vary, affecting the nature and quality of the data collected. Instructions must be delivered in a clear and standard manner and should include an opportunity for participants to clarify expectations. Furthermore, compliance with task protocol should be monitored throughout the assessment and standardized instructions for responding to parental noncompliance should be developed.

2.2. Assessment room design

Once the paradigm is developed and the psychological constructs, target behaviors, and video tracking variables are identified, the assessment room set-up must be designed. If the investigator is exploring video tracking variables related to regions of interest (ROIs), then room design decisions include selecting locations and sizes of ROIs in the video image, with respect to the room size and shape as well as fixed features of the room. Ultimately, the size and shape of the room(s) will significantly impact and inform the number and placement of ROIs. Furniture and task stimuli must also be selected, and placement of these items within ROIs must be considered carefully, as the location of these items will impact child movement. Ideally, it is important to consider all aspects of a video tracking protocol (EthoVision XT variables, tasks, stimuli, ROIs, room size, shape and fixed features) in an iterative manner.

The details of each ROI (size, shape and location) should be defined so that movements in that area represent a single behavior of interest, thus creating optimal measurement specificity. The size of an ROI directly affects specificity of a metric. For example, an ROI meant to measure parent-child interaction rather than independent play must be large enough to encompass both the child and the parent, but small enough to exclude non-interactive behaviors. If the ROI is too small, the child could be measured outside of the ROI when interacting with the parent, thus decreasing specificity. If the ROI is too large, behaviors

outside the scope of interactions with the caregiver (e.g., independent play) could be captured in the variables, also decreasing specificity.

ROI placement and spacing can also impact measurement specificity. If regions that are meant to represent different behaviors/constructs (e.g., a social interaction region with parent and a non-social activity region with toys) are placed too close to one another, a child could be tracked by the video tracking system in one region (e.g., sitting in the social interaction region), while actually conducting activity that is meaningful to the other region (e.g., playing independently with toys in the nonsocial activity region). Therefore, it is important to have enough space between regions so that resulting latencies, durations, and frequencies of entries into each region are distinct and meaningful (see Fig. 1, right image, as an example of a room with sufficient space for distinct center toy and parent ROIs.)

ROI location with respect to fixed features of the room (e.g., wall-mounted video recording equipment, shelving units, windows, mirrors, doors) must also be considered to optimize metric specificity. Such features can significantly influence children's movement and location, making the underlying motivation for being in a ROI unclear. For example, if a parent ROI is placed near a window or mirror, the child could enter that region to interact with the parent or to look in the mirror or out the window. If a shelving unit is part of the parent region, the child could use it to facilitate toy play with the parent or independently. In either case, the child's location would contribute to measurement in the parent ROI. If placing a ROI near a potentially confounding fixed feature of the room is unavoidable, it is important to control that fixed feature of the room as much as possible (e.g., covering windows or mirrors with shades, blocking off shelving units).

Standardizing the protocol across different recording rooms poses additional challenges. Different room sizes can produce different distances between ROIs across rooms, impacting video tracking variables such as latencies and frequencies to enter each ROI, as well as distance measures. Larger rooms or square/rectangular rooms will have a greater distance between ROIs and potentially more distance between parent and child (see Fig. 1). In contrast, smaller rooms or rooms with unconventional features will restrict the placement of distinct ROIs, potentially leading to overlapping regions, which could impact the accuracy and meaning of resulting ROI variables (see Fig. 1, left image). One way to control for differences in room size and shape across multiple sites is to standardize placement of ROIs, furniture and stimuli, but customize the size of the ROIs proportional to each room's size and configuration (see Fig. 2).

ROI location must also be considered with respect to the overhead camera's placement and field of view. If an ROI is placed in a region of the room outside of the overhead camera's view, then child movements and behaviors in that ROI will not be tracked.

To inform decisions about ROI location and size, measure several ROI sizes and conduct pilot recordings of children engaged in the task. Using the video tracking software, the ROI iterations can then be applied to the recordings and compared to manual coding of the behavior of interest. Review of the recordings and manual coding will demonstrate the degree to which the behaviors of interest for each ROI iteration are captured. Decisions on

the minimum and maximum sizes of any ROI will be determined by behaviors of interest, which will be constrained by the room's size and shape, neighboring ROIs, furniture and stimuli.

Finally, the location of the assessment room within the lab space is important. The room should be as free of external noises and distractions as possible. Children's attentional focus and engagement in tasks can be disrupted by outside noises (whether from outside the building, hallways, or behind a one-way observation mirror) with subsequent impact upon movement and location in the room.

In summary, when designing a protocol for video tracking human movement, it is critical to consider the following factors: 1) sample characteristics, including age, language, and developmental levels; 2) the constructs of interest from both a psychological/behavioral perspective (i.e., human coding outcomes) and a movement/activity perspective (i.e., video tracking outcomes); 3) the stimuli and furniture selection and placement; and 4) the number, size, and locations of ROIs, with consideration for room size, shape, location and fixed features of the room(s). All of these factors will interact and impact the quality and interpretation of the data.

3. Recording equipment and set-up

When designing a human video tracking protocol, in addition to designing a developmentally appropriate task with appropriate physical room set-up, it is important to consider the choice of camera equipment and lens and the placement of the camera in the assessment room ceiling. Prior to initiating data collection, video recordings must be evaluated to ensure that the overall video quality is sufficient for the parameters of the experiment. Careful consideration of camera resolution, camera stability, color reproduction, image distortion, and room lighting should be conducted. It is particularly important to monitor the stability of the camera placement in the ceiling throughout the study because once the software's arena is set, any changes to the view of the assessment room will impact data accuracy. All of these factors contribute to the software's ability to reliably and accurately track the participant during the task.

3.1. Equipment selection and distortion

In this section, we discuss the impact of equipment choice and placement on data quality with a particular focus on image distortion and highlight the importance of camera stability in producing quality recordings and data. Our recommendations are based on our experience of video recording with Gigabit Ethernet (GigE) camera (Basler acA1300–60gc) with a varifocal lens (Computar ½ inch 4.5–12.55 mm f1.2), recorded at 30 frames per second (fps) using Noldus Media Recorder software (Media Recorder, Noldus Inc. Netherlands). While we down-sampled the frame rate in subsequent automated analysis, we recommend recording at 30 fps to facilitate manual review of the play session. We also recorded video and audio from wall mounted side cameras, which helped corroborate the movements captured by the automated software tracking, and helped assess participants' understanding of task instructions.

Here, we provide illustrations of potential limitations and errors related to camera height and offer suggestions to optimize recordings. When using a video tracking system for a parent-child interaction task, it is important to capture the entire assessment room in the field of view of the camera so that the participants can be tracked throughout the experiment. Cameras that satisfy the requirement of capturing the entire assessment room will have a wide-angle lens. However, wide-angle lenses are subject to spherical (also known as barrel) distortion, where image magnification decreases with distance from the optical axis. In other words, this distortion follows a radial pattern from the center (Lee, Lee, & Choi, 2009). All images are also subject to perspective distortion. Perspective distortion results from the projection of three-dimensional space onto a two-dimensional image, whereas spherical distortion arises from fixed optical properties of the lens. Perspective is dictated by the distance of objects from the camera, such that closer objects appear magnified and farther objects appear smaller. Altering the perspective in any way, whether by moving the camera closer to the participant or tilting the camera on its axis, will also alter the relative distances of objects to the camera. Therefore, the degree of overall distortion varies with camera height relative to child's height. This change in perspective will also impact spherical distortion. A distorted image creates inaccuracy in the tracked position of the participant, which translates to downstream effects on the accuracy of dependent variables such as distance measures or entries into ROIs. Therefore, it is important to understand both the degree of distortion and its effects relative to the specific parameters of the experiment. It is recommended that the camera be placed high enough to minimize spherical distortion and that the axis of the camera be placed perpendicular to the arena surface to prevent perspective distortion.

3.1.1. Illustration of image distortion pattern—To characterize the distortion pattern of our ceiling mounted camera, we used a ~5 ft × ~5 ft fabric checkerboard calibration grid in the room to compare where points lie on the grid in the image with where points should be in the real physical space, based on the distances of the dimensions of each square on the grid. Due to a slight manufacturer defect in the fabric, each "square" within the calibration grid measured 7.25 × 7.5 in. with the shorter edges consistently in the same direction (true total dimensions 58 × 60 in.).

