Special Issue Article

Sociological Methods & Research
2023, Vol. 52(3) 1201–1238

DOI: 10.1177/00491241221147495
journals.sagepub.com/home/smr
Sage

3D Social Research: Analysis of Social Interaction Using Computer Vision

Yoav Goldstein[1], Nicolas M. Legewie[2], and Doron Shiffer-Sebba[3]

Abstract

Video data offer important insights into social processes because they enable direct observation of real-life social interaction. Though such data have become abundant and increasingly accessible, they pose challenges to scalability and measurement. Computer vision (CV), i.e., software-based automated analysis of visual material, can help address these challenges, but existing CV tools are not sufficiently tailored to analyze social interactions. We describe our novel approach, "3D social research" (3DSR), which uses CV and 3D camera footage to study kinesics and proxemics, two core elements of social interaction. Using eight videos of a scripted interaction and five real-life street scene videos, we demonstrate how 3DSR expands sociologists' analytical toolkit by facilitating a range of scalable and precise measurements. We specifically emphasize 3DSR's potential for analyzing physical distance, movement in space, and movement rate – important aspects of kinesics and proxemics in interactions. We also assess data reliability when using 3DSR.

[1] Stanford Center of Professional Development, Stanford University, Stanford, CA, USA
[2] Faculty of Economics, Law, and Social Sciences, University of Erfurt, Erfurt, Germany
[3] Institute for Policy Research, Northwestern University, Evanston, USA

Corresponding Author:
Nicolas M. Legewie, Faculty of Economics, Law, and Social Sciences, University of Erfurt, Erfurt, Germany.
Email: nicolas.legewie@uni-erfurt.de

Keywords
social interaction, computer vision, computational methods, video data, video data analysis, microsociology, digital research methods, kinesics, proxemics

Introduction

Video data, as early ethnomethodologists recognized (e.g., Cicourel 1974; Erickson 1977; Mehan 1979), offer novel opportunities for sociological research; they contain detailed and multimodal (i.e., audio and visual) information on social interaction, they can be re-watched and re-analyzed, and they can often be shared as primary data with colleagues and readers. Together, these strengths allow "study[ing] complex constructs and interactions in real, complex and dynamic […] environments" (Asan and Montague 2014:165). Sociologists and other social scientists manually analyze video data, either self-recorded or from third parties, to study a range of social interactions such as violence in groups (Collins 2008), crime and policing (Willits and Makin 2018), interrogations (Alison et al. 2013), family interactions (Golann, Mirakhur, and Espenshade 2019), situational social ties (Pallotti, Weldon, and Lomi 2020), military negotiations (Klusemann 2009), teamwork (Waller and Kaplan 2018), healthcare provision (Hunziker et al. 2011), classroom instruction (Congdon, Novack, and Goldin-Meadow 2018), and political speech (Mendelberg, Karpowitz, and Baxter Oliphant 2014), among many others (for an overview, see Nassauer and Legewie 2022).

While increasingly popular, such manual video-based sociological research faces challenges in precision and scalability. Even short video segments of social interactions can take hours to analyze, placing a burden on research teams and budgets. Moreover, some aspects of social interactions are difficult to measure by human coders, for example, movement rate of specific individuals or distances between individuals or objects (e.g., Hoeben et al. 2021). These challenges hinder social scientists from using the full potential of video data to study social interaction. This article proposes "3D social research" (3DSR), a new computer vision (CV) approach that combines human body-recognizing algorithms with 3D cameras that measure absolute position in space.

CV is a subfield of computer science that analyzes signals of images and videos through machine learning and other computational methods. Typical CV tasks include detecting the presence of objects or actions in space and/

or time through the classification of video frames or other images. 3DSR utilizes widely available 3D cameras together with the open-source OpenPose software (Cao et al. 2021) to locate and track individuals' physical bodies, and body parts nested within bodies, in 3D space over time. It thereby allows for studying larger samples and analyzing hard-to-measure aspects of video-captured social interactions in granular detail.

3DSR can be employed in a wide range of research scenarios. In terms of data, the approach can be used on any 3D footage. Currently, such footage will most likely be obtained through self-recording by positioning a small camera with connection to a computer (for recent examples of such studies, albeit using 2D cameras, see Dietrich and Sands 2021; Golann et al. 2019). However, 3D camera technology is proliferating. 3D camera models aimed at the consumer market are available and affordable, and many mobile phones have built-in 3D cameras. As data storage capacities also increase, it seems likely that individuals will increasingly capture and share 3D videos, which would allow applying 3DSR to found video data, as well.

In terms of the research process, 3DSR's measures of human motion and physical distance can be used as either operationalizations of phenomena to be explained, or explanations for other outcomes, using either qualitative video analysis or quantitative models. As many aspects of interactions, including the nature of speech, lie outside the scope of 3DSR and must be analyzed with other manual or computational tools, 3DSR can be combined with other data types and analytical approaches, depending on the research question. 3DSR, therefore, is not a self-contained method for studying video data in isolation. Rather, it is a flexible, scalable, and precise toolkit for measuring human position and movement in 3D space, which researchers can incorporate into their knowledge of the larger theoretical context and research setting.

The article proceeds as follows: first, we review the literature on interaction research, highlighting relevant video-based studies and important limitations of manual video analysis. Second, we provide a short methodological primer on CV and then discuss CV's potential to fill these gaps, and specifically the ways 3DSR advances the existing CV toolbox. Third, we describe 3DSR in detail and present the data we use in this article: eight videos of a scripted interaction and five randomly selected real-life videos of street scenes. Fourth, we demonstrate 3DSR's potential, highlighting the range of measurements it facilitates within kinesics and proxemics, two core aspects of social interaction analysis. Specifically, we focus on physical distance, movement in space, and movement rate, for which 3DSR is particularly useful. Fifth, we assess 3DSR's data reliability and find that reliability varies with camera type and filming position. In the conclusion, we argue that despite its limitations 3DSR can advance sociological empirical and theoretical insights that relate to human interaction.

Recent Advances, Challenges, and Opportunities in Social Interaction Analysis

Social Interaction Analysis

Video data introduce new possibilities for the analysis of social interaction, which has been a core endeavor in the social sciences. Video data are detailed, multimodal, and infinitely re-watchable. Researchers can use such data to examine social interactions step-by-step and frame-by-frame, access first-hand recordings of situations and events that they did not observe in situ, and analyze the same situation multiple times (for an overview of advantages of video data, a discussion of analytical perspectives, and reflections on limitations of video data, see Nassauer and Legewie 2019, 2021, 2022; also see DeCuir-Gunby, Marshall, and McCulloch 2012; Derry et al. 2010; LeBaron et al. 2018; Lindegaard and Bernasco 2018; Makin, Willits, and Brooks 2021; Pauwels 2010). For brevity and clarity, we focus on three facets of social interaction: individuals' physical distances, how individuals move in space, and individuals' movement rates. These facets, which are overlapping yet analytically distinct, are highly relevant for studying proxemics (i.e., how people use physical and personal space; Hall 1966) and kinesics (the study of human body movement; Birdwhistell 2010), which are core aspects of social interaction with broad relevance for foundational sociological theories (e.g., Collins 2004; Goffman 1971). Before discussing the challenges that video-based social interaction analysis presents and the opportunities CV offers to address those challenges, we review how social scientists have studied kinesics and proxemics through video analysis and direct ethnographic observations without CV or other computational methods.

Physical Distance. Physical distance refers to the distance between individuals or between individuals and objects. A large body of research has examined individuals' physical distance during interactions to study children's gender socialization (Gansen 2017), the significance of personal space, and how it affects situational dynamics (e.g., Dabbs Jr and Stokes III 1975; Goffman 1971; Hall 1966; Sobel and Lillith 1975). One strain of this literature focuses on situational dynamics of race and discrimination (Brown 1981;

Goff, Steele, and Davies 2008; Hendricks and Bootzin 1976). Brown (1981), for example, finds that personal space in conversational dyads in a shopping mall is invaded more frequently if both members of the dyad are Black. Hendricks and Bootzin (1976), as well as Goff and colleagues (2008), find that White respondents maintained a greater distance from Black respondents during conversations.

Physical distance can also reveal dynamics of rhythmic entrainment, i.e., how individuals fall into shared movement patterns during interactions. Rhythmic entrainment serves fundamental social functions, such as inducing affect and emotion (Trost, Labbé, and Grandjean 2017), and is thus a crucial part of interactions (Collins 2004:28; Durkheim 2008:231–32; Merker, Madison, and Eckerdal 2009). In a study of convenience store robberies, for instance, Nassauer (2018) analyzes how robbers and store clerks often keep a fairly constant distance while circling each other in standoff-like situations, falling into an almost dance-like routine. Rhythmic entrainment of body movements also matters for collective action and group dynamics: Jackson et al. (2018) show that cognitive arousal increases the positive effect of synchronization on cooperation.

Distances between individuals are also useful for studying social ties. Scholars of intimate relations, for example, conceive of intimacy not only in terms of the substance of relationships, but also individuals' proximity (Zelizer 2005). Hall (1966) describes different categories of distance that tend to signal different types of social ties in the U.S., from "intimate distance" (0–0.5 m), through "personal distance" (0.5–1.2 m) and "social distance" (1.2 – 3.7 m), to "public distance" (3.7 m or more). In social network research, survey data is often used over direct observations of social ties (Borgatti et al. 2009:895), but video data provide a unique resource for studying situational social ties in detail. For example, Willis (1966) uses the distance at which individuals start a conversation as a descriptor of their relationship. The author compares initial speaking distance between different role relations, age, gender, and race. The study shows that parents address their children at a similar initial speaking distance as they do strangers, while friends initiate conversations at a closer distance. Physical distance can thus help theorists unravel network patterns as they manifest through interactions in space.

Of course, the meaning of distance also stems from its social context. For instance, during the Covid-19 pandemic, many countries instituted physical distance of around 1.5 m as a guideline for safe navigation of public spaces. Hoeben et al. (2021) use video data from CCTV cameras in Amsterdam to study violations of social distancing guidelines in public spaces. They find that after initial widespread compliance, levels of compliance – operationalized as distance maintained between individuals – quickly wane. Physical contact (i.e., extremely small distances) can also vary in meaning, signaling violence (e.g., Adang 2018; Collins 2008; Nassauer 2019), intimacy, solidarity, and emotional support (e.g., Argyle and Dean 1965; Lindegaard et al. 2017). In developmental psychology, scholars often examine the moment in which toddlers or children first touch an object (Włodarczyk et al. 2018).

Movement in Space. Movement in space refers to where individuals move during an interaction and how much physical space they occupy. Space is an important but often overlooked category of sociological analysis. It can help explain phenomena such as the formation of social relations (Small 2009) and power structures (Harvey 2014; Lefebvre and Nicholson-Smith 1991). For social interaction research, movement in space is a core element because it is foundational to interaction rituals (Collins 2004; Goffman 1971, 1982), and can drive situational dynamics. For example, comportment, or how individuals move bodies and body parts in space, is central for "doing" gender (West and Zimmerman 1987).

Movement in space can also reveal other structures of inequality and power. For instance, the use of physical space can be associated with dominance, as when male-presenting gym patrons seek to maintain dominance over the social space by practicing regionalization of "male" and "female" spaces in the gym (Hertzog and Lev 2019:845; Johansson 1996:32). Robbers can claim dominance during robberies through body postures, which are determined by body part movement in space, where some weapons are more conducive to achieving dominant postures than others (Mosselman, Weenink, and Lindegaard 2018). On the group level, the symbolic value of space can play an important role, for example, in incursions into protest spaces by police that protesters perceive as symbolically charged. Nassauer (2021) shows that such incursions can form part of interaction chains that turn peaceful protests into violent altercations.

Movement in space can also reveal dynamics of collective behavior. Faria et al. (2010) study jaywalking as a case of rule-breaking in everyday life and focus on individuals' movement in space. The authors analyze the sequence in which individuals waiting at an intersection in a UK city cross the street during a red light. They find that a person was up to two times more likely to start crossing if their nearest neighbor crossed. Such local social influence typically results in wave-like patterns of crossing pedestrians. Philpot and Levine (2021) use a similar approach to analyze local social influence in evacuation behavior during subway train emergencies. Movement in space is thus a crucial component in many social processes at the core of foundational sociological theories.

Movement Rate. Movement rate refers to the rate with which individuals or groups move their entire body or constituent body parts. Movement rates can indicate situational engagement. For instance, Collins' (2004) theory of interaction ritual chains marks "emotional energy" as a core tenant of successful interactions, which in turn can lead to heightened group identification (Collins 2004:108–9). Movement rate, either slow or fast, can potentially indicate situational engagement that can generate emotional energy in certain contexts, which could shed further light on social interactions such as courtship (McFarland, Jurafsky, and Rawlings 2013). Lagged movement rate can be used to understand how individuals anticipate one another's movement in different contexts (Tavory and Eliasoph 2013), and movement rate can also illuminate comparisons between different groups or types of interactions, such as adults and children (Thorne 1993).

Movement rates can also provide measurements of situational dominance. For instance, Nassauer (2018) argues that situational dominance was a crucial factor in successful armed robberies in convenience stores, finding that clerks hardly moved at all during these interactions. In a similar vein, Klusemann (2009) analyzes an 8-h-video of the military negotiation between Generals Mladic and Karremans prior to the Srebenica massacre. The author describes how, over the course of the interaction, Mladic gained situational dominance and Karremans adopted a submissive demeanor, including stiffness and reduced body movement.

In addition to situational engagement and dominance, movement rates of discrete body parts can indicate embodied cultural capital (Bourdieu 1984:218; Harvey 2022). For example, in his study of schools educating children of different class backgrounds, Harvey (2022) shows that some embodied behaviors signal cultural capital through their speed as children are taught to shake hands with a certain vigor.

Finally, movement rates can measure prosocial behavior in some contexts. In their study of evacuation behavior during subway train emergencies, for example, Philpot and Levine (2021:9–10) distinguish between running and walking as indicators of different reactions to an explosion in a subway car. This movement rate-based distinction helps the authors disentangle the situational dynamics of the evacuation scenario, with running representing a more dangerous and challenging behavior.

In short, movement rate of either entire bodies or discrete body parts, just as physical distance, and movement in space, can help generate insights into a wide range of important sociological theories from cultural capital to prosocial behavior.

Challenges and Opportunities in Video-Based Social Interaction Analysis

Two Challenges in Using Video Data for Social Interaction Research. Due to advantages such as re-watchability and shareability, video data have been increasingly popular in studying movement in space, movement rate, and physical distance. Yet, non-computational video-based sociological research faces two fundamental challenges. First, video data's richness makes their analysis time and resource intensive. This is true for qualitative studies, in which researchers analyze a relatively small number of videos by observing fine-grained processes frame by frame (e.g., Klusemann 2009; Nassauer 2018), but also quantitative studies where researchers code relatively large samples for predetermined characteristics to employ statistical analysis (e.g., Hoeben et al. 2021; Levine, Taylor, and Best 2011). For instance, the New Jersey Families Study (Golann et al. 2019; for a similar study, see Ochs and Kremer-Sadlik 2013) placed self-activating video cameras in participants' homes for 2 weeks to capture family interactions and study how families cultivate their pre-school children's skills. Producing hundreds of hours of video data, the authors state that, "one of the key challenges […] is figuring out how to manage and reduce the vast amount of data collected" (2019:398). In a recent quantitative study of social distancing during the Covid-19 pandemic, Hoeben et al. (2021) had "more than 20,000 h of recordings" available but "coded only a small subset of the available data, because manual coding of video behavioral data is very labor-intensive" (Hoeben et al. 2021:4; for a similar issue, see Gentrup et al. 2020). While selective coding and analysis of video data are plausible approaches, larger datasets can quickly overwhelm research teams.[1] Indeed, it can take between 7 and 88 s to identify one or more objects of interest in a single frame (Jain and Grauman 2013; Su, Deng, and Fei-Fei 2012). Identifying complex behaviors (e.g., running vs. pacing) in videos that are comprised of thousands of images will thus be prohibitively time-consuming beyond small sample sizes. In short, video data analysis faces a scalability problem: even short video segments of multiple individuals interacting can take hours to analyze. Hence, video data coding that is more time efficient can offer substantial benefits for both qualitative and quantitative video-based research.

A second challenge for video data analysis is precise measurement of elements such as distance or movement. Researchers are forced to rely on rough estimates or ingenious, but laborious approximations (e.g., Hoeben et al. 2021:4; Stickle et al. 2020:6). In addition, camera perspectives can bias observer estimates of distance (Hoeben et al. 2021:4). Precise measurement is even more challenging when it comes to complex behaviors such as mimicking behavior, rhythmic entrainment, and contagion (Collins 2004; Hatfield et al. 2014). Such behaviors constitute a synchronization of movement, micro-expressions, emotional states, and/or pitch by participants, often subconsciously (Collins 2004:79). They can occur as consequences of situational dynamics, but also impact situational outcomes. Despite these behaviors' frequent occurrence and important role in social interaction, their analysis has been limited, in part because they are difficult to operationalize with precision (Collins 2004:79).

In sum, both quantitative and qualitative researchers interested in kinesics and proxemics in social interaction could benefit from tools that help dissect interaction dynamics with greater precision and measure key elements related to movement and distance at scale.

Computer Vision's Potential for Video-Based Interaction Analysis. CV refers to software-based automated analysis of visual material (for brief introductions to CV in the social sciences, see Torres and Cantú 2021; Zhang and Peng 2021; for an in-depth introduction, see Webb Williams, Casas, and Wilkerson 2020). CV aims to mimic human visual perception for the purpose of analyzing recorded images (or videos) by applying various algorithms and transformations (Klette 2014). CV thus "automates discovery from [visual] data" (Molina and Garip 2019:28) by transforming the pixels of an image into a vector of numbers and using that numeric information to detect patterns. These patterns then help classify images or videos in meaningful ways. To achieve this, many CV models rely on machine learning algorithms such as neural networks, which use visual data to adjust their parameters iteratively in a training process. The greater the variability and size of the dataset, the higher the chances of the model to generalize and perform well on more difficult tasks (Khan and Al-Habsi 2020).

CV is being applied in such fields as agriculture (Tian et al. 2020), safety and security (Fang et al. 2020), self-driving vehicles (Kohli and Chadha 2020), robotics (Sünderhauf et al. 2018), social media (Wu et al. 2019), human–machine interaction (Ke et al. 2018), and medicine (Goldstein, Schätz, and Avigal 2022; Khanam, Al-Naji, and Chahl 2019), among many others. Typical CV tasks include detecting the presence of objects or actions in space and/or time through classification of video frames or other images (tasks often referred to as localization and segmentation; Khan and

Al-Habsi 2020:1444; Voulodimos et al. 2018:7). Objects, in CV jargon, refer to anything from a chair to a human or other animal. Actions can be isolated, such as "sitting," or compound, such as "playing cricket." CV can be applied to any type of object or action, some of which are especially relevant for video data analysis in the social sciences. For instance, CV can answer questions such as: "in which frames in the video, and in which sections of those frames, can we see a person giving a speech?" In practice, CV applications often entail combinations of several tasks.

Yet, CV is not preferable in every video-based analysis of social interactions. Questions that rely on contextual information outside the captured video cannot be tackled by CV algorithms.[2] Some aspects of human interaction, such as ironic gestures, may also be too complex for CV algorithms in the foreseeable future. In short, human coding may be preferable for many tasks, even those CV can in principle perform. Nonetheless, existing CV tools offer solutions for the challenges presented above (for a more extensive review, see Nassauer and Legewie 2022).

Despite its potential for video-based interaction research, to date, CV has scarcely been used in sociological analyses of elements such as physical distance, movement in space, and movement rate in video data. One primary reason for this lack of application is that CV models have only recently become adept at tracking human bodies.

Some noteworthy exceptions have used CV tools to study social interaction. Dietrich and Sands (2021) employ CV techniques to study racial avoidance in micro-interactions, analyzing videos from 2D New York traffic cameras in a field experimental design. They find that, on average, pedestrians pass Black confederates at greater distances compared to White non-Hispanic confederates. In another study, Dietrich (2021) uses motion detection to measure movement rate as an indicator of inter-party interactions after votes in the U.S. House of Representatives. The author shows that inter-party interactions have decreased since the end of the 1990s, and that sparse inter-party interactions are predictive of party votes in later sessions. Bernasco et al. (2022) employ CV tools to analyze CCTV video data of public spaces during the Covid-19 pandemic. Compared to Hoeben et al.'s (2021) related study, which relies on human coding, Bernasco and colleagues analyze a much larger sample of videos to draw conclusions on social distancing violations, thus demonstrating the potential of CV tools for video data analysis of social interactions at scale. Neumann et al. (2022) study body language in political campaign videos in the U.S. They employ automatic pose detection and find that, in line with gender stereotypes, male candidates move their hands more assertively compared to female candidates.

The above studies contributed important insights into their respective research fields by developing case-specific innovative CV tools for analyzing video. However, to enable broader use of CV, researchers would benefit from a generalizable, flexible, and precise tool for measuring aspects such as physical distance, movement in space, and movement rate. For instance, both Dietrich and Sands (2021) and Bernasco et al. (2022) went to great lengths to develop and validate strategies for reliably measuring physical distance between pedestrians.

