From Motion to Meaning: A Systematic Review of Motion Tracking in XR-based Experimental Psychology

M M Rushadul Mannan, Mark Roman Miller

Abstract—Immersive technologies such as Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) are being increasingly used in experimental psychology because they enable controlled environments while simultaneously producing high-frequency motion-tracking data from headsets, controllers, and body trackers. These motion traces provide continuous behavioral signals that may reflect affective, behavioral, and cognitive processes during human–human, human–computer, and human–virtual human interactions. Despite the growing use of motion measures in XR (Extended Reality) experiments, there is limited synthesis of how such data are collected, transformed, analyzed, and interpreted. This paper presents a systematic literature review of studies collected from the ACM Digital Library, IEEE Xplore, PsycINFO, PubMed, and Web of Science that use motion-tracking data from XR devices in experimental psychology, focusing on affect, behavior, cognition, or their intersections. The review examines the devices used, the body parts tracked, and the data formats employed in these studies. More importantly, it analyzes how raw motion-tracking data streams are transformed into kinematic or trajectory-level features and the analytical methods used to link them to psychological constructs. Our findings highlight substantial variability in device configurations and motion data extraction methods, but we observe a common pattern of using descriptive statistical measures to summarize motion patterns or applying machine learning models to learn from processed motion features. Based on these observations, we outline practical recommendations for selecting tracking setups, collecting and validating motion data, and designing analysis pipelines in experimental psychological studies. This review aims to support researchers seeking to incorporate motion-tracking data into XR-based psychological experiments and to promote more consistent and interpretable use of motion measures in immersive research.

Index Terms—Extended Reality (XR), Motion Data Analysis, Experimental Psychology, Systematic Literature Review

1 Introduction

Understanding human psychological processes has traditionally relied on a combination of self-report, performance outcomes, and physiological indicators. While these approaches have contributed substantially to theory development in cognitive, affective, and social psychology, they often provide either retrospective accounts of experience or temporally discrete behavioral patterns [9, 24]. As a result, they may not fully capture how psychological processes unfold dynamically during ongoing interaction with the environment. In contrast, non-verbal behavioral signals provide a temporally rich and observable basis for studying psychological processes, making it possible to examine how perceptual, cognitive, and motor components of behavior are coordinated as actions evolve [13]. Patterns of bodily movement including orientation, locomotion, postural adjustment, gesture production, and action timing have long been recognized as important behavioral indicators in experimental psychology, motor control research, and human-computer interaction (HCI) [32]. These behavioral signals reflect the dynamic coupling between the individual and its environment, providing insight into processes such as attention allocation, motor planning, social coordination, and more. From this perspective, motion trajectories are not merely motor outputs but integral components of affective, behavioral and cognitive processing.

Within this broader shift, the concept of behavioral tracing has emerged as a valuable methodological approach [12, 20, 36, 66]. Rather than focusing solely on end-state performance or categorical responses, behavioral tracing examines the temporal evolution of action trajectories, enabling researchers to infer intermediate psychological states such as hesitation, exploration, conflict resolution, emotion or adaptation. These approaches share the premise that psychological processes can be understood as continuous dynamical phenomena, observable through structured patterns of movement over time. Such perspectives have motivated calls for richer behavioral data streams that capture the unfolding of human action in construct specific contexts rather than relying exclusively on simplified laboratory tasks.

Recent advances in extended reality (XR) technologies, including virtual reality (VR), augmented reality (AR), and mixed reality (MR), have significantly expanded opportunities for large-scale behavioral tracing in controlled yet immersive environments. Modern XR platforms also provide continuous tracking of head position and orientation, hand or controller movements, and, in some configurations, full-body kinematics and gaze direction [48, 65]. Compared to traditional motion capture setups, XR platforms offer relatively accessible, portable, and cost-effective means of collecting high-resolution motion data while maintaining experimental control over environmental stimuli and task context. This capability has led to increasing adoption of XR methodologies in experimental psychology to investigate constructs such as spatial navigation [4], motor learning and rehabilitation [77, 81], interaction and presence [47, 61, 72], stress and emotional regulation [11], and cognitive workload [64]. The ability to manipulate environmental affordances and social cue while simultaneously recording behavioral responses at fine temporal scales positions XR as a powerful methodological bridge between laboratory experimentation and ecologically valid behavioral observation.

Despite these opportunities, the use of motion tracking data in psychological research raises important methodological and interpretive challenges. Movement trajectories recorded in XR environments are inherently behavioral signals rather than psychological variables, and their relevance to constructs such as affect, cognition, or behavioral regulation depends critically on how they are transformed and interpreted. Raw motion streams typically consist of continuous spatial coordinates sampled at high frequency, which must be transformed into interpretable motion measures such as velocity profiles, path deviation, variability metrics, or higher-order descriptors so that meaningful inference can be made by statistical analysis or predictive modeling. However, recent studies demonstrate substantial heterogeneity in how researchers perform these transformations, reflecting differences in theoretical assumptions, task demands, hardware constraints, and analytical preferences. In the absence of consolidated methodological guidance, this diversity can complicate the comparison of findings

• *M M Rushadul Mannan is with Illinois Institute of Technology. E-mail: mmannan@hawk.illinoistech.edu.*

• *Mark Roman Miller is with Illinois Institute of Technology. E-mail: mmiller30@illinoistech.edu.*

Manuscript received xx xxx. 202x; accepted xx xxx. 202x. Date of Publication xx xxx. 202x; date of current version xx xxx. 202x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.202x.xxxxxxx

across studies and may lead to inconsistent interpretations of similar behavioral signals.

A further challenge arises from the inherently multivalent nature of behavioral data that a single movement pattern may support multiple plausible interpretations depending on context, developmental stage, and concurrent environmental conditions. Such interpretive flexibility underscores the need for theoretically grounded frameworks that link motion measures to psychological constructs in a principled manner. Without such frameworks, there is a risk that behavioral metrics become detached from the processes they are intended to represent, leading to ambiguous conclusions. These interpretive challenges are compounded by practical considerations related to data collection and representation. Differences in hardware configuration, sampling rate, tracking volume, sensor placement, and pre-processing strategies can all influence the structure of recorded motion signals. As XR motion tracking becomes more widely adopted in experimental psychology, there is a growing need to examine how these methodological decisions shape the validity of behavioral inference.

Previous work has reviewed behavioral tracing in VR, methodological considerations for immersive psychological research, and broader findings from psychological experimentation in VR [1,71,80]. However, these works do not systematically examine how continuous motion traces are transformed into quantitative measures and interpreted as evidence of affective, behavioral, or cognitive processes. A systematic review of current studies can help identify common patterns, methodological decisions, and future scope for improving the reliability and interpretability of motion based measures.

Therefore, this review addresses how motion tracking data from XR devices have been used in experimental psychology research. Specifically, the review aims to:

- Identify the computational approaches used to transform raw motion data into interpretable behavioral indicators.
- Summarize key methodological characteristics such as hardware selection, tracked body points, sampling properties in association with psychological constructs.
- Provide evidence based recommendations for future experimental design, data analysis, and reporting practices in XR motion based psychological experiments.

2 Method

The systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) guidelines [34]. This framework guided the definition of the research objectives, inclusion and exclusion criteria, search strategy, study selection process, data extraction procedures, synthesis methods, and the reporting of results. Literature was searched across the following electronic databases: IEEE Xplore, ACM Digital Library, APA PsycInfo, PubMed, and Web of Science. Studies published up to January 27, 2026, the date on which the search was conducted, were considered for inclusion.

The search strategy was designed to identify literature that extracts and analyzes motion-derived measures and relates to some psychological constructs with a focus on XR tracking data. Considering this, search terms were organized into 4 groups representing motion data, immersive technology, psychological constructs and measurement. The individual search terms can be summarized in one big query that can be described as: (motion OR motor OR movement OR gesture OR "motion capture" OR tracking OR posture) AND ("virtual reality" OR VR OR immersive OR "augmented reality" OR AR OR "head-mounted display" OR HMD OR “VR controller” OR “Virtual Reality Controller” OR “handheld controller” OR MoCap OR "inertial measurement unit") AND (psychology OR behavior OR behaviour OR cognition OR cognitive OR attention OR emotion OR affect OR engagement OR stress OR "mental state" OR interaction) AND (measure OR metric OR analysis) . The terms had to be included in the title or abstract of an article.

The PICOS model, recommended approach by PRISMA-P, was used to establish the selection criteria that defines the patient problem or population (P), intervention (I), comparison (C), outcome (O) and study design (S) to answer the research question. The following PICOS criteria were used in our selection process:

- P: Studies were included if they involved human participants of any age performing tasks that required physical movement, where the movement data were examined in relation to psychological construct. Studies involving robots, virtual humans, or avatars without direct human input were excluded.
- I: Studies involved the use of immersive XR technologies in which motion data were collected through integrated systems such as head-mounted displays (HMD), controller or/and body trackers. External motion-capture infrastructures such as optical motion capture systems, depth camera and standalone Kinect setups were excluded. Studies using additional sensing modalities not integrated by default into the XR device, such as external eye-tracking systems or supplementary tracking technologies, were considered only when the primary motion data analyzed were derived from the XR device itself.
- C: No explicit comparison condition was required for the selection process. Studies were included if they examined the relationship between motion derived measures and psychological constructs in an immersive environment.
- O: Studies were included if motion-derived features were used to quantitatively infer, predict or examine associations with psychological construct within the domains of affect [60], behavior [2] and cognition [46]. Affect included emotional or affective states, cognition included cognitive processes or states, and behavior referred to semantically meaningful actions within the task that were interpreted in relation to a psychological or social construct. Studies could address a single domain or constructs spanning multiple domains. Quantitative analytical approaches such as statistical testing, modeling, or machine learning methods were required to establish relationships between motion measures and psychological outcomes.
- S: Empirical studies published in English from 2015 onward in peer-reviewed journals or conference proceedings were included. Other publication types, such as meta-analyses, literature reviews, letters, commentaries and editorials were excluded.