The calibration grid was placed in each of four quadrants of an assessment room (10 feet, 6 in. × 10 feet) by aligning the inner corner of the grid with the center of the room. This process was repeated at both the floor level and at a raised level of 29 in. to estimate perspective distortion as affected by the height of the participant being tracked relative to the camera. Twenty-nine inches was chosen as an approximation of the height of the tracker based on the average height of the participants being tracked in the ABC-CT study (children between 6 and 11 years of age who might sit, stand, or kneel on the floor). A static image of the calibration grid in each quadrant of the room, at both the floor and raised levels, was captured with the overhead camera. The four images were then merged using the photo-editing program GNU Image Manipulation Program (Version 2.8; Kimball & Mattis, 2012) to obtain a single composite image. Fig. 3 shows the marked composite image at the floor and 29-in. raised levels.

Points on the calibration grid were selected to quantify and characterize distortion (marked with a green cross on the composite image, Fig. 3). Pixel coordinates of selected points were obtained from original, unedited images rather than the merged composite image to avoid any inaccuracies introduced by image editing. The center pixel of the image was marked with a blue cross and deviated a few inches from the center of the room.

To estimate the distortion pattern of the wide-angle lens and demonstrate relative differences in distortion at the floor and raised levels, a conversion factor was calculated between pixels on the image and real-world measurements. The calibration grid was positioned under the camera such that the center of a single square of the pattern was aligned with the center pixel of the image, where distortion was negligible. The conversion factor was then computed from a comparison of the known real-world dimensions of the square and measured pixel dimensions.

The distortion of a wide-angle lens worsens radially from the center (Lee et al., 2009); thus, maximal distortion is expected at locations furthest from the center in the corners of the room. Distortion error between the measured distance in pixels on the image and the known distance based on real world measurements was calculated as the proportion $\frac{|D_{IMAGE} - D_{REAL}|}{D_{REAL}} \times 100$ (Edmund Optics Inc., 2019).

The analysis of the calibration grid allowed for better characterization of lens distortion from the overhead camera. In our analysis of maximal distortion, we used the known 10.5-inch diagonal of a single square of the calibration grid as a representative short distance to examine in the periphery and center of the room. The distortion was characterized by the percent error between the measured distance on the image and the known distance. At floor level, the maximal distortion in the corners of the room caused an underestimate of the distance by 4.7 in. over the 10.5 in., a 45% error. Near the center of the room, at floor level, the error was less than 5%. At the raised level, the maximal distortion in the corners of the room caused an underestimate of the distance by 6.1 in. over the 10.5 in., a 58% error. Near the center of the room, at the raised level, the error was less than 5%.

To further illustrate the relationship between height and distortion and the overall distortion pattern in the room, percent distortion was calculated for each of the selected points on the calibration grid with respect to distance from the center pixel. The distortion at each point was then plotted against the true distance from the center pixel in inches as shown in Fig. 4. The distortion curves for both the floor and raised levels are shown. A polynomial trendline was applied as an approximate model to aid in visualizing the distortion relationship (Lee et al., 2009).

In our recording environment, distortion increases moving radially away from the center pixel of the camera and is increased at the raised level for equal distance from the center pixel. The effect of height of the marker on distortion suggests that to minimize distortion the camera should be placed as high above the participant as possible.

3.2. Camera stability

Data quality can also be affected by unstable camera placement. A camera that is mounted in the ceiling can shift for several reasons, including: physical features of the room and building, such as weak room walls, loose ceiling tiles, and pipes in the ceiling; sudden movements, such as opening and closing doors harshly; and other disruptive events such as replacing ceiling tiles near the camera or throwing an object at the camera. Any movement of the camera will introduce inaccuracy in the position of the participant relative to software-defined ROIs. Therefore, it is important to mount the camera in a stable manner in the ceiling. Section 5.3 provides suggestions on how to detect and correct for shifts resulting from an unstable camera.

4. EthoVision XT arena construction

To analyze movements recorded from an overhead camera, a digital representation of the arena must be constructed. We defined our arena as the area on the image where the participant is to be tracked by the software. At the most basic level, arena construction includes grabbing a static image of the assessment room, calibrating the image with the real-life measurements of the room, and then defining the boundaries of the arena on the image. The arena may also include more complex features such as specific ROIs and points, depending on the defined variables of interest. ROIs refer to areas, or zones, delineated by the researcher, which can be used to calculate variables related to location and time. Points refer to specific fixed locations anywhere in the room that can be used to generate variables such as distance.

Due to the distortive nature of the recording equipment outlined in Section 3.1, any characteristic of the assessment room that will be defined in the arena must be physically measured. For video tracking human movement, the participants' location can be detected using a color marker such as a red T-shirt. When setting up the arena in EthoVision XT, calibration lines, ROIs, and points should be delineated at the height of the participant's marker. This section (Arena construction) is composed of two parts. The first illustrates the importance of calibrating and measuring zones at the anticipated height of the tracked marker, and the second provides suggestions on how to achieve this.

4.1. Impact of marker height on total distance moved

As demonstrated in Section 3.1, measurement accuracy is impacted by distortion and is worse when the marker is closer to the lens. Therefore, in this section we characterize the effect of marker and calibration height on the EthoVision XT variable "Total Distance Moved" (TDM), defined as the length of the vector connecting two sample points calculated using Pythagoras' theorem (Noldus et al., 2001).

The EthoVision XT 11.5 reference manual (Noldus Information Technology, 2015a,2015b) recommends calibrating at the height of the tracker on the participant. To demonstrate this point, we calibrated the room at a height of 29 in. To illustrate the impact of tracking above and below calibration height, we used EthoVision XT to track a red marker at three heights along a standardized path: 1) 1 inch (floor level); 2) 29 in. (calibration height); and 3) 41 in.

(12 in. above calibration height). We used the EthoVision XT dependent variable TDM as the outcome measure used to determine the height at which tracking accurately matched the measured distance of a standardized path. We expected that the TDM of the marker at 29 in. would show the smallest percent difference from the actual measured distance because the marker height is equal to the calibration height.

To delineate a standardized path and measure its true distance, sample points were created using nine 7.25-inch diameter circles cut from white card stock, and marked with a red dot using the photo editing software GNU Image Manipulation Program (Version 2.8; Kimball & Mattis, 2012). The circles were taped on the floor to create a standardized path and the distance between each red dot on the circles was measured. The true distance of the path, measured as the sum of the distances between each red dot along the path, equaled 670.56 cm (see Fig. 5).

We created markers at three heights: 1 in., 29 in., and 41 in.. To create the floor (1-inch) level marker, we used a black 1-inch thick foam disc, and marked the center point with red tape. To create the 29-inch marker, we placed a small piece of red tape on top of a 28-inch table leg and mounted this table leg on top of the red center of the 1-inch disc. The 41-inch marker was created by stacking the 28-inch table leg on top of 13 1-inch black discs. These markers were moved to each step in the standardized path and left there for several seconds for recording and tracking from the overhead camera. Fig. 6 illustrates the marker at each height on the first step of the standardized path.

To prepare the recording for processing in the EthoVision XT software, the recording was trimmed into three separate recordings, one for each marker height, using Solveig MM Video Splitter software (Version 7; Solveig Multimedia, 2019). The three recordings were then trimmed again so that transitions between steps in the standardized path were eliminated and so that each step in the path was 1 frame in length. This was done to standardize the trimmed videos and to reduce errors in the TDM variable.

TDM was then calculated by running the trimmed videos through EthoVision XT 11.5. Arena settings consisted of horizontal and vertical calibration lines marked and measured at 29 in. No other arena settings were used. Detection settings were set to track the color red. There were no additional settings applied and/or edits made to tracks. The TDM of the tracking marker at the three heights was compared to the true distance of the path (670.56 cm), and absolute and percent differences were calculated. For each marker height, Table 1 provides a summary of EthoVision XT’s calculation of TDM, the absolute difference between the measured distance and EthoVision XT’s TDM variable, and the percent differences.

These results indicate that EthoVision XT’s distance measures are most accurate when tracking occurs at the height of the calibration line and that inaccuracy increases as tracking moves away from the calibration height. These results also demonstrate that percent difference between the true and calculated distance measure is magnified as marker height increases above the calibration line due to worsening distortion. The results suggest that to ensure that a marker is accurately tracked in a ROI, or that a point of interest is on the same

plane as the marker, ROIs and points should be measured at the anticipated height of the marker. To determine the anticipated height of the marker for human movement tracking, it is important to consider the range of heights of the participants as well as the potential postures (sitting, standing, lying down) they may assume during the task.

