2

How Can Concepts of Ethology Be Applied to Large-Scale Digital Data?

Guillaume Dumas, Sophia Frangou, Heidi Keller, Daniel P. Lupp, Virginia Pallante, Tomáš Paus, and Kim A. Bard

Abstract

The ethological approach is used to study naturally occurring behavior. In the modern world, many such behaviors are connected to, and recorded by, a wide array of digital services (e.g., social networking, information search, closed-circuit television). How can ethological concepts be applied to help us characterize the environment in which humans live? What aspects of the ethological approach can guide us to obtain measures captured directly from digital data generated by our everyday activities? What kinds of models do we need to understand how human behaviors/activities can be inferred from the physical and built environment? This chapter explores the bidirectional nature of these relationships; namely, how individuals create their environment, and how the environment shapes the individual. It discusses how to proceed from observation and data sampling to knowledge extraction and causal inference. The complementary nature of common and specific are addressed as well as the challenge of integrating niches at both physical and social levels. Finally, all these concepts and associated methods are illustrated through a hypothetical study.

Reflecting on Observation

> What we observe is not nature itself, but nature exposed to our method of questioning. —Werner Heisenberg (1958)

The early ethologists Karl von Frisch, Konrad Lorenz, and Nikolaas Tinbergen, relied on observations as their core method of inquiry. After careful observation, they would fully describe the behaviors of interest. Then, as a second step, they would contemplate the function of these behaviors, assessed through a process of classification of possible behaviors, field experiments, and comparison of behaviors within and across contexts or species. Tinbergen (1963) argued that behavior could be explained on four levels: ontogenetic, phylogenetic, proximate (immediate cause), and ultimate (evolutionary reasons). By systematizing their observations and focusing on the more fixed behaviors, they could easily replicate their findings. While watching, they wondered about the parameters of the behavior and conducted field experiments from which they could determine proximal causes. For example, one of Tinbergen's experiments involved understanding how digger wasps locate their home burrow after flying away in search of food (Tinbergen 1972). He conducted a series of experiments that involved placing a pinecone at the burrow entrance as the wasps were leaving and then moving it to a nearby location while the wasps were away. Upon their return, the wasps flew to the relocated pinecone rather than to their burrow entrance. In this way, Tinbergen discovered that digger wasps use landmarks to identify their burrow.

Another important study by Tinbergen focused on courtship in stickleback fish (1952), where he identified the specific behaviors of the male to which its prospective mate responded. While this may seem anecdotal, this led to the creation of one of the first ethograms (i.e., a comprehensive list, inventory, or description of the behavior of an organism) (Figure 2.1a), the key tool of ethology. Eibl-Eibesfeldt adopted these same principles of detailed observation coupled with experimental causal inference when beginning the field of human ethology, focusing on the structure of human behavior from recorded observations of people living in diverse settings around the world (Eibl-Eibesfeldt 1989). Since then, the field has grown considerably to encompass a wide range of activities, highlighted by the International Society for Human

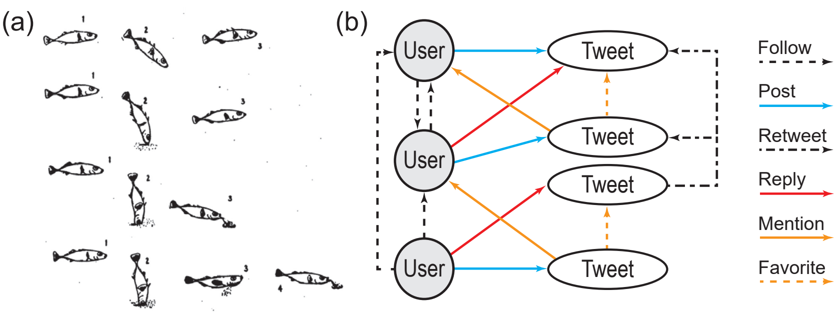

Figure 2.1 Two types of ethograms: (a) A traditional one showing four levels of intensity during the excavation of sand by the three-spined stickleback (*Gasterosteus aculeatus*) (Tinbergen 1951). Intensity ranges from minimal (top) to maximal (bottom); numbers indicate the sequence of behaviors: (1) swimming, (2) digging sand for a nest pit, (3) losing sand through the gills, and (4) spitting out the sand. (b) An example of a digital ethogram based on a user–tweet interaction model (after Belkaroui et al. 2015), showing the six canonical behaviors (arrows) on the social media network Twitter/X.

Ethology. Recently, the interdisciplinary field of computational ethology has emerged at the crossroad of physical computer and life sciences. Here, the idea is to leverage recent progress in machine learning to create ethograms through automatic detection and analysis of behavior while still using individual-level behaviors as the main unit of observation (Anderson and Perona 2014). In this chapter, we focus on digital ethology, which relies on large-scale digital data coupled with geocoding of physical and social environments. In this respect, individual-level behaviors are aggregated at the level of geospatial units, thus guaranteeing better safeguarding of the privacy of individuals while allowing researchers to examine human–environment bidirectional relationships.

Ethologists observe behavior and ponder about its function. So, when Tinbergen watched the digger wasps, he saw them fly in a pattern over the burrow as they emerged, as well as before they reentered the burrow. After extensive observation, he was able to draw the pattern in which they flew. He wondered why they flew like that, under those circumstances. Through field experiments, he determined the proximate cause: the digger wasps encoded features of the landscape that identified their own burrow. Behavior is what is observed, and the construct is either an explanatory or a functional mechanism. The documentation of behaviors allows the ethologist to determine ethograms. Current ethologists develop ethograms of select behaviors to answer specific questions. The constellation of behaviors found to be related to a specific outcome measure might therefore constitute a construct, such as environmental variables related to an increased risk for depression. Interestingly, ethology can also gather information from the constraints of those behaviors (see Figure 2.1a).

Digital ethology poses unique challenges that must be managed if we are to generate an ethogram. Figure 2.1b provides an illustration of how a digital ethogram could be generated using the "tweets" with "influence" as the construct of interest. Like a typical ethogram, Figure 2.1b shows the observed activity of Twitter/X users and the behavioral patterns that arise from their interactions. In this example, constraints are built into the Twitter/X platform; similar to the epigenetic landscape of Waddington (1957), constraints may also be embodied in the physical environment through space, resources, and risks. Pallante et al. (this volume) demonstrate how such constraints influence the probability of engaging in a specific behavior, such as resolution or reconciliation. Digital ethology could thus gather information about area-level constraints on certain behaviors or activities coming from other domains, such as the built environment (e.g., accessing "resources" such as food when stores are not nearby). A digital ethogram would contain selected digital data thought to represent area-level aggregates of behaviors, naturally occurring variations in these behaviors, and assumptions about their functional significance and underlying mechanisms.

The creation of an ethogram implies a process of reduction and simplification aimed at controlling and describing the observations. The boundaries that delimit a behavioral pattern underpin the quantitative approach in ethological studies: the ethogram is meant to be the coding scheme that quantifies the observations, which necessarily leads to a reduction of the variability observed. For the early ethologists, ethograms refer to the complete set of behaviors. In typical modern-day studies, especially in humans, it is nearly impossible to construct an ethogram of all behaviors. Thus, ethograms must necessarily be comprised of select behaviors of interest. Notwithstanding, ethograms are descriptions of (parts of) the observed behavioral repertoire, aimed at capturing and explaining the variability of this repertoire in time and space.

Early in the development of an ethogram, *ad libitum* observations of behaviors are required (Altmann 1974). This means that an inductive approach must be taken, usually by nonsystematically recording the behavioral patterns observed in a group of animals. This helps the researcher to become familiar with and gain insights into the behavioral repertoire of the species. The *ad libitum* nature of such a sampling technique relies on the ability to collect as many observations as possible when behaviors, individuals, and time sessions are chosen without restrictions. The behaviors classified into the ethogram are those that can be clearly described, follow a specific pattern, are limited and repeated over time, and are usually performed by several individuals in the colony. The recording of behaviors is conducted by naming and describing the specific behavioral patterns observed. These notes will turn into items of the ethogram.

