AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

ISSN: 1082-989X

https://doi.org/10.1037/met0000651

A Framework for Studying Environmental Statistics in Developmental Science

Nicole Walasek[1, 2], Ethan S. Young[1], and Willem E. Frankenhuis[1, 3, 4]
[1] Department of Psychology, Utrecht University, Netherlands
[2] Developmental Psychology, Behavioral Science Institute, Radboud University, Netherlands
[3] Department of Criminology, Max Planck Institute for the Study of Crime, Security and Law, Freiburg, Germany
[4] Department of Evolutionary and Population Biology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Netherlands

OPEN DATA

OPEN MATERIALS

Abstract

Psychologists tend to rely on verbal descriptions of the environment over time, using terms like “unpredictable,” “variable,” and “unstable.” These terms are often open to different interpretations. This ambiguity blurs the match between constructs and measures, which creates confusion and inconsistency across studies. To better characterize the environment, the field needs a shared framework that organizes descriptions of the environment over time in clear terms: as statistical definitions. Here, we first present such a framework, drawing on theory developed in other disciplines, such as biology, anthropology, ecology, and economics. Then we apply our framework by quantifying “unpredictability” in a publicly available, longitudinal data set of crime rates in New York City (NYC) across 15 years. This case study shows that the correlations between different “unpredictability statistics” across regions are only moderate. This means that regions within NYC rank differently on unpredictability depending on which definition is used and at which spatial scale the statistics are computed. Additionally, we explore associations between unpredictability statistics and measures of unemployment, poverty, and educational attainment derived from publicly available NYC survey data. In our case study, these measures are associated with mean levels in crime rates but hardly with unpredictability in crime rates. Our case study illustrates the merits of using a formal framework for disentangling different properties of the environment. To facilitate the use of our framework, we provide a friendly, step-by-step guide for identifying the structure of the environment in repeated measures data sets.

Keywords: environmental statistics, development, time-series analysis, unpredictability, theory

Supplemental materials: https://doi.org/10.1037/met0000651.supp

Psychologists often study interactions between people and their environments. Thus, they describe both the individual and the environment across time. The “environment” often refers to anything external to the individual, including the physical as well as the social world. Consider parental warmth. What is the range and average level of parental warmth (distributional properties)? How does it change across time (dynamic properties)? Psychological studies often involve assumptions or claims about such properties. However, definitions of distributional and dynamic properties tend to be ambiguous (Bringmann et al., 2022; Young et al., 2020).

To illustrate, suppose a researcher investigates the effects of parental warmth on emotional adjustment. They measure parents’ supportive behaviors during a task (Luby et al., 2016). If this researcher aims to capture the level of parental warmth, they might compute a simple average across measurements. However, if they are also interested in “consistency” in parental warmth, a simple average is not

Nicole Walasek https://orcid.org/0000-0003-4411-319X
Ethan S. Young https://orcid.org/0000-0002-8232-0184
Willem E. Frankenhuis https://orcid.org/0000-0002-4628-1712

Willem E. Frankenhuis’ contributions have been supported by the Dutch Research Council (V1.Vidi.195.130) and the James S. McDonnell Foundation (https://doi.org/10.37717/220020502). We thank Sinead English, Marc Mangel, Meriah DeJoseph, Karen Smith, and Judith Dubas for helpful feedback on an earlier draft.

This research has been uploaded as a preprint to the Open Science Framework (https://osf.io/preprints/psyarxiv/fr87n).

Data for this article are publicly available from the New York City Open Data project (https://shorturl.at/korP4) and the U.S. Census Bureau (https://data.census.gov/mdat/#/). Additionally, all used data and code can be found at https://osf.io/acrw4/.

The data are available at https://osf.io/acrw4/.

The experimental materials are available at https://osf.io/acrw4/.

Correspondence concerning this article should be addressed to Nicole Walasek, Department of Psychology, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands. Email: walasek.nicole@gmail.com

enough. They should quantify variability in warmth across time: similar warmth scores imply lower variability, thus greater stability. The standard deviation or range offers two ways to quantify consistency. There are, of course, other options. The main point is that a single research question often affords many different possibilities for quantifying the environment. This ambiguity creates challenges. It leads to constructs being loosely connected to measures and inconsistent data analyses approaches across studies.

Here, we argue that the integration of research across disciplines (e.g., psychology, biology, anthropology), studies, and methods (e.g., different statistical models) will benefit from stating environmental properties in statistical terms. By statistically defining environmental properties, we translate verbal descriptions (e.g., level) to mathematical descriptions (e.g., mean or median). Almost always, multiple mathematical descriptions are suitable for one verbal description. By specifying which one we are using, we reduce ambiguity and increase comparability across findings from different disciplines, studies, and methods. Without choosing an appropriate statistical definition, our construct and research question will remain ambiguous.

The need for statistical definitions comes into sharp focus when talking about change. Many studies explore the challenges that environmental change poses to developing individuals. To identify these challenges researchers study how environmental change shapes such outcomes as health and cognition (Farkas et al., 2022). Indeed, predictable and unpredictable changes may produce different developmental outcomes (Baram et al., 2012; Gee, 2021; Ugarte & Hastings, 2023; Werchan et al., 2022). Documenting such effects is key to understanding both functional and maladaptive responses (Hartley, 2022).

Take, for example, the construct of "unpredictability." In evolutionary developmental psychology, unpredictability is typically defined as "random variation in harshness over space or time," where "harshness" refers to rates of disability and death (Ellis et al., 2009). This verbal definition is ambiguous. It is consistent with multiple statistical formalizations of change over time. For example, one researcher might compute variance in harshness across the measurement period (e.g., Li et al., 2018, 2022). Another might track abrupt shifts in the mean or variance in harshness (i.e., changepoints; described in Young et al., 2020). A third might measure how well current levels of harshness predict future levels (i.e., autocorrelation; e.g., Burgess & Marshall, 2014; Marshall & Burgess, 2015; Werchan et al., 2022). Which statistical definition captures "random variation in harshness" best? And, do different definitions relate to developmental outcomes in the same way? If we want to test specific hypotheses about unpredictability, verbal definitions alone may serve as a starting point, but are rarely sufficiently clear and precise.

The Need for a Framework

In our view, hypothesis testing requires statistically defining distributional and dynamic properties of the constructs under investigation. Doing so creates a structured space for science to progress. Clear definitions serve as the building blocks for exploring and testing questions related to individual-environment interactions across time. Of course, some verbal definitions are precise enough to narrow the possible range of statistical definitions. However, most verbal definitions are consistent with many definitions because verbal description allows for different interpretations (Frankenhuis & Walasek, 2020). In such cases, identifying a set of competing statistical definitions may seem problematic for a theory, but we disagree. Instead, identifying ambiguity is a valuable opportunity to refine theory by exploring competing hypotheses (Platt, 1964). Ignoring ambiguity carries costs (Frankenhuis et al., 2023). Ambiguous definitions weaken the match between theory and methods within studies (e.g., the mapping of constructs to measures), and lead to inconsistencies across studies (e.g., different operationalizations of the same construct). Statistical definitions increase precision and transparency, making it easier to cumulatively build on each other's work (Roisman, 2021). But, before we can compare and evaluate different statistical definitions, we need to organize existing definitions of distributional and dynamic properties. To this end, we present a framework for studying environmental statistics in developmental science (Table 1).

Our Framework

Building on existing work in the social and biological sciences, we integrate familiar approaches with elements borrowed from other fields, such as biology, anthropology, ecology, and economics, which have been using similar frameworks for decades (Bernardi & Hutter, 2007; Ehlman et al., 2023; Hammel, 2005; Marshall & Burgess, 2015; Vasseur & Yodzis, 2004; Vinton et al., 2022; Warlaumont et al., 2022). Our framework organizes statistical definitions of distributional and dynamic properties of an individual's environment. This allows us to reduce the complexity of person-level time series (e.g., repeated measures) to a few summary values describing various statistical properties of the environment.

Our framework offers a collection of univariate, interpretable summary statistics (see Table 1). Other, more complex, approaches to quantifying dynamic properties of longitudinal data (Hoffman, 2008; Nordgren et al., 2020), do not have this feature. These approaches extract coefficients from complex multilevel models as indicators of variability across time. Although these coefficients are rich, they are difficult to interpret. Our approach is agnostic to modeling choices; researchers compute environmental statistics first (e.g., mean, standard deviation, autocorrelation, slope), and may then enter them as predictors into a model. The first step allows researchers to quantify an individual's environment, such as the typical within-person range in harshness exposure. The second step allows researchers to relate environmental properties to outcomes. Our environmental statistics also allow for measurement and validity analyses, such as relating different statistics to each other to understand their interdependence (e.g., the correlations between different statistical descriptions of unpredictability, such as standard deviation, autocorrelation, and entropy).

We distinguish between distributional and dynamic statistics (Jebb et al., 2015; Ram & Gerstorf, 2009). Distributional statistics (mean, median, range, standard deviation) are time-unstructured; the value of a particular statistic does not depend on the order of the data. Even if we shuffle the data, the standard deviation always has the same value. In contrast, dynamic statistics are time-structured; the order of individual data points matters. Dynamic statistics describe how data changes over time. Some dynamic statistics describe the magnitude of changes in mean and variance. For example, a linear slope describes the steepness of mean changes over time. Changepoints describe the number of abrupt shifts in the mean, variance, or both. Other dynamic statistics quantify regular or irregular changes over time. For example, sample entropy quantifies whether sequences of similar values are

Table 1
Glossary of Environmental Statistics

Statistic	Type	Theoretical range	Interpretation
Standard deviation	Distributional	$[0, +\infty]$	Average fluctuations around mean level
Range	Distributional	$[0, +\infty]$	Difference between highest and lowest value; quantifies the absolute range of the data
Interquartile range	Distributional	$[0, +\infty]$	Range of the most common values (middle 50%); a large interquartile range indicates a large range
Slope (linear model)	Dynamic	$[-\infty, +\infty]$	Linear association between time and dimensions of the environment; indicates the trend across time. A positive (negative) slope indicates an increase (decrease) across time
Period	Dynamic	$[0, +\infty]$	Length of a cycle, if any is present; the presence of a cycle indicates that similar values occur every cycle
Autocorrelation	Dynamic	$[-1, 1]$	The extent to which past values correlate with current values; high absolute autocorrelation indicates high predictability
Entropy	Dynamic	$[0, +\infty]$	Sample entropy for time series indicates the extent to which similar sequences in values are followed by additional similar values; low entropy indicates long sequences of similar levels
Spectral coefficient (color of noise)	Dynamic	$[-\infty, +\infty]$	The extent to which noise in dimensions of the environment is autocorrelated across time lags; is often called color of noise
Number of changepoints in mean	Dynamic	$[0, +\infty]$	How often the average level changes across the measurement period
Number of changepoints in variance	Dynamic	$[0, +\infty]$	How often variance in levels changes across the measurement period

Note. The first column states the environmental statistics. The second column states whether this statistic describes a distributional or dynamic property of environmental data. The third column describes the theoretical range of each statistic. "$-\infty$" means any negative number and "$+\infty$" any positive number. In the fourth column, we provide an interpretation of each statistic.

followed by additional similar sequences. Thus, there is a wide variety of possible distributional and dynamic statistics (see Table 1 for a selection of statistics). Researchers must select statistics according to their research question, hypothesis, or theory.

Environmental Statistics and Measurement

Environmental statistics may also inform measurement. Measurement in psychology is notoriously difficult because we often cannot directly observe and quantify what we aim to measure. For example, we cannot measure intelligence in the same way as weight or height. Measurement includes many interdependent steps, such as defining the construct of interest, identifying suitable indicators, and actually measuring data from which we can derive these indicators. In recent years, researchers have raised concerns about various aspects of the measurement process, such as a lack of guiding theory, conceptual ambiguity, insufficient validation, and questionable measurement practices (Bringmann et al., 2022; Eronen & Bringmann, 2021; Flake & Fried, 2020; Frankenhuis et al., 2023; Hodson, 2021; Meier, 2023). Our framework, of course, does not offer solutions to all of these challenges. However, it can aid measurement when the goal is to conceptualize and quantify environmental stability and change.

An important role of statistical definitions is to clarify and narrow the conceptual definition of a construct. When different statistical definitions are equally suited to capture the construct, the conceptual definition may not (yet) be precise enough (Borsboom et al., 2021; Frankenhuis & Tiokhin, 2018; Smaldino, 2020). We illustrate this using the case of unpredictability, defined as "random variation in harshness over space or time" (Ellis et al., 2009). Narrowing this definition can aid measurement. For example, we might define unpredictability in statistical terms (e.g., autocorrelation) and directly quantify it in a data set (e.g., of crime rates). Alternatively, we may define unpredictability as a composite of statistical definitions (e.g., autocorrelation and entropy). Each of these individual statistics then serves as an indicator capturing an aspect of unpredictability. Such a construct would be formative, not reflective, because the indicators would actually constitute unpredictability, rather than reflect it (Bollen & Lennox, 1991; Coltman et al., 2008; Edwards & Bagozzi, 2000).

Environmental statistics may be well-suited as indicators of formative constructs. Which statistical definitions are appropriate in such cases depends on the research question and construct definition. While indicators of formative constructs do not need to be highly correlated (e.g., socioeconomic status [SES]), this is a requirement for reflective constructs (e.g., intelligence; Blotenberg et al., 2022). We think that environmental statistics are not well-suited to serving as indicators of reflective constructs. However, this might be possible, if independent raters can agree which environmental statistics reflect a given latent construct. Changes in one statistic would also need to reflect changes in other statistics (Fleuren et al., 2018). Regardless of formative or reflective constructs, our framework helps to identify conceptual ambiguity and promotes subsequent refinement of constructs, indicators, and measures.

The Benefits of an Environmental Statistics Framework

Our framework has four major benefits. First, the framework increases conceptual clarity by providing formal definitions of distributional and dynamic environmental properties, such as unpredictability. Second, the framework provides guidance by offering tools to compute these statistical definitions and explore their relations. Third, the framework provides common ground for the integration of findings across different empirical studies. Fourth, the framework provides methods for leveraging the time-series nature of longitudinal data.

A crucial step in understanding how the physical and social environment influence development is to quantify how the environment changes within individuals across time. At present, we often use data from single time periods or means across multiple time periods to capture the environment. The reasons for this are often pragmatic: to capture how the environment unfolds across time we need appropriate statistics for quantifying temporal dynamics and many repeated measures. Both of these are often scarce in developmental studies. Our framework can help by providing concrete tools to compute environmental statistics of dynamic change within individuals. At the same time large-scale longitudinal and cross-sectional studies with many repeated measures have grown in number or are actively being conducted over the past few years (Ehlman et al., 2023). Similarly, ecological momentary assessment (EMA) and experience sampling methods (ESM) studies have drastically increased in number because of advancements in smartphone logging technology. EMA and ESM studies monitor the development of behavior, emotions, and mood in real time (Trull & Ebner-Priemer, 2009). Studies using this methodology provide a large number of repeated measures collected across days and weeks allowing us a glimpse into the everyday lives of individuals. All of these various types of data sources present unique opportunities for the application of our framework. Another opportunity for applying our framework lies in computing environmental statistics from secondary data. Such secondary data includes both individual-level data collected as part of previous studies or environment-level data, such as administrative data provided by the local government (Hatzenbuehler et al., 2020; Kievit et al., 2021; Miller et al., 2018; O'Brien et al., 2015). Environment-level data refers to data about the environment of an individual or a group of people. In this way, our framework also facilitates the combination of both individual-level and environment-level data within the same study. Studies that include statistics from both types of data as predictors may afford separating their effects on development, cognition, or behavior.

Our framework complements existing work using dynamic systems theory to study how individuals interact with other individuals, different contexts, or environmental factors (using both subjective and objective measures; Granic & Hollenstein, 2003; Olthof et al., 2020; Ugarte & Hastings, 2023). Dynamic systems theory uses differential or difference equations to describe changing and interacting systems (often called "complex systems"). It typically focusses on a system of multiple, interacting variables in densely sampled data (i.e., minutes, hours, days). We focus on capturing distributional and dynamic statistics of a single environmental variable (e.g., harshness) sampled over longer timescales (i.e., weeks, months, years). So, our framework is well-suited to simultaneously exploring the effects of different environmental variables on developmental outcomes, but not to analyze the coupling of these variables with each other.

Environmental Unpredictability: A Case Study

We use the case of unpredictability to illustrate the four benefits of our framework. In some psychological research, "unpredictability" refers to a person's subjective perception of, or psychological response to, their environment (e.g., in Raab et al., 2022). In the current article, "unpredictability" refers to an objective property of the environment, which individuals may detect and adjust to. For example, children adjust how they navigate their attention to the objective predictability of their auditory environment (Werchan et al., 2022).

In this case study, we focus on unpredictability, defined as random (stochastic) variation in harshness across space, time, or both (Ellis et al., 2009). The term "harshness," in this context, refers to the risk of disability and death (Brumbach et al., 2009; Ellis et al., 2009). In nonhuman animals, resource scarcity and predator density are indicators of harshness. In humans, poverty and crime rates are often used to measure harshness (Brumbach et al., 2009; Young et al., 2020), because they are consistently correlated with morbidity and mortality (Eberly et al., 2022; Sundquist et al., 2006). That said, the points we make in our case study of harshness apply to unpredictability in other dimensions of the environment just the same (e.g., rainfall, temperature).

We illustrate our framework using existing, publicly available crime records in New York City (NYC, United States) as indices of harsh environmental conditions. The data span 15 years from January 2006 until December 2020. For different regions in NYC, we compute and compare a range of statistical definitions of unpredictability. In addition, we use publicly available survey data to derive measures of unemployment, poverty, and educational attainment for different regions in NYC. We explore associations between the NYC survey variables and different statistical definitions of unpredictability in NYC crime data. All our code and data are available at https://osf.io/acrw4/.

Method

In this section, we present a case study of environmental unpredictability in the form of a friendly, step-by-step guide for applying our framework. We preface our guide by emphasizing the importance of ensuring the validity and reliability of input measures. When measures are imprecise or unreliable, researchers may inadvertently interpret random error rather than true changes across time. In addition, the validity and reliability of some measures might change with participant age or across groups. For example, measures of child temperament and personality at young ages may not be valid at older ages. The general point is that measures in our framework should capture the construct of interest, minimize measurement error (e.g., are reliable), and measure the same construct over time and across groups. This increases the likelihood that our framework quantifies real dynamics in the construct. Thus, whenever possible, we recommend using well-validated measures and quantifying measurement invariance (DeJoseph et al., 2022; Flake & Fried, 2020; Putnick & Bornstein, 2016).

Environmental Unpredictability: A Case Study

Using our framework, we want to explore whether unpredictability in crime is associated with socioeconomic outcomes. Specifically, we explore three questions: is an unpredictable threat of assault associated with (a) unemployment, (b) poverty, and (c) educational attainment? Our analyses are exploratory. We draw no inferential conclusions.

Step 1: Specifying the Construct and Selecting Appropriate Statistics

We conceptualize unpredictability as random variation in harshness over space or time (Ellis et al., 2009). Multiple statistical definitions of unpredictability fit this definition.

We explore six different definitions of unpredictability that feature in the literature: the standard deviation, changepoints in mean, changepoints in variance, autocorrelation, entropy, and color of noise. The standard deviation describes the average deviation from the mean in assault rates. In Table 1, we have referred to the standard deviation as a distributional statistic because it summarizes the distribution of the data and not how they change across time (it is time-unstructured). Nonetheless, we include the standard deviation as an unpredictability statistic, because previous work has used the standard deviation of a model's residuals to quantify unpredictability (e.g., Li et al., 2018, 2022). Changepoints describe abrupt shifts in the mean or variance (Haynes et al., 2016; Killick & Eckley, 2014; Young et al., 2020). We assume time-series that are characterized by a large number of changepoints to be less predictable. The autocorrelation indicates how much past values predict current values. Suppose we have monthly measures of assault rates across 1 year. The autocorrelation then corresponds to the correlation of this time series with itself lagged by 1 month (Burgess & Marshall, 2014; Marshall & Burgess, 2015). This is called a lag-1 autocorrelation. Lag-2 autocorrelation computes autocorrelation for a lag of 2 months. Entropy of a time series quantifies the extent to which similar sequences in assault rates are followed by additional similar sequences (Pincus, 1991; Richman & Moorman, 2000). Higher entropy values imply that a time series is complex with few redundancies and thus highly unpredictable (e.g., a random sequence of zeros and ones); lower entropy implies redundancy and scope for predictability (e.g., a sequence in which zeros and ones alternate). Color of noise summarizes the extent to which noise in assault rates is autocorrelated across time lags (Burgess & Marshall, 2014; Frankenhuis et al., 2019; Marshall & Burgess, 2015; Ruokolainen et al., 2009; Vasseur & Yodzis, 2004). Noise is what is left of the data after subtracting systematic patterns, such as trend and season. Although theoretically any positive or negative number is possible (see Table 1), color of noise values commonly range between −1 and 3 (Ruokolainen et al., 2009). Noise can change randomly across time ("white noise," color of noise around 0). It can change slowly, resulting in long runs of above or below average conditions ("red" and "brown noise," color of noise between 1 and 2). Or, it can change rapidly but predictably ("blue noise," color of noise between 0 and −1). Table A1.1 in the online supplemental material 1 outlines in detail how each statistic is computed.

Table 2 describes what type of research questions can be asked with different unpredictability statistics. This level of specificity allows researchers to ask nuanced questions about the association between unpredictability and socioeconomic outcomes.

Step 2: Finding a Data Set

Our data need to fulfill two goals. First, they need to be suitable for studying our research question. Second, they need to meet the criteria for applying our framework and any subsequent analyses. We first describe the input and outcome data and then outline how these data meet our criteria.

The Data. To explore unpredictability in assault rates, we chose publicly available crime records from boroughs and public use micro-data areas (PUMAs) in NYC as input to our framework. NYC consists of five boroughs: The Bronx, Brooklyn, Manhattan, Queens, and Staten Island. Each borough is further divided into community districts for local governance. PUMAs are nonoverlapping geographical areas with at least 100,000 inhabitants which largely overlap with community districts (Figure 1; the online supplemental material 2 shows the overlap between PUMAs and community districts).

Our crime data are part of the NYC Open Data project (https://opendata.cityofnewyork.us/). The database contains data related to business, governance, education, environment, safety, and health. We focus on the New York City Police Department Arrests Data (Historic; https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u). This data set records every arrest by the New York City Police Department dating back to 2006 and continues to be updated every quarter. Each entry holds information about the type of crime, the arrest location, and the time of arrest. We only use information about crimes related to assault resulting in 644,684 entries across 15 years, from January 2006 until December 2020. Table A1.2 in the online supplemental material 1 lists all offenses that we included under "assault" and their description provided by the police department. Each assault can be linked to its respective borough and PUMA using publicly available shapefiles from the NYC department of city planning (https://www.nyc.gov/site/planning/index.page). Shapefiles are regularly updated to account for changes in boundaries. We used shapefiles from 2000 and 2010 to match the geographical boundaries of the survey data.

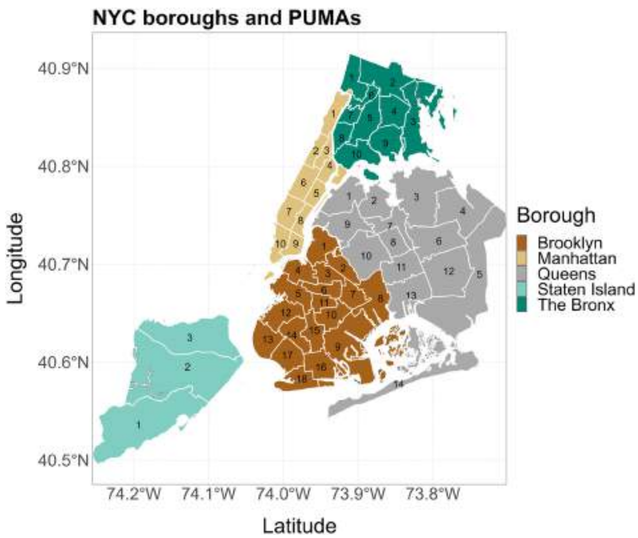
As outcome data, we chose publicly available data from the American Community Survey (ACS). The U.S. Census Bureau developed the ACS to collect data about social, economic, housing, and demographic characteristics in American communities. For this purpose, they introduced PUMAs as data collection units. Data are released in a 1- and 5-year format. The 5-year format (ACS-5) represents the average characteristics across 60 months of data collection and is considered the most reliable estimate.

Table 2

Overview Over Unpredictability Statistics and Specific Research Questions

Type of statistic	Research question
Standard deviation	Are large fluctuations around the mean level of assault rates associated with socioeconomic outcomes?
Changepoints	Are frequent changes in the mean (variance) of assault rates associated with socioeconomic outcomes?
Autocorrelation	Is low predictability in assault rates across years (or months, weeks, etc.) associated with socioeconomic outcomes?
Entropy	Is inconsistency in sequential assault rates associated with socioeconomic outcomes?
Color of noise	Is low predictability in noise of assault rates across years, days, and months associated with socioeconomic outcomes?

Figure 1
NYC Boroughs and PUMAs



Note. Colors indicate boroughs and numbers indicate PUMAs which are separated by white borders. Officially, two digits precede PUMA labels to differentiate between boroughs. To increase visibility these have been omitted. The horizontal axis shows the longitude and the vertical axis the latitude. Borough and PUMA boundaries are based on shapefiles from 2010 (available in our Open Science Framework repