

Human Behavior: Real and Digital

From “Digital Ethology: Human Behavior in Geospatial Context,”
edited by Tomáš Paus and Hye-Chung Kum. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 978026254813

Leveraging Video Footage for Ethological Observation of Human Behavior

Virginia Pallante, Lasse Suonperä Liebst, Peter Ejbye-Ernst,
Camilla Bank Friis, and Marie Rosenkrantz Lindegaard

Abstract

Originating in biology, the ethological approach to studying human behavior has increasingly spread across various disciplines, including the social sciences. In addition to offering biologically proximate and evolutionary explanations, ethology provides a methodological framework for systematically observing and analyzing human behavior in natural face-to-face settings. This chapter discusses the relevance of using the ethological approach for the study of human behavior, particularly by leveraging video recordings of public behavior for ethological observation. This prospect is demonstrated through an outline of recent video-observational research on violent and bystander helping behaviors. Further avenues are discussed to advance video-based human ethology.

A Video-Based Human Ethology

The use of digital data has the potential to reshape how social science is fundamentally conducted, as the digital footprint left on digital and social media platforms provides unique insight into human behavior (Blok and Pedersen 2014; Zhang et al. 2020). Digital data takes many forms and shapes. In this chapter, we argue that the use of visual digital data, especially video recordings of behavior in public places, offers a unique but as yet underutilized potential to examine human behavior. While people's online presence and digital footprint bear witness to many aspects of human social life, a great deal of human behavior remains nondigital in nature and leaves none or only a shallow digital footprint and may thus be better observed *in situ* (Molotch and Boden 1993). Here, video data offers great potential as it opens a window into the nondigital

social world, where people behave bodily, are co-present, and often interact face to face (Nassauer and Legewie 2022).

By capturing our daily routines and rare encounters, public cameras provide a versatile tool for conducting detailed and unobtrusive field observational research on human behavior. Within social science, however, systematic naturalistic observation—whether video-based or conducted on-site—has been surprisingly underutilized compared to self-reports in qualitative interviews and surveys (Reiss Jr 1992). This indirect approach to studying human behavior allows understanding of people’s motivations for their actions, but it provides only a coarse-grained picture of how people actually behave. As such, for decades there has been a call for wider use of naturalistic observation techniques within the social sciences (Baumeister et al. 2007). Reflecting this, Erving Goffman (1971), an extremely influential and early pioneer of the study of interpersonal behavior, suggested that the subfield of micro-sociology should be practiced as “interaction ethology,” a kind of human ethology with a particular focus on the *interactional* aspect of social behavior. His realization was that ethologists had developed the most detailed methodological skill set and procedures for systematically observing human interaction *in situ*, and that this should be taken as a methodological model for how micro-sociology should be conducted. Recently, Goffman’s vision has begun to show its methodological potential in the social sciences. While Goffman needed to rely on on-site observations of human behavior decades ago, high-quality video recordings (captured, e.g., by surveillance cameras and smartphones) are now available to scholars (Gerrard and Thompson 2011). When such video data is utilized for ethological observation, it opens groundbreaking possibilities for the study of human behavior (Nassauer and Legewie 2022; Philpot et al. 2019).

First, the sampling of human behavior through video recordings may dramatically increase the sample size of rare events, which an on-site observer may never (e.g., terrorist attacks) or only rarely witness (e.g., street fights) (Lindgaard 2022). Second, observation and quantification of human behavior through video recordings have higher reliability and precision than on-site observations. This is because the event and subsequent behavior can (and often needs to) be observed many times, in slow motion, to cross-validate records between observers. As such, evidence suggests that the dynamic or interactional part of social encounters cannot be reliably captured with on-site observations (Morrison et al. 2016). For example, while ethnographic participant observation has excellent ecological validity, its reliability in capturing micro-interactional details is low.

Finally, video recordings are a highly unobtrusive data source. In many countries, recording devices, such as surveillance systems installed by police and municipalities in public settings, are an accepted part of the natural environment. What the videos reveal, therefore, is people’s unstructured behavior, unaffected by the observer. The unobtrusiveness of video data further has the benefit that even dangerous human behavior may be observed without

exposing the observer to direct threat, as could happen in on-site observations (Lindegaard et al. 2020).

In sum, there is little doubt that if Goffman had lived to experience the digital era, he would have embraced video data, just as several of his students did (e.g., Collins 2008). This current chapter follows in Goffman's footsteps and highlights the value of a video-based human ethology to study face-to-face interaction. With this ambition, it must be acknowledged that other attempts have been made to develop video-based approaches for the micro-sociological study of social interaction, such as the Video Data Analysis approach (Nassauer and Legewie 2018). This latter approach does not, however, take its point of departure in ethology and therefore has no separate interest in biological considerations of proximate and evolutionary explanations or cross-species comparisons. In addition, the current approach puts a stronger emphasis on intercoder reliability tests and has an explicit ambition to test the generalizability of behavioral hypotheses and, as such, in quantitative and large-*N* applications.

The exciting prospect of video-based human ethology is that many of the assumptions regarding human interpersonal behavior embedded within the social sciences can be checked against systematic observational evidence (Mortensen and Cialdini 2010). Often, such reality testing is not a priority within the social sciences, reflecting a weak interest in replicable testing (Makel and Plucker 2014) and the fact that the available methods offer low reliability and validity for examining human behavior. This has put the social sciences in a puzzling position where, as sociologist John Levi Martin (2017:118) summarizes, "probably more is known about interactions between chimpanzees than interactions between humans." We believe that a video-based human ethology is one way to address these issues.

Video Observation as a Method

Video-based human ethology can be applied to the study of various human behaviors, and here we will focus on one area where its methodological value has been clearly demonstrated: the study of interpersonal violence. Traditionally, this field of study relied almost exclusively on self-reported data and laboratory experiments—despite the limitations of these methods for examining violent behavior (for a review, see Philpot et al. 2019): Self-reports of violence and other crimes are subject to social desirability and recollection bias, likely exacerbated by the distress of these events. Laboratory experiments are limited by the practical and ethical circumstances that actual violence cannot be realistically simulated.

The growing availability of video data offers a way to overcome this methodological impasse in studying actual, unstructured violence. In analyzing these data, we largely follow a procedure developed and applied within human ethology (Eibl-Eibesfeldt 1989; Jones et al. 2018). This involves a strong

emphasis on inductive observation of the behavior under study to inform the construction of an ethogram or behavioral inventory with detailed behavioral definitions. In ethology, ethograms are taken as the point of departure to study the behavioral repertoire of a species (Lehner 1998). The development of an ethogram is the product of nonsystematic *ad libitum* observations to select which behaviors to include, particularly those that are more discernible, delimited, and repeated over time (Altmann 1974). These behaviors can range from social interactions to individual activities, postures, or movements. This phase includes testing and revising the inter-reliability of the ethogram by comparing the ratings of two or more independent observers. Once high agreement is reached, the ethograms are applied systematically to observe and code the behavior of interest.

Ethograms are often refined and validated in subsequent studies; the aim is to develop a standardized ethogram of a given category of behavior. To illustrate, consider bystander behavior at violent public events. Initially developed by Levine et al. (2011), the ethogram of bystander behavior has been applied and validated in a number of studies (Ejbye-Ernst 2022; Liebst et al. 2019; Philpot et al. 2020). The resulting standardized ethogram includes bystander behaviors such as “pacifying gesturing,” “calming touches,” “blocking contact,” “holding, pushing, or pulling an aggressor away from the conflict,” and “consoling a victim of aggression.” An example of an ethogram for nonviolence includes face-touching behaviors (Liebst et al. 2022) and was developed during the COVID-19 pandemic to examine the potentially adverse (self-inoculation) effects of mask-wearing. This ethogram describes fine-grained distinctions between whether a person touched a mucosa area (e.g., the T-zone of eyes, nose, and mouth), which is the main entry point for viral infection.