4.2. Measuring calibration lines, regions of interest (ROIs), and points

Setting up the arena features (i.e., calibration lines, ROIs, and points of interest) at the height of the marker requires delineating the arena features in the assessment room using a structure that is equal to the desired height. The structures used to outline the arena features in the assessment room are then used to calibrate and draw features in EthoVision XT. The following methods make two assumptions: 1) the arena consists of the entire assessment room (i.e., everything visible to the overhead camera), and 2) the tracking method is color marker detection for human participants wearing a specified color shirt.

4.2.1. Overview—To measure and mark arena features in the assessment room, we set up physical indicators at the anticipated marker height. To create physical indicators, we used a combination of wall markings with painter's tape and columnar structures made from Styrofoam. By charting out the ROIs in the room with these indicators, we ensured that the features of the arena drawn in EthoVision XT were exact in size and location.

4.2.2. Choosing a marker height—EthoVision XT detects the center-point of a specified color marker on the participant. For children wearing a colored T-shirt, the marker is tracked at roughly the center of the child's torso. The height chosen to set the arena features should be equal to the center of the detected color on the shirt. The anticipated tracking height for human participants can be difficult to identify because marker detection varies during tracking due to differences in children's physical height, changes in physical postures (e.g. standing, sitting, laying down), and the location of the marker on the child (e.g. shoulders, back, chest).

Concerning posture, if participants are predicted to spend most of their time sitting on the floor, experimenters may set up the arena at a lower height than if their sample was expected to spend more time standing. Regarding location of the marker, experimenters should consider where on the shirt the center of the marker most often tracks and how the marker location might change across participant postures. For participants wearing a colored shirt, the marker is often tracked in the middle of the torso when participants are leaning or lying down, but on the shoulders when the participant is sitting upright. This may be a limitation for studies that are interested in capturing limb movements, such as reaches and leans into ROIs. Researchers interested in measuring reaching and leaning movements may opt to use a long-sleeve colored shirt as the marker. To reduce variation in the location of the marker detection, consider localizing the marker on the child (e.g., using red tape on the shoulders rather than having participants wear a colored shirt).

4.2.3. Determining calibration lines—The first step in measuring the arena features is to decide on the location and number of calibration lines. It is recommended that more than one calibration line be used to minimize drawing error, and that lines should be horizontal

and vertical to compensate for pixels that are not perfectly square (Noldus Information Technology, 2015a,2015b). For each calibration line, two indicators are set at the chosen height on opposing walls. The true physical distance between the indicators is measured and documented, then applied in software to a still image recorded from the overhead camera. We found that the most efficient way to set up physical indicators for calibration lines is to place painter's tape on the wall, ensuring that it is large enough to be seen from the overhead camera (see Fig. 7). If the indicators on the wall fall outside the view of the overhead camera, two columnar structures can be placed opposite one another at the edge of the camera's visibility (see Fig. 7).

4.2.4. Setting up ROIs and points of interest—The location, number, and size of the regions and points of interest will be specific to each project's paradigm, dependent variables, room configuration, and participant sample. When needed, a combination of wall markings and columnar structures may be used to set up indicators that delineate the boundaries of the chosen ROIs and the location of points of interest. Only one columnar structure is needed to create indicators for points of interests. For ROIs, the more indicators used to outline the regions, the more accurately the region will be drawn in software. Once the indicators are set, a recording that captures the indicator configuration for all ROIs and/or points of interests is acquired using the overhead camera (see Fig. 8).

4.2.5. Drawing the arena—Next, recordings are imported into software. Using EthoVision XT arena setting tools, images of the arena features are "grabbed" from the recordings and the ROIs are drawn using lines and/or shapes to connect the physical indicators (see Fig. 9). If real room measurements and physical indicators of arena features are not used to develop the arena, the arena features may be drawn disproportionate to the desired real-world size.

5. Pre-acquisition considerations

Arena setup is important to ensure that the ROIs and points of interest accurately capture the dependent variables of interest. However, an ideal arena setup does not guarantee good data. Important considerations before software acquisition of track data are detection settings, frame shifts, and review of the recordings for circumstances that compromise ROIs and/or points of interest. We explain these factors below, and provide suggestions for improving and correcting for any potential threats to a clean track.

5.1. Color marker detection

EthoVision XT 11.5 includes two types of detection of color marks: marker assisted identification and color marker tracking. Marker assisted identification is specifically designed for rodents, whereas color marker tracking is designed for a broader array of species (Noldus Information Technology, 2015a,2015b). Color marker tracking does not collect information on the shape and size of the subject; rather, it detects a predefined color, estimates a center point of the detected color marker, and records the position of the participant based on this point.

EthoVision XT allows the user to adjust the thresholds for color detection to a limited portion of the color spectrum by adjusting hue, saturation, and brightness to match the chosen color marker. In general, colors should be chosen with respect to the lighting in the room to ensure that they don't appear too dark due to dim lighting, or too washed out due to bright lighting.

It is important to control the colors present in the environment (e.g., toys, clothing, accessories). Given that our color marker was red, we controlled color in the environment in the following ways: 1) by selecting toys that did not contain red or similar hues (orange, pink); 2) by not allowing extraneous items into the assessment room; and 3) by requesting that the parent not wear red, pink, or orange clothes. To ensure that the selected toys would not be tracked with our detection settings, we conducted color testing procedures. This was done by placing all the toys in the assessment room, recording an overhead video, and then analyzing the video in EthoVision XT to see if any of the toys were detected. If a toy was tracked with the red detection settings, it was replaced. If a parent wore red, pink, or orange, we covered larger items or removed smaller items before s/he entered the assessment room (e.g., if a parent was wearing a red shirt, we covered it with a study-provided sweatshirt or blanket; if the parent was wearing red socks or shoes, we had them remove or cover them).

Next, the type of color marker needs to be determined. Characteristics of the participants and detection accuracy goals should be considered in tandem when selecting a marker for human tracking. Color markers on humans could include articles of clothing (shirts, socks, shoes, hats) or a localized marker (a sticker or tape). To track children with ASD, we chose red, short-sleeved, tag-less T-shirts as our color marker. First, EthoVision XT allows for a minimal marker size measured by pixels; therefore, the size of the red shirts served as a discriminant compared to the size of other solid-color objects in the room, allowing for further detection accuracy (i.e., tracking the child's shirt rather than a smaller, extraneous red item in the room). If it was important to the research project to track reaching and leaning or other arm movements, then a long-sleeved shirt or a color marker on the arms would be appropriate. Second, markers placed below the waist were excluded because these are likely to be hidden from the overhead camera (e.g., if the child sits at the table or sits on his/her knees). Furthermore, children with ASD often exhibit sensory sensitivities to textures like tags and to the sensation of being touched on the head as well as difficulty with transitions. Therefore, we chose tagless T-shirts as they seemed least likely to trigger sensory sensitivities; could not be as easily removed by participants as hats, socks, or shoes; and could be tracked above the waist. Overall, short-sleeved T-shirts were ideal for both ease of application and for visibility of the marker from the overhead camera. However, for our study, we allowed some flexibility in standardization of the marker for children who resisted wearing the study-provided red T-shirt. In these cases, the participant was allowed to wear a personal red shirt, or red vinyl tape was applied to the shoulders and back of the child's non-red personal shirt. In our study, 3 children opted to have red tape applied to their own shirt, and these participants were successfully tracked.

EthoVision XT has the capacity to track up to 16 colors simultaneously in an arena. This can be a useful function if the project goals support tracking multiple participants in an assessment space. For example, to characterize natural interactions between a child and a

social partner (e.g., parent, peer, sibling), it may be best to not limit the interaction to a particular region of the room. To track multiple participants, follow the above-stated steps for single participant tracking (i.e., select a color, control the environment for color, select a type of marker). However, for multiple participant tracking, select distinct colors for each participant so that there is no overlap in the portion of the color spectrum selected. This will prevent the marker from mistakenly switching participants if one participant becomes obscured (e.g., a parent blocking the child by reaching over or hugging the child). Also, having two or more color markers can pose an added challenge to controlling the environment for the marker colors. The investigator will need to ensure that all participants are devoid of marker colors that are not their own, and that all objects in the room do not contain the colors, or related hues of the colors being tracked. Depending on the experiment, this consideration may limit color selection for the shirts. For example, children's toys are very commonly blue or green, so it may be easier to choose a different shirt color than eliminate all blue and green toys.

