Human Behavior in Geospatial Context

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

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

1

Human Brain and Behavior in Geospatial Context

Why and How

Tomáš Paus

General Background

From conception onward, the individual is developing, maturing, working, playing, and aging in their[1] context. As illustrated in Figure 1.1, multiple layers of environment (context) surround an individual across space and time: from the uteroplacental circulation connecting the fetus and their mother before birth, to the influence of their caregivers, extended family, and peers during childhood, adolescence, and adulthood. This "proximal" context (light gray) is embedded in larger geospatial units, such as specific neighborhoods, cities, or countries (dark gray). All environmental influences unfold in time throughout the individual's lifespan. Needless to say, the different layers interact, in a bidirectional manner, with each other. Thus, for instance, a pregnant person responds to signals generated by the fetus, and vice versa (Fowden et al. 2022; Kolle et al. 2020; Menon 2019), the pregnant person interacts with their partner, and vice versa (Khaled et al. 2021; Saxbe et al. 2018), and the caregiver interacts with the child, and vice versa (Carollo et al. 2023; Paquette and St. George 2023). At the same time, the individual and those in their proximal context (e.g., caregivers and peers) act as both recipients and co-creators of their area-level environment along all its dimensions, including physical environment (e.g., air quality), built environment (e.g., parks and transportation network), and social environment (e.g., social cohesion). Different aspects of the environment change over time in an interdependent fashion (e.g., air quality, vehicular traffic, lack of green space, demographic characteristics), often

[1] Throughout this chapter, "they" (and its derivations) is used as a gender-neutral third-person pronoun.

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

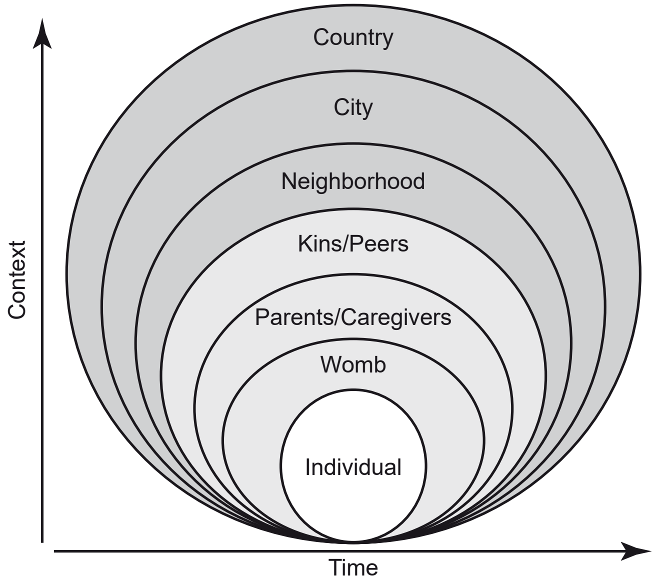

Figure 1.1 Conceptualization of the multiple layers that comprise the contextual environment of an individual across space and time.

reflecting the resources and policies in place at different levels of geospatial granularity (e.g., country, city, neighborhood). Both within and across countries, the lack of environmental justice is reflected in disproportional exposures of marginalized communities to various combinations of adverse environments and, in turn, their *combined* health effects (Van Horne et al. 2023).

For those of us interested in understanding the forces that shape the human brain and behavior, from conception onward, the complexity of this multilayered "exposome" (Munzel et al. 2023; Wild 2005) is staggering. The field of population neuroscience emerged to face this challenge; it brings together epidemiology, genetics, and neuroscience to gain insights into factors underpinning the interindividual variability in the structure and function of the human brain (Paus 2010, 2013, 2016). Owing to the ease of characterizing the individual's genome and the advances in our understanding of related biological processes, initial studies focused on the genetic side of the equation. Working mostly in the context of international consortia, such as ENIGMA (Thompson et al. 2014) and CHARGE (Psaty et al. 2009), we have learned a great deal about the molecular architecture of various quantitative traits derived from magnetic resonance images of the human brain (Grasby et al. 2020; Satizabal et al. 2019; Shin et al. 2020), but efforts on the environment front lags behind. This is understandable given the difficulty of characterizing an individual's environment. Published studies in this area have addressed a handful of factors—one at the time—from the different context layers illustrated in Figure 1.1, such as intrauterine environment (e.g., exposure to maternal cigarette