Based on these selection criteria, literature extracted through keyword searches on titles and abstracts were initially screened. Duplicate records were removed prior to further analysis and the remaining records were screened in three phases. In the first phase, titles were reviewed to remove clearly irrelevant articles, including papers retrieved because of proximity-based matches in advanced database searches but unrelated to the scope of the review. In the second phase, abstracts were screened to determine whether the studies involved XR technologies, included motion-derived measures, and involved human participants. This stage primarily addressed the population, intervention, and study design components of PICOS by excluding studies that did not involve human XR-based empirical research. In the third phase, the full texts of the remaining articles were reviewed to assess the outcome component of PICOS. Specifically, we examined whether XR-derived motion measures were used to infer, predict, or analyze psychological constructs within the affective, behavioral, or cognitive domains. Articles that did not establish a relevant relationship between motion measures and these ABC outcomes were excluded at this stage.

The first author conducted the primary screening and filtering of the retrieved records. Borderline cases were not resolved solely by the first author. When uncertainty arose regarding whether a study satisfied the inclusion criteria, it was discussed with the senior author. The eligibility criteria and the specific reason for uncertainty were reviewed jointly. Ambiguity often occurred when determining whether a reported behavior represented a psychologically meaningful construct. For example, payment decisions in a virtual trust experiment, the amount of speech during an interaction, or path choices used to operationalize social behavior were considered within scope, whereas outcomes focused primarily on ergonomics, fatigue reduction, or interface efficiency were generally excluded. The senior author made the final decision regarding inclusion or exclusion for some borderline cases.

For each included study, the first author extracted information relevant to the review objectives, including study characteristics, XR devices used and tracked body segments, motion features, sampling rates when reported, analytical methods, psychological constructs, and evaluation strategies. These fields were defined based on the review questions and were applied consistently across the included studies. The included studies were categorized at two levels. First, studies were classified into the broad psychological domains of affect, behavior, and cognition. Second, studies were grouped by their more specific research context into Affective XR, Social Interaction, Clinical/Rehabilitation, Cognitive Assessment, XR Behavior, Locomotion/Navigation, and Motor Behavior/Learning. This two-level categorization captured both the broader psychological focus and the more specific application context of XR-derived motion analysis. Studies whose categorization was unclear or that appeared to span multiple themes were discussed with the senior author. The senior author also monitored the overall extraction and categorization process, and ambiguous cases were resolved through discussion, with the senior author making the final decision when necessary.

Because the screening, data extraction, and categorization were not independently performed by two reviewers, formal inter-rater reliability statistics were not calculated. The senior author instead reviewed ambiguous and borderline cases and provided oversight throughout the process. We acknowledge that this approach does not provide the same level of reliability assessment as independent double screening or double coding and therefore represents a methodological limitation of the review.

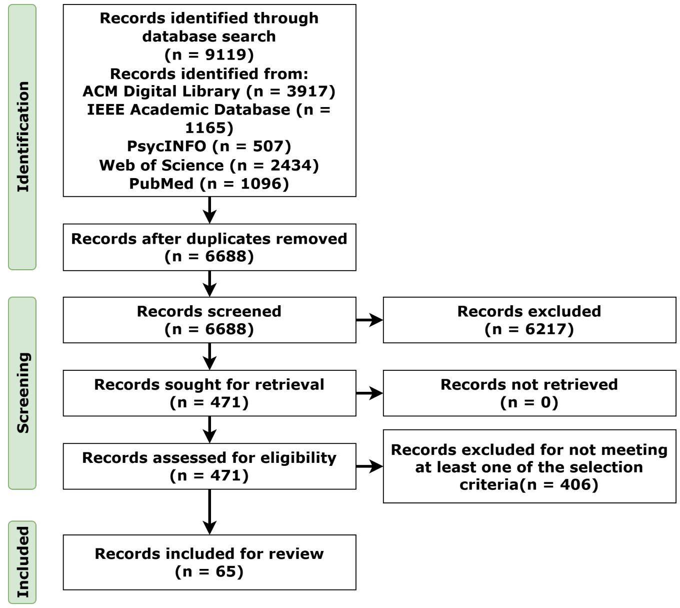
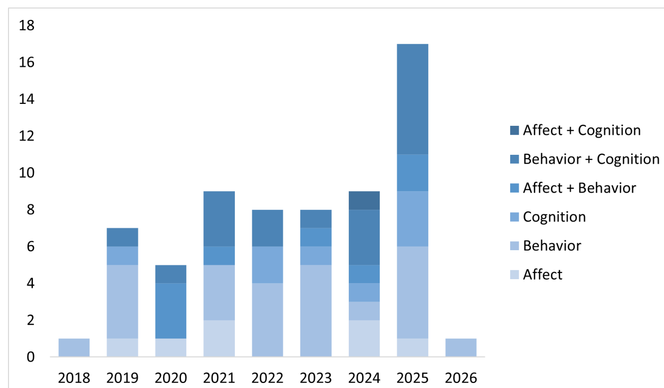
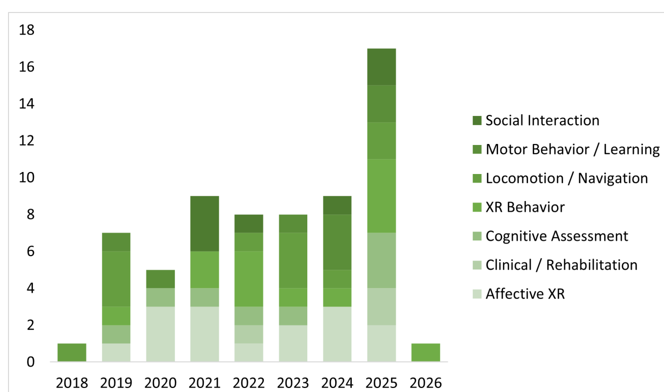


Figure 1: Process and Results of the Literature Review Screening Following PRISMA



(a) Distribution across years segmented by psychological construct.



(b) Distribution across years segmented by primary research domains.

Figure 2: Overview of the included studies showing temporal trends, psychological construct and domain distribution.

3 Results

3.1 High Level Summary

Based on the search strategy, we extracted a total of 9,119 papers. After removing duplicates, 6,688 papers remained. Applying the PICOS selection criteria and conducting multiple rounds of screening resulted in a total of n = 65 studies that met the inclusion criteria and were included in the final analysis (see Figure 1). The selected publications spanned the period from 2018 to 2025, although the initial search included studies published as early as 2015. This distribution reflects the relatively recent and growing interest in the use of motion-tracking data from immersive technologies in experimental psychology. The studies were published across a diverse range of venues, including major human–computer interaction and extended reality conferences (e.g., CHI, IEEE VR, ISMAR, VRST), as well as interdisciplinary journals (e.g., PLOS One, Journal of Experimental Psychology: Applied, Nature Scientific Reports, Frontiers in Psychology, and more).

At a high level, the reviewed papers covered multiple research domains. A substantial portion of studies focused on affective state recognition [n = 15, $\approx$ 23%], general behavioral analysis in immersive environments [n = 13, $\approx$ 20%], navigation behavior [n = 11, $\approx$ 17%], motor behavior and learning [n = 8, $\approx$ 12%], social interaction dynamics [n = 7, $\approx$ 11%], and, cognitive workload and attention assessment [n = 8, $\approx$ 12%] using movement data. A small but notable subset of studies addressed clinical and rehabilitation [n = 3, $\approx$ 5%] applications. Figure 2 provides a temporal overview of the reviewed literature, showing an exponential increase in experimental psychology using XR motion data over time. Figure 2a shows that reviewed literature is concentrated in behavior-focused work, often at the intersection of behavior and cognition, with comparatively fewer studies addressing purely affective or purely cognitive outcomes with distributed growth across multiple domains rather than a single dominant area (see Figure 2b). Detailed study characteristic (Domain, psychological construct, sample size, tracking setup, tracked body parts, sampling rate and experiment duration) is provided in the *Supplemental Materials Table 1*.

3.2 Study Samples

Across the reviewed literature, the datasets were primarily derived from controlled XR experiments involving human participants. Among the studies, sample sizes ranged from 8 [7, 57] to 937 [21], with a median sample size of 30, indicating that most studies relied on relatively small to moderate experimental samples. Although a few large-scale datasets were identified [11, 16, 21, 35, 67], the majority of studies remained within the range typical of laboratory-based immersive research, reflecting the practical constraints of XR data collection.

The participant populations represented in the review were diverse but remained centered on adult experimental samples. In addition to healthy adult participants, the literature included studies involving specialized populations such as children [67], athletes [78], and clinical groups [4, 62, 72]. Domain-level patterns further suggest that clinical and rehabilitation studies form a smaller but distinct subset of the literature, while affective, cognitive, and social-interaction studies make up a larger proportion of the current evidence base. Overall, the reviewed datasets indicate that XR motion-tracking research is still largely shaped by controlled, small-sample experimental designs, while a smaller number of studies are beginning to move toward larger-scale and more ecologically varied data collection.

3.3 Motion Data Collection

Across the 65 reviewed studies, motion data were collected primarily using consumer-grade virtual reality systems (summary in Table 1), with a clear concentration around the HTC Vive ecosystem ($\approx 42\%$) and Meta/Oculus Quest head-mounted displays ($\approx 34\%$). A smaller number of studies employed alternative platforms such as Pico devices, console-based systems (e.g., PSVR) or custom motion-capture integrations. Motion sensing was most commonly implemented through built-in headset tracking and handheld controllers that provide six-degree-of-freedom (6-DoF) movement capture, while only a minority of studies incorporated additional trackers (e.g., foot, pelvis, or full-body sensors). [n = 62, $\approx$ 95%] studies analyzed head motion, while [n = 34, $\approx$ 52%] analyzed hand, controller, arm, or other upper-limb movement. Lower-limb or foot motion was examined in [n = 8, $\approx$ 12%] studies, and only [n = 5, $\approx$ 8%] used whole-body or multi-segment tracking. Because studies could track more than one body segment, these categories were not mutually exclusive. This distribution indicates that the spatial representation of behavior was typically restricted to head-centric or upper-body interaction data, reflecting both technological affordances of these XR devices and task design constraints.

The technical properties of the motion data varied considerably across the reviewed studies. Sampling rates ranged widely (approximately 0.1 Hz [73] – 600 Hz [29], with most studies operating in the 60–90 Hz range), suggesting differences potential sensitivity to micro-movement dynamics. Information on experimental task duration was inconsistently documented across studies. However, available descriptions suggest that motion data were typically recorded during relatively short, task-specific sessions rather than across the entire experimental period (e.g., while participants completed questionnaires or survey measures). Overall, the reviewed literature demonstrates a methodological trend toward accessible standalone VR hardware and minimal tracking setups, which may enhance ecological validity and scalability but simultaneously constrain the richness of behavioral inference. This imbalance highlights an emerging gap between the theoretical ambition to capture embodied psychological processes and the practical limitations of current experimental motion-tracking implementations. Overall, the reviewed studies point to a clear methodological shift toward the use of accessible, standalone VR hardware and relatively minimal tracking configurations. Such systems are typically less expensive and easier to deploy than multi-sensor laboratory motion-capture setups.

3.4 Motion Data Transformation

Across the reviewed literature, most studies transformed headset, controller or full-body trajectories into a set of derived motion measures that could be interpreted psychologically. A useful taxonomy established from the reviewed papers is: (1) Descriptive Statistical Reduction, where continuous motion data are collapsed into scalar summaries such as time, distance, velocity, angular rotation, or variability; (2) Event-based Transformation, where meaningful motion units such as turns, reactions, or interaction episodes are detected first; (3) Trajectory-level Transformation, where the geometric or temporal structure of the movement path is preserved; (4) Feature Engineering Approach, where higher-order descriptors such as jerk, entropy, or cross-correlation are computed; and (5) Predictive Feature Transformation, where motion data are converted into a representation suitable for machine-learning or probabilistic modeling.

This progression is important because it reflects increasing preservation of temporal structure: descriptive summaries provide interpretable global measures, but they also discard the sequential organization of the movement, whereas trajectory and engineered-feature approaches retain more information about how the movement unfolded over time.

Summary of motion features, transformation approach, analysis method and key analysis for each study is provided in the *Supplemental Materials Table 2*.

3.4.1 Descriptive Statistical Reduction

The most common approach was to reduce motion recordings to a small set of interpretable summary measures. These included task or movement duration, distance traveled, path length, average speed, angular rotation, and movement variability. Such measures were widely used in studies of locomotion, navigation, motor performance, and social interaction because they provided direct descriptions of how much or how quickly participants moved [16, 18, 40, 43, 47, 48, 52, 57]. Continuous hand and gaze trajectories were also converted into movement time, throughput, effective width, and error rate in a three-dimensional selection task [75].

Head motion was similarly summarized using measures such as mean viewing direction, angular rotation, angular speed, and rotational variability. These measures were used to examine visual exploration [10], cybersickness [37], affect [14], embodiment [30], and behavioral intention [21, 39]. In auditory and communication tasks mean gaze angle, difference between target head orientation angle, temporal derivatives of head angle as speed, and root mean square of speed were used to estimate orienting behavior during listening or localization [15, 19].

Clinical and gait-oriented studies used stride time, stride velocity, stride width, foot clearance, and the proportions of straight, turning, or spinning steps [4, 76]. Apart from these approaches, motion streams were also discretized into summaries in studies exploring behavioral patterns such as during dyadic collaboration [50] and training with cognitive impairment [54], asymmetric collaboration [49]. These studies demonstrate the dominance of descriptive summaries used for motion data transformations in experimental psychology.

3.4.2 Event-based Transformation

A second group of studies first identified meaningful events or task periods and then calculated motion measures within them. Events were commonly defined using movement thresholds, task markers, or changes in behavioral state. In one visual-exploration study by Wirth *et al*. [78], a head turn was defined as a vertical-axis rotation exceeding 125°/s, after which head-turn frequency, excursion, and velocity variability were calculated. Similarly, thresholds for eye, neck, arm, and stepping movements were used to estimate movement onset, perceptual latency, response time, and maximum displacement [77].

Tong *et al*. [69] explored people's trust in AI guidance during evacuation, head rotation data were transformed using an event based approach by converting yaw rotation into participants view direction. It was then used to measure participants' trust on AI's suggested path thereby transforming raw head orientation data into a temporally localized signal indicative of trust. Several other studies also employed event-like structures to transform raw motion data into behavioral metrics. For example, reaction occurrence, reaction intensity, and reaction velocity were used to measure presence in an unfamiliar environment [28]; fixation and saccade extraction during simulation events were used to recognize human behavior [23, 25]; head-turn deviation prior to passing a ball was used to assess exploratory activity [59]; and measures such as counts, time ratios, and head-direction bias were used to characterize interactions with pedestrians and vendors [74]. These event-based transformations are particularly useful in experimental psychology because they can map relevant motion data onto discrete behavioral patterns.

Table 1: Overview of motion data collection practices across reviewed studies

Device	No. of Paper	Data Collection	Research context
HTC Vive ecosystem	27	Predominantly head + controller; few multi-tracker studies	Commonly used in behavioral and mixed-construct studies involving spatial navigation, locomotion tracking, and controlled laboratory interaction paradigms
Meta / Oculus Quest ecosystem	22	Mostly upper-body (Head and hand)	Commonly used in cognitive and affective task-based experiments emphasizing response actions, social presence, or interaction dynamics in portable setups
Custom VR setup	8	Mostly single segment; few full-body tracking	Commonly used in exploratory or interdisciplinary studies examining diverse psychological constructs with non-standardized motion configurations
HoloLens	2	Head-centric	Used mainly in perception and cognition oriented experiments focusing on spatial awareness and environmental interaction
Pico Neo 3 Pro Eye	1	Head-centric	Emerging use in cognitive or interaction-based experimental contexts using lightweight standalone systems
Consolve VR (PSVR)	1	Controller motion	Rarely used, typically in behavior or performance-driven task settings with constrained interaction environments
Not reported	4	Mixed tracking (Head or Hand)	Limits interpretation of construct alignment and raises reproducibility concerns in psychological motion-tracking research

3.4.3 Trajectory-level Transformation

A smaller but analytically richer group of studies preserved the temporal structure of continuous motion data rather than relying only on descriptive summaries. These studies transformed movement while retaining the geometry or temporal organization of the tracking path. Yu *et al*. [81], for example, moved beyond task completion time by deriving refinement time proportion, refinement space, normalized path error, and normalized jerk error. They also fitted participant-level learning curves using a power function to examine how motor adaptation unfolded over time.

Trajectory-level measures were also used in navigation and avoidance research. Patotskaya *et al*. [52] analyzed turning points, minimum distance between agents, and clearance area during interactions in constrained environments. Mousas *et al*. [43] used trajectory deviation and path length to examine behavioral changes during self-avatar interactions, whereas Tian *et al*. [68] analyzed route trajectories and escape distance to characterize evacuation behavior in an indoor route-finding task.

Dynamic trajectory comparison was particularly prominent in movement-training and cognition-related studies. Dietz *et al*. [7] transformed the positional and rotational data of body joints into a movement-accuracy score using Dynamic Time Warping (DTW). By directly comparing participants' movements with expert reference trajectories, the study examined how perceptual conditions influenced movement performance. Lustig *et al*. [35] used maximal cross-correlation coefficients and time-lag analyses between hand trajectories and head yaw to quantify disruptions in head–hand coordination under increased cognitive demand.