Although video records allow for the application of the ethological method to humans, the recorded social contexts can differ compared to animal studies. Ethological research frequently focuses on closed animal communities, where repeated interactions among the same individuals are possible, and where kin and social relationships are known or can be determined through repeated observations of the same subjects. By contrast, public security cameras record public spaces where people are present for only a limited amount of time, and typically no repeated observations of the same person are possible. Even though video records may thus fail to document some of the dynamics that occur between affiliated individuals during recurrent interactions, they provide a realistic insight into what is at the core of urbanized human ecology: a social structure organized between interacting strangers (Christakis 2019).

Bystander Helping in the Wild

In our violence research, we have specifically utilized video observation to examine the role of bystanders in violent incidents. For decades, the leading

theory of bystander behavior within the social sciences has been the so-called “bystander effect” hypothesis (Darley and Latane 1968). This theory posits that people lose their moral compass when present in crowds, and thus remain passive and apathetic when witnessing someone in need of help. In other words, in crowds, the responsibility for taking action is diluted among those present, which, in turn, inhibits the helping likelihood. The bystander effect hypothesis was initially developed to explain the case of Kitty Genovese, who was raped and murdered in public in New York in 1964, while 38 bystanders allegedly remained passive. To study bystander passivity, field experiments were conducted: researchers staged emergencies in public places and then documented how the likelihood of intervention decreased when co-present with additional bystanders.

Despite being initiated by real-life violent events, research soon became uncoupled from the reality it set out to explain, due to the experimental approach used. The staged emergencies were often very trivial in nature (e.g., people dropping coins in an elevator), thus questioning the generalizability of these results to actual violent events, such as the Kitty Genovese case. Stressing this concern, a meta-analysis showed that the bystander effect was attenuated in experiments that simulated emergencies with some level of danger, albeit none simulating direct violence (Fischer et al. 2011). For the most dangerous situations included, the analysis indicated that additional bystanders offered welcome support, making the intervention more—not less—likely. The problem remained, however, that no meta-analysis is better than the studies included. Without analyzing any violent studies, it could not provide ecologically valid insights into how bystanders act in actual violent events.

The field of bystander studies encapsulates the concern of Tinbergen (1963:411) that researchers “skipped the preliminary descriptive stage that other natural sciences had gone through, and so was soon losing touch with the natural phenomena.” A reality check is needed, based on detailed naturalistic observations of real-life bystander behavior (Lindegard 2022). The first video-based study of this kind was conducted by Levine et al. (2011) who, in direct contradiction to the bystander effect narrative, showed that bystanders play an active and effective role in regulating violent events. Building on this insight, Philpot et al. (2020) conducted a video-based study to investigate whether bystanders intervened in 219 street violence assaults captured on video in the Netherlands, the United Kingdom, and South Africa. They found that in nine out of ten situations, at least one, and typically four, bystanders did something to help the victim (Philpot et al. 2020). Furthermore, it was found that the likelihood of victims receiving help increased with the number of bystanders present. In other words, intervention is the norm, and there is safety in numbers. This is the reality of real-life bystander behavior outside the artificiality of the experimental setting. Characteristically, this was also the case in the Kitty Genovese case: historical analysis has documented that bystanders actually tried to intervene, although unsuccessfully (Manning et al. 2007).

These insights represent only the beginning of an ecologically valid and detailed understanding of bystander behavior. With the use of video-based ethological methods, a range of additional insights into bystander behavior has been revealed:

1. Regarding the causes of individual bystander intervention, the video data offer mixed results concerning the number of bystanders present, in contrast to the uniformly robust result that social relationship ties between bystanders and conflict victims dramatically increase the likelihood of individual intervention. Friends help friends (Liebst et al. 2019; Lindegaard et al. 2017). This conclusion is consistent with social psychological and evolutionary theory, which stresses that individuals have stronger empathic feelings toward in-group members with whom they have interdependent social ties (de Waal and Preston 2017; Stürmer et al. 2006).
2. Further, the level of violent danger is a very influential predictor of intervention. This suggests that people act when it really matters, especially when events become explicitly aggressive and dangerous (Lindegaard et al. 2021).
3. Bystander intervention is not a single act, as often portrayed in experimental settings. Instead, it is an intervention trajectory involving various actions that follow a specific behavioral pattern. Bystanders who intervene tune into the aggression level of the conflict, and stopping the fight requires consistent insistence and preparedness to scale up the intervention intensity (Ejbye-Ernst et al. 2021).
4. Relatedly, bystander intervention is not merely performed by an individual but is typically carried out in collaboration with others. This is because the violent conflict may require the actions of several individuals acting in concert to be stopped (Bloch et al. 2018; Levine et al. 2011; Weenink et al. 2022).
5. Bystander interventions may take place during all phases of the conflict, including in its aftermath where bystanders may provide consolation to victims of aggression (Bloch et al. 2018; Lindegaard et al. 2017). This behavior is similar to what has been documented among human children (Verbeek 2008) and nonhuman primates (de Waal and van Roosmalen 1979).
6. Given the high bystander intervention rate, there might be a concern that intervening bystanders may be victimized themselves when helping others. In general, however, the likelihood of bystander victimization is low (around 5%), and if victimization occurs, it is often relatively non-severe (Liebst et al. 2020).
7. Bystander intervention is actually effective in terminating violence, especially when performed as forceful interventions rather than as mere expressions of disapproval (Ejbye-Ernst 2022).

Prospects and Challenges

Video-based human ethology holds great promise, but the journey ahead is replete with a plethora of possibilities and unaddressed issues. Numerous questions emerge from the fact that although humans are a great ape and should be studied as such (Turner and Maryanski 2018), we are also different—not in kind but in degree (Darwin 1871)—from other animals and primates. This difference has important methodological and theoretical implications. *Homo sapiens*, like every other species, has unique characteristics that must be considered. Critical specificities for humans include evolved cognitive skills, which enable advanced capacities for collaboration, symbolic communication, and cultural learning (Tomasello and Herrmann 2010; cf. Bard et al. 2021). Specialized methods and theories have been developed to grasp these human social qualities, which cannot be fully captured through ethological observation of nonverbal behavioral displays (Geertz 1973).

The limitations of human ethological observation may be further magnified by the technology of public security cameras, which typically do not capture sound and thus do not permit content analysis of verbal communication. Considering that ethology is the study of behavior, and that humans often use speech when they interact (Austin 1975), video data is not optimal for examining verbal human behavior. With respect to the study of violence, this is crucial because during the initial phase of conflicts or in low-intensity disputes, verbal exchanges often unfold prior to the use of physical force (Friis et al. 2020). Thus, the inherently dispute-related nature of many violent crimes, involving mutual verbal insults and retaliations (Felson 1982), cannot be fully grasped with public security cameras.

One way to overcome this limitation is to analyze how verbal behavior is often expressed in conjunction with nonverbal cues (Eibl-Eibesfeldt 1989) and to use this to make some rough inferences from observations of nonverbal behavior to their verbal counterparts. Alternatively, scholars are using video data recorded by devices that capture sound, such as mobile phones and body cameras (Friis et al. 2020; Sytsma et al. 2021). Finally, the lack of sound may be compensated for by triangulating with other verbal data sources (e.g., interview data in combination with video-observational data). This could provide insights into the cultural, motivational, and meaningful content of social life, which aids in understanding why people do what they do (Friis 2022; Small and Cook 2021).

Furthermore, video data combined with additional information on the locations of public security cameras may offer a fruitful basis for explaining the behavioral data captured on camera. For instance, Sampson and Raudenbush (2004) combined systematic observations with census data, police records, and surveys to examine whether racial stigma, the economic context, and the actual observation of social disorder shape how people perceive social disorder. Similarly, in the analysis of criminal events, participant observation and

interviews with people in specific locations may help to contextualize and explain observed interactions. These data sources provide information on the characteristics of the neighborhood as well as the people living in the area, including the subjective motivations that underpin the observed behaviors (Lindegaard and Bernasco 2018).