Interobserver reliability is a major concern for modern-day ethologists, especially since the behaviors under observation are not usually fixed (i.e., with fixed releasers and fixed forms). Thus, after the development phase of an ethogram and before all observations are coded, a second independent observer is trained to apply the ethogram to evaluate if the definitions of the behaviors included are clear enough to allow for their coding. This may lead to a modification of the original ethogram. Once agreement is reached at an acceptably high level (e.g., Bakeman 2023), the ethogram can be applied for the data collection. Modern-day ethologists use observational methodologies of observing systematically with well-defined ethograms, coding schemes specifying how observations are collected, as well as inter- and intra-observer reliability, etc. (For a review of observational methodology, see Bakeman and Quera 2011.)

Despite technical methods, one may still question how "reliability" differs from "subjectivity," especially when qualitative and quantitative approaches are compared (for an overview of terms, see Appendix 2.1). This discussion is about the quality of assessment procedures, including observations. Reliability and validity are the main characteristics of quality in quantitative approaches (Mays and Pope 2000). While quantitative methodology tries to objectify subjectivity, qualitative methodology tries to represent the subjective meaning systems of the research participants. Because of this subjective component, qualitative methods need alternative validation methods. Just as in quantitative methods, different methods exist, but the rigor of a qualitative study is mostly represented by its trustworthiness, defined by the confidence in the data.

From Data Sampling to Knowledge Extraction

You cannot see things till you know roughly what they are. —C. S. Lewis (1943)

Data collection—the first practical step in the journey—is open to discussions about validity and reliability. Classic ethology relies on prolonged observations of activity in humans or nonhuman species from which the observer can begin to identify distinct patterns (i.e., behavior) to build an ethogram. Even in purely observational studies, the influence of the observer on the activity being observed, and hence the validity of observations, is often a matter of debate. This issue becomes even more complex when one attempts to access internal states (e.g., motives, salience attributions) that can be reported or inferred in human studies but only conjectured in nonhuman species. In all instances, whether research is considered "qualitative" or "quantitative," the observer remains part of the observed in terms of the behaviors selected for observation, the instruments used to record behavior, and the attribution of function and cause. The reliability (i.e., consistent reproducibility) of behavioral measures is also beset with difficulties as simply having a quantitative index of behavior is not sufficient.

The availability of technological platforms (e.g., video cameras, sensors) can minimize the perceived presence of the observer and aid in measurement reliability as they can provide quantitative estimates of behavior that minimize inter- and intra-rater variability. The role of the observer, however, remains integral to the process. For example, Twitter/X users are aware that their behavior is being observed and exploit this to make their behavior accessible to a large number of observers. Similarly, the use of video or CCTV data to study human behavior, such as conflict resolution, requires significant observer input even though the observed are usually not aware that they are being recorded at the time. Often, members of a lab sit together in front of videotapes and discuss jointly what they observe and what the observed behavior could mean. Other methods involve triangulation and respondent validation. In data science, this triangulation weights the validity of a given approach by comparing it to other data, other technical approaches, or both (Oppermann 2000). In epidemiology, a similar triangulation is used (Lawlor et al. 2016), but validity can also be inferred from counterfactual reasoning (Höfler 2005). In both qualitative and quantitative research, it is essential that the research process, including the subjective perception of the researcher, is made transparent and conscious.

Digital ethology needs to adapt those different approaches to data collection and evaluation that address the unprecedented scale of the data and their heterogeneity. Similar to other forms of ethology, the goal of digital ethology is constant: to observe and ponder what is meaningful at different functional levels for the problem under consideration. Whether it is precision medicine, understanding a mechanism, or identifying external stressors, all the previous scientific knowledge will partially constrain the search space of constructs and measures. In this sense, before any data have been collected, some choices are already made, explicitly or even unconsciously. As there are no definitive answers to these issues, awareness and transparency are essential.