5.2. Trimming circumstances from recordings

In human behavioral research, it is important to monitor task administration in real-time and to quality review the task post-administration to ensure that protocols are being adhered to and that high quality, reliable and valid data are being collected. Deviations from or interruptions to the protocol can alter the meaning of the ROIs and the resulting EthoVision XT-generated variables. To ensure quality data are collected during a human video tracking protocol, we recommend: 1) Identifying troubleshooting procedures in the event that the task or video tracking are disrupted; 2) Creating a form to note all deviations and circumstances that arise, and how they were handled; 3) Monitoring tasks in real-time; 4) Reviewing all recordings after the task is completed; and 5) Removing unstandardized circumstances (e.g., bathroom breaks) that interrupt and threaten the integrity of the video tracking variables with video trimming software.

Participant behaviors can be difficult to control and standardize during a free-play task. Behaviors or occurrences that deviate from the task protocol can affect the meaning of the ROIs. For example, if a parent moves substantially within or out of a designated "parent ROI," there is no longer a reliable "parent ROI" from which to gather data. If the child leaves the room or shuts the lights, s/he cannot be tracked. Therefore, human video tracking sessions need to be monitored in real time and reviewed after administration. To support real-time monitoring and post-administration quality review, sessions can be recorded with side cameras with audio and video.

Monitoring the task in real-time will allow for circumstances that might alter the meaning of ROIs to be immediately fixed. A brief interruption to the task can reset the ROI quickly and regain accurate data. In deciding whether or not to interrupt the session to fix an error, it may be important to consider the tolerable length of time for such an interruption to occur. If, for example, a parent steps out of the parent ROI briefly, there may not be significant data loss relative to the total task time. However, if the parent never returns to the parent ROI or takes too long to return to it, the entire session may be invalidated. To aid in real-time monitoring of the task, two research assistants should be assigned to a session: one to administer the

protocol, and one to record and monitor the session. In this way, the camera operator can see and hear the parent and child during the task and alert the examiner to intervene if a circumstance arises that could compromise tracking or the meaning of ROIs.

During quality review, side camera and overhead recordings can be used together to identify deviations or potential threats to accurate video tracking and data quality. These circumstances should be trimmed from the video prior to running it through the software to preserve the integrity of the video tracking metrics.

For ease of processing, the beginning and end of the video tracking session can also be trimmed. Introductions to the task and assessment room and the time it takes participants to leave at the end of the task can vary in length of time. Therefore, it is important to identify events that marks the true beginning of the task (e.g., the examiner closing the door, the parent and/or child assuming a position in the room) and the true end (e.g., the examiner opening the door to end the play session) and to trim the video at those events.

5.3. Identifying and correcting frame shifts

Finally, as mentioned in Section 3.2, camera stability is critical to producing accurate video recordings and location data. If the camera moves during the course of a study, this will present as an apparent “frame shift” in the video recording where the position or perspective of the field of view will change. In such instances, it is no longer appropriate to use the original drawn ROIs for data acquisition. Therefore, it is important to monitor all recordings for potential frame shifts.

One way to detect frame shifts is to choose a fixed feature of the assessment room (e.g., trim on the floor, walls, door, power outlets, light switches) that is visible from the overhead recording, and compare the location of that fixed feature across two video images. See Fig. 10 for an illustration of a frame shift. A frame shift can be subtle, and thus hard to detect. One option for systematically checking for and detecting frame shifts is to capture a still frame image of the start of every recording before the child enters the room, and sort through those images chronologically and in succession. Another option is to compare the still frame images of the recordings by overlaying them in a photo editing program.

5.3.1. Methods of correcting frame shifts—One way to correct a frame shift is to repeat setup of real world indicators in the room and acquire new images after the frame shift occurred. New room features can be drawn using these physical indicators that are accurate to the new position of the camera. However, repeating real-world measurements in the assessment room can be a cumbersome task. Alternatively, existing reference images for drawing room features can be altered in a photo editing program to accommodate the frame shift.

We provide an example of how to correct a frame shift using the photo-editing program GNU Image Manipulation Program (Version 2.8; Kimball & Mattis, 2012) for a single ROI. First, we took a still frame image from an overhead video of the room after the frame shift occurred, and an image from a video of physical indicators for an ROI (see Fig. 10). This allowed us to align the pre-frame shift image of the delineated ROI with an image of the

assessment room after the frame shift occurred, producing a new image of the ROI's physical indicators that was then used to redraw the ROI in the software.

We overlaid the two images in software and made one image transparent (see Fig. 10). At this point, transformations should be made to align the original image of physical indicators to the frame-shifted image. The necessary transformations (rotations, translations) can be estimated using the fixed features of the room such as the trim on the floor, walls, door, or power outlets. This process will require trial and error in the absence of any advanced algorithms or other tools. Transformations are restricted to two-dimensional rotation and translational movements. If the camera is significantly tilted from its original axis, consider using a photo editing program that offers perspective editing. Perspective editing should be used with caution, however, as it is more difficult to estimate accurately and may distort the image unrealistically. For this example, a rotation of −1.15 degrees was applied, followed by a translation of 58 pixels to the left and 4 pixels down as shown in Fig. 11. Any transformations made should be recorded to use when repeating the process for other ROIs and/or points of interests.

The image is cropped to the position of the image after the frame shift as shown in Fig. 11. This produces a final image of the physical indicators that corrects for the frame shift. This process should be repeated for each arena feature. Once all arena features are corrected, the corrected images can be put into an Audio Video Interleave (AVI) video format to beimported into EthoVision XT and used to set up a new arena. All recordings post-frame shift should then be run through the new arena settings.

The above method is best used in cases of small/minor frame shifts. For large frame shifts, it is possible that some of the physical indicators will be cropped out of the final image after transformation. In these cases, it would be more appropriate to repeat setup of the physical indicators in the room and acquire new recordings. Physical readjustment of the camera can also be considered in extreme cases, although further arena adjustments will be necessary.

6. Post-acquisition considerations

Once circumstances have been trimmed and the data has been acquired, it is important to ensure that tracks accurately reflect the participant's movements and that the software is only tracking movements relevant and meaningful to the dependent variables. In support of this, two critical factors should be considered: track smoothing settings and track deviations. Part 1 introduces track smoothing methods offered in EthoVision XT and identifies potential ways to determine appropriate settings. Part 2 highlights the importance of reviewing tracks after data acquisition and identifying and correcting potential track deviations.

6.1. Track smoothing

EthoVision XT offers track smoothing methods both during acquisition (e.g., track noise reduction) and after acquisition (e.g., Lowess and Minimal Distance Moved). Smoothing options are applied to reduce various sources of noise that can impact dependent variables. Lowess is a type of post-acquisition smoothing that eliminates small body movements called "wobble" that could impact EthoVision XT's distance measures. A Lowess smoothing

parameter of 10 is recommended to increase stability of the center point during locomotion (Noldus Information Technology, 2015a,2015b).

After applying Lowess smoothing, the Minimal Distance Moved (MDM) feature can also be considered. Two MDM methods are offered in EthoVision XT: 'Direct Path' (DP) and 'Along the Path' (ATP). The DP method takes into account the shortest distance between two points. In this iterative process, if the distance between two points is smaller than the threshold, then the second point takes the value of the first point. If the distance is greater than the threshold, then the second point retains its value. ATP smoothing takes into account the cumulative distance along the path. For example, if the distance between two points is shorter than the threshold, but adding the distance to a third point increases beyond the threshold, then the first two points are grouped together and the third one retains its value (Noldus Information Technology, 2015a,2015b). Compared to the DP method, ATP most resembles the raw path taken by the participant.

When tracking animal behavior, the MDM feature is intended to reduce small movements while the animal sits still (e.g., breathing) or to reduce system noise. A short MDM setting might be appropriate for small animals, but might overestimate movement for larger color markers in humans (e.g., red shirt). Human movements such as adjusting posture, reaching or leaning momentarily, may result in futile movements of the center-point of the marker, and impact dependent variables such as frequencies in and out of ROIs. Therefore, the investigator may consider adapting the MDM feature to reduce nonmeaningful movements.