When adopting ethological insights to study human behavior, it is necessary to consider the extent to which its methodological aspects are distinct from its theoretical aspects. For Goffman, inspiration from ethology should be methodological, not theoretical. In his view, the ethological application of a “Darwinian frame” leads to “very unsophisticated statements,” but “if we politely disattend this feature of ethology, its value for us as a model stands clear.” (Goffman 1971:xvii). While Goffman seems to be referring to a reductionist evolutionary perspective that was prevalent at the time, contemporary ethology and evolutionary theories today are interested in questions central to sociological reasoning (Meloni 2014): prosociality, empathy, and how social relationships create group structures and influence conflict management strategies (de Waal 2000; de Waal and Preston 2017). As such, recent micro-sociology, inspired by Goffman, is engaging in fruitful dialogues with evolutionary and biological schools of thought, which support rather than erode the importance of sociological mechanisms (Heinskou and Liebst 2016; Lindegaard et al. 2017; Turner and Maryanski 2018).

A precondition for this type of interdisciplinary exchange is cross-species comparisons of behaviors between humans and nonhuman primates (Turner and Maryanski 2018), and in this area, we lack human *adult* ethological data. The limited evidence available is biased toward human children (Verbeek 2008), leading to the constrained conclusion that “other primates are *mentally* like human children” (de Waal 1989:249). Using a video-based human ethological approach, Lindegaard et al. (2017) conducted the first study on human adult post-conflict consolation behavior, comparing the observed patterns with those of chimpanzees (Lindegaard et al. 2017). We strongly recommend that future research examine other human adult behaviors with a view toward cross-species comparisons.

Video-based interaction ethology offers a way to compare human behavior in different conflict phases, conflict types, and cultural contexts. Behavioral variations are found in different steps of conflicts—for example, affiliative touching is more frequent in the aftermath than before or during robberies (Lindegaard et al. 2017; Philpot et al. 2022)—and intervention behavior is more physical at the end of the conflict than at the beginning (Ejbye-Ernst et al. 2021). Further, cultural comparisons allow us to theorize about the mechanisms of the observed behavior and may help us understand the extent to which human nature is universal (Brown 1991). Cross-cultural comparisons between South Africa, the Netherlands, and the United Kingdom, for instance, reveal similar bystander intervention frequencies in street violence events (Philpot et al. 2020).

Bystander intervention in street fights might be, however, conflict-type specific and thus not generalizable to other kinds of conflict (e.g., robberies, partner violence, war atrocities). For example, in an analysis of bystander intervention during armed robberies in the Netherlands, we found that bystanders only intervened in a minority of robberies (unpublished data), in contrast to street violence (Philpot et al. 2020). Further, when bystanders do intervene in robberies, the risk of victimization is much higher during armed robberies than in street fights (Liebst et al. 2020). Such examples highlight variations in bystander behavior across different types of conflict, potentially related to differences in conflict dynamics, cultural settings, and causal mechanisms, and underscore the need for further research in this area.

The primary strength of the ethological approach is its focus on detailed and naturalistic description (Lorenz 1973) yet integrating this with a focus on explaining causal mechanisms remains challenging. For Goffman, this was less of a concern in his vision of interaction ethology, as he deliberately refrained from moving beyond description to test causal hypotheses (Verhoeven 1993). Recent Goffman-inspired research using video data, however, argues that such a step should and could be taken, given the strength of video observation is how it allows one to “study if there is causality at the microlevel” (Nassauer and Legewie 2018:163). Considering the studies mentioned above, the issue is that most rely on cross-sectional, between-subject (or between-situation) designs—a weak approach for testing causality. One solution could be to employ field-experimental methods more extensively, as it is commonly done in ethology (Cuthill 1991), despite the obvious ethical limitations with respect to how violence or danger may be simulated in field experiments.