Motion trajectory analysis was also used to demonstrate a clear separation between go and stop trials, thereby confirming the capture of response inhibition associated with the task [62]. Other studies employed trajectory analysis to identify startle responses [61], characterize movement associated with cybersickness [51], and examine gaze dynamics [44]. These studies are methodologically valuable because they preserve the temporal structure of motion while capturing movement correction, coordination, and path geometry that would be lost in simple mean-based summaries.

3.4.4 Feature Engineering Approach

The reviewed literature also includes a conceptually practical form of transformation: feature engineering, in which raw motion data are converted into higher-order descriptors intended to capture characteristics such as regularity, smoothness, complexity, and search organization. Castillo *et al*. [72], for example, compared adaptive and non-adaptive VR systems by dividing controller trajectories into one-second windows and transforming them into dimensionless jerk. This measure was subsequently mapped to fatigue estimates using a polynomial model. Krasovsky *et al*. [27] similarly derived jerk and maximum tangential velocity to examine motor learning.

Entropy-based features were used to represent the regularity and complexity of movement. Reinhardt *et al*. [55] temporally normalized controller trajectories and transformed them into sample entropy, treating movement irregularity as an indicator of mental workload. Related entropy-based measures were used by Pradhan *et al*. [53] for trajectory prediction and by Kim *et al*. [26] to characterize behaviors associated with Autism Spectrum Disorder. Robitaille and McGuffin [58] combined maximum speed, maximum acceleration, movement duration, and path curviness to identify fast body movements, thereby representing motion episodes through both kinetic and geometric properties.

Other studies developed features to describe movement composition, postural complexity, and signal characteristics. Masoner *et al*. [38] converted raw head-position coordinates into Euclidean displacements between consecutive samples and summarized movement magnitude and variability to examine postural sway and affordance perception. They also introduced Effort-to-Compress (ETC) as a complexity-based motion measure. Cho *et al*. [5] normalized absolute changes in yaw, pitch, and roll by the total head movement within each time frame, producing relative axis-specific movement proportions for examining behavioral patterns associated with ADHD. Ferrarotti *et al*. [11] derived head-displacement magnitude and movement direction from consecutive samples for stress assessment, whereas Elvitigala *et al*. [8] used acceleration along the x, y, and z axes together with motion median frequency to detect engagement in VR gaming.

Feature engineering was also used to represent task-specific interaction behavior. Patotskaya *et al*. [52] extracted turning points by identifying when the participant's medial center of mass shifted laterally beyond a predefined threshold from a straight trajectory. Positional stability and movement synchrony were used to examine interaction behavior in other studies [33, 42]. Tseng *et al*. [70] applied rule-based transformations to derive motion measures related to effort and fatigue, while Nataraj *et al.* [45] derived features such as agency and contact timing to evaluate task performance. Miller *et al.* [41] used sample-level binning and nonlinear functions to examine motion–affect relationships. Head motion was also transformed into gaze-related patterns to examine environmental effects [17] and into proxemic measures to characterize interpersonal behavior [40].These studies show that engineered features often arise when authors want to capture not only "how much" movement occurred, but also "how organized," "how smooth," "how complex," or "how reactive" that movement was.

3.4.5 Predictive Feature Transformation

Another set of papers transformed raw motion data into feature vectors for predictive modeling. In these studies, the main contribution was not a single interpretable motion metric, but the conversion of one or more motion streams into features suitable for classification, regression, or probabilistic state estimation. Setu *et al*. [64] proposed a predictive framework for assessing cognitive states during multitasking by incorporating a broad set of gaze measures and HMD position and orientation as model inputs, providing a clear example of multidimensional behavioral encoding. Tong *et al*. [69] examined trust in AI guidance during evacuation by transforming each motion sequence into relative motion changes and dividing it into ten temporal segments. Summary statistics, including the sum, mean, and standard deviation, were extracted from these segments and used to train K-Nearest Neighbors (KNN), Convolutional Neural Network (CNN), Random Convolutional Kernel Transform (ROCKET), and Time Series Transformer (TST) models.

Predictive representations also frequently combined eye and head movement. Khatri *et al*. [25] investigated consumer behavior in a virtual retail environment using preprocessed eye-movement features, including fixations and saccades, together with raw head-motion data represented in both three-dimensional space and two-dimensional planes. These features were used as inputs to classification models, including Support Vector Machines (SVMs), to infer personality traits. Similarly, Hu *et al*. [23] combined fixation duration, fixation rate, saccade duration, saccade amplitude, fixation dispersion, and absolute head velocity. These measures were first examined statistically and then incorporated into a CNN-based model for task recognition.

Motion-based feature vectors were also used in affect-related research. Kusano *et al*. [29] extracted the mean and variance of peak intervals from head-motion signals, treating them as heartbeat-like rhythmic patterns in movement. These features were used as inputs to an SVM with a radial basis function kernel, and model performance was evaluated using leave-one-subject-out cross-validation. Vlasiu *et al*. [73] directly used raw motion data as predictive features and applied classification methods including SVMs, decision trees, and random forests to investigate mental-disability-related patterns.

More complex feature-transformation pipelines were used for continuous state estimation. Brambilla *et al*. [3] transformed head and hand motion into features such as linear and angular velocity, linear and angular acceleration, grip pressure, and interaction-related variables. Following feature selection, these measures were incorporated into a probabilistic state-space pipeline using Distributed Kalman Filtering (DKF) and Hidden Markov Models (HMMs) to support continuous stress tracking. Setu *et al*. [63] used both raw and processed motion features to classify cybersickness and other affective and cognitive states using Gated Recurrent Units (GRUs), Long Short-Term Memory networks (LSTMs), and Multilayer Perceptrons (MLPs).

Across these studies, the computational model generally appeared at the end of the transformation chain. Raw motion streams were first converted into structured feature representations, after which machine-learning or probabilistic models were applied for classification, prediction, or continuous psychological-state estimation.

4 Discussion

4.1 Motion as Context-Dependent Psychological Evidence

A central idea of this review is that motion *alone* is not inherently meaningful. Its meaning depends on why the participant had an opportunity to move, what part of the body could express the construct, and how the signal was transformed. Head and hand position, rotation, path deviation, jerk, clearance distance, and other kinematic measures are only behavioral traces. Their psychological meaning emerges only when they are connected to a task context, theoretical rationale, and explicit process of transformation and interpretation.

The same motion feature can support different interpretations depending on what the participant is doing and why. For example, increased head movement has been interpreted as visual exploration in sports [78], stress-related agitation [3, 11, 29], cybersickness-related instability [37], and task-irrelevant movement in attention-related conditions [5]. Similarly, slower movement may indicate caution, fatigue, uncertainty, motor impairment, or strategic control depending on whether the task involves interaction, search, or rehabilitation [4, 52, 72, 76].

Transforming raw movement into higher-order features does not remove this ambiguity. Measures such as entropy, normalized path error, refinement time proportion, dimensionless jerk, movement complexity, head-turn frequency, and turning point provide more informative descriptions than raw coordinates [23, 38, 52, 55, 72, 81]. However, these features are not self-interpreting. Dimensionless jerk may represent fatigue in a rehabilitation task [72], but jerk may also capture urgency, repeated correction, or sensorimotor difficulty in another context [56]. Similarly, increased movement entropy may reflect cognitive workload when behavior becomes less regular [55]. Similarly, increased movement entropy may reflect cognitive workload when behavior becomes less regular [53], but it may also arise because the task encourages broader visual scanning.

The reviewed studies showed stronger psychological interpretations when the experiment design established why a motion pattern should represent a particular construct. For example, trajectory deviation, clearance distance, and turning behavior were interpreted as social avoidance because the studies manipulated cues such as avatar gaze, emotional expression, and proximity [43, 52]. Without these manipulations, similar trajectory differences could reflect route preference, scene geometry, collision boundaries, or navigation uncertainty. A similar process was noticed in studies of visual exploration. Head rotation, angular speed, scanning behavior, and region-specific movement were interpreted as indicators of attention, exploration, and engagement because the task required participants to inspect different parts of immersive content [10, 14, 39]. These interpretations were often supported by questionnaires, interviews, or task-performance measures. These examples demonstrate that psychological meaning depends on alignment among the construct, experimental task, tracked body segment, motion transformation, and validation strategy. A similar interpretive logic is visible in studies on postural and locomotor behavior [4, 38, 57], spatial cognition [31, 82], and several other reviewed papers.

4.1.1 Patterns in Motion Tracking

In the reviewed literature, the selection of tracked body segments was closely related to both the psychological construct and the behavioral demands of the experiment. Head position and rotation were the most commonly examined motion signals, partly because they are recorded by nearly all XR headsets without requiring additional equipment. However, head motion was especially prominent in studies of exploration, navigation, affective response, postural behavior, and broad spatial cognition [11, 14, 15, 19, 22, 29, 38, 51, 59, 69, 73, 78, 82]. In these contexts, head orientation or displacement was often treated as an indicator of where participants directed attention, how extensively they explored the environment, or how they physically responded to visual and auditory stimuli.