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

smoking during pregnancy; Muller et al. 2013; Toro et al. 2008), family environment (e.g., family socioeconomic status; Noble et al. 2015), population density (Xu et al. 2022a), as well as variations in the physical (e.g., air pollution; Sukumaran et al. 2023), built (e.g., green space; Kardan et al. 2015) and social (e.g., income inequality; Parker et al. 2017) environments across neighborhoods, cities, and/or countries. Although encouraging, major gaps remain. The Ernst Strüngmann Forum on Digital Ethology, convened in Frankfurt am Main, Germany, in July 2022, brought together scholars and experts to address a number of conceptual and practical gaps in this area.

As pointed out above, the scarcity of multidimensional data that can be used to characterize an individual's environment in an integrated fashion represents the key challenge for studying relationships between the multilayered, multi-domain environment and individual-level outcomes, such as brain development and aging. The Forum addressed this challenge in two ways. *Conceptually*, it called for adopting an ethological approach whereby human behavior is observed, or inferred, in the "wild"; that is, without influencing the observed individual (e.g., by asking them questions). *Practically*, it called for focusing on data sources that either exist or can be readily harnessed at an aggregate level, with different area-level (spatial) granularity (e.g., neighborhood, city, country). The ethological framework presented in Dumas et al. (Chapter 2) underpins the name to this Forum. By "digital ethology," we mean the observation of human behavior through its digital manifestations, such as the use of a search engine, a payment card, or through posting on social media. This behavior leaves "digital footprints" that are particularly relevant for characterizing the social environment of a given area-level unit, as discussed by Weigle et al. (Chapter 4). Human behavior is also reflected and constrained by the surrounding physical and built environments, as outlined by Lovasi et al. (Chapter 3). Finally, variations in individual-level outcomes as a function of the multidimensional area-level environment can best be studied using large datasets; the practicalities as well as legal and ethical considerations are addressed by Medeiros et al. (Chapter 5). Finally, Chapters 6 through 12 provide primers to many of the concepts and strategies that underpin digital ethology.

A Case Study: Inequalities in Area-Level Environment and Brain Health[2]

Social, economic, and political conditions produce *health inequalities* within and across countries (Metzl and Hansen 2018; Scambler 2012; Stuart and Soulsby 2011). In high-income countries, for instance, individuals are more likely to experience poor mental health if they grow up in households with low

[2] This section is a modified version of an article published in *Frontiers in Neuroimaging* (Paus et al. 2022).

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

income (Bjorkenstam et al. 2017) or affluence (Elgar et al. 2015; Rajmil et al. 2014), live in areas with high deprivation (Kivimaki et al. 2020), or experience inequalities in income distribution (Mangalore et al. 2007). Certain communities are disadvantaged more than others (Waldron 2018). This is especially true for Indigenous (Ogilvie et al. 2021) and racialized (Castro-Ramirez et al. 2021) communities, which are at higher risk for mental-health problems and simultaneously experience a lower likelihood of receiving evidence-based treatment (Castro-Ramirez et al. 2021). At the area level, our physical, built, and social environments combine to create ecosystems in which we live and work. Together, these ecosystems, as well as the structures and systems that produce them, contribute to what has been termed "social and structural determinants of health" (Diderichsen et al. 2001; Vandenbroucke 1990).