An alternative to testing causality in aggression and bystander behaviors is to match subjects with themselves under different study conditions, a powerful method for controlling both observed and unobserved confounders (Dawkins 2007). Typically, this involves observing the same subjects under different situations, which is often not feasible with public video data. Recently, however, a few studies have shown that subjects can be measured several times *within* unfolding situations, allowing for a fixed-effect panel regression approach that is considered a robust approximation to causality (Listl et al. 2016). This was done, for example, in the cited video-based study that examined danger levels as a predictor of intervention (Lindegaard et al. 2021): By following the same individuals throughout the unfolding situation, we established which level of danger caused the bystanders to intervene. The success of video-based human ethology hinges on how its descriptive and causal-explanatory potentials are united. This should be a priority in future work.

While one of the strengths of video-based human ethology is its high ecological validity and reliability, a potential weakness is its generalizability, often due to working with nonrepresentative samples of low statistical power (Taborsky 2010). Manually coding behavior second by second is very labor-intensive. To increase sample sizes, computer programs could be used for

automatic behavior annotation, creating observational datasets much larger than would otherwise be possible; this approach may also be considered digital ethology (Anderson and Perona 2014). Specifically, computer vision scientists have worked for decades on training algorithms to detect automatically different kinds of behavior in video clips (Jain et al. 2015). Instead of relying on costly and potentially biased human observers, such as municipal employees or law enforcement agents, computer vision tools could identify relevant study situations from large pools of video clips. Although using computer vision tools to identify conflict situations might yield numerous false positives, filtering out these erroneous clips would still significantly reduce the time required for sampling relevant situations compared to human observers. An example of integrating computer vision tools in video-based interaction ethology is our research project investigating social distancing behavior during the COVID-19 pandemic. We were able to measure automatically when people failed to keep the recommended distance from each other on the street. Using computer vision enabled us to analyze the behavior of over half a million individuals across thousands of hours of footage (Bernasco et al. 2022), a task that would have been impossible with human coders.

Closing Remarks

While video-based human ethology shows significant potential for future research on human behavior, its primary development has been within the domain of interpersonal violence. In our view, many other fields could benefit from utilizing video observations. Broadly speaking, the use of this approach may be one means to make the social sciences a more high-consensus and rapid-discovery science, similar to what has been achieved within the natural and medical sciences. Compared to these disciplines, a limitation of the social sciences is that they are to a lesser degree propelled forward by innovations in research hardware and technologies (Collins 1994). For example, Galileo's brilliance was not only his novel ideas but how he made use of research hardware and technologies: lenses assembled into microscopes and telescopes that led to a series of groundbreaking discoveries. The social sciences have embraced such research hardware to a lesser extent, but this is likely to change with the advent of a more hardware-driven and computational science that harnesses the potential of digital data, simulations, and artificial intelligence (Sallach 2003). For the micro-sociological study of interpersonal behavior, video technology is specifically suggested to hold potential for scientific advancement, given its possibility to map the micro-world of human behavior (Collins 1994).

While this application remains to be fully embraced within academia, the groundbreaking potential of video data has already proven its worth outside academia. In a certain sense, living in contemporary society implies being a

video-trained human ethologist, given our massive exposure to video-recorded content. Video data allow us to see behavioral realities that cannot and should not be unseen, whether as scholars or citizens (Goold 2006). Poignant examples include the murder of George Floyd in 2020, where security and witness footage drove the global outrage over the atrocity we all observed, or the video documentation of war crimes in Syria and Ukraine. Video technology, ever-present in contemporary society, is already revolutionizing our perception of the world.

Acknowledgments

This study was funded by the Dutch Research Council (NWO VI.Vidi.195.083) by means of a grant awarded to Marie Rosenkrantz Lindegaard. This project has also received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant (101029161) awarded to Virginia Pallante.