The road from data to knowledge passes through information before reaching it and, hopefully, heads off in the direction of wisdom. This road rarely follows a straight line, however, and to avoid getting lost en route, a map is useful. Data science is an informed exploratory process (Huber 1996). Intelligent data exploration requires a clear hierarchy of analysis plans, from the definition of which data structure (e.g., variable, relationship, model) should be considered to the choice of plan (or plan type, in the case of higher-level decisions) that is appropriate. Thereafter, one is faced with the problem of deciding between different plausible but not always consistent results produced by different analytical procedures (e.g., least squares, resistant line, number of data partitions). In optimization, the "units" or "domains" of analysis may vary by the predefined outcome. If a massive dataset is to be analyzed without the direct supervision of a human user, then a representation of the process conducted becomes an even more necessary component of the result. This is one of the key distinctions between planning and other forms of search: the aim is to generate a sequence (or more complex combination) of operations, not simply a result.

It all starts with data and, in the case of digital ethology, with *big data*. Big data is an umbrella term used to describe datasets whose size and structure are so large and complex that conventional computational tasks become unfeasible. The term is commonly associated with vast amounts of data, although it should not be constrained by such a narrow definition. Doug Laney (2001) defines big data with respect to the three Vs:

1. Volume, which refers to the size of the dataset in multiple dimensions (i.e., in the number of records or the number of recorded variables).
2. Velocity or the speed at which data is gathered and processed.
3. Variety, which describes the heterogeneity in the structure of data gathered.

Laney's definition serves as a basis for many alternative definitions that often add additional Vs (e.g., veracity). Still, consensus is lacking on one specific definition. Independently of the definition to which one subscribes, a key point to keep in mind is that big data does not solely refer to size but entails various aspects of complexity within the data and data collection. Thus, for instance, it is entirely possible to have a big dataset with comparatively low volume, but high velocity and variety.

How then do we move from (big) data to knowledge? Data are usually considered as raw measures that have not yet been contextualized. Although the choice of recording one measure rather than another is already contextual, the switch from data to information is usually when those measures are contextualized at the moment of analysis. In this sense, raw data are "dry" and one needs to "rehydrate the data" (Claudia Bauzer Medeiros, pers. comm.) to be able to interpret it. Information emerges by moving from "raw" data to "minimally processed" (e.g., satellite images cleaned from artifacts) or "pre-processed" data (e.g., engineered features extracted from satellite images such as roads, sidewalks, and building types). The move from information to knowledge is then linked to the interpretation of the information and the generation of meaningful claims. In a sense, knowledge generation cannot happen solely based on analyses of information. It needs an outcome; that is, explanatory theories must be generated about a specific phenomenon. Big data might be considered as a catchall ethogram for all possible behaviors, but a more explicit ethogram is needed to answer specific questions and extract laws that govern the associated phenomenon.

With big data come big analyses, and the technological progress in computing has enabled the fast-paced development of artificial intelligence methods such as deep learning. A key issue in machine learning is the generalizability of the results to another context. In this respect, metadata are critical to understanding the link "who was collecting, how, why, and where?" (For further discussion, see Lovasi et al., this volume.) It is important for the sampling to be as representative as possible of the population of interest. Still, it is hard to guarantee the representativeness of the sample regarding the whole population, as the whole population is not a *representation*; it is a *description* (see Medeiros et al., this volume). Knowledge extraction is thus not a monolithic activity; it can come from imposing or discovering structure.

What happens when no structure at all is imposed? Here, bioinformatics offers some clues. Indeed, technological developments in genetics have driven a move from the traditional single-gene approach to a polygenic and even "omnigenic" perspective (Boyle et al. 2017). This new way of viewing the genome emphasizes the interdependence of genes and leverages new digital tools to measure holistic effects at the molecular level. In digital ethology, progress in artificial intelligence may induce a similar shift in ethogram construction, from considering a few discrete canonical behaviors to embracing all observable behavioral patterns. The challenge becomes one of interpretability, since algorithms may detect and use behaviors that are imperceptible to the eye of a human observer. The result could outsmart traditional human ethology in prediction, but at the price of having a clear inference of underlying mechanisms. Thus, beyond prediction, a clear challenge for digital ethology in knowledge generation remains the inference of causality.

Inferring Causality

> You are smarter than your data. Data do not understand causes and effects; humans do. —Judea Pearl and Dana Mackenzie (2018)

Humans and, in particular, scientists often like causal explanations. The concept of causality has been debated for centuries by diverse disciplines that have emphasized different aspects of causality. For instance, sociology, anthropology, and psychology place significant emphasis on the *context* in which causal relationships are examined. Causality thus comes in different flavors and requires different approaches to assess it. In ethology, there are at least four different types: ontogenetic (caused by the development), phylogenetic (caused by the evolution of species), proximate (caused by immediate physiological or environmental factors), and ultimate (associated with goals and function from an evolutionary point of view; Tinbergen 1963). In a broader scientific context, three general classes of causality are usually specified: direct, structural, and logical (Craver 2007). Direct and structural causality require, most of the time, an experiment (an empirical intervention with planned perturbation of the system) or at least a quasi-experiment (the use of events that occur independently from the research planning but could nevertheless be exploited to infer how those events causally impact the system). While direct causality is associated with physical events and mechanisms in time (e.g., an earthquake destroyed a city), structural causality is associated with physical objects and mechanisms in space (e.g., the transportation system constrains the growth of the city). Finally, logical causality is independent of time and space since it relies on abstract propositions, reasoning, and implication.

In the case of big data, when trying to determine a causal link between variables A and B (i.e., "the presence of property A causes the likely presence of property B"), one invariably stumbles across known problems. For instance, when analyzing a given dataset D for causal links, one wishes to determine whether property A is necessary and sufficient for the (likely) presence of property B. In this context, both "necessary" and "sufficient" are needed: without "necessary" we cannot deduce that A is the cause of B, and without "sufficient" we cannot deduce that B is the result of A. Though this is the method to determine causal links, it would be incorrect to infer from this that a causal link is a necessary and sufficient condition. Indeed, mathematically speaking, necessity and sufficiency is a characterization of equivalence, not implication or causation. So how can it be that we determine causation (itself a form of implication) by checking equivalence? Here, it is important to note that necessity and sufficiency are both determined with respect to the given dataset D. In other words, property A is necessary and sufficient for property B within the dataset D. This does not, however, mean that A is necessary and sufficient for B in all other datasets. In practical terms, this is the same as saying that there might be other settings where something other than A causes B as well.

This apparent discrepancy is precisely the gap between the *closed-world assumption* (CWA) and *open-world assumption* (OWA) (Reiter 1978). Under the CWA, inference is made with respect to a given dataset D: a statement is considered true if and only if it is true in D. For example, for a dataset containing three records of people—Anna, Bob, and Catherine—the statement "David is not a person" would be considered true. Under the OWA, a statement is considered valid[1] or true if and only if it can be proven to be true; that is, it is true for all possible datasets. In the previous example, "David is not a person" cannot be proven so it can be both true and false depending on context.

What does this have to do with the actual analysis of causality? Unless the dataset being analyzed provides an accurate description of the entire universe of discourse, there will always be a limitation to the relationships we can determine as causal via necessity and sufficiency (inference of probabilistic equivalence using the CWA) and "actual" causation (inference of probabilistic implication using the OWA), since we will never manage to prove that nothing other than A can cause B. In a different context, it might well be the case that C causes B. To summarize, necessity and sufficiency within a dataset only provide a tool for revealing possible causal links; they do not define causality. Though we will likely never be able to close completely the gap between the CWA and OWA, analysis of multiple datasets and access to big data describing a more complete view of the relevant data can narrow the gap.

In biomedical research, causality is usually considered when a change in one parameter (cause) within a system is associated with a change in another parameter or the wider state of a system (effect).[2] We can consider two subtypes of causality. First, idiographic causality concerns itself with causal relationships for specific units or events such as a group (considered as singular entities, regardless of how they are defined), an individual, or specific event (Molenaar 2004). For example, the idiographic approach to depression would focus on the cause of depression in a single individual without requiring or even being concerned as to whether the same causal factors may or may not apply to other people. Second, nomothetic causality is concerned with factors that are generalizable to other contexts (i.e., individuals, groups, events), such as the causes of depression whenever and wherever it occurs.