As described elsewhere (Paus 2016), there are countless permutations of the physical, built, and social environments that surround us in space and time. We both "receive" and "create" our environments (Kendler et al. 2003), thus co-determining what air we breathe, how many steps we take, how hot or cold we are, as well as what and who we see, hear, and interact with during our commutes. Together with our genes, these "external exposures" contribute to "internal" environments that exist in our body: on body surfaces (e.g., microbes on our skin and in the gut), in the lungs (e.g., particulate matter), circulating blood (e.g., toxins, micronutrients, inflammatory molecules), and the brain (e.g., stress- and reward-related neurotransmitters, cumulative engagement of specific neural circuits).

As pointed out above, the use of aggregate-level (spatial) data, produced from multiple locations and time points, is one strategy for characterizing physical, built, and social environments surrounding the individual. In turn, linking such aggregate-level data with individual-level information about a person's health in general, and brain health in particular, provides the first step toward understanding these relationships. Below, the basic steps in this process are reviewed, which are covered in depth in Chapters 6–12.

Geospatial Mapping of Area-Level Environments

Geospatial science and related tools enable spatial analysis and visualization of the external environments in which we spend considerable amount of our lives (e.g., our residence, place of work, school, recreation or a commute path) and an evaluation of their impact on our health. Datasets can be created at different levels of spatial granularity matching the goals of a given study and availability of relevant data. In Canada, for example, geographic units include six-digit postal codes, Canadian Census geographic units such as dissemination areas (400 to 700 persons), and census tracts (2,500 to 8,000 persons), or larger areas such as city districts. The spatial unit used to link geospatial datasets to health data varies; depending on the study and actions necessary to protect confidentiality of study participants, this can be as precise as the exact street address

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

or a postal code (half of a city block in dense urban areas), or as coarse as a city district, a county, a province/state or a country. The temporal dimension depends on the type of data; it may range from data sampled monthly (e.g., air quality), annually (e.g., public transportation), or up to every five years (e.g., the Canadian Census).

Spatiotemporal datasets can be created using existing tools and databases provided by large GIS-based (geographic information systems) organization and companies, such as ESRI, DMTI Spatial, Google Earth Engine, as well as open sources (e.g., Open Street Map), government sources (e.g., Statistics Canada), and academic organizations. In Canada, we have acquired, curated, and disseminated geospatially coded information about the physical and built environments through the Canadian Urban Environmental Health Research Consortium, CANUE (Brook et al. 2018). Metrics derived from different sources can be combined to ask, for example, questions about the relationship between socioeconomic indicators (e.g., household income) and the built environment (e.g., access to parks), and thereby used to assess inequity in the spatial distribution of environmental good or hazards. Figure 1.2 illustrates inequality in the access to parks and recreation (derived from Open Street Map data) across areas with a high level (top 20%) of material deprivation (derived from Canadian Census data; Pampalon et al. 2012).

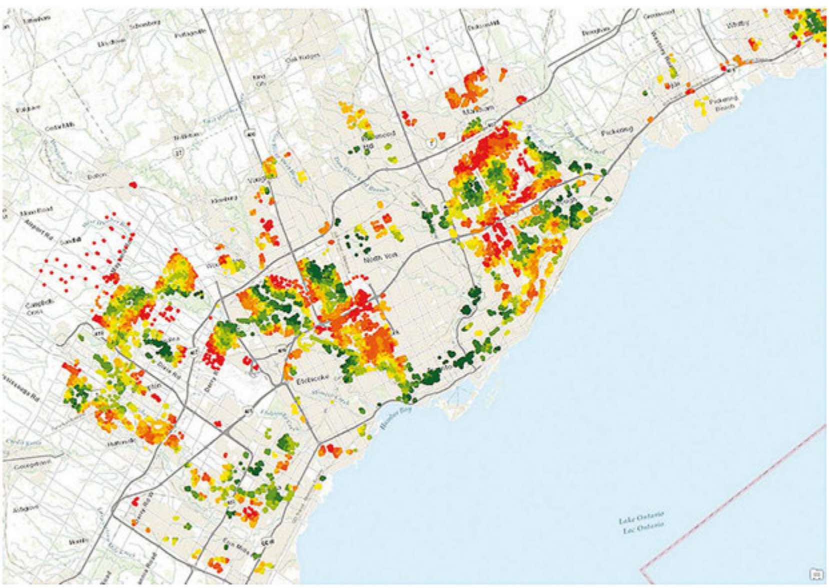

Figure 1.2 Material deprivation and access to parks and recreation in the Greater Toronto Area. All colored areas represent postal codes characterized by high (top 20%) material deprivation (Pampalon et al. 2012). Green indicates postal codes in the highest 10% density of park and recreational amenity within 1 km; red indicates postal codes in the lowest 10% (Source: Open Street Map).