We propose a new standardized method, 3DSR, which can be applied to different research settings without substantial tailoring. 3DSR thus offers a foundational tool for video-based research on kinesics and proxemics in social interactions. It extends other CV tools by providing comprehensive data on individuals' position and movement during social interactions, down to specific body parts. This allows precise measurements of distances and locations, but also affords researchers considerable flexibility for extracting the elements important for their research. Below, we illustrate the capabilities of 3DSR by showcasing a range of potential measurements for physical distance, movement in space, and movement rate.

Researchers can employ 3DSR in various ways and at different points in the research process. For instance, after data collection researchers can use 3DSR to filter a large corpus of video data for relevant videos (or sections of videos) based on criteria such as a minimum number of co-present individuals or duration of interactions. As we show below, 3DSR also enables extracting measures from video data that are difficult to obtain through manual coding, such as precise physical distances or movement rates. These precise measures could, again, be used for both quantitative and qualitative analyses.

Data and Methods

3D Social Research

To implement 3DSR, we used widely available 3D cameras in combination with an existing skeleton tracking CV model called "OpenPose" (Cao et al. 2021). OpenPose is widely used and compatible with most operating systems. 3D cameras are increasingly common and are now built into many mobile phones. Thus, 3DSR is an accessible resource for researchers aiming to self-record videos of real-life interactions, and may allow analysis of videos recorded by third parties in the near future.

OpenPose identifies the locations of skeleton "keypoints" – joints and body parts such as ears, hips, and hands – in 2D color images or video. The number of keypoints depends on the exact OpenPose variant; 3DSR employs an OpenPose model that tracks 21 keypoints, which we list in Figure 1.

OpenPose provides the pixel location of each keypoint on the 2D image using (x, y) coordinates.[3] For example, in figure 1 it might place the thorax keypoint in the pixel "(517, 431)".[4] Additionally, OpenPose provides a confidence score ranging between 0 and 1000 for each keypoint, marking the estimated certainty that the keypoint resides in the identified pixel. Say the above keypoint was identified with a confidence level of 600, OpenPose would mark the thorax output above as "(517, 431, 600)."

OpenPose can track multiple individuals within the same frame. Other software packages (e.g., RealSense's Skeleton Tracking SDK) also track human movement, but OpenPose is more efficient, consistent, and accurate when

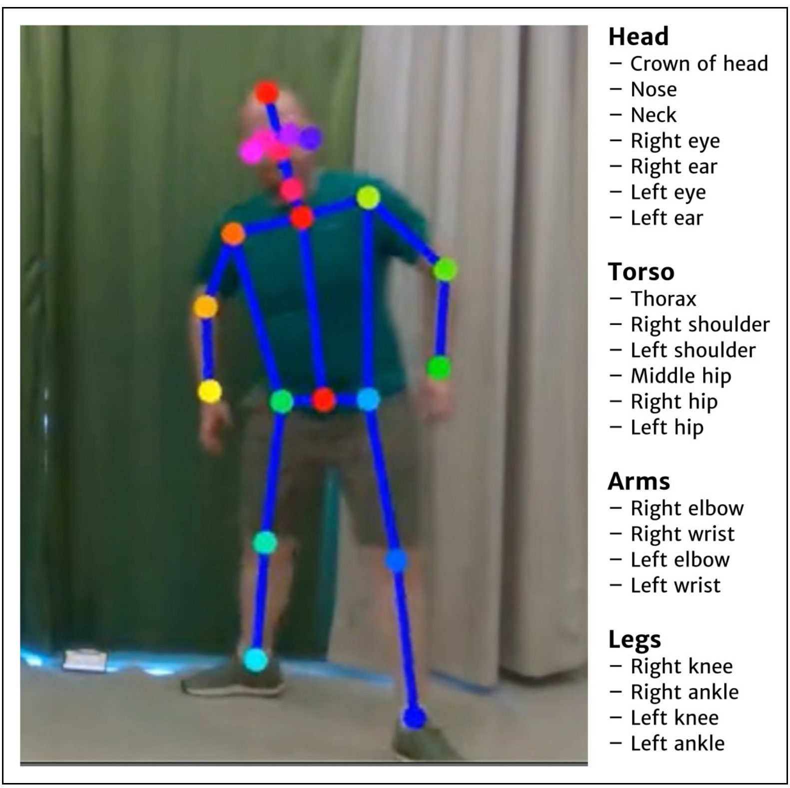

Figure 1. Overview of 21 keypoints.

following multiple people (Cao et al. 2021; Raaj et al. 2019). Raaj et al.'s (2019) version of OpenPose improves the algorithm's tracking by (largely) resolving the substantial challenge of following individuals across frames. The expansion assigns a consistent ID to each individual throughout the video (though some limitations remain, which we discuss in the reliability section below).

To measure individuals' location not merely in a 2D image (as OpenPose currently allows), but in 3D space, we use depth cameras (or "3D cameras"). In addition to the traditional color data encapsulated within each pixel in a video, 3D cameras detect the distance of every pixel from the camera lens and express it as a depth value. Recent applications of CV, such as self-driving cars, use this technology to measure distances between cars and objects. 3D cameras are extremely valuable in CV fields such as object recognition, human motion capture, human–computer interaction, and 3D reconstruction (Grzegorzek et al. 2013). They operate using several technologies. One method uses stereo triangulation of two cameras, which is similar to the human perception system. Another method known as "light imaging detection and ranging" (or LIDAR) uses a laser beam to illuminate a target and measure the time of the returned reflected light. Because the speed of light is a known constant, the camera can calculate the actual distance of the 3D point hit by the ray (Horaud et al. 2016).[5] In 3DSR, we transform 2D coordinates (x, y) into 3D coordinates (x, y, z) by matching color images with respective depth images, adding each keypoint's distance from the camera. So, depending on the specific camera, the 2D thorax coordinate above may transform into (0.48, 0.46, 1.48), marking a point in space that is 0.48 m right of camera center, 0.46 m above it, and 1.48 m away from it. This 3D coordinate inherits OpenPose's confidence score (in the above example, 600).

In our implementation of 3DSR, we recorded 3D videos with RealSense cameras and associated RealSense software, which automatically separates each frame into a depth and a color image.[6] We input the color images into OpenPose frame by frame to obtain skeleton keypoints' pixel coordinates on the image. Our resulting dataset includes coordinates for a set of keypoints for every individual in every frame. In total, a 100 s video produces 3,000 frames (because our cameras capture 30 frames per second). When specifying 21 keypoints, this produces 62,000 keypoints per individual. The resulting tables, which 3DSR outputs, include a row for each video frame for each individual, and columns for each keypoint's X, Y, and Z coordinates, as well as the OpenPose confidence value.

We prepared a Github repository and Jupyter Notebook that allow interested readers to run our entire data processing and analysis pipeline in an online environment.[7] These resources provide access to our data, Python and R code, figures, and tables, and can be used as the basis to explore 3DSR's capabilities and incorporate 3DSR in future research. We also include in the repository a list of citations for the Python and R packages we drew on to build 3DSR.

Data Collection

We demonstrate 3DSR with eight scripted videos that were designed to test its capabilities, filmed using Intel's RealSense D415 and L515 cameras. D415 is a "passive stereo" camera, which uses the constant angle between two lenses to triangulate depth, while L515 uses LIDAR. Each technology has advantages and disadvantages, and other alternatives exist. We compared these cameras as they were widely available and relatively affordable.

To systematically assess the potential and challenges of using 3DSR for sociological research, we filmed eight videos displaying two people enacting the same scripted sequence of actions.[8] For instance, in every video at around 1:25 min, person 2 touches the shoulder of person 1 (appendix B in Supplementary Material). The script was intended to include a range of body positions and movements common in social interactions, but also situations that may be challenging for 3DSR to manage (e.g., persons circling each other, hugging).

As Table 1 shows, the eight videos differ in the type of camera (D415 vs. L515), lighting (dark vs. light), angle of filming (right angle facing the interaction vs. side view), and distance from the camera (close vs. far). We chose these elements as previous research suggests they produce variation in 3D camera data quality (Breitbarth, Hake, and Notni 2021; Murcia et al. 2021). For simplicity, we base all figures on data from one video, "D415 (setting: far, light)," which proved most reliable (see reliability section below, as well as appendix C in Supplementary Material).

To assess how 3DSR performs in an uncontrolled setting, we supplement these scripted interactions with a random sample of five videos from a corpus of 400 street scene recordings, which we filmed on public walkways for a project that analyzes distances between individuals during the Covid-19 pandemic.[9] We randomly selected videos that contain two individuals for most of their duration to make these real-world videos largely comparable to the eight scripted videos. The recordings, which were filmed at different moments of the Covid-19 pandemic, depict individuals walking on a narrow path. Again, figures throughout the paper are based on one of the five videos for simplicity.

The main purpose of our research design is to demonstrate the range of 3DSR's potential for sociological research in controlled and uncontrolled

Table 1. Characteristics of the Eight Scripted Videos.

	Close, lit	Close, lit, side	Far, dark	Far, dark, side	Far, lit
L515	✓	✓	✓	✓	✓
D415	✓		✓		✓

settings, and to systematically assess issues of reliability by comparing similar (scripted) videos that vary on key dimensions. Rather than providing an in-depth demonstration of a single 3DSR measure that may be used for a specific research project, we instead showcase a broad array of measures with possible applications in a variety of subfields.

3DSR's Potential for Social Interaction Research

To demonstrate 3DSR's potential in sociological research, we focus on kinesics and proxemics, two core aspects of social interaction. More specifically, we illustrate 3DSR's capabilities for studying physical distance, movement in space, and movement rate. Scalable and precise measurement of each of these aspects, as well as other potential measures, derives from 3DSR's core capability to locate each individual's keypoints in physical space multiple times per second. We demonstrate these applications below through a range of measurements using both scripted interactions and real-world videos.

Physical Distance

Since 3DSR tracks individuals' locations over time, it can measure physical distances between individuals throughout an interaction. Figure 2A displays the distance between the two individuals in our scripted video, operationalized through the distances between their respective thorax keypoints over frames. The blue line smooths out noise from misclassified keypoints. As expected from our script, the distance between the two rapidly decreases as they take their starting positions, remains constant at around 1.2 m throughout a relatively long period, and then fluctuates as the two circle one another. Around frame 4000, before filming ends, they hug – the distance between them approaching zero.

Figure 2B shows the corresponding measure for the street-scene video. In accordance with the fact that the two individuals in the scene walked approximately parallel throughout the video, the distance between their thorax

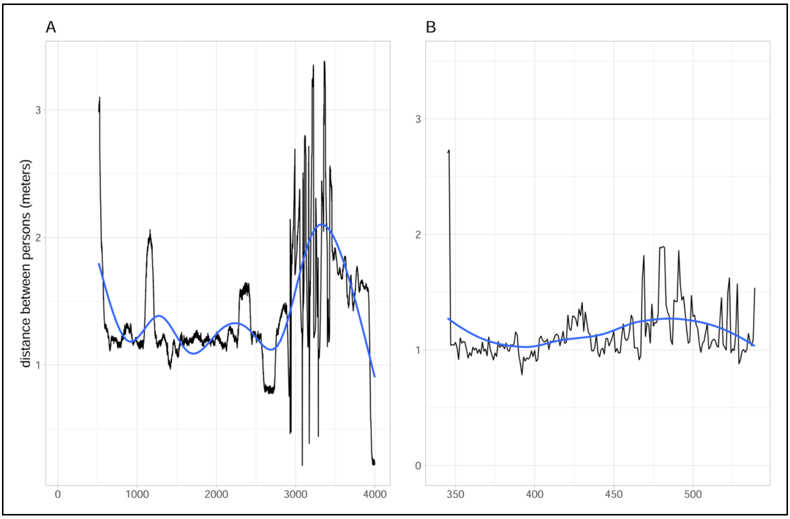

Figure 2. Distance between two individuals' thorax keypoints over frames. Note: Extreme values stemming from keypoint misclassification are cropped from this figure.

keypoints remains roughly constant, around 1.2 m. One could also measure physical distances between different body parts, for example, the distance between one person's hand and another's eyes, as an operationalization of breaching personal space or of excitement, or someone placing their hand on their chest as an indicator of self-calming behavior.

3DSR produces data that is highly flexible; its output can be used not only for quantitative analyses, but also for qualitative inquiry. For instance, Figure 3 shows the continuous physical distance between the two individuals during the interaction as an animated graph. This type of visualization greatly facilitates reconstructing sequences and flow in social interactions. It can allow qualitative researchers to focus on certain aspects of the interaction without being distracted by the full breadth of information present in raw video data. The resulting findings can later be combined with insights from other analytical procedures.

Using such continuous measures of physical distance, 3DSR facilitates a range of analyses, such as assessment of interactional space. With 3DSR's output, we can use distance categorizations such as Hall's (1966) classification of "intimate distance" (0 – 0.5 m), "personal distance" (0.5–1.2 m),

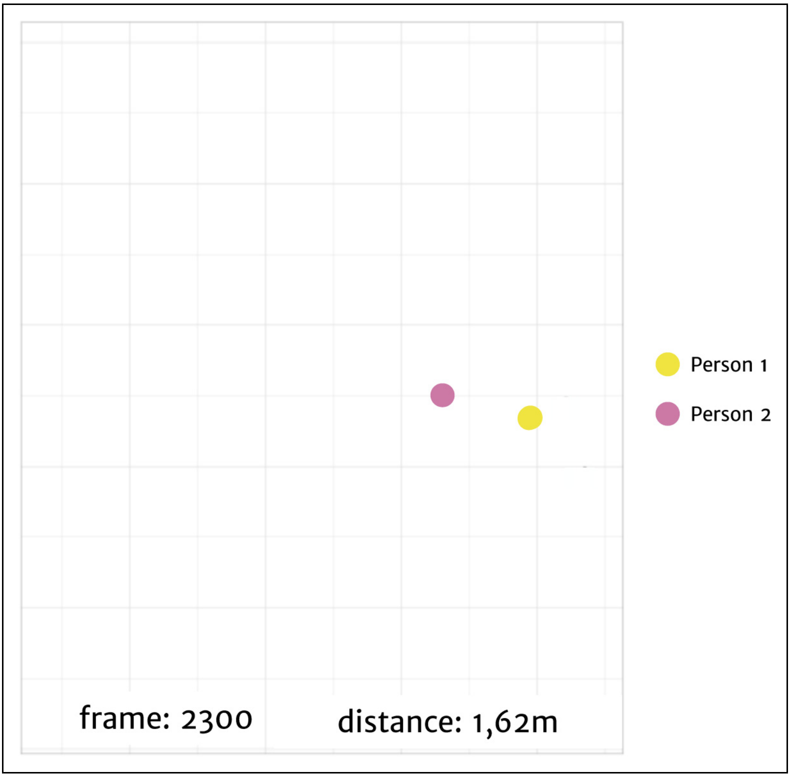

Figure 3. Continuous position of individuals.
Note: Since it is not possible to display animated figures in print, we provide Figure 3 online: http://doronshiffersebba.com/3dsr/.

“social distance” (1.2 – 3.7 m), and “public distance” (3.7 m or more), and assess the frequency with which individuals’ physical distance falls in each category over the course of an interaction. Figure 4A shows the relative frequency the two individuals in our scripted interaction were found in each of the distance categories. We see that during the interaction, personal distance dominated (more than 50 percent of frames), followed by social distance (around 40 percent), with some intimate distance (about 3 percent of frames). Public distance between the two individuals was altogether absent during the interaction. By adjusting the distance threshold, we could use other distance categorizations, such as the social distancing guideline

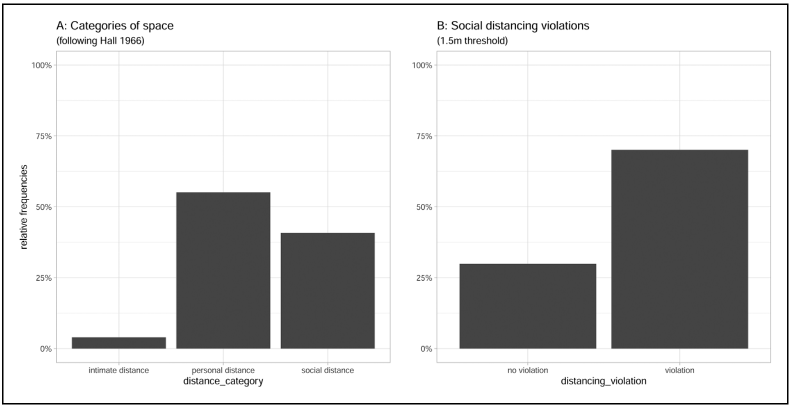

Figure 4. Frequency of physical distance categories in video frames.

during the Covid-19 pandemic (Bernasco et al. 2022; Hoeben et al. 2021). Figure 4B shows the frequency of frames in which the two individuals were closer than 1.5 m, which constituted a violation of social distancing guidelines during the Covid-19 pandemic in many contexts. Of course, it might also be useful to analyze in which frames social distancing was breached, e.g., to understand which other physical behaviors occurred prior to or concurrent with breach of distancing. 3DSR enables such analyses by allowing distance measures frame by frame.

Scholars studying situational dynamics such as rhythmic entrainment, i.e., how individuals in a situation fall into a common movement pattern, may find these capabilities particularly useful. 3DSR's ability to track individuals' physical distance over time facilitates such analyses and, through its increased precision, opens new possibilities for cross-case comparisons, for example, by identifying different types of entrainment. 3DSR measures of physical distance can also help operationalize how people move in response to others. Finally, through tracking physical distance, 3DSR can be used to identify instances of specific distances such as direct physical contact.

Movement in Space

3DSR allows mapping individuals' positions over the course of an interaction by providing the absolute location of every detected body part for each individual. Figure 5 exhibits one way of visualizing and analyzing such movement: a bird's eye view of two people's nose keypoint position over time,

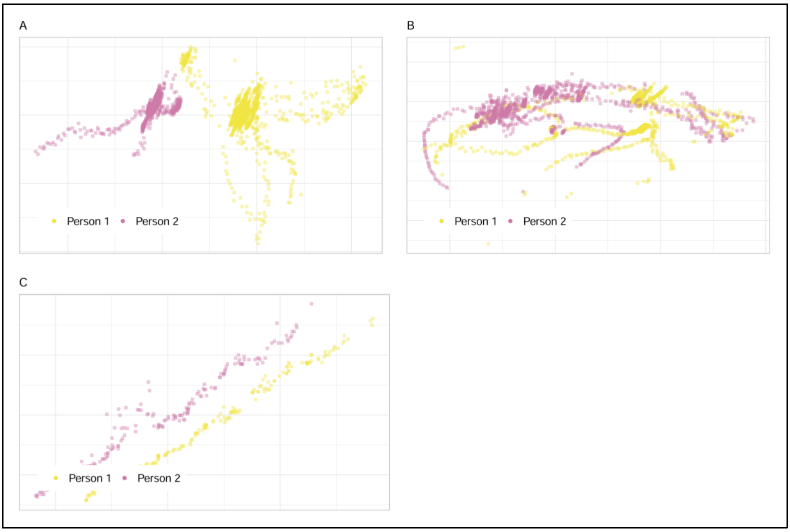

Figure 5. Bird's eye view of individuals' movement in space over time.

separately for two sections of the video. In the first half of the interaction (figure 5A), participants faced one another. In the second, shorter section (figure 5B), individuals circled one another. Figure 5 thus reveals the spatial separation of the individuals in the first half of the interaction (5A), where each individual moved in their own discrete region. Person 1 limited themselves to a 1.2 m^2 area of the room, while person 2 traversed around 1.5 m^2. In the second half (5B), on the other hand, the two enter one another's areas, walking in each other's footsteps. (Figure 3, shown above, shows the animated version of this analysis, without keeping track of individuals' prior locations).

Figure 5C shows a corresponding measurement from the street-scene video. In it, we see two individuals walking on two mostly parallel lines. 3DSR can provide the precise location of multiple individuals' 21 keypoints over time.

Individuals' movement in space, both their entire bodies and discrete body parts, can serve to illuminate a plethora of sociological phenomena. For instance, precise mapping of movement in space can help understand how team members collaborate in work environments, how members of different demographic groups move in social spaces depending on their social status, or how political candidates employ space to project confidence and a connection with the audience during television debates. 3DSR provides a flexible approach for collecting and analyzing data on such movement.

Movement Rate

3DSR can measure individuals' movement rates by computing the distance each skeleton keypoint "travels" between frames; not only individuals' entire bodies, but also discrete body parts such as hands or necks. Figure 6A depicts the mean movement rate of individuals' keypoints from the scripted video over frames. Specifically, we sum the distance every keypoint traveled from the previous frame for all non-missing keypoints in a set of two consecutive frames and divide the sum by the number of keypoints.

Movement rates can reveal different behaviors and states. For instance, lower rates of body movement, such as those during the first 2500 frames of the scripted video, indicate different behaviors from the faster movement rates around frame 3000. Indeed, in our script individuals stay mostly stationary in the first half of the video but move around the room in the second half. Difference across individuals in movement rates, such as the difference between the two individuals around frame 1000 in Figure 6A, can indicate distinct interactional roles. Indeed, that portion of the script had one individual lean away from the other, while their counterpart stood still.