Given the primary dependent variables of interest in our ABC-CT study (i.e., frequency, duration and latency to core ROIs), two sources of futile center-point movements were identified. First, we observed that the center point jumps from one shoulder to the other when a participant's head is positioned upright. When the center point moves to one shoulder, EthoVision XT tends to track to the midpoint of that shoulder because one shoulder is seen as a contiguous area representing the participant. A slight change in orientation or movement may cause the center point to jump to the midpoint of the other shoulder. Second, sudden between-ROI movements such as leaning to reach for objects or to adopt a comfortable position may result in the center-point jumping from shoulder to shoulder, or top to bottom of the torso, thus over-estimating movement. For the ABC-CT study, we sought to reduce shoulder to shoulder center-point jumps that could impact frequency, latency and duration in ROIs. We adjusted the MDM setting for participants' estimated biacromial breadth (i.e., shoulder-to-shoulder width), which is approximately 29.5 in. for 6- to 11-year-old males and females (McDowell, Fryar, & Ogden, 2009). Shoulder-to-shoulder center-point jumps were estimated to be approximately half the biacromial breadth (about 15 cm); therefore, we applied an ATP MDM of 15 cm. An alternative way to determine optimal MDM profiles may be achieved through a comprehensive coding scheme. Steps for this process might include: identifying events to code (e.g., entries in ROI); recording an integrated visualization of EthoVision XT trials with center-point marker and no MDM; and using a behavioral coding software to mark frequencies of child entries into ROIs. Then, reliability of human and automated coded events can be assessed across different MDM settings. This range can demonstrate which MDM profiles achieve excellent agreement (ICC > .75) between human and automatic coding.

One consequence of selecting a higher MDM is that meaningful smaller movements will also be removed from the data. If small postural movements, leaning or reaching are meaningful, then investigators may consider using a short MDM threshold or none at all. If small movements are important, but shoulder-to-shoulder center-point jumps are not desirable, other options to control futile movements should be explored. For example, one could consider selecting a smaller or more localized color marker such as tape or a sticker on the participant's shirt. In this way, shoulder-to-shoulder jumps will be eliminated; however, this approach could also lead to missing data or inaccurate tracking if the marker is obscured. Three other options to reduce center-point jumps include: 1) Adjusting the sample rate; 2) Applying dilation and erosion in EthoVision XT's Detection Settings to make the two shoulders as one (Noldus Information Technology, personal communication, June 26, 2018); and 3) applying a Maximal Distance Moved filter, which is available in more recent versions of EthoVision XT. Although the EthoVision XT 11.5 reference manual and online help offer guidelines for optimal sample rates for different animal species, there are currently no guidelines for human participants. To determine the ideal sample rate for human participants, we recommend testing several sample rates on a video recording of a human participant in EthoVision XT. Then, plot the dependent variable of interest (e.g., Total Distance Moved) against sample rate. The ideal range of sample rates will align with a plateau in the graph (Noldus Information Technology, 2015a,2015b). Ultimately, decisions about track smoothing need to be guided by the investigator's definition of meaningful versus futile movements, the primary dependent variables, and core hypotheses.

6.2. Track deviations

To ensure that tracks accurately reflect the participant's movements, EthoVision XT trials should also be reviewed for errors in the tracked data (i.e., track deviations) such as the participant marker disappearing or detection of something other than the participant. This is crucial in the beginning of data collection because track errors affecting many sample points indicate a problem in the experimental setup, camera setup, arena settings, trial control settings, and/or detection settings. Identifying errors related to experimental and camera setup are especially important because these errors will persist across all trials if left uncorrected.

Some errors relating to EthoVision XT project settings (i.e., trial control, arena, detection) can be corrected at any time by creating a new project, applying the corrected settings, and re-acquiring trials. Other track errors can be addressed by using EthoVision XT's track editing feature, but should be limited to track deviations that cannot be fixed by re-running the data under new settings. The use of the track editor is time intensive, laborious, and difficult to implement in a standardized way. Thus, solutions that minimize the use of the track editor should be prioritized over immediate use of this feature. If researchers opt to use the track editor tool, protocols should be developed, and training conducted, to ensure reliability between editors and to make sure edits are applied similarly across participants and tracks.

Unique to the track editor, it is possible to review the raw data of the participant's tracked path. Raw EthoVision XT data includes the participant's position in the arena (x, y

coordinates) and the participant's size (cm^2). Track deviations are reflected in the raw data in two ways: 1) missing data (i.e., no values exist for participant size and position), and 2) inaccurate data (i.e., the values for participant position and size do not accurately reflect real-life position and size). Lastly, track deviations can be identified by reviewing the statistics for unexpected results. For example, in our results, we would not expect to see values greater than one for "Frequency in Arena" because the participant enters and remains in the room until the play session ends.

In our experience, missed samples and detection errors were the most common causes for track deviations. The main cause of detection errors were regions outside the view of the overhead camera. Some less common and less severe sources of detections errors include the use of one detection setting for all assessment sites; absence of the color marker as a result of the participant refusing the red shirt and/or removing it during the session; the marker being covered by an object; and items/features in the assessment room being detected by EthoVision XT as red. The performance variable "subject not found" is a good indicator for track deviations caused by detection errors; however, this variable cannot be reviewed in isolation because high percentages of "subject not found" may be the result of missed samples (i.e., if the samples are missing, then the participant cannot be detected).

Missing and incorrect raw data can be corrected using EthoVision XT track editing tools. We applied these tools most often to track deviations caused by participants moving out of camera view. Since our ROIs extended conceptually into these out of range regions, we decided that if the participant's location could be unambiguously determined using the overhead and side camera recordings, then we could edit the track to reflect correct ROI placement without compromising data integrity. To maintain data integrity, we created a reproducible method and criteria for editing tracks, and every track was reviewed by a second person. We used the side and overhead recordings and knowledge of each room's layout to correct the marker. When these three tools did not provide sufficient information, we did not correct the track deviations. A trial that contained more than 30 s of track deviations that could not be corrected was deemed invalid and excluded from analyses.

7. Discussion

Behavioral phenotyping assays have been applied to autism rodent models to identify biological mechanisms underlying core behavioral symptoms of autism and inform treatments for autism. Similar methods to assess behavior in humans have the potential to reliably inform clinical interventions, but traditionally rely on labor-intensive manual coding. As part of the Autism Biomarkers Consortium on Clinical Trials, a video tracking protocol was applied to children with autism and typical development during a parent-child play interaction to explore its utility as an objective, automated measure of autism. If applied accurately, such an automated approach has the potential to fill a need for translational behavioral biomarkers, and advance research targeting the biological mechanisms underlying the behavioral phenotype of ASD.

In this report, we provide methodological guidelines to consider when utilizing an automated video tracking system, Noldus EthoVision XT, to observe and measure human

movement and position. At each step in the process of designing, acquiring and processing video tracking data, there are a range of issues that the investigator should consider, including participant factors such as ages, cognitive and language abilities, and developmental levels; environmental factors including room sizes, shapes and configurations; and the psychological domains and constructs being investigated. Our recommendations are born from our experience in applying video tracking to a play-based parent-child play interaction with school-aged children with ASD and typical development in the context of a multi-site study; however, these guidelines and considerations can be applied when designing a range of tasks and interactions for video tracking human behaviors.

7.1. Summary of guidelines

Although manual coding of behavior during tasks is a well-established method of quantifying both animal and human behavior, and automated approaches have been applied to experiments measuring animal behavior (e.g., Ahern. Modi, Burkett and Young, 2009; Nadler et al., 2004), automated tracking of human movement during lab-based interactions is relatively unexplored (Cohen et al., 2010). Therefore, when planning to apply automated tracking to human movement, it is important to consider the constructs and behaviors of interest, not only in relation to the participants' age range and developmental level, but also with respect to how well the constructs are reliably and validly captured with automated movement tracking. For example, sociability measures in mice focus on proximity and movement (Yang et al., 2011), but humans' also have the ability to interact using gesture, language, and other symbolic behaviors and rely more heavily on visual cues (whereas rodents rely primarily on olfaction and touch). The props used, such as toys and furniture, will also influence participants' movements and behaviors, such as the degree of proximity necessary for successful cooperative play or social interaction. Once the task and resulting variables are defined, the assessment room must be carefully selected and designed with consideration for the room shape, size, and configuration in order to optimize the specificity of the measures. In the case of multi-site studies, standardizing the room setup and controlling environmental site differences are critical to generating reliable and valid data.