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

In addition to sourcing and creating data about physical and built environments from existing databases (see Table 1 in Paus 2016), one can also derive relevant measures from new data streams such as high-resolution satellite and street-level imagery combined with machine-learning techniques (Weichenthal et al. 2019). For example, Google Street View allows investigators to assess different features of the built environment using panoramic street-level images taken mostly by camera-equipped cars, while recent satellite technology provides daily coverage of most inhabited areas on Earth at a resolution of only a few meters. These geocoded images can be rated for various features, such as signs of physical disorder (e.g., litter, graffiti), physical decay (e.g., poor conditions of sidewalks), type of stores, traffic, or street walkability (Less et al. 2015; Odgers et al. 2012); this approach does have, however, some limitations (Curtis et al. 2013). In turn, computer vision and machine-learning algorithms can exploit these image data to generate indirect indices of the social environment (e.g., psychosocial stress) and physical environment (e.g., air or noise pollution) in a manner similar to that used by others to derive measures characterizing living environment, health, and crime (Suel et al. 2019).

As summarized in Table 1.1 (social environment), a wealth of data speak to basic (often self-reported) measures of socioeconomic factors (e.g., education, employment, immigration, household spending habits, volunteering, and giving) collected by governmental agencies (e.g., census) and national surveys. One can, however, also use data from digital streams (e.g., search engines, social media) to generate new measures of the social environment that are relevant for attitudes vis-à-vis health and health interventions (e.g., vaccination), as well as social cohesion, social support and role models and, most recently, for the emerging issues related to environmental anxiety (Hickman et al. 2021; Soutar and Wand 2022; To et al. 2021; Usher 2022).

Once properly curated, all aggregate-level data (e.g., see Table 1.1) should be described using comprehensive metadata and coded to different geographic units (e.g., postal codes, dissemination areas, and other census geographies), as has been previously done by CANUE.

Linkage with Individual-Level Data

Ultimately, what we are interested in doing is to link aggregate-level "exposures" described above to individual-level "outcomes." In this section, two examples illustrate how this can be achieved using administrative health databases and data acquired in research cohorts.

Administrative Data

In the recent past, we have all seen the power of mapping administrative data related to COVID-19 (across countries, provinces/states, or cities) and communicating these numbers to the public. In Canada, administrative health data

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

Table 1.1 Examples of measures, with the corresponding sources of raw geospatially coded data and examples of the new types of data to be derived.

Physical and Built Environment	**Social Environment**
Air quality (NO_2, O_3, SO_2, PM2.5)[1]	Demographic (b)[6]
Greenness (greenest pixel, tree canopy)[1]	Households (c)[6]
Nighttime light[1]	Socioeconomic (d)[6]
Noise[2]	Water quality concerns[7]
Public transportation[3]	Composting and recycling behavior[7]
Proximity to roads[4]	Involvement in outdoor activities[7]
Proximity to retail outlets and sales of alcohol, tobacco, cannabis, gambling[4]	Caregiving and care receiving[8]
Green roads[5]	Social identity[8]
Facility index[5]	Giving, volunteering, and participating[8]
Cumulative opportunities (a)[5]	Victimization[8]
	Social media and search engine use by youth: frequency and time of day[9]
	Social media and search engine use by youth: content[9]
	Built environment predictors of psychosocial stress[10]
	Built environment predictors of social cohesion[10]