Figure 6B measures the movement rate of individuals in the street-scene video. Individuals only enter the picture around frame 350 so the movement rate is omitted prior to that frame. From frame 350, we see a similar pattern across the two individuals during the first half of the video, where both move at a similar rate across frames. In the second half of the video, person 1 increases their pace slightly while person 2 maintains a steady pace. In light of Figure 5C, which showed the pair walking in parallel, this figure may indicate a loss of rhythmic entrainment whereby two interacting individuals no longer match their movement patterns. Very different movement rates might have indicated more distinct roles across individuals or a form of interactional failure, depending on the study's setting.

In some cases, it may be more informative to examine movement rates of discrete body parts. 3DSR enables such fine-grained analysis, as illustrated in Figure 7A. The figure depicts the average keypoint movement rate by individual over frames, separated into keypoint groups: arms, head, legs, and torso. Specifically, each keypoint group averages certain keypoints (e.g., the

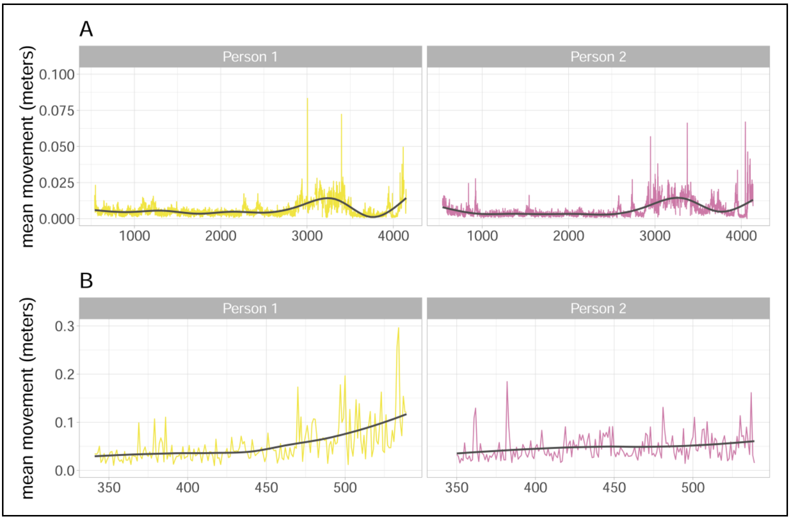

Figure 6. Comparison of aggregate movement of all keypoints over frames.
Note: We use a 5-frame moving median in order to achieve a more reliable distance measure. Extreme values stemming from keypoint misclassification are cropped from this figure.

keypoint group "legs" averages an individual's knee and ankle keypoints). Within each keypoint group, we calculate the mean movement in a way analogous to the total movement measured above. This body part movement rate shows that person 1 (left pane) moved their legs (light blue line) more than other body parts around frame 3000, but otherwise moved body parts at similar rates. Person 2 (right pane), meanwhile, moved their arms (red line) at relatively higher rates toward frame 4000. Both individuals moved their arms more in the first few seconds of the interaction. By isolating keypoint groups 3DSR provides a precise yet robust measure for a multitude of interactional behaviors. Depending on the research focus, researchers can compare body part movement rates across individuals, within individuals, across different moments of an interaction, or between multiple interactions.

Figure 7B measures keypoint group movement rates by individual in the street-scene video. Like figure 6B, movement rates are omitted for each body part until around frame 350 when individuals enter the picture. The two individuals display a similar pattern in that both increase the movement rates of some body parts during the video: their heads (green line) and arms

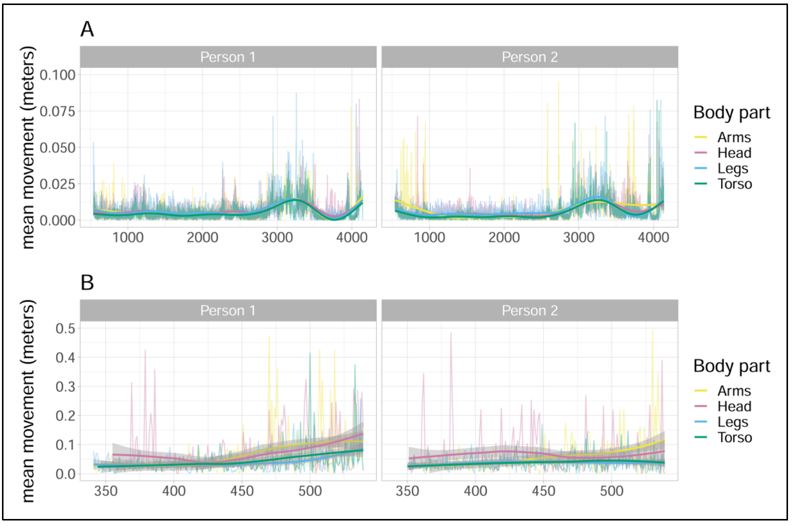

Figure 7. Comparison of aggregate movement of body parts over frames.
Note: We use a 5-frame moving median in order to achieve a more reliable distance measure. Extreme values stemming from keypoint misclassification are cropped from this figure.

(red line) move faster starting around frames 450–500. However, the two individuals also display some differences. Namely, person 1's legs and torso movement rates largely follow their arms and head, increasing throughout the interaction. Meanwhile, person 2's legs and torso sustain a more stable pace. Such measures can be particularly useful either in contexts where some body parts' movement are constrained, for example, while sitting on a chair or standing behind a podium, or in studies of specific elements of communication, such as frequency of hand gestures (e.g., Alibali and Nathan 2007).

Across physical distance, movement in space, and movement rate, then, 3DSR offers granularity and precision at scale, allowing flexible measurements for tackling sociological questions. The potential applications are numerous and can facilitate new study designs for exploring core areas of social inquiry.

3DSR and Data Quality

3DSR is not without its limitations. It is better suited for some videotaped interactions over others, and several filming features impact data reliability.

This section alerts researchers who seek to collect video data for analysis with 3DSR of its most important challenges.

Reliability Measures

In broad terms, there are three challenges to data reliability when using 3DSR: missing keypoint data, where some or all keypoints of a person are missing in a given frame; keypoint measurement error, where keypoints receive an erroneous value or are labeled incorrectly (e.g., a left eye is labeled as a right eye); and inconsistent labeling of individuals across frames, where an individual is labeled "person 1" in one frame and "person 2" in another. All three challenges, examined thoroughly in Appendix C in Supplementary Material, stem from either some or all keypoints being outside the frame; an extreme distance from the camera (either closer than 2.5 m or further than 10 m); or occlusion, where some or all person keypoints are blocked from the camera's view by the person's own body, another person, or an object. In other words, these reliability issues emerge in specific recording scenarios and may, in some cases, be minimized through careful planning of the recording setup. Moreover, depending on the research question and theoretical lens, certain reliability issues may not matter for analysis; e.g., inconsistent labeling of individuals is irrelevant if the researcher is only interested in the physical distance between two people. Table 2 presents indicators of reliability of 3DSR, both in the scripted videos (top panel) and street-scene videos (bottom panel).

Table 2 shows that data reliability differs across videos and that data are generally more reliable in the wild than in our scripted videos, which were designed to test 3DSR's capabilities (we provide a more elaborate discussion of Table 2 in Appendix C in Supplementary Material). Table 3 shows a systematic comparison of video features using only the (comparable) scripted videos by averaging values within each dimension. For instance, when comparing camera models, we only compared scenarios captured by both cameras ("settings: far, light" and "settings: close, light"; four videos).

Table 3 indicates that the largest gap in missingness was between different distances from the camera, with a farther distance resulting in fewer missing keypoints. Similarly, distance had the greatest impact on mean confidence values, with a 214-point difference between close (380) and far (594) filming distances. For misclassification, however, camera model had the greatest impact, with D415 resulting in fewer misclassified keypoints than L515. Inconsistent labeling of individuals (or "PID switches") appear influenced by both distance from the camera and camera model.

Table 2. Reliability Comparison.

Video	% keypoints missing	% all keypoints missing	N of misclassifications	Confidence (mean)
Scripted videos				
D415 (setting: close, light)	51.88	0.04	10	331.57
D415 (setting: far, dark)	22.74	0.00	80	554.49
D415 (setting: far, light)	19.11	0.00	66	598.50
L515 (setting: close, light, side)	37.91	0.00	168	428.83
L515 (setting: close, light)	47.34	0.01	177	313.32
L515 (setting: far, dark)	30.92	0.00	658	576.02
L515 (setting: far, dark, side)	32.79	0.00	1029	573.53
L515 (setting: far, light)	36.83	0.00	543	588.72
Street scene videos				
Street Scene I	13.31	0.00	818	744.52
Street Scene II	22.42	0.03	264	748.33
Street Scene III	15.61	0.00	214	739.06
Street Scene IV	17.31	0.00	124	744.90
Street Scene V	27.18	0.04	428	629.35

Table 3. Reliability Comparison by key Recording Feature.

Comparison	Group	% missing (mean)	% all missing (mean)	*N* of misclassifications (mean)	Mean confidence (mean)	*N* of PID switches (mean)
Camera	D415	35.50	0.02	38.00	465.03	11.00
	L515	37.37	0.00	355.50	508.78	5.00
Distance	close	44.89	0.02	89.00	380.20	10.50
	far	27.97	0.00	304.50	593.61	5.50
Lighting	dark	26.83	0.00	369.00	565.26	6.50
	light	27.97	0.00	304.50	593.61	5.50
Angle	front	34.41	0.00	413.00	502.43	7.00
	side	40.07	0.01	603.00	443.42	9.00

Note: All comparisons were made holding other dimensions constant; as a result, four to six out of eight videos were used for each comparison. PID switches refer to instances of changes in an individual's label by the OpenPose algorithm.

Optimizing Filming Setup

Our analyses suggest that 3DSR contains a reliability tradeoff across video features: aspects of the video that minimize missingness and misclassification tend to introduce more inconsistent labeling of individuals. If a given research scenario lends itself to resolving identifier switches (or if they do not matter for the research question), using the D415 camera model from a far distance and a frontal angle proved the most reliable of the options investigated. From a strictly technical perspective, switches and missingness are best resolved with careful design of the filmed scenarios, parallel to how one would proceed with most self-recorded video data, even if only intended for manual analysis (Nassauer and Legewie 2021:154–57, 2022:111–44): minimizing all types of occlusion (by objects or other people) and making sure the selected frame fully captures the people filmed throughout the video. Filming from an elevation may also minimize occlusion, which may be especially helpful for interactions involving several individuals. Researchers could even potentially triangulate footage from several 3D cameras simultaneously to minimize occlusions. However, such an approach will likely introduce new technological challenges, such as matching individuals across different videos (for an approach in the field of behavioral biology, see Francisco, Nührenberg, and Jordan 2020).