The analytical value of head tracking was less straightforward in studies involving action execution, motor coordination, or rehabilitation. In these contexts, the reviewed studies more often incorporated hand, limb, torso, foot, or whole-body movement. Studies of cognitive and sensorimotor performance, for example, used hand movement together with head motion to examine coordination, response execution, and temporal coupling [7, 35, 55, 81]. Similarly, research on locomotion, balance, movement adaptation, and rehabilitation incorporated lower-limb or broader body tracking because the relevant behavioral response could not be represented adequately through head movement alone [4, 57, 72, 75, 76, 81].

A similar pattern was seen in social interaction research. Head movement was often combined with hand movement, interpersonal distance, gaze, or speaking behavior in the studies that examined coordination, engagement, or interpersonal behavior [17, 31, 33, 49, 50]. These additional signals captured different components of the interaction, such as gesture, proxemic regulation, joint attention, and movement synchrony. The reviewed studies therefore suggest that multiple body part tracking can most informative when each body segment represented a distinct part of the behavioral process.

Overall, the review showed a general alignment between the behavioral expression of a construct and the body segment selected for measurement. The predominance of head tracking reflects both its psychological relevance and its technical accessibility. However, because other body segments often require additional sensors and calibration, the most frequently tracked segment is not necessarily the most direct indicator of a given construct.

4.1.2 Motion Data Collection and Quality Control

The reviewed literature highlights that the interpretability of motion features depends not only on which signals are recorded, but also on how measurement decisions defined and validated. For instance, some studies [77, 78] transformed continuous recordings into discrete events using predefined thresholds. Examples included neck rotations exceeding (6°), hand or step displacements greater than 10 cm, and head turns exceeding (125°/s). Such thresholds helped distinguish intentional or task-relevant movement from small fluctuations.

Other studies reduced positional variability by aligning hand trajectories to the exact starting position of each trial to remove between-participant variability before extracting motion data [62]. These transformations were particularly relevant when absolute coordinates were affected by room setup, participant height, controller placement, or starting posture. The literature therefore showed that motion quality was not limited to whether a tracker produced valid coordinates, it also concerned whether the recorded coordinates were comparable across participants, trials, and conditions.

The literature also included several examples of proxy measurement. Head orientation was sometimes used as an indirect indicator of gaze or attention when eye tracking was unavailable [5, 21, 40, 78]. Likewise, HMD sway was used as an approximation of postural stability or center-of-pressure-related behavior [38, 79]. These approaches expanded the range of constructs that could be studied using standard XR hardware, but they also introduced uncertainty about whether the proxy captured the same behavioral process as the measure it replaced. In such cases, measurement quality depended not only on technical accuracy but also on the conceptual correspondence between the recorded signal and the intended construct.

Tracking failures and lack of preprocessing information was present across the reviewed studies. Some papers described calibration procedures, data alignment, filtering, or threshold selection in detail, whereas others provided limited information about missing frames, device disconnection, or synchronization across multiple streams. This inconsistency made it difficult to determine whether unusual motion patterns reflected participant behavior, application events, or technical artifacts.

Taken together, these findings show that motion-data quality in XR is multidimensional. It includes signal stability, cross-participant comparability, synchronization, and the validity of any proxy or derived feature. The reviewed literature also indicates that quality-control decisions were closely connected to later interpretation: preprocessing and event definitions did not merely clean the data, but shaped the behavioral patterns that became available for psychological interpretation.

4.1.3 Approaches to Motion Data Analysis

Descriptive summaries such as completion time, path length, mean speed, movement range, angular displacement, and variability remain useful in XR research. They are interpretable, relatively robust, and often sufficient when the question concerns an overall behavioral difference between conditions [4, 43, 47, 48, 57, 75, 76]. However, such aggregates can obscure the temporal and structural properties of motion data. To address this limitation, some studies have benefited from trajectory-level transformations and engineered metrics, such as normalized path error [81], entropy [55], movement complexity [38], dynamic time warping [7], temporal lag structures [35], and more. When the construct of interest involves processes such as corrective adjustments, interpersonal coordination, movement adaptation, affective behavior, among others, reducing motion to a single summary value may not always be theoretically adequate.

Temporal segmentation also emerged as an important analytical approach. Some studies compared early and late responses, identified discrete movement episodes, or examined localized trajectory events rather than treating the entire session as a single unit [28,52,69,77,78]. This is theoretically important because the meaning of motion may change over time. Early instability may indicate surprise or initial adaptation, while later instability may reflect fatigue. A pause before an action may indicate deliberation, whereas a pause after a failed interaction may indicate confusion or correction.

Construct-relevant filtering was used to improve interpretation in some studies. In an ADHD classroom task, separating task-relevant yaw rotations from pitch and roll movements helped distinguish visual orientation from unrelated movement [5]. Similarly, region-based analysis, filtered rotation velocity, and event-specific windows made local behavioral patterns more visible [14, 49, 69]. However, window length should be justified because long windows can conceal local effects, while very short windows may exaggerate variability.

Predictive models were conceptualized as another level of transformation. In studies employing predictive approaches [8,23,25,29,56,64], raw motion streams were rarely analyzed contextually. Instead, predictive models were trained directly on raw motion streams. While this strategy can preserve fine-grained temporal structure and potentially improve predictive performance, it also introduces a risk of construct drift, and accurate predictions are achieved without clearly identifying which movement patterns are psychologically meaningful. This issue is particularly relevant in research on affective states, cognition, and personality traits, where behavioral indicators may be subtle and context-dependent.

4.2 Recommendations for Construct-Aligned XR Motion Research

The findings above suggest that motion-based psychological research should follow a follow a sequence that begins with the construct and proceeds through interpretation to the motion measure. Researchers should begin with the psychological construct, design the task context in which that construct can affect behavior, track the body segments through which the response can become observable, preserve the relevant temporal and spatial structure, and validate the final interpretation against plausible alternatives. Table 2. includes the potential map between different psychological constructs with useful motion measures along with recommended tracking points and data transformation techniques.

4.2.1 Align the Construct, Context, and Tracked Body Segment

A motion feature is useful only when the task gives the participant a meaningful opportunity to express the construct through that part of the body. Therefore, the XR application should provide the context required to stimulate the construct and the tracked segment should provide the expected behavioral pathway.

At first, researchers should distinguish confirmatory from exploratory hypotheses before collecting XR motion data. Confirmatory studies should preregister a small set of theory-driven motion measures and hypotheses, while broader feature exploration should be clearly labeled as secondary or exploratory. This hybrid approach supports discovery without weakening the validity of confirmatory inference.

Tasks should be designed in which the psychological construct can produce observable and meaningful movement. For example, a study of visual exploration should distribute relevant information across the environment, while an avoidance task should provide genuine contexts to approach, maintain distance, or change direction. These contexts should remain comparable across conditions because target number, object placement, field of view, locomotion technique, and controller design can independently shape movement. Relevant application events

Table 2: Construct-aligned recommendations for body-segment selection, temporal segmentation, motion transformation, and interpretation in XR research.

Psychological Construct or Task Context	Recommended Body Segments	Most Informative Task Periods	Useful Motion features	Main Interpretation Risks
XR Behavior	Head and hands; gaze: when visual attention is central	Salient events, environmental transitions, object interactions, and opportunities for exploration	Head-turn events; movement range; interaction frequency; spatial and temporal region-based movement	More movement does not necessarily indicate stronger presence or engagement and may instead reflect familiarity, scene layout, or interface demands.
Cognitive Assessment	Head and gaze; hands: when a manual response is required	Information search, pre-response period (hesitation, confusion etc), response execution, post response period, movement correction, and post-error periods	Head–eye coordination; response latency; entropy; temporal lag; movement in sequential event; ; movement jerk	Pauses may reflect workload alone rather than cognitive processing, confusion, interface difficulty, or waiting.
Affective XR	Head and hands; full body when expressive movement is expected	Stimulus onset, salient emotional events, immediate response, and recovery	Response latency; peak speed; movement range; variability; nonlinear temporal models; region-based head orientation; movement jerk	Motion may reflect startle, workload, exploration, or social response rather than affect alone.
Social interaction	Head; hands; gaze;	Speaking turns, transition events, joint-attention episodes, gestures, and approach events	Relative orientation; interpersonal distance; synchrony; gesture frequency; lagged coordination	Synchrony and pauses depend on conversational role, turn-taking, and task structure.
Locomotion/Navigation	Head or torso position; feet for gait; gaze: for visual search	Route choices, intersections, turns, obstacle approaches, and arrival	Path length; curvature; speed profile; route deviation; clearance; trajectory entropy	Scene layout, locomotion method, visibility, and collision boundaries may independently produce movement differences.
Motor behavior/learning	Hand (task performing/dominant mainly), arm, leg, or joint	Preparation, movement correction, refinement, and completion	Path error; refinement space; peak velocity; smoothness; trial-to-trial change	Task difficulty, fatigue, controller mapping can influence performance.
Clinical/Rehabilitation	Head; gaze; feet or torso when physical locomotion is involved	Visual acceleration, turning, locomotion onset, stopping, and recovery	Head-rotation speed and variability; head-eye, hand-hand; hand-feet/lower limb coordination; event-related change; temporal models	High movement may reflect compensatory behavior, whereas low movement may reflect symptom avoidance.

should also be logged so that construct-related responses can be distinguished from motion caused by menus, failed inputs, re-centering, or other interface effects. Also, include sufficient familiarization through test sessions to reduce interface-learning effects.