We highlighted several issues that can impact measurement specificity during human video tracking, including camera distortion, camera stability, room calibration, track smoothing, and track deviations, and suggested methods for characterizing and/or addressing these issues.

When distortion effects are present, it is important that they be characterized and methods for correcting them explored. An example of a post-data acquisition correction method is the calibration grid that we used to illustrate distortion in Section 3.1.1. One limitation of our illustration is that the center pixel of the calibration grid was not the center of distortion, and therefore, resulted in over-estimating distortion in one half of the room and underestimating distortion in the opposite half of the room. In our case, this estimation error was negligible. Algorithmic methods for distortion correction (e.g., Shah & Aggarwal, 1996; Santana-Cedrés et al., 2015) may also be applied to pre-processed videos, but we have not evaluated these approaches in the context of video tracking.

Camera placement is important to establish prior to collecting video tracking data. Camera placement can impact the stability of the recording equipment and should be monitored throughout data collection. We found that even when cameras were securely mounted on a stable ceiling fixture that limited translational and rotational movement, perceptible camera movement occurred over the course of our study. Camera shifts should be monitored, and if they occur, should be corrected.

For room calibration, determining the best marker height for human tracking can be challenging, particularly if participants' heights vary widely and if participants assume postures that create variability in the height of the marker being tracked. Thus, it may be useful to record participants' heights, measure and calibrate the room at several representative heights, and test the best calibration height.

Track smoothing settings, such as Minimal Distance Moved (MDM), may need to be adjusted for the sample and the behaviors being studied. For the ABC-CT study, we tested MDM based on biacromial breadth of school-aged participants with color tracking on their torso. Investigators should adapt track smoothing settings to gather the data that will best support their research goals. For example, if hyperactivity is the construct of interest for participants with ADHD, then sudden or smaller movements may not be considered futile.

Finally, recordings should be reviewed for tracking deviations and corrected whenever possible to preserve the integrity of the tracks and to enhance the accuracy of the data generated by EthoVision XT. Corrections can occur through real-time monitoring as well as post-administration quality review of side camera recordings and overhead videos.

7.2. Limitations and future directions

The purpose of this paper is to provide recommendations on how to apply automated video tracking to human movement so as to optimize the chances of producing meaningful, objective data. Our direct experience in applying this method is constrained to the context of the ABC-CT, which employed a play-based task for parents and school-aged children with ASD and typical development; therefore, the examples we provide are based on that task, which was designed to elicit social approach and withdrawal behaviors and exploratory play. However, our guidelines for designing a task, selecting stimuli, setting up the room, selecting variables, and setting up the recording equipment can be flexibly applied to other contexts, constructs, behaviors and tasks. Furthermore, while our methodological recommendations describe our experience with EthoVision XT, the experimental considerations we describe are largely applicable to other automated tracking methods, such as binary tree analysis (Buchanan & Fitzgibbon, 2006), probabilistic tracking methods (Rodriguez et al., 2017), deep learning approaches (Mathis et al., 2018) and radio frequency tracking technologies (Kritzler, Lewejohann, Krüger, Martin, & Sachser, 2006).

As this is a methodological paper, we have not included data analyses and results from the video tracking protocol used in the ABC-CT study. Thus, while we believe the data produced from applying video tracking to lab-based tasks with human participants (in our case, parent-child play interactions) holds promise, we are not yet able to provide a definitive conclusion about the clinical utility of this approach. Ongoing analyses explore

group differences and relations between EthoVision XT-generated outcomes and standard clinical outcomes collected from participants of the ABC-CT. As this is still an exploratory method, we recommend that researchers collect information about children's skills (language, social, cognitive, adaptive, and motor) from traditional standardized measures with known reliability and validity, such as parent reports and clinician-administered assessments, and include that information in their analysis plans.

To date, there have been no methodological guidelines on how best to apply automated video tracking to naturalistic tasks with human participants; therefore, the initial learning curve for applying this method was time- and labor-intensive. We have developed guidelines to improve the process for collecting optimal human movement data using video tracking technology and reduce the human labor and time involved in designing and setting up a video tracking protocol. Future research should explore methods for correcting issues such as camera distortion and room calibration so as to optimize the accuracy of human movement tracking data. Future research is also needed to assess the clinical meaningfulness and predictive ability of automated video tracking measures with standard clinical outcomes. Such additional research efforts will help to optimize human movement tracking protocols for ASD research that can capture automated, objective measures of autism behaviors, such as an autism severity composite, that has translational value for behavioral phenotyping assays of autism rodent models

7.3. Conclusions

Naturalistic behavioral tasks can be effectively used with children with ASD representing a wide range of ages, developmental levels, cognitive and language abilities, and symptom severity. However, traditional manual coding methods used to quantify observed human behaviors are time- and labor-intensive. Automated tracking of human behavior during laboratory-based, naturalistic tasks offers an objective, quantitative, and scalable alternative to manual coding. This paper is the first to offer methodological guidelines for designing, recording, acquiring, and evaluating video tracking data of human movement patterns. Future research can apply these guidelines to a variety of ages, behaviors, and clinical populations to more fully explore the optimal procedures for video tracking human movement and to determine its utility in advancing translational and biological research in ASD.

Acknowledgements

Support for the Autism Biomarkers Consortium for Clinical Trials (ABC-CT) was provided by NIMHU19 MH108206 to James McPartland. This research was also supported in part by the Marcus Foundation and NICHDP50HD093074.

A special thanks to all of the families and participants who join with us in this effort. In addition, we thank our external advisory board, NIH scientific partners, and the FNIH Biomarkers Consortium. We acknowledge Noldus Information Technology, Inc. for their feedback on this manuscript.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

References

Adamson LB, Bakeman R, & Deckner DF (2004). The development of symbol-infused joint engagement. Child Development, 75(4), 1171–1187. 10.1111/j.1467-8624.2004.00732.x. [PubMed: 15260871]

Adamson LB, Bakeman R, Deckner DF, & Romski M (2009). Joint engagement and the emergence of language in children with autism and Down syndrome. Journal of Autism and Developmental Disorders, 39(1), 84–96. 10.1007/s10803-008-0601-7. [PubMed: 18581223]

Adamson LB, Deckner DF, & Bakeman R (2010). Early interests and joint engagement in typical development, autism, and Down syndrome. Journal of Autism and Developmental Disorders, 40(6), 665–676. 10.1007/s10803-009-0914-1. [PubMed: 20012678]

Ahern TH, Modi ME, Burkett JP, & Young LJ (2009). Evaluation of two automated metrics for analyzing partner preference tests. Journal of Neuroscience Methods, 182(2), 180–188. [PubMed: 19539647]

Berry D, Blair C, Willoughby M, Granger DA, & Mills-Koonce WR (2017). Maternal sensitivity and adrenocortical functioning across infancy and toddlerhood: Physiological adaptation to context? Development and Psychopathology, 29(1), 303–317. 10.1017/S0954579416000158. [PubMed: 27065311]

Buchanan A, & Fitzgibbon A (2006). Interactive feature tracking using K-D trees and dynamic programming. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) 10.1109/CVPR.2006.158.

Cohen IL, Gardner JM, Karmel BZ, & Kim SY (2014). Rating scale measures are associated with Noldus EthoVision-XT video tracking of behaviors of children on the autism spectrum. Molecular Autism, 5(15), 1–16. 10.1186/2040-2392-5-15. [PubMed: 24410847]

Edmund Optics Inc (2019). Distortion (n.d.), Retrieved fromhttps://www.edmundoptics.com/resources/application-notes/imaging/distortion/.

El-Kordi A, Winkler D, Hammerschmidt K, Kästner A, Krueger D, Ronnenberg A, … Fischer J (2013). Development of an autism severity score for mice using Nlgn4 null mutants as a construct-valid model of heritable monogenic autism. Behavioural Brain Research, 251, 41–49. [PubMed: 23183221]

Hong W, Kennedy A, Burgos-Artizzu XP, Zelikowsky M, Navonne SG, Perona P, & Anderson DJ (2015). Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. Proceedings of the National Academy of Sciences, 112(38), E5351–E5360.

Kimball S, & Mattis P (2012). GNU image manipulation program (version 2.8) [software]. Retrieved fromhttps://www.gimp.org/.