(a) Travel times (walking, public transport) to jobs, leisure, and shopping, as well as health, medical, and social services
(b) population (total and densities), proportions (by age, sex, ethnicity, marital status, mobility/migration status, religion, mother tongue)
(c) household size, total housing units, proportion rented, type of dwelling
(d) household income, unemployment rate, proportion below poverty line, proportion (by age/sex) in labor force

Sources:
1 Landsat
2 CANUE
3 OpenStreetMap (OSM)
4 DMTI Spatial
5 OSM and CANUE
6 census
7 Household and the Environment Survey (Canada)
8 The General Social Survey (Canada)
9 newly derived measures from raw data streams (e.g., Twitter/X, Google search engines),
10 newly derived measures from raw data streams (satellite and street view imagery)

(i.e., data captured during the course of providing services or running programs) are made available for research use by provincial governments and other agencies, often in close partnership with academic organizations (Lucyk et al. 2015). In all provinces, these data are longitudinal and population-based, covering all residents who have received health care and social services (e.g., education), from birth onward. This creates comprehensive and important data for the population of interest, such as youth.

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

In the province of Ontario, for example, administrative health data are curated and made available for research by the Institute for Clinical Evaluative Sciences (ICES), a not-for-profit research institute made up of a community of research, data, and clinical experts that provide a secure and accessible inventory of Ontario's health-related data. Behind a firewall, ICES provides access to coded and linkable databases containing, for example, the Ontario Mental Health Reporting System. Just in the City of Toronto, these data are available for about 270,000 adolescents and youth (12–22 years of age). In addition to health data, many of the provincial custodians of administrative data provide access to other linked datasets, such as education, workplace or justice data (e.g., Population Data BC). When linking administrative data with geospatial datasets containing area-level characteristics of the physical, built, and social environment, one would typically use the residential six-digit postal codes (Canada) and relevant geographies (e.g., dissemination blocks) reported in the administrative data for each individual. Postal code-indexed geospatial datasets are linked in the secure environments controlled by the custodian of the individual-level health data (Boyd et al. 2013; Kum and Ahalt 2013; Pencarrick Hertzman et al. 2013). Here, ethical and legal guidance is necessary to provide assurance to data stewards that this form of data linkage and access can be done in a privacy-preserving and transparent manner that respects all applicable legal, regulatory, and ethical requirements. Ongoing efforts address issues relevant for ensuring public trust, such as transparency of the current practices and systems of governance, and understanding public opinion regarding the use of "big data" in the service of population health (Aitken et al. 2016; O'Brien et al. 2019; Schmit et al. 2021).

Cohort Studies

One of the key advantages of administrative health data is their population-wide coverage. By definition, these data show only the tip of the "health iceberg"; namely, individuals with health issues significant enough to enter the health-care system. This is where community-based cohort studies come in as a complementary source of information, with longitudinal birth cohorts being most valuable. For example, birth cohorts are well suited for investigating relationships between brain health (individual-level data) and context (aggregate-level characteristics of the environment) for several reasons:

1. Many birth cohorts, such as ALSPAC (Boyd et al. 2013), Generation R (Tiemeier et al. 2012) and Northern Finland Birth Cohorts (Rantakallio 1988), ascertained their participants (pregnant women) in a relatively small geographic area.
2. Each cohort includes a relatively large sample size of individuals (~10,000).

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

3. Brain (e.g., mental) health of cohort members is assessed using a number of instruments, often on a continuous scale.

The combination of the first two features makes it likely that a reasonable number of participants live in each geospatial unit, hence providing sufficient statistical power to investigate these relationships. The third feature (assessment) permits the capture of "subclinical" mental-health problems. Finally, additional deep-phenotyping of cohort members through, for example, cognitive assessment, neuroimaging, blood-based biomarkers (e.g., inflammation), genotyping and epigenotyping provides rich information suitable for detailed modeling of exposure–outcome relationships and their mediators and moderators (Paus 2013).