Tracked segments should match the expected behavioral pathway. Navigation may require head or torso trajectories because they capture route choice and spatial movement, with foot tracking added when gait or obstacle crossing is central. Motor-learning studies should prioritize the task-performing limb because it directly reflects execution, correction, and refinement. Social-interaction studies may combine head orientation, gestures, gaze, and interpersonal distance to capture attention and coordination, while cognitive-workload studies may benefit from head–eye coordination and hand movement to represent information search and response execution.

4.2.2 Data Quality Control

Motion-data quality should be considered part of experimental design rather than a purely technical post-processing step. Tracker placement, room calibration, lighting, occlusion, sampling frequency, synchronization, filtering, and threshold selection can influence the extracted features and their interpretation. Pilot testing should be done to evaluate key analytical assumptions such as filtering, thresholds, or segmentation to ensure that the task and environment elicit the expected movement patterns.

Also, a short standardized calibration task can help distinguish individual movement characteristics from task effects. Participants may be asked to perform simple head rotations, reaches, still-standing periods, or short walking movements before the main experiment. These data can provide estimates of natural movement range, baseline speed, handedness, body scale, tracker noise, and controller offset.

Unusual motion segments should be treated as candidates for investigation rather than automatically removed. We distinguish three sources of anomalies.

- Tracking artifacts: include loss of tracking, controller disconnection, recentering, and physically implausible positions or rotations.
- Application-induced segments: include menu interactions, loading screens, frozen virtual objects, and failed input recognition.
- Participant-generated segments: may include hesitation, strategic pauses, fatigue, avoidance, corrective movements, rest, and waiting for another participant.

Only the first category should generally be considered for removal or interpolation. These decisions should be made during preprocessing and supported by evidence collected during the experiment, such as system logs, researcher notes, screen recordings, or video recordings of the participant, or participant feedback. The original raw recordings should remain unchanged, while artifact flags, excluded intervals, and any interpolated values should be clearly documented in the processed dataset. Application-induced intervals should remain in the raw data but may be excluded from construct-specific analysis or modeled as a separate state. Participant-generated anomalies should usually be retained because they may represent the behavior of interest.

4.2.3 Meaningful Segmentation of Motion Data

A complete XR session rarely reflects a single psychological state. Instruction, familiarization, stimulus exposure, active movement, decision-making, correction, waiting, and recovery periods may involve different behavioral processes. Collapsing these phases into one summary can dilute meaningful effects. Researchers should therefore define theoretically meaningful analysis windows, such as stimulus onset to first response, locomotion initiation and stopping, pre-decision hesitation, initial and corrective movement, target acquisition, speaking and listening periods etc.

The interpretation of motion should depend on when it occurs. For example, cybersickness-related motion may be most informative during visual acceleration, sharp turns, or locomotion onset rather than during instruction screens. Similarly, a pause before a difficult decision may indicate cognitive processing, whereas a pause during collaboration may represent turn-taking, and a pause following failed input may indicate an interface problem.

Segmentation doesn't have to be exclusively temporal. Motion can also be divided according to spatially meaningful regions, orientations, or relationships to virtual objects and characters. For example, time spent looking away from a social character, repeatedly entering specific viewing angles, or maintaining greater interpersonal distance may indicate gaze avoidance, discomfort, or reduced engagement. Spatial segmentation can therefore should be considered for psychologically meaningful behavior.

4.2.4 Data Transformation

Descriptive statistics and advanced temporal models should be treated as complementary rather than competing approaches. Descriptive measures such as mean, range, variability, completion time, and peak speed provide interpretable information about what changed. Event-based, trajectory-level, and nonlinear approaches can show when and how the change occurred.

To retain the temporal information, event-aligned windows can be helpful. Trajectory-level transformations are useful when the spatial organization of movement matters. Path curvature, route deviation, refinement space, and dynamic time warping can distinguish movement strategies. Nonlinear models may be particularly beneficial when the relationship between motion and psychological outcomes changes across time or movement intensity. For example, moderate head movement may reflect exploration, whereas very low movement may indicate disengagement or avoidance and very high movement may indicate instability. Such relationships may be missed by linear models.

However, motion analysis may not always require the finest temporal analysis. Following the bandwidth–fidelity trade-off described by Cronbach and Gleser, session-level summaries provide broader coverage of behavior but may combine psychologically different processes, whereas segment-level measures are more specific but may overlook effects expressed across the wider session [6]. The appropriate level of analysis should therefore be guided by the construct and task rather than by an assumption that finer-grained measures are always better.

A useful strategy is therefore to evaluate transformations progressively: first test interpretable summaries, then add event-based, trajectory-level, relational, or nonlinear representations when it is theoretically relevant. More complex approaches should be retained only when they capture meaningful temporal structure or improve prediction beyond simpler models. Overall, the aim of transformation is not to preserve every recorded sample, but to preserve the temporal and spatial information needed to represent the behavioral process under investigation.

4.2.5 Validate Psychological Interpretations

Motion features differ in how directly they represent the construct being studied. Measures such as walking speed, path length, head rotation, and hand velocity describe observable physical behavior. In contrast, concepts such as hesitation, exploration, avoidance, and coordination are interpretations of how that behavior unfolds within a task. Constructs such as stress, fatigue, presence, workload, and emotion are even further removed from the recorded signal because they cannot be observed directly. To strengthen a motion-based claim to validate a hypothesis, it should be supported by convergent evidence. Experimental manipulations, questionnaires, physiological measures, observer coding, interviews, and replication across tasks or environments can help determine why the explanation is most plausible.

Researchers should make the inferential claims explicit by reporting what changed, where and when in the task the change occurred, and why the tracked body segment was relevant to the proposed explanation. For example, the statement "participants with higher presence scores showed head orientation towards the their notepad while taking order instead of eye contact with the virtual customer when the customer showed negative attitude" is more defensible than claiming that head orientation is an indicator of avoidance behavior. The former identifies the observed behavior, its temporal context, and its association with an external measure, whereas the latter assumes that the same movement has a fixed psychological meaning across tasks and environments.

4.3 Limitation

The limitations of this work arise from both the methodological weaknesses of the reviewed studies and the design of the review itself. In this review screening and coding were not conducted through fully independent double coding, and therefore formal inter-rater reliability was not calculated. Borderline eligibility decisions were discussed with a senior author, but the lack of independent coding may still have introduced subjective judgment during study selection and data extraction. No formal quality assessment or risk-of-bias tool was applied, so the synthesis does not distinguish between stronger and weaker studies on the basis of methodological quality. The ABC-based classification also reflects a working definition of psychological outcomes rather than an exhaustive representation of all psychological constructs.

Within the reviewed literature, some studies used very low sampling rates, including one that recorded motion at 0.1 Hz, which may be too low to capture meaningful motion changes. Details of data processing were also frequently missing, particularly for filtering, anomaly handling, segmentation, and synchronization with task events. Together with the limited availability of raw data, these gaps make it difficult to reproduce the reported findings.

The participant samples also limit how broadly the findings can be applied. Many studies relied on relatively small samples, often consisting mainly of university students. However, motion can be influenced by age, physical ability, body dimensions, prior XR experience, and cultural background, independently of the psychological construct under study. In addition, several studies relied heavily on session-level descriptive statistics. Although these summaries can be useful, they may overlook when and where meaningful behavioral differences occur within a task.

5 Conclusion

This review examined how motion data obtained from XR devices have been used as behavioral evidence in experimental psychology how they offer valuable opportunities to capture fine-grained behavioral traces in immersive environments. However, the reviewed studies also demonstrate that motion signals are not psychological variables in themselves; their interpretive value depends on how they are theoretically framed, experimentally elicited, and analytically transformed into meaningful motion measures.

The reviewed literature demonstrates considerable variability in motion data collection practices and transformation approaches, ranging from basic kinematic summaries to more advanced trajectory-level analyses. While this diversity reflects the methodological flexibility afforded by XR technologies, it also underscores the need for clearer and more consistent methodological guidance. Future research would benefit from adopting an iterative and integrative approach that maintains alignment between psychological construct specification, motion data acquisition decisions, and analytical strategies.

Overall, XR motion tracking holds considerable promise for advancing behavioral tracing in experimental psychology. Realizing this potential will require a construct-centered approach in which movement data are systematically linked to psychological theory and task context, enabling more robust and interpretable insights into how affect, behavior, and cognition unfold in action.

Acknowledgments

Large language model (e.g., ChatGPT) was used to assist with grammar checking and minor language editing during manuscript preparation.