Another aspect of causality refers to its *nature*, which is conventionally considered in terms of deterministic or probabilistic and necessary or sufficient (Khemlani et al. 2014). Deterministic causal relationships require that a change in a specific parameter (parameter A) is *always* followed by a change in another specific parameter (parameter B). In probabilistic causality, "always" is replaced by "frequently"; in other words, a change in the parameter A increases the probability that a change in parameter B will occur. Such "causes" are referred to as "risk factors" or "enabling conditions" with the latter avoiding assumptions about the desirability of the effect. A causal relationship is considered necessary when a change in parameter B can never happen unless there is a change in parameter A. A causal relationship is considered sufficient when a change in parameter A can cause a change in parameter B, although changes

1 Note: one refers to validity under the OWA as opposed to truth.

2 "Cause" and "effect" may be named differently in other fields; for example, "exposure" and "outcome" in epidemiology or "independent variable" and "dependent variable" in psychology.

in parameter B can occur through changes in other parameters (Rothman and Greenland 2005).

A further important dimension of causality is the criteria that need to be met to consider an association between changes in parameters A and B as cause and effect. The most influential set of considerations in the context of medicine was established by the British statistician and epidemiologist, Austin Bradford Hill (1965), and are

- strength of association,
- consistency of association,
- specificity of association,
- temporality, biological gradient (dose-response relationship),
- biological plausibility,
- coherence with previous knowledge,
- experimental evidence (e.g., clinical trials, intervention studies including natural experiments), and
- analogy (i.e., testing that there are analogous causal mechanisms in certain animal models and humans).

Modern medicine and epidemiology tend to rely increasingly on counterfactual reasoning and related approaches to infer causality (Höfler 2005).

In the case of digital ethology (and many other fields), the problem of causality becomes more complicated because causality may be bidirectional, where changes in parameters A and B can reiteratively influence each other. For instance, our social and built environments form ecosystems that contribute to what has been termed "social and structural determinants of health." Thus, we both "receive" and "create" our environments, while codetermining what air we breathe, how many steps we take, how hot or cold we are, and what and who we see, hear, and interact with during our commutes (Paus et al. 2022). Ultimately, causality also moves across levels of organization, from the emergence of collective dynamics to the downward causation when individuals tune their behavior in response to estimates of collectively computed macroscopic properties (e.g., social inequality; Flack 2017).

Common and Specific

Brian: You're all individuals!
Followers: Yes, we're all individuals!
Brian: You're all different!
Followers: Yes, we are all different!
Dennis: I'm not.
—Monty Python, from the *Life of Brian* (1979)

The availability of large-scale digital data has the potential to enable the interrogation of behaviors across diverse cultural, ethno-racial, and socioeconomic human groups. Also of interest is how human behaviors may relate to behaviors observed in other species, such as nonhuman primates. Behavioral patterns (either human or nonhuman) are typically assigned to different constructs that are theoretically defined (e.g., attachment).

Differences between groups may arise at the theoretical meaning of a construct. If constructs are theoretically deemed to be similar, differences may arise in the behaviors assigned to the construct in distinct groups (or species) or the measures developed to assess the construct in these distinct groups (or species). Further, even though the construct and context remain constant, there may still be different measures for assessing this construct representing preferences or conventional practices among researchers. Establishing equivalence of constructs and measures is a prerequisite for comparative studies and a complex task in itself because there is no universally agreed definition of what constitutes equivalence and how it can be established.

We advocate for the scheme provided by Hui and Triandis (1985), who consider equivalence between constructs at the conceptual, functional, item, and scalar levels. Conceptual equivalence requires that a construct has the same meaning across groups (or species). Functional equivalence requires that constructs have similar nomological properties across groups (i.e., same predictors, consequences, and correlates). Conceptual equivalence is conventionally established through a process of building a theoretical consensus, whereas establishing functional equivalence involves statistical strategies that aim to identify common patterns of associations between constructs and their nomological properties across groups (or species). Item equivalence and scalar equivalence can only be considered for constructs that are conceptually and functionally equivalent. Item equivalence refers to the instruments used to assess a construct and their goodness of fit for that construct. Finally, scalar equivalence requires that the same instrument yields similar results when used in different groups. Item and scalar equivalence can be assessed by a variety of methods including reliability coefficients, examination of the internal structure of an instrument, measurement invariance across groups, or using tools from item response theory.