Generalizability to Real-Life Videos

Hypothetically, all reliability issues can be resolved with a different placement of the camera. Therefore, the implications of reliability issues for studies' inferences are context dependent – they are tied to the influence of the camera's position on the scenario filmed. The scripted scenario we used included sequences challenging to 3DSR, producing largely lower quality data than the street-scene videos. Still, these use cases are very different from many potential research contexts, some of which may exacerbate challenges. Many people in a confined space could produce a great deal of occlusion and hence lower data reliability. The presence of furniture or other large objects, or a contorted space, could all lead to greater occlusion. Real-life situations can also be unpredictable, and interactions may move to locations outside the video frame.[10] On the other hand, some real-life scenarios may involve fewer challenges. If two or more individuals move in a wide space without getting near one another, missing keypoint data and identifier switches will likely be less frequent. The same is true of individuals engaging in a stationary conversation, for instance during political debates. Since candidates often remain behind a podium, full occlusion would be virtually non-existent.

A further question is whether 3DSR works on diverse skin tones, body types, and people of different gender presentation. Since 3DSR builds on the OpenPose library, it inherits possible biases from the models used in that library. OpenPose was trained on the COCO and MPII datasets (Lin et al. 2014). It is, therefore, used best when applied to images that resemble those data. Though one should not necessarily expect lower reliability when OpenPose is used with non-White individuals or different body types, this question requires explicit investigation through further research, as COCO and other image datasets have been shown to produce biased results along elements such as gender (Wang et al. 2019).

Conclusion

Recent developments in video technology and dissemination offer new opportunities for using rich video data for studying foundational phenomena in sociology. Luckily, computer scientists have simultaneously developed tools, such as computer vision (CV), that enable sociologists to make use of these data at scale. CV models can and have been used in the analysis and data management in research fields as diverse as violence, public behavior, urbanization, education, child development, and others. In this article, we introduce 3DSR, a CV approach we developed, which offers precise measurement of individuals' location in space, allowing the study of foundational sociological theories in fields such as culture and education, among many others. We show how 3DSR enables a range of possible measurements, which expand the analytic toolkit for scalable and precise measurement of core aspects of social interaction.

3DSR can be incorporated into both qualitative and quantitative video-based research. For instance, researchers can use 3DSR to filter large video corpora into relevant samples of videos, or they can measure certain aspects of an interaction, such as the precise distance between individuals throughout an interaction. Such automated measurement could free up time for more in-depth analyses in qualitative applications or facilitate the analysis of larger samples for quantitative video-based research. Measures extracted from video data using 3DSR can be used to operationalize the central phenomenon of interest or to explain outcomes in combination with other analytical approaches, both qualitative and quantitative.

3DSR presents new analytical lenses to existing theories, including typologies of interaction space (Hall 1966), situational dominance (Klusemann 2009; Nassauer 2021), and social avoidance (Dietrich and Sands 2021; Hendricks and Bootzin 1976). 3DSR is particularly useful for micro-sociological theories involving interpersonal interactions, as it enables measurement of movement of discrete body parts that can help understand the role of unconscious physical behavior, emotional energy, mimicking behavior, and more. Indeed, 3DSR might be of interest to sociological domains that pay attention to proxemics and kinesics but do not currently make much use of video data. For example, it can help scholars of intimate relations (Zelizer 2005) identify such ties based on physical distance, it can reveal patterns of comportment to researchers investigating how gender is "done" (West and Zimmerman 1987), and it can help identify moments of socialization among children, such as which children are allowed to be proximate or touch others (Gansen 2017). 3DSR introduces a new level of granularity into interaction analysis, analogous to a powerful microscope that enables measuring a new level of reality.

Challenges to applying 3DSR remain, both technological and conceptual, and 3DSR is not appropriate for every project. Some research questions focus on aspects of interactions that lie outside the scope of 3DSR, such as the nature of speech. These require alternative tools, whether manual or computational. Gender and racial bias in algorithms (Benjamin 2019; Crawford and Paglen 2021) is another challenge that requires additional examination (Wang et al. 2019; Wang, Narayanan, and Russakovsky 2020). Still, the benefits of using 3DSR can be substantial. As we show in this article, careful partnering of appropriate technologies (cameras, CV models), research contexts (number of individuals), and filming techniques (angles, lighting) can produce reliable data to tackle a wide array of sociological questions. We hope scholars consider using these new tools for investigating existing and new areas of research.

Acknowledgements

We are indebted to Anna Kochnev–Goldstein for crucial assistance with data collection. We are also grateful to Annette Lareau, Peter Harvey, Blair Sackett, Ashleigh Cartwright, Judith Eckert, the participants of the 2021 Computer Vision in the Social Sciences Workshop, and the SMR anonymous reviewers for suggestions that improved the article in countless ways. Further, we thank the developers of the software packages we used in this research project. For a comprehensive list of packages and their citations, visit https://github.com/yoavgoldstein1/3D-Social-Research.

Authors' Note

All authors contributed equally.

We prepared a Github repository and Jupyter Notebook that allow interested readers to run our entire data processing and analysis pipeline in an online environment. These resources provide access to our data, Python and R code, figures, and tables, and can be used as the basis to explore 3DSR's capabilities and incorporate 3DSR in future research. To see the repository, visit https://github.com/yoavgoldstein1/3D-Social-Research.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Fritz Thyssen Stiftung Traveling Grant and a, Google Cloud Research Grant.

ORCID iDs

Yoav Goldstein https://orcid.org/0000-0002-0896-773X
Nicolas M. Legewie https://orcid.org/0000-0001-9605-1299
Doron Shiffer-Sebba https://orcid.org/0000-0002-3417-3941

Supplemental Material

The supplemental material for this article is available online.

Notes

1. This is also true for non-video image data (Torres and Cantú 2021).
2. This also applies to analyses relying on manual coding.
3. For a more detailed description of how OpenPose works, see Cao et al. (2021). For a less technical introduction, see Hua (2019). We also provide a brief description in Appendix A in Supplementary Material.
4. The (0,0) point is located at the top left corner of the image.
5. See Wagner (2018) for a useful primer on depth camera technologies.
6. Intel has since discontinued the RealSense camera line. However, RealSense cameras are still available for purchase at the time of publishing, other brands provide similar capabilities, and even smartphones now include 3D cameras.
7. To see the repository, visit https://github.com/yoavgoldstein1/3D-Social-Research
8. An example video can be watched at http://doronshiffersebba.com/3dsr/
9. Data collection was approved by an Institutional Review Board.
10. Again, this caveat applies to any video recording that uses a fixed camera location, whether intended for manual or computational analysis (Nassauer and Legewie 2022).

11. The same is true for non-CV video analysis. Nassauer and Legewie (2021, 2022) speak of "optimal capture" as an important dimension of data quality; if individuals move out of frame, the interaction is not captured fully and the video provides less valid data for analyzing the situation.
12. In more technical terms, we created two distance measures for each keypoint: (1) the distance between the keypoint k(i) in a given frame t(i) for a given person p(i) and that same keypoint k(i) for the same person p(i) and the preceding frame t(i-1) (so p(i)k(i)t(i) - p(i)k(i)t(i-1)). And (2) the distance between the keypoint k(i) in a given frame t(i) for a given person p(i) and that same keypoint k(i) for the other person p(j) and the preceding frame t(i-1) (so p(i)k(i)t(i) - p(j) k(i)t(i-1)). We then compared those two distance measures, and defined as misclassifications if (1) is larger than (2).

References

Adang, Otto M. J. 2018. "A Method for Direct Systematic Observation of Collective Violence and Public Order Policing." *Sociological Methods & Research* 47(4):761-86. doi: 10.1177/0049124116661578.

Alibali, Martha W. and Mitchell J. Nathan. 2007 "Teachers' Gestures as a Means of Scaffolding Students' Understanding: Evidence From an Early Algebra Lesson." Pp. 349-365 in *Video Research in the Learning Sciences*, edited by R. Goldman, R. Pea, B. Barron, and S. J. Derry. New York/London: Routledge.

Alison, Laurence J., Emily Alison, Geraldine Noone, Stamatis Elntib, and Paul Christiansen. 2013. "Why Tough Tactics Fail and Rapport Gets Results: Observing Rapport-Based Interpersonal Techniques (ORBIT) to Generate Useful Information from Terrorists." *Psychology, Public Policy, and Law* 19(4):411-31. doi: 10.1037/a0034564.

Argyle, Michael and Janet Dean. 1965. "Eye-Contact, Distance and Affiliation." *Sociometry* 28(3):289. doi: 10.2307/2786027.

Asan, Onur and Enid Montague. 2014. "Using Video-Based Observation Research Methods in Primary Care Health Encounters to Evaluate Complex Interactions." *Journal of Innovation in Health Informatics* 21(4):161-70. doi: 10.14236/jhi.v21i4.72.

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code. Medford*. Medford, MA: Polity.

Bernasco, Wim, Eveline Hoeben, Dennis Koelma, Lasse Suonperä Liebst, Josephine Thomas, Joska Appelman, Cees Snoek, and Marie Rosenkrantz Lindegaard. 2022. "Promise Into Practice: Application of Computer Vision in Empirical Research on Social Distancing." *Sociological Methods & Research* (online first). doi: 10.31219/osf.io/ex9fy.

Birdwhistell, Ray L. 2010. *Kinesics and Context*. Philadelphia: University of Pennsylvania press.

Borgatti, Stephen P., Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. 2009. "Network Analysis in the Social Sciences." *Science* 323(5916):892-95. doi: 10.1126/science.1165821.

Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Cambridge: Harvard University Press.

Breitbarth, Andreas, Cornelius Hake, and Gunther Notni. 2021. "Measurement Accuracy and Practical Assessment of the Lidar Camera Intel RealSense L515." P. 1178213 in *Optical Measurement Systems for Industrial Inspection XII*. Vol. 11782. International Society for Optics and Photonics.

Brown, Clifford E. 1981. "Shared Space Invasion and Race." *Personality and Social Psychology Bulletin* 7(1):103-8.

Cao, Zhe, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2021. "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43(1):172-86. doi: 10.1109/TPAMI.2019.2929257.

Cicourel, Aaron V. 1974. *Cognitive Sociology: Language and Meaning in Social Interaction*. New York: Free Press.

Collins, Randall. 2004. *Interaction Ritual Chains*. Princeton: Princeton University Press.

Collins, Randall. 2008. *Violence: A Micro-Sociological Theory*. Princeton: Princeton University Press.

Congdon, Eliza L., Miriam A. Novack, and Susan Goldin-Meadow. 2018. "Gesture in Experimental Studies: How Videotape Technology Can Advance Psychological Theory." *Organizational Research Methods* 21(2):489-99. doi: 10.1177/1094428116654548.

Crawford, Kate and Trevor Paglen. 2021. "Excavating AI: the Politics of Images in Machine Learning Training Sets." *AI & SOCIETY* 36. doi: 10.1007/s00146-021-01162-8.

Dabbs, James M.Jr and Neil A. StokesIII. 1975. "Beauty Is Power: the Use of Space on the Sidewalk." *Sociometry* 38(4):551-57.