References

[1] J. N. Bailenson, C. DeVeaux, E. Han, D. M. Markowitz, M. Santoso, and P. Wang. Five canonical findings from 30 years of psychological experimentation in virtual reality. *Nature Human Behaviour*, 9(7):1328–1338, 2025. 2

[2] A. Bandura et al. Social foundations of thought and action. *Englewood Cliffs, NJ*, 1986(23-28):2, 1986. 2

[3] S. Brambilla, G. Boccignone, N. A. Borghese, and L. A. Ripamonti. Stress assessment in virtual reality horror games using players' behavioural and physiological data. *IEEE Transactions on Games*, 2025. 6

[4] J. Chen, V. P. Basavatia, K. T. Kwei, and S. K. Agrawal. Navigation in virtual reality floor mazes: Added cognitive demand and its effects on gait and balance in parkinson's disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 34:382–393, 2025. 1, 4, 6, 7

[5] Y. J. Cho, J. Y. Yum, K. Kim, B. Shin, H. Eom, Y.-j. Hong et al. Evaluating attention deficit hyperactivity disorder symptoms in children and adolescents through tracked head movements in a virtual reality classroom: The effect of social cues with different sensory modalities. *Frontiers in Human Neuroscience*, 16:943478, 2022. 5, 6, 7

[6] L. J. Cronbach and G. C. Gleser. *Psychological Tests and Personnel Decisions*. University of Illinois Press, Urbana, IL, 2 ed., 1965. 9

[7] D. Dietz, F. Berger, C. Ou, F. Chiossi, G. Graeber, A. M. Butz et al. Vreflect: Evaluating the impact of perspectives, mirrors and avatars in virtual reality movement training. In *Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2025. 3, 5, 6, 7

[8] D. S. Elvitigala, R. Somarathna, Y. Yan, G. Mohammadi, and A. Quigley. Towards using involuntary body gestures for measuring the user engagement in vr gaming. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–3, 2022. 5, 7

[9] K. A. Ericsson. Protocol analysis. *A companion to cognitive science*, pp. 425–432, 2017. 1

[10] Y. Feng, D. C. Duives, and S. P. Hoogendoorn. Wayfinding behaviour in a multi-level building: A comparative study of hmd vr and desktop vr. *Advanced Engineering Informatics*, 51:101475, 2022. 4, 6

[11] A. Ferrarotti, S. Baldoni, M. Carli, and F. Battisti. Stress assessment for augmented reality applications based on head movement features. *IEEE Transactions on Visualization and Computer Graphics*, 30(10):6970–6983, 2024. 1, 3, 5, 6

[12] J. B. Freeman and N. Ambady. Mousetracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior research methods*, 42(1):226–241, 2010. 1

[13] J. J. Gibson. *The ecological approach to visual perception: classic edition*. Psychology press, 2014. 1

[14] M. Goel, P. Srivastava, M. Agrawal, R. Singhal, R. Chand, and A. R. Baijesh. Head-movement analysis of 360° affective experience. In *2021 IEEE 7th International Conference on Virtual Reality (ICVR)*, pp. 231–239. IEEE, 2021. 4, 6, 7

[15] A. Gulli, F. Fontana, E. Orzan, A. Aruffo, and E. Muzzi. Spontaneous head movements support accurate horizontal auditory localization in a virtual visual environment. *Plos one*, 17(12):e0278705, 2022. 4, 6

[16] A. Guth, J. J. Good, E. Han, J. Bailenson, and T. C. Peck. An evaluation of movement data analysis techniques for virtual reality. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1565–1574. IEEE, 2025. 3, 4

[17] E. Han, C. DeVeaux, M. R. Miller, G. M. Harari, J. T. Hancock, N. Ram et al. Alone together, together alone: The effects of social context on nonverbal behavior in virtual reality. *PRESENCE: Virtual and Augmented Reality*, 33:425–451, 2024. 6, 7

[18] J. Hartbrich, S. A. Arboleda, S. Göring, and A. Raake. The effect of hand visibility in ar: Comparing dexterity and interaction with virtual and real objects. *IEEE Transactions on Visualization and Computer Graphics*, 2025. 4

[19] M. Hartwig, V. Hohmann, and G. Grimm. Speaking with avatars-influence of social interaction on movement behavior in interactive hearing experiments. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 94–98. IEEE, 2021. 4, 6

[20] J. M. Hausdorff. Gait dynamics, fractals and falls: finding meaning in the stride-to-stride fluctuations of human walking. *Human movement science*, 26(4):555–589, 2007. 1

[21] F. Herrera and J. N. Bailenson. Virtual reality perspective-taking at scale: Effect of avatar representation, choice, and head movement on prosocial behaviors. *new media & society*, 23(8):2189–2209, 2021. 3, 4, 7

[22] V. Holzwarth, J. Schneider, J. Handali, J. Gisler, C. Hirt, A. Kunz et al. Towards estimating affective states in virtual reality based on behavioral data. *Virtual reality*, 25(4):1139–1152, 2021. 6

[23] Z. Hu, A. Bulling, S. Li, and G. Wang. Ehtask: Recognizing user tasks from eye and head movements in immersive virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(4):1992–2004, 2021. 4, 6, 7

[24] D. Kahneman. *Thinking, fast and slow*. macmillan, 2011. 1

[25] J. Khatri, J. Marín-Morales, M. Moghaddasi, J. Guixeres, I. A. C. Giglioli, and M. Alcañiz. Recognizing personality traits using consumer behavior patterns in a virtual retail store. *Frontiers in psychology*, 13:752073, 2022. 4, 6, 7

[26] B. Kim, D. Jeong, M. Jeong, T. Noh, S.-I. Kim, T. Kim et al. Vista: User-centered vr training system for effectively deriving characteristics of people with autism spectrum disorder. In *Proceedings of the 28th ACM symposium on virtual reality software and technology*, pp. 1–12, 2022. 5

[27] T. Krasovsky, M. Kafri, M. H. Aharoni, and A. Cheshin. Virtual agents for studying the effect of emotional feedback on motor learning: a feasibility study. In *Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents*, pp. 1–10, 2024. 5

[28] E. Kukshinov, K. S. Nathan, N. Bowman, B. Rooney, and L. E. Nacke. Walking into the unknown: Manipulating knowledge of the external environment to promote presence in vr. *International Journal of Human–Computer Interaction*, pp. 1–22, 2025. 4, 7

[29] H. Kusano, Y. Horiguchi, Y. Baba, and H. Kashima. Stress prediction from head motion. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 488–495. IEEE, 2020. 4, 6, 7

[30] S. Langener, R. Klaassen, J. VanDerNagel, and D. Heylen. Immersive virtual reality avatars for embodiment illusions in people with mild to borderline intellectual disability: user-centered development and feasibility study. *JMIR Serious Games*, 10(4):e39966, 2022. 4

[31] J. Laroche, L. Vuarnesson, A. Endaltseva, J. Dumit, and A. Bachrach. [re] moving bodies–a shared diminished reality installation for exploring relational movement. *Frontiers in Psychology*, 12:539596, 2021. 6, 7

[32] M. L. Latash. *Neurophysiological basis of movement*. Human Kinetics, 2008. 1

[33] A. Léchappé, R. Johnstone, A. Milliat, J. H. Williamson, M. Chollet, and J. R. Williamson. Understanding social interactions in reality versus virtuality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1–18, 2025. 5, 7

[34] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. Ioannidis et al. The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *Bmj*, 339, 2009. 2

[35] A. Lustig, M. Wilf, I. Dudkiewicz, and M. Plotnik. Higher cognitive load interferes with head-hand coordination: virtual reality-based study. *Scientific Reports*, 13(1):17632, 2023. 3, 5, 6, 7

[36] J. S. Magnuson. Moving hand reveals dynamics of thought. *Proceedings of the National Academy of Sciences*, 102(29):9995–9996, 2005. 1

[37] A. Maneuvrier, N.-D.-T. Nguyen, and P. Renaud. Predicting vr cybersickness and its impact on visuomotor performance using head rotations and field (in) dependence. *Frontiers in Virtual Reality*, 4:1307925, 2023. 4, 6

[38] H. Masoner, A. Hajnal, J. D. Clark, C. Dowell, T. Surber, A. Funkhouser et al. Complexity of postural sway affects affordance perception of reachability in virtual reality. *Quarterly journal of experimental psychology*, 73(12):2362–2375, 2020. 5, 6, 7

[39] M. Metwaly and A. J. Quinn. 360 video viewing with virtual reality headsets: Connecting user head movements to intentions. In *Proceedings of the 16th ACM Multimedia Systems Conference*, pp. 1–11, 2025. 4, 6

[40] M. R. Miller, C. DeVeaux, E. Han, N. Ram, and J. N. Bailenson. A large-scale study of proxemics and gaze in groups. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 409–417. IEEE, 2023. 4,

6, 7

[41] M. R. Miller, H. Jun, and J. N. Bailenson. Motion and meaning: Sample-level nonlinear analyses of virtual reality tracking data. In *IEEE International Symposium on Mixed and Augmented Reality Adjunct, ISMAR 2021 Adjunct, Bari, Italy, October 4-8, 2021*, pp. 147–152. IEEE, 2021. doi: 10.1109/ISMAR-ADJUNCT54149.2021.00039 5

[42] M. R. Miller, N. Sonalkar, A. Mabogunje, L. Leifer, and J. Bailenson. Synchrony within triads using virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–27, 2021. 5

[43] C. Mousas, A. Koilias, D. Anastasiou, B. Rekabdar, and C.-N. Anagnostopoulos. Effects of self-avatar and gaze on avoidance movement behavior. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 726–734. IEEE, 2019. 4, 5, 6, 7