Jahromi LB, Kasari CL, McCracken JT, Lee LS-Y, Aman MG, McDougle CJ, … Posey DJ (2009). Positive effects of methylphenidate on social communication and self-regulation in children with pervasive developmental disorders and hyperactivity. Journal of Autism and Developmental Disorders, 39(3), 395–404. 10.1007/s10803-008-0636-9. [PubMed: 18752063]

Kas MJ, Glennon JC, Buitelaar J, Ey E, Biemans B, Crawley J, … Noldus LPJJ (2014). Assessing behavioural and cognitive domains of autism spectrum disorders in rodents: Current status and future perspectives. Psychopharmacology, 231(6), 1125–1146. 10.1007/s00213-013-3268-5. [PubMed: 24048469]

Kasari C, Gulsrud AC, Wong C, Kwon S, & Locke J (2010). Randomized controlled caregiver mediated joint engagement intervention for toddlers with autism. Journal of Autism and Developmental Disorders, 40(9), 1045–1056. 10.1007/s10803-010-0955-5. [PubMed: 20145986]

Kasari C, Gulsrud A, Paparella T, Hellemann G, & Berry K (2015). Randomized comparative efficacy study of parent-mediated interventions for toddlers with autism. Journal of Consulting and Clinical Psychology, 83(3), 554–563. 10.1037/a0039080. [PubMed: 25822242]

Kritzler M, Lewejohann L, Krüger A, Martin R, & Sachser N (2006). An RFID-based tracking system for laboratory mice in a semi natural environment. 37th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC).

Lee SH, Lee SK, & Choi JS (2009). Correction of radial distortion using a planar checkerboard pattern and its image. IEEE Transactions on Consumer Electronics, 55(1), 27–33. 10.1109/TCE.2009.4814410.

Mathis A, Mamidanna P, Cury K, Abe T, Murthy VN, Mathis MW, … Bethge M (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. Nature Neuroscience, 21, 1281–1289. 10.1038/s41593-018-0209-y. [PubMed: 30127430]

McDowell MA, Fryar CD, & Ogden CL (2009). Anthropometric reference data for children and adults: United States, 1988–1994. National Center for Health Statistics. Vital Health Statistics, 11(249), Retrieved from https://www.cdc.gov/nchs/data/series/sr_11/sr11_249.pdf.

Moy SS, Nadler JJ, Perez A, Barbaro RP, Johns JM, Magnuson TR, Piven J, & Crawley JN (2004). Sociability and preference for social novelty in five inbred strains: An approach to assess autistic-like behavior in mice. Genes, Brain and Behavior, 3(5), 287–302. 10.1111/j.1601-183X.2004.00076.x.

Moy SS, Nadler JJ, Young NB, Perez A, Holloway LP, Barbaro RP, … Crawley JN (2007). Mouse behavioral tasks relevant to autism: Phenotypes of 10 inbred strains. Behavioural Brain Research, 176(1), 4–20. 10.1016/j.bbr.2006.07.030. [PubMed: 16971002]

Nadler JJ, Moy SS, Dold G, Trang D, Simmons N, Perez A, … Crawley JN (2004). Automated apparatus for quantitation of social approach behaviors in mice. Genes, Brain and Behavior, 3(5), 303–314. 10.1111/j.1601-183X.2004.00071.x.

Noldus LPJJ, Spink AJ, & Tegelenbosch RAJ (2001). EthoVision: A versatile tracking system or automation of behavioral experiments. Behavior Research Methods Instruments & Computers, 33(3), 398–414. 10.3758/BF03195394.

Noldus Information Technology (2015a). EthoVision XT version 11.5: Reference manual. Available fromhttps://www.noldus.com/downloads.

Noldus Information Technology (2015b). Noldus EthoVision XT (Version 11.5) [Software]. Available from https://www.noldus.com/animal-behavior-research/products/ethovision-xt.

Rodriguez A, Zhang H, Klaminder J, Brodin T, Andersson PL, & Andersson M (2017). ToxTrac: A fast and robust software for tracking organisms. Methods in Ecology and Evolution, 9(3), 460–464. 10.1111/2041-210X.12874.

Ruff HA, & Lawson KR (1990). Development of sustained, focused attention in young children during free play. Developmental Psychology, 26(1), 85–93. 10.1037/0012-1649.26.1.85.

Santana-Cedrés D, Gomez L, Alemán-Flores M, Salgado AM, Esclarín J, Mazorra L, … Alvarez L (2015). Invertibility and estimation of two-parameter polynomial and division lens distortion models. SIAM Journal on Imaging Sciences, 8(3), 1574–1606. 10.1137/151006044.

Shah S, & Aggarwal JK (1996). Intrinsic parameter calibration procedure for a (high distortion) fish-eye lens camera with distortion model and accuracy estimation. Pattern Recognition, 29(11), 10.1016/0031-3203(96)00038-6.

Silverman JL, Yang M, Lord C, & Crawley JN (2010). Behavioural phenotyping assays for mouse models of autism. Nature Reviews Neuroscience, 11(7), 490. [PubMed: 20559336]

Simon P, Dupuis R, & Costentin J (1994). Thigmotaxis as an index of anxiety in mice: Influence of dopaminergic transmissions. Behavioural Brain Research, 61(1), 59–64. 10.1016/0166-4328(94)90008-6. [PubMed: 7913324]

Solveig Multimedia (2019). SolveigMM video splitter (version 7) [software] (n.d.), Available fromhttp://www.solveigmm.com/en/products/video-splitter/.

Wöhr M, & Scattoni ML (2013). Behavioural methods used in rodent models of autism spectrum disorders: Current standards and new developments. Behavioural Brain Research, 251, 5–17. 10.1016/j.bbr.2013.05.047. [PubMed: 23769995]

Yang M, Silverman JL, & Crawley JN (2011). Automated three-chambered social approach task for mice. Current Protocols in Neuroscience, 56(1), 8–26. 10.1002/0471142301.ns0826s56.

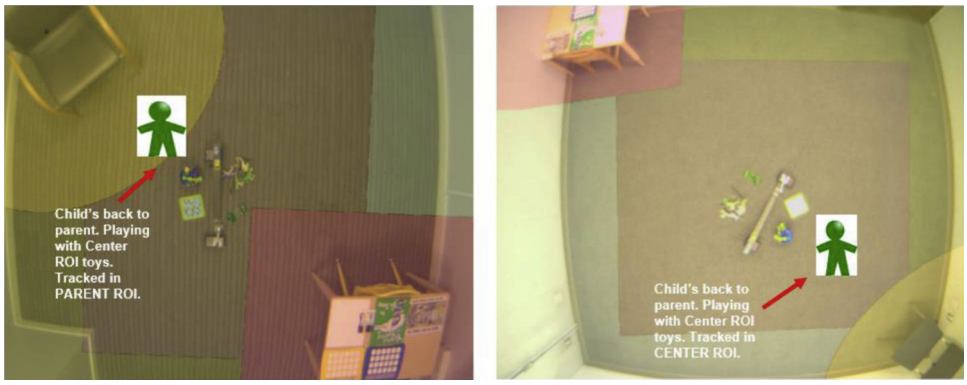

Fig. 1.

Arena with same-sized regions of interest across two different assessment rooms.

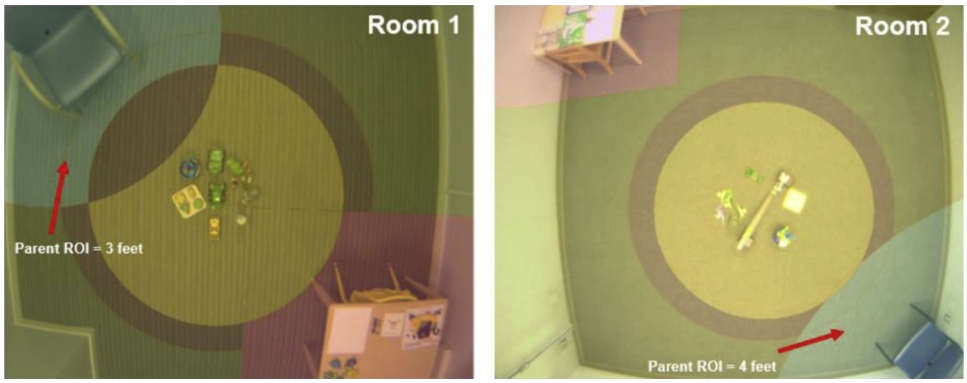

Fig. 2.

Arenas of two different assessment rooms, using site-specific sized regions of interest. The room on the left is smaller than the room on the right.

Fig. 3.