DeCuir-Gunby, Jessica T., Patricia L. Marshall, and Allison W. McCulloch. 2012. "Using Mixed Methods to Analyze Video Data: a Mathematics Teacher Professional Development Example." *Journal of Mixed Methods Research* 6(3):199-216. doi: 10.1177/1558689811421174.

Derry, Sharon J., Roy D. Pea, Brigid Barron, Randi A. Engle, Frederick Erickson, Ricki Goldman, Rogers Hall, Timothy Koschmann, Jay L. Lemke, Miriam Gamoran Sherin, and Bruce L. Sherin. 2010. "Conducting Video Research in the Learning Sciences: Guidance on Selection, Analysis, Technology, and Ethics." *Journal of the Learning Sciences* 19(1):3-53. doi: 10.1080/10508400903452884.

Dietrich, Bryce J. 2021. "Using Motion Detection to Measure Social Polarization in the U.S. House of Representatives." *Political Analysis* 29(2):250-59. doi: 10.1017/pan.2020.25.

Dietrich, Bryce J. and Melissa Sands. 2021. “Seeing Racial Avoidance on City Streets.” *Working Paper*.

Durkheim, Émile. 2008. *The Elementary Forms of Religious Life*. Oxford: Oxford University Press.

Erickson, Frederick. 1977. “Some Approaches to Inquiry in School-Community Ethnography.” *Anthropology & Education Quarterly* 8(2):58-69.

Fang, Weili, Peter E. D. Love, Hanbin Luo, and Lieyun Ding. 2020. “Computer Vision for Behaviour-Based Safety in Construction: a Review and Future Directions.” *Advanced Engineering Informatics* 43:100980. doi: 10.1016/j.aei.2019.100980.

Faria, Jolyon J., Stefan Krause, and Jens Krause. 2010. “Collective Behavior in Road Crossing Pedestrians: the Role of Social Information.” *Behavioral Ecology* 21(6):1236-42. doi: 10.1093/beheco/arq141.

Francisco, Fritz A., Paul Nührenberg, and Alex Jordan. 2020. “High-Resolution, Non-Invasive Animal Tracking and Reconstruction of Local Environment in Aquatic Ecosystems.” *Movement Ecology* 8(1):27. doi: 10.1186/s40462-020-00214-w.

Gansen, Heidi M. 2017. “Reproducing (and Disrupting) Heteronormativity: Gendered Sexual Socialization in Preschool Classrooms.” *Sociology of Education* 90(3):255-72.

Gentrup, Sarah, Georg Lorenz, Cornelia Kristen, and Irena Kogan. 2020. “Self-Fulfilling Prophecies in the Classroom: Teacher Expectations, Teacher Feedback and Student Achievement.” *Learning and Instruction* 66:101296. doi: 10.1016/j.learninstruc.2019.101296.

Goff, Phillip Atiba, Claude M. Steele, and Paul G. Davies. 2008. “The Space Between Us: Stereotype Threat and Distance in Interracial Contexts.” *Journal of Personality and Social Psychology* 94(1):91.

Goffman, Erving. 1971. *Relations in Public: Microstudies of the Social Order*. London: Allen Lane.

Goffman, Erving. 1982. *Interaction Ritual. Essays on Face-to-Face Behavior*. New York: Pantheon.

Golann, Joanne W., Zitsi Mirakhur, and Thomas J. Espenshade. 2019. “Collecting Ethnographic Video Data for Policy Research.” *American Behavioral Scientist* 63(3):387-403. doi: 10.1177/0002764218820566.

Goldstein, Yoav, Martin Schätz, and Mireille Avigal. 2022. “Chest Area Segmentation in 3D Images of Sleeping Patients.” *Medical & Biological Engineering & Computing* 60(8):2159–2172.

Grzegorzek, Marcin, Reinhard Koch, Andreas Kolb, and Christian Theobalt (eds). 2013. *Time-of-Flight and Depth Imaging. Sensors, Algorithms and Applications: Dagstuhl Seminar 2012 and GCPR Workshop on Imaging New Modalities*. Berlin, Heidelberg: Springer.

Hall, Edward Twitchell. 1966. *The Hidden Dimension*. Vol. 609. Garden City, NY: Doubleday.

Harvey, David. 2014 "Spacetime and the World (2005)." Pp. 46-50 in *The People, Place, and Space Reader*, edited by Jen Jack Gieseking and William Mangold. Routledge.

Harvey, Peter Francis. 2022. "'Make Sure You Look Someone in the Eye': Socialization and Classed Comportment in Two Elementary Schools." *American Journal of Sociology* 127(5):1417-59.

Hatfield, Elaine, Lisamarie Bensman, Paul D. Thornton, and Richard L. Rapson. 2014. "New Perspectives on Emotional Contagion: a Review of Classic and Recent Research on Facial Mimicry and Contagion." *Interpersona: An International Journal on Personal Relationships* 8(2):159-79. doi: 10.5964/ijpr.v8i2.162.

Hendricks, Michael and Richard Bootzin. 1976. "Race and Sex as Stimuli for Negative Affect and Physical Avoidance." *The Journal of Social Psychology* 98(1):111-20.

Hertzog, Esther and Assaf Lev. 2019. "Male Dominance Under Threat: Machoism Confronts Female Defiance in Israeli Gyms." *Journal of Contemporary Ethnography* 48(6):836-66. doi: 10.1177/0891241619834662.

Hoeben, Evelien M., Wim Bernasco, Lasse Suonperä Liebst, Carlijn van Baak, and Marie Rosenkrantz Lindegaard. 2021. ""Social Distancing Compliance: a Video Observational Analysis" Edited by H. Seale." *PLOS ONE* 16(3):e0248221. doi: 10.1371/journal.pone.0248221.

Horaud, Radu, Miles Hansard, Georgios Evangelidis, and Clément Ménier. 2016. "An Overview of Depth Cameras and Range Scanners Based on Time-of-Flight Technologies." *Machine Vision and Applications* 27(7):1005-20. doi: 10.1007/s00138-016-0784-4.

Hua, Quan. 2019. "[CVPR 2017] OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields." *Towards Data Science*. Retrieved June 28, 2021 (https://towardsdatascience.com/cvpr-2017-openpose-realtime-multi-person-2d-pose-estimation-using-part-affinity-fields-f2ce18d720e8).

Hunziker, Sabina, Anna C. Johansson, Franziska Tschan, Norbert K. Semmer, Laura Rock, Michael D. Howell, and Stephan Marsch. 2011. "Teamwork and Leadership in Cardiopulmonary Resuscitation." *Journal of the American College of Cardiology* 57(24):2381-88. doi: 10.1016/j.jacc.2011.03.017.

Jackson, Joshua Conrad, Jonathan Jong, David Bilkey, Harvey Whitehouse, Stefanie Zollmann, Craig McNaughton, and Jamin Halberstadt. 2018. "Synchrony and Physiological Arousal Increase Cohesion and Cooperation in Large Naturalistic Groups." *Scientific Reports* 8(1):127. doi: 10.1038/s41598-017-18023-4.

Jain, Suyog Dutt and Kristen Grauman. 2013. "Predicting Sufficient Annotation Strength for Interactive Foreground Segmentation." Pp. 1313–20 in *Proceedings of the IEEE International Conference on Computer Vision*.

Johansson, Thomas. 1996. “Gendered Spaces: the Gym Culture and the Construction of Gender.” *Young* 4(3):32-47.

Ke, Qiuhong, Jun Liu, Mohammed Bennamoun, Senjian An, Ferdous Sohel, and Farid Boussaid. 2018 “Computer Vision for Human–Machine Interaction.” Pp. 127-45 in *Computer Vision for Assistive Healthcare*, edited by M. Leo and G. M. Farinella. London: Academic Press.

Khan, Asharul Islam and Salim Al-Habsi. 2020. “Machine Learning in Computer Vision.” *Procedia Computer Science* 167:1444-51. doi: 10.1016/j.procs.2020.03.355.

Khanam, F. T. Z., A. Al-Naji, and J. Chahl. 2019. “Remote Monitoring of Vital Signs in Diverse Non-Clinical and Clinical Scenarios Using Computer Vision Systems: a Review.” *Applied Sciences* 9(20):4474. doi: 10.3390/app9204474.

Klette, Reinhard. 2014. *Concise Computer Vision: An Introduction into Theory and Algorithms*. 2014 edition. London: Springer.

Klusemann, Stefan. 2009. “Atrocities and Confrontational Tension.” *Frontiers in Behavioral Neuroscience* 3, doi: 10.3389/neuro.08.042.2009.

Kohli, Puneet and Anjali Chadha. 2020 “Enabling Pedestrian Safety Using Computer Vision Techniques: A Case Study of the 2018 Uber Inc. Self-Driving Car Crash.” Pp. 261-79 in *Advances in Information and Communication. Vol. 69, Lecture Notes in Networks and Systems*, edited by K. Arai and R. Bhatia. Cham: Springer International Publishing.

LeBaron, Curtis, Paula Jarzabkowski, Michael G. Pratt, and Greg Fetzer. 2018. “An Introduction to Video Methods in Organizational Research.” *Organizational Research Methods* 21(2):239-60. doi: https://doi.org/10.1177/1094428117745649.

Lefebvre, Henri and Donald Nicholson-Smith. 1991. *The Production of Space*. Malden, Mass: Blackwell.

Levine, Mark, Paul J. Taylor, and Rachel Best. 2011. “Third Parties, Violence, and Conflict Resolution: the Role of Group Size and Collective Action in the Microregulation of Violence.” *Psychological Science* 22(3):406-12. doi: 10.1177/0956797611398495.

Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick 2014. “Microsoft Coco: Common Objects in Context.” Pp. 740–55 in *European conference on computer vision*. Springer.

Lindegaard, Marie Rosenkrantz, Lasse Suonperä Liebst, Wim Bernasco, Marie Bruvik Heinskou, Richard Philpot, Mark Levine, and Peter Verbeek. 2017. ““Consolation in the Aftermath of Robberies Resembles Post-Aggression Consolation in Chimpanzees” Edited by A. Siegel.” *PLOS ONE* 12(5): e0177725. doi: 10.1371/journal.pone.0177725.

Lindegaard, Marie Rosenkrantz and Wim Bernasco. 2018. “Lessons Learned from Crime Caught on Camera.” *Journal of Research in Crime and Delinquency* 55(1):155-86. doi: 10.1177/0022427817727830.

Makin, David A., Dale W. Willits, and Rachael Brooks. 2021. “Systematic Social Event Modeling: a Methodology for Analyzing Body-Worn Camera Footage.” *International Journal of Social Research Methodology* 24(2):163-76. doi: 10.1080/13645579.2020.1766775.

McFarland, Daniel A., Dan Jurafsky, and Craig Rawlings. 2013. “Making the Connection: Social Bonding in Courtship Situations.” *American Journal of Sociology* 118(6):1596-649. doi: 10.1086/670240.