[44] H. Murakami and T. Inamura. Inferring user state from gaze dynamics in a vr throwing task: Toward adaptive and user-centered rehabilitation support. In *Companion Proceedings of the 27th International Conference on Multimodal Interaction*, pp. 235–239, 2025. 5

[45] R. Nataraj, D. Hollinger, M. Liu, and A. Shah. Disproportionate positive feedback facilitates sense of agency and performance for a reaching movement task with a virtual hand. *PLoS One*, 15(5):e0233175, 2020. 5

[46] U. Neisser. *Cognitive psychology: Classic edition*. Psychology press, 2014. 2

[47] M. Nelson, A. Koilias, S. Gubbi, and C. Mousas. Within a virtual crowd: Exploring human movement behavior during immersive virtual crowd interaction. In *Proceedings of the 17th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 1–10, 2019. 1, 4, 7

[48] M. G. Nelson, A. Koilias, D. Kao, and C. Mousas. Effects of speed of a collocated virtual walker and proximity toward a static virtual character on avoidance movement behavior. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 930–939. IEEE, 2023. 1, 4, 7

[49] N. Numan and A. Steed. Exploring user behaviour in asymmetric collaborative mixed reality. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2022. 4, 7

[50] D. Osei Tutu, S. Habibiabad, W. Van den Noortgate, J. Saldien, and K. Bombeke. When action speaks louder than words: exploring non-verbal and paraverbal features in dyadic collaborative vr. *Sensors*, 25(17):5498, 2025. 4, 7

[51] S.-Y. Park and D.-K. Koo. The impact of virtual reality content characteristics on cybersickness and head movement patterns. *Sensors*, 25(1):215, 2025. 5, 6

[52] Y. Patotskaya, L. Hoyet, A.-H. Olivier, J. Pettré, and K. Zibrek. Avoiding virtual humans in a constrained environment: Exploration of novel behavioural measures. *Computers & Graphics*, 110:162–172, 2023. 4, 5, 6, 7

[53] V. Pradhan, S. Rossi, and P. Cesar. Exploring entropy-based solutions for trajectory prediction in virtual reality. In *Proceedings of the 17th International Workshop on IMmersive Mixed and Virtual Environment Systems*, pp. 15–21, 2025. 5

[54] J. Qu, S. Zhu, Y. Shen, Y. Zhang, and L. Bu. Understanding the impact of longitudinal vr training on users with mild cognitive impairment using fnirs and behavioral data. In *2024 IEEE conference virtual reality and 3D user interfaces (VR)*, pp. 568–578. IEEE, 2024. 4

[55] D. Reinhardt, S. Haesler, J. Hurtienne, and C. Wienrich. Entropy of controller movements reflects mental workload in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 802–808. IEEE, 2019. 5, 6, 7

[56] A. J. Ries, C. Callahan-Flintoft, A. Madison, L. Dankovich, and J. Touryan. Decoding target discriminability and time pressure using eye and head movement features in a foraging search task. *Cognitive research: principles and implications*, 10(1):53, 2025. 6, 7

[57] A. Ríos, M. Palomar, and N. Pelechano. Users' locomotor behavior in collaborative virtual reality. In *Proceedings of the 11th ACM SIGGRAPH conference on motion, interaction and games*, pp. 1–9, 2018. 3, 4, 6, 7

[58] P. Robitaille and M. J. McGuffin. Increased affect-arousal in vr can be detected from faster body motion with increased heart rate. In *Proceedings of the ACM SIGGRAPH symposium on interactive 3D graphics and games*, pp. 1–6, 2019. 5

[59] C. D. Rojas Ferrer, H. Shishido, I. Kitahara, and Y. Kameda. Read-the-game: System for skill-based visual exploratory activity assessment with a full body virtual reality soccer simulation. *PloS one*, 15(3):e0230042, 2020. 4, 6

[60] J. A. Russell. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145, 2003. 2

[61] J. Schirm, G. Tullius, and J. Habgood. Towards an objective measure of presence: examining startle reflexes in a commercial virtual reality game. In *Extended abstracts of the annual symposium on computer-human interaction in play companion extended abstracts*, pp. 671–678, 2019. 1, 5

[62] P. A. Schroeder, E. Collantoni, V. Meregalli, E. Rabarbari, C. Simonazzi, J. Svaldi et al. Persistent avoidance of virtual food in anorexia nervosa-restrictive type: Results from motion tracking in a virtual stopping task. *International Journal of Eating Disorders*, 57(3):624–634, 2024. 4, 5, 7

[63] J. N. Setu, J. M. Le, R. K. Kundu, B. Giesbrecht, T. Höllerer, K. A. Hoque et al. Mazed and confused: A dataset of cybersickness, working memory, mental load, physical load, and attention during a real walking task in vr. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1048–1057. IEEE, 2024. 6

[64] J. N. Setu, J. M. Le, R. K. Kundu, B. Giesbrecht, T. Höllerer, K. A. Hoque et al. Predicting and explaining cognitive load, attention, and working memory in virtual multitasking. *IEEE Transactions on Visualization and Computer Graphics*, 2025. 1, 6, 7

[65] M. Slater and M. V. Sanchez-Vives. Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3:74, 2016. 1

[66] M. J. Spivey and R. Dale. Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, 15(5):207–211, 2006. 1

[67] L. Storli and H. Lorås. The association between 7-to-10-year-old children's leisure-time physical activities and their motor behavior in a virtual playground environment. *Perceptual and motor skills*, 133(1):57–85, 2026. 3, 4

[68] P. Tian, Y. Wang, Y. Lu, Y. Zhang, X. Wang, and Y. Wang. Behavior analysis of indoor escape route-finding based on head-mounted vr and eye tracking. In *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pp. 422–427. IEEE, 2019. 5

[69] F. Tong, R. Wang, D. Xu, and A. Lu. Exploring people's trust towards ai with human motions in a virtual evacuation study. In *2025 International Conference on Metaverse Computing, Networking and Applications (MetaCom)*, pp. 88–93. IEEE, 2025. 4, 6, 7

[70] W.-J. Tseng, S. Huron, E. Lecolinet, and J. Gugenheimer. Fingermapper: Mapping finger motions onto virtual arms to enable safe virtual reality interaction in confined spaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2023. 5

[71] M. Vasser and J. Aru. Guidelines for immersive virtual reality in psychological research. *Current opinion in psychology*, 36:71–76, 2020. 2

[72] J. F. Villada Castillo and J. E. Muñoz Cardona. Adaptive immersive virtual reality exergame for post-stroke rehabilitation: Design and implementation. In *Proceedings of the 2025 International Conference on Information Technology for Social Good*, pp. 70–78, 2025. 1, 4, 5, 6

[73] A. R. Vlasiu, M. E. Frincu, C. Daniel, D. Mocan, and M. Gabor. Classifying mental disabilities from virtual reality behavioral data using machine learning. In *2024 E-Health and Bioengineering Conference (EHB)*, pp. 1–4. IEEE, 2024. 4, 6

[74] M. Volonte, Y.-C. Hsu, K.-Y. Liu, J. P. Mazer, S.-K. Wong, and S. V. Babu. Effects of interacting with a crowd of emotional virtual humans on users' affective and non-verbal behaviors. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*, pp. 293–302. IEEE, 2020. 4

[75] U. Wagner, M. N. Lystbæk, P. Manakhov, J. E. S. Grønbæk, K. Pfeuffer, and H. Gellersen. A fitts' law study of gaze-hand alignment for selection in 3d user interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2023. 4, 6, 7

[76] N. Wasaki, K. Hiranai, and A. Takahashi. Age-related differences in the effect of mental fatigue on obstacle crossing in virtual reality. *Scientific Reports*, 15(1):9527, 2025. 4, 6, 7

[77] G. B. Wilkerson, J. C. Lansey, C. N. Noblett, and C. E. Sarris. Test-retest reliability of immersive virtual reality measures of perceptual-motor performance. *Perceptual and Motor Skills*, 130(6):2484–2504, 2023. 1, 4, 7

[78] M. Wirth, S. Kohl, S. Gradl, R. Farlock, D. Roth, and B. M. Eskofier. Assessing visual exploratory activity of athletes in virtual reality using head motion characteristics. *Sensors*, 21(11):3728, 2021. 4, 6, 7

[79] T. L. Wu, A. Gomes, K. Fernandes, and D. Wang. The effect of head tracking on the degree of presence in virtual reality. *International Journal of Human–Computer Interaction*, 35(17):1569–1577, 2019. 7

[80] H. E. Yaremych and S. Persky. Tracing physical behavior in virtual reality: A narrative review of applications to social psychology. *Journal of experimental social psychology*, 85:103845, 2019. 2

[81] D. Yu, M. Cibulskis, E. S. Mortensen, M. S. Christensen, and J. Bergström. Metrics of motor learning for analyzing movement mapping in virtual reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pp. 1–18, 2024. 1, 5, 6, 7

[82] H. Zhang, B. Andrade, X. Wang, I. Aburabee, and S. Yuan. A study of spatial cognition in the rural heritage based on vr 3d eye-tracking experiments. *Heritage Science*, 12(1):1–19, 2024. 6