Social Inequality and Mental Health

To close this section, let us consider a hypothetical example illustrating how one can use aggregate-level information about the physical, built, and social environments to unpack the relationship between poverty and mental health. As pointed out by Diderichsen et al. (2001), and represented in Figure 1.3, social stratification—with poverty being but one example of social, economic, and political inequalities—generates a vicious circle: Disadvantaged persons

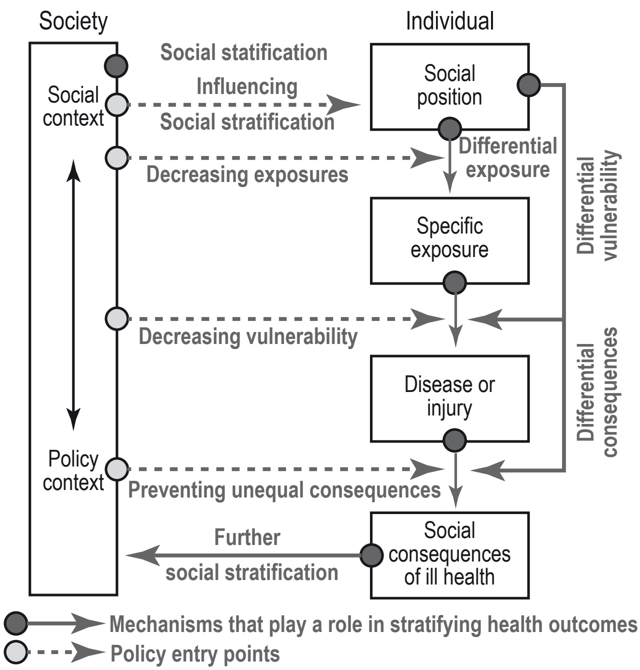

Figure 1.3 From structural inequalities to ill health (Diderichsen et al. 2001).

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

are more likely to be exposed to harmful or deprived physical (e.g., air pollution), built (e.g., access to food stores), and social (e.g., lack of social support) environments as well as to population-level challenges (e.g., heat waves, SARS-CoV-2). These exposures lead to an increased vulnerability to other exposures (e.g., victimization), and both the exposures and vulnerabilities combined precipitate (mental) illness. This vicious circle is closed by the illness leading to further social stratification (e.g., lost educational and employment opportunities). Having extensive multi-domain area-level datasets that can be used to characterize the physical, built, and social environments would enable us to test a variety of possible pathways (and their combination) leading from social stratification to brain health; decomposition analysis is but one method that can be used to quantify contributions of various factors to the observed outcomes (O'Donnell et al. 2008).

Looking Forward

As discussed by Lovasi et al. (Chapter 3) and outlined in the environmental justice framework for exposure science (Van Horne et al. 2023), complexities of the multilayered relationships between the individual and their environment require not only top-quality data and conceptual and analytical approaches but also meaningful engagement with communities and their policy makers, as well as development and implementation of adequate strategies by funders, academic institutions, and journal editors working in this research field. Innovative methods should be employed to address one of the main limitations of observational strategies, namely the difficulty of making causal inferences (see Dumas et al., Chapter 2). For example, dense time-series of multiple exposures and outcomes offer an opportunity for estimating Granger causality (Imran et al. 2023). Pseudo-experimental design can explore causality and directionality in cases of discrete events that affect local environment; note that events such as forest fires may impact not only physical (air quality; Khraishah et al. 2022) but also built (loss of infrastructure) and social (evacuation) environments. To begin to address this issue, at least for certain environments, Mendelian Randomization (Smith and Ebrahim 2003) could be used. For example, using genetic variants associated with biomarkers of low-grade inflammation (Liu et al. 2019; Xu et al. 2022b), one can test a mechanistic path by which air pollution affects brain-related outcomes (Fani et al. 2021). Finally, as reflected in the diversity of the Forum participants, this enterprise requires experts from wide-ranging domains, including geospatial and data science, behavioral and brain science, epidemiology and public health, ethics and law, as well as urban planning. Finding a common language and purpose will allow us to work together toward the understanding of how humans transform their environments and how environments shape human brain and behavior.

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