Mehan, Hugh. 1979. “““What Time Is It, Denise?”: Asking Known Information Questions in Classroom Discourse.”.” *Theory into Practice* 18(4):285-94.

Mendelberg, Tali, Christopher F. Karpowitz, and J. Baxter Oliphant. 2014. “Gender Inequality in Deliberation: Unpacking the Black Box of Interaction.” *Perspectives on Politics* 12(1):18-44.

Merker, Bjorn H., Guy S. Madison, and Patricia Eckerdal. 2009. “On the Role and Origin of Isochrony in Human Rhythmic Entrainment.” *Cortex* 45(1):4-17. doi: 10.1016/j.cortex.2008.06.011.

Molina, Mario and Filiz Garip. 2019. “Machine Learning for Sociology.” *Annual Review of Sociology* 45(1):27-45. doi: 10.1146/annurev-soc-073117-041106.

Mosselman, Floris, Don Weenink, and Marie Rosenkrantz Lindegaard. 2018. “Weapons, Body Postures, and the Quest for Dominance in Robberies: A Qualitative Analysis of Video Footage.” *Journal of Research in Crime and Delinquency* 55(1):3-26. doi: 10.1177/0022427817706525.

Murcia, Harold, David Sanabria, Dehyro Méndez, and Manuel G. Forero. 2021. “A Comparative Study of 3D Plant Modeling Systems Based on Low-Cost 2D LiDAR and Kinect.” Pp. 272–81 in *Mexican Conference on Pattern Recognition*. Springer.

Nassauer, Anne. 2018. “How Robberies Succeed or Fail: Analyzing Crime Caught on CCTV.” *Journal of Research in Crime and Delinquency* 55(1):125-54. doi: https://doi.org/10.1177/0022427817715754.

Nassauer, Anne. 2019. *Situational Breakdowns: Understanding Protest Violence and Other Surprising Outcomes*. New York: Oxford University Press.

Nassauer, Anne. 2021. “‘Whose Streets? Our Streets!’: Negotiations of Space and Violence in Protests.” *Social Problems* 68(4):852–869.

Nassauer, Anne and Nicolas M. Legewie. 2019. “Analyzing 21st Century Video Data on Situational Dynamics—Issues and Challenges in Video Data Analysis.” *Social Sciences* 8(3). doi: 10.3390/socsci8030100.

Nassauer, Anne and Nicolas M. Legewie. 2021. “Video Data Analysis: A Methodological Frame for a Novel Research Trend.” *Sociological Methods & Research* 50(1):135-73. doi: 10.1177/0049124118769093.

Nassauer, Anne and Nicolas M. Legewie. 2022. *Video Data Analysis: How to Use 21st Century Video in the Social Sciences*. London: Sage.

Neumann, Markus, Erika Fowler, and Travis Ridout. 2022. “Body Language and Gender Stereotypes in Campaign Video.” *Computational Communication Research* 4(1):254–274.

Ochs, Elinor and Tamar Kremer-Sadlik, eds. 2013. *Fast-Forward Family*. Berkeley: University of California Press.

Pallotti, Francesca, Sharon Marie Weldon, and Alessandro Lomi. 2020. “Lost in Translation: collecting and Coding Data on Social Relations from Audio-Visual Recordings.” *Social Networks* 69:102–112.

Pauwels, Luc. 2010. “Visual Sociology Reframed: an Analytical Synthesis and Discussion of Visual Methods in Social and Cultural Research.” *Sociological Methods & Research* 38(4):545-81. doi: 10.1177/0049124110366233.

Philpot, Richard and Mark Levine. 2021. “Evacuation Behavior in a Subway Train Emergency: A Video-Based Analysis.” *Environment and Behavior* 54(2):383–411.

Raaj, Yaadhav, Haroon Idrees, Gines Hidalgo, and Yaser Sheikh. 2019. “Efficient Online Multi-Person 2d Pose Tracking with Recurrent Spatio-Temporal Affinity Fields.” Pp. 4620–28 in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Small, Mario Luis. 2009. *Unanticipated Gains: Origins of Network Inequality in Everyday Life*. Oxford: Oxford University Press.

Sobel, Robert S. and Nancy Lillith. 1975. “Determinants of Nonstationary Personal Space Invasion.” *The Journal of Social Psychology* 97(1):39-45.

Stickle, Ben, Melody Hicks, Amy Stickle, and Zachary Hutchinson. 2020. “Porch Pirates: Examining Unattended Package Theft Through Crime Script Analysis.” *Criminal Justice Studies* 33(2):79-95. doi: 10.1080/1478601X.2019.1709780.

Su, Hao, Jia Deng, and Li Fei-Fei. 2012. “Crowdsourcing Annotations for Visual Object Detection.” in *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence*.

Sünderhauf, Niko, Oliver Brock, Walter Scheirer, Raia Hadsell, Dieter Fox, Jürgen Leitner, Ben Upcroft, Pieter Abbeel, Wolfram Burgard, Michael Milford, and Peter Corke. 2018. “The Limits and Potentials of Deep Learning for Robotics.” *The International Journal of Robotics Research* 37(4–5):405-20. doi: 10.1177/0278364918770733.

Tavory, Iddo and Nina Eliasoph. 2013. “Coordinating Futures: Toward a Theory of Anticipation.” *American Journal of Sociology* 118(4):908-42.

Thorne, Barrie. 1993. *Gender Play: Girls and Boys in School*. New Brunswick, NJ: Rutgers University Press.

Tian, Hongkun, Tianhai Wang, Yadong Liu, Xi Qiao, and Yanzhou Li. 2020. “Computer Vision Technology in Agricultural Automation —A Review.” *Information Processing in Agriculture* 7(1):1-19. doi: 10.1016/j.inpa.2019.09.006.

Torres, Michelle and Francisco Cantú. 2021. "Learning to See: Convolutional Neural Networks for the Analysis of Social Science Data." *Political Analysis*: 30(1):113-131.

Trost, W. Johanna, Carolina Labbé, and D. Grandjean. 2017. "Rhythmic Entrainment as a Musical Affect Induction Mechanism." *Neuropsychologia* 96:96-110. doi: 10.1016/j.neuropsychologia.2017.01.004.

Voulodimos, Athanasios, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. 2018. "Deep Learning for Computer Vision: A Brief Review." *Computational Intelligence and Neuroscience* 2018:1-13. doi: 10.1155/2018/7068349.

Wagner, Daniel. 2018. "Depth Cameras for Mobile AR: From IPhones to Wearables and Beyond." *Medium.* Retrieved June 21, 2021 (https://medium.com/@DAQRI/depth-cameras-for-mobile-ar-from-iphones-to-wearables-and-beyond-ea29758ec280).

Waller, Mary J. and Seth A. Kaplan. 2018. "Systematic Behavioral Observation for Emergent Team Phenomena: Key Considerations for Quantitative Video-Based Approaches." *Organizational Research Methods* 21(2):500-15. doi: 10.1177/1094428116647785.

Wang, Angelina, Arvind Narayanan, and Olga Russakovsky. 2020. "REVISE: A Tool for Measuring and Mitigating Bias in Visual Datasets." *ArXiv:2004.07999 [Cs].*

Wang, Tianlu, Jieyu Zhao, Mark Yatskar, Kai-Wei Chang, and Vicente Ordonez. 2019. "Balanced Datasets Are Not Enough: Estimating and Mitigating Gender Bias in Deep Image Representations." Pp. 5310–19 in *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

Webb Williams, Nora, Andreu Casas, and John D. Wilkerson. 2020. *Images as Data for Social Science Research: An Introduction to Convolutional Neural Nets for Image Classification*. Cambridge: Cambridge University Press.

West, Candace and Don H. Zimmerman. 1987. "Doing Gender." *Gender & Society* 1(2):125-51.

Willis, Frank N. 1966. "Initial Speaking Distance as a Function of the Speakers' Relationship." *Psychonomic Science* 5(6):221-22.

Willits, Dale W. and David A. Makin. 2018. "Show Me What Happened: Analyzing Use of Force Through Analysis of Body-Worn Camera Footage." *Journal of Research in Crime and Delinquency* 55(1):51-77. doi: 10.1177/0022427817701257.

Włodarczyk, Aleksandra, Elsner Claudia, Schmitterer Alexandra, and E. Wertz Annie. 2018. "Every rose has its thorn: Infants' responses to pointed shapes in naturalistic contexts." *Evolution and Human Behavior* 39(6):583-593.

Wu, Carole-Jean, David Brooks, Kevin Chen, Douglas Chen, Sy Choudhury, Marat Dukhan, Kim Hazelwood, Eldad Isaac, Yangqing Jia, Bill Jia, Tommer Leyvand, Hao Lu, Yang Lu, Lin Qiao, Brandon Reagen, Joe Spisak, Fei Sun, Andrew Tulloch, Peter Vajda, Xiaodong Wang, Yanghan Wang, Bram Wasti, Yiming Wu, Ran Xian, Sungjoo Yoo, and Peizhao Zhang. 2019 "Machine Learning at Facebook: Understanding Inference at the Edge." Pp. 331-44 in *2019 IEEE International Symposium on High Performance Computer Architecture (HPCA)*. Washington, DC, USA: IEEE.

Zelizer, Viviana A. Rotman. 2005. *The Purchase of Intimacy*. Princeton, N.J.: Princeton Univ. Press.

Zhang, Han and Yilang Peng. 2021. "Image Clustering: An Unsupervised Approach to Categorize Visual Data in Social Science Research. *"Sociological Methods & Research* (online first). https://doi.org/10.1177/00491241221082603

Author Biographies

Yoav Goldstein is a deep learning software engineer and researcher who develops applications in the fields of computer vision, autonomous driving, virtual and augmented reality, mobile, deep learning, and gaming. He holds a BSc degree in electrical and computer engineering from the Hebrew University of Jerusalem and MSc degree in computer science from the Open University, researching implementations of computer vision and machine learning in medical fields.

Nicolas M. Legewie is a postdoctoral research fellow at the faculty of economics, law, and social sciences, University of Erfurt. He teaches and writes about social inequality and mobility, social networks, and digitalization, as well as research design, digital social science research, research ethics, and video data analysis. Recent publications include "Grass Roots Efforts For Integrating Refugees – Evidence from a Randomized Field Experiment" (*Journal of Ethnic and Migration Studies*) and "Video Data Analysis: How to use 21st century video in the social sciences" (*SAGE*).

Doron Shiffer-Sebba is a postdoctoral fellow at the Institute for Policy Research, Northwestern University. His research focuses on inequality, stratification, economic sociology, and urban sociology, using qualitative, quantitative, and computational methods. His current work investigates the role of families in economic inequality, and particularly how extended family networks organize to transmit advantages across generations.