Checkerboard pattern illustrating distortion at floor level (left) and 29-in. level (right). NOTE: Green marks are used as a visual aid to show the points selected for distortion analysis. The blue cross marks the center of the room. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

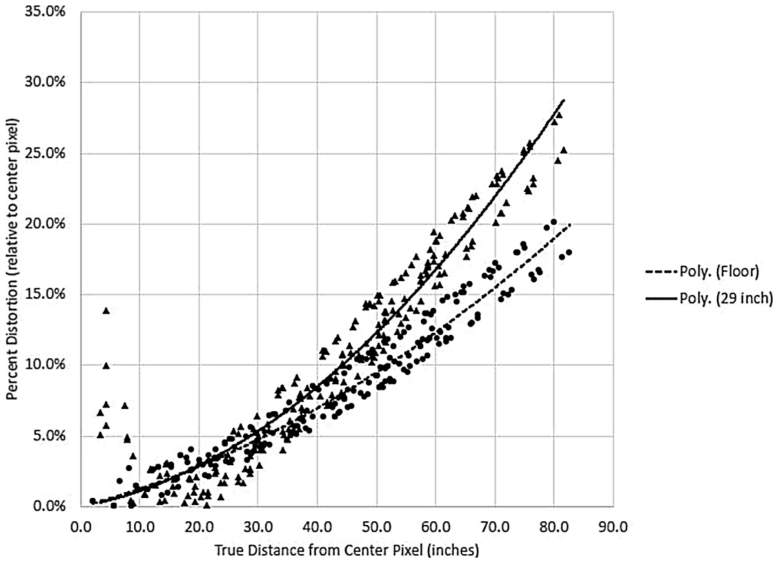

Fig. 4.

Percent distortion at each point plotted against true distance from center pixel of the checkerboard pattern placed at the floor and 29-inch levels.

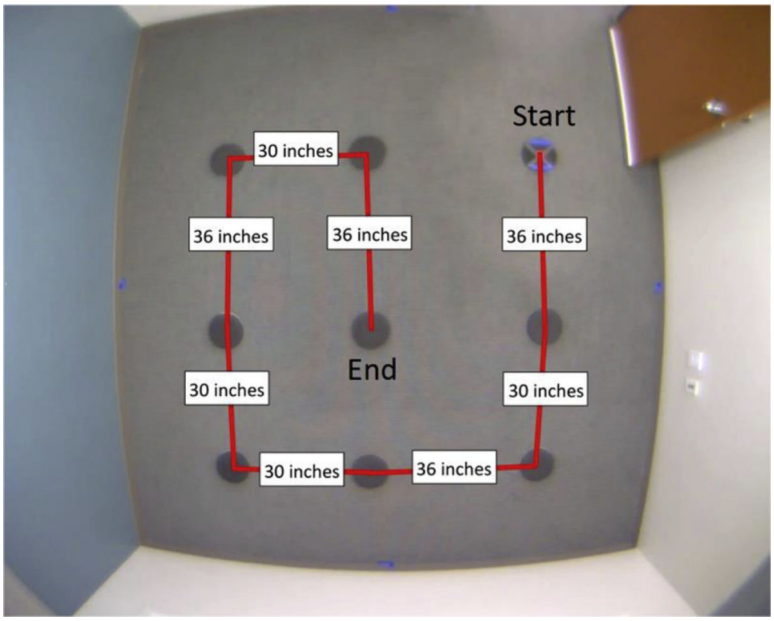

Fig. 5.
The standardized path and true measured distance between each step in the path.

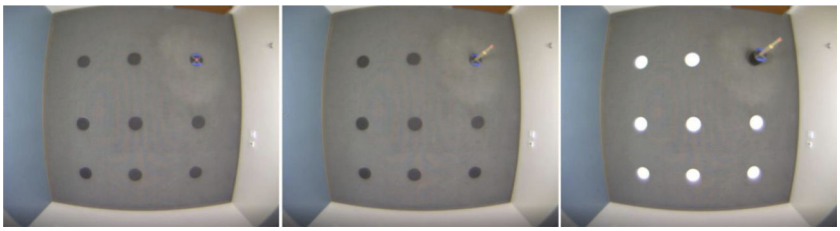

Fig. 6.

The marker disc set up on the first step of the standardized path at heights of 1 inch (left), 29 in. (middle) and 41 in. (right).

Fig. 7.

Example of calibration points on the wall (left) and calibration points created using two columnar structures (right).

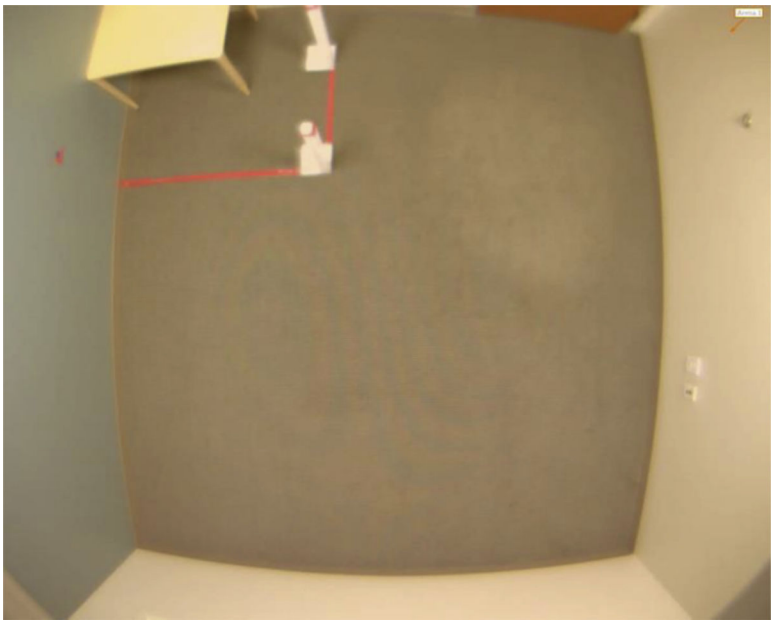

Fig. 8.

Example of the table ROI measured in the assessment room and delineated with physical indicators at the chosen height (i.e., the red tape at the top of the columns and on the wall). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

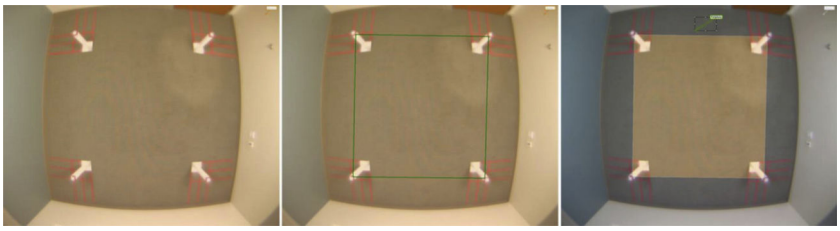

Fig. 9.

Example of ROI from the grabbed image to the fully constructed zone in EthoVision XT (left to right). Note that the parallel red floor markings were used to correctly position the Styrofoam columns. The zones were delineated by drawing lines in EthoVision XT, connecting the indicators (i.e., red tape) on top of the Styrofoam columns. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

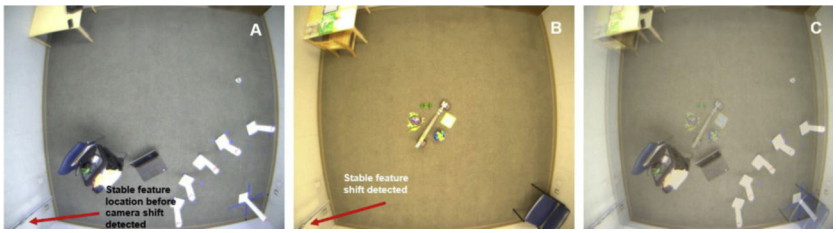

Fig. 10.

Example of an original arena with calibration markers (A), frame shift (B) and overlaid images (C) before applying transformations.

Fig. 11.

Steps to fixing a frame shift, including rotation (A), translation (B) and final cropped image (C).

Table 1

Total Distance Moved (TDM) for the tracking markers at the Floor, 29-inch, and 41-inch levels.

Trial	EthoVision XT Calculated TDM (cm)	Absolute difference (cm)	Percent difference
Floor_1 frame	550.81	119.76	17.86
29 in_1 frame	703.93	33.37	4.98
41 in_1 frame	796.19	125.63	18.73