For example, attachment is a construct that refers to a child's relationships to their social partners and their embeddedness in their social world (e.g., Keller and Chaudhary 2017). It is crucial for a child's development of trust—both in themselves as well as in others—and sense of self. Historically, research on attachment has focused on Western middle-class families, often described as WEIRD (western, educated, industrialized, rich, democratic; Henrich et al. 2010). In these contexts, attachment typically unfolds within the framework of a nuclear family, where there is usually one primary caregiver, often the mother, engaging in exclusively dyadic interactions with the child. These interactions, characterized by distal communication such as face-to-face interaction, language, and play with toys, are structured to foster psychological autonomy and self-consciousness in the child from an early age (Keller 2021).

This WEIRD perspective does not represent, however, the diverse nature of attachment across different sociocultural contexts (Henrich et al. 2010; Keller and Bard 2017). In many non-WEIRD societies, including traditional farming, hunter-gatherer, and fishing communities, childcare involves a more extensive network of caregivers, which may include up to 20 people, both related and unrelated (Keller 2021). The mother, while often a central figure, may be one among many caregivers or even play a marginal role. In these settings, children's interactions are mainly proximal, involving bodily based communication processes emphasizing rhythm and synchrony. These societies prioritize the development of a communal self, teaching children to be integral and responsible members of their community, and often have hierarchical social structures that influence communication and interaction rules. This contrasts sharply with the WEIRD model of fostering individual autonomy and self-reliance (Keller and Chaudhary 2017; Morelli et al. 2017).

Digital ethology, with its potential for analyzing large-scale digital data capturing a wide array of behaviors, offers a unique opportunity to examine how the construct of attachment is expressed and understood differently across cultures. By exploring behavior patterns in digital communication, digital ethology can reveal how attachment and socialization strategies are expressed across various cultures. This approach can also be relevant in examining the formation and expression of multiple cultural identities, especially in a globalized world where migration plays pivotal roles (Garcia Coll and Marks 2011). Nevertheless, it is essential to be aware of the potential for an even more narrow bias toward the "digital WEIRD" subpopulation in digital ethology (i.e., the part of the WEIRD population that is accustomed to digital technologies). This means ensuring that digital ethology does not simply reinforce the attachment models based on research conducted on Western societies, but instead captures the rich diversity of attachment expressions globally. To obtain a more representative and comprehensive understanding of global behaviors, it is imperative to analyze digital interactions not only through the lens of advanced technologies prevalent in Western societies (e.g., expensive smartphones) but also through technologies and platforms used widely in non-WEIRD contexts. This includes focusing on more popular tools in developing countries (e.g., affordable mobile models) and exploring messaging apps and social platforms that are available as globally as possible (e.g., apps that are avoiding censorship). Thoughtfully applying the framework proposed by Hui and Triandis (1985) becomes particularly relevant in this context. This framework emphasizes ensuring that the construct of attachment is inclusively and consistently defined across cultures (conceptual equivalence), as its meaning can vary significantly. It is crucial for researchers to also verify that the role and significance of attachment behaviors are comparable across different groups (functional equivalence). This includes adapting measurement tools, like questionnaires or digital analysis algorithms (item equivalence), to suit each cultural context and ensuring these tools yield consistent results (scalar equivalence) across various cultures. This approach, especially challenging in digital ethology due to the diversity of online platforms and communication styles, demands careful construction, adaptation, and validation of research methods. This process allows researchers to draw more reliable conclusions, recognizing the richness of cultural variations while maintaining scientific rigor and comparability of data.

Conclusion

Digital ethology is grounded in the established core methods of observation and knowledge extraction of traditional ethology yet it faces burgeoning challenges associated with large-scale digital data. Major challenges are associated with causality, especially when humans are bidirectionally coupled to their environment: "…enough people participating in an individual activity can result in structural change and vice versa" (Lovasi et al., this volume, p. 33). This becomes obvious when certain behaviors have no meaning at the individual level (e.g., Gini index or synchronization phenomena). Thus, at the methodological level, we need to develop "collective" ethograms and mathematical tools to account properly for these niche constructions at ecological and social levels (Krakauer et al. 2020). In addition, at the legal and ethical level, we should keep in mind that data ownership can go beyond individuals, for instance, in the case of Indigenous communities where communal structures override individual claims.

Acknowledgments

We are thankful to the organizers of the Ernst Strüngmann Forum and all our colleagues for the pleasant and stimulating "mind jazz" over the week. We especially thank the Forum participants who took part in our group's discussions: Beate Ritz, Claudia Bauzer Medeiros, Fabio Kon, Gina S. Lovasi, Henning Tiemeier, Hye-Chung Kum, José Balsa-Barreiro, Michael Brauer, Michele C. Weigle, Nitesh V. Chawla, Steven Bedrick, and Sven Sandin. GD wants to thank Alexandra Glavaschi for providing a high-definition scan of the original Tinbergen's drawing in Figure 2.1.

Appendix 2.1: Glossary

Ascertainment bias, differential recording of outcomes or imbalance screening for outcomes among exposed individuals compared to unexposed individuals.

Convergent validity, often measured by applying different tests and observational methods that intend to measure the same construct with the same individual or groups of individuals and test the consistency or interrelationship.

Discriminant validity assesses how much tests/other methods that are not intended to measure the construct in question, deviates/differs from assessments intended to measure the construct. Reliability and validity belong together.

Dissemination area, the smallest standard geographic area for which all census data are disseminated, usually a small area composed of one or more neighboring dissemination blocks (400–800 inhabitants).

Internal consistency means that individuals/groups respond consistently across items measuring the same construct. If you have a questionnaire measuring one construct, you can, for example, split the items and correlate the two sets. Challenges are, for example, the quality of formulation and preciseness of the items and the extent to which they measure the construct.

Inter-rater reliability in which two trained raters observe the same situation, or the same videotape. Their agreement is statistically assessed, most simply in percentage, more usually with a Cohen's Kappa coefficient.

Outcome (variable) is an event or metric that captures a construct or a predicted behavior. It is measured as categorical (nonparametric statistics), ordinal (nonparametric statistics), or continuous (parametric statistics) values.

Reflexivity means sensitivity to the ways in which the researcher and the research process have shaped the collected data, including the role of prior assumptions and experience, which can influence even the most avowedly inductive inquiries. Personal and intellectual biases need to be made plain at the outset of any research reports to enhance the credibility of the findings.

Reliability refers to the consistency of a measurement. Three types of consistency are usually considered: over time (test–retest reliability), across items (internal consistency), and across different observers/coders (inter-rater reliability).

Respondent validation, or "member checking," includes techniques in which the investigator's account is compared with those of the research subjects to establish the level of correspondence between the two sets. Participants' reactions to the analyses are then incorporated into the study findings.

Sampling Frame:

Test–retest reliability means measuring the same construct/variable at two different points in time on the same individual or group of individuals and testing the correlation of the two measurements. One challenge in this method is the potential for learning effects; for example, if the same items are used, participants might remember their previous responses, which can influence the consistency of the construct over time.

Triangulation compares the results from either two or more different methods of data collection (e.g., interviews and observation) or, more simply, two or more data sources (e.g., interviews with members of different interest groups). The researcher looks for patterns of convergence to develop or corroborate an overall interpretation.

Validity refers to the extent to which a measure represents the variable or construct intended to measure. There are also different kinds and different ways to define validity, most often it is convergent and discriminant validity.

This is a section of doi:10.7551/mitpress/15532.001.0001

Digital Ethology

Human Behavior in Geospatial Context

Edited by: Tomáš Paus, Hye-Chung Kum

Citation:

Digital Ethology: Human Behavior in Geospatial Context
Edited by: Tomáš Paus, Hye-Chung Kum
DOI: 10.7551/mitpress/15532.001.0001
ISBN (electronic): 9780262378840
Publisher: The MIT Press
Published: 2024

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open

The MIT Press

Series Editor: J. R. Lupp
Editorial Assistance: A. Gessner, C. Stephen
Lektorat: BerlinScienceWorks

The book was set in TimesNewRoman and Arial.

Library of Congress Cataloging-in-Publication Data is available.

ISBN: 978-0-262-54813-7

10 9 8 7 6 5 4 3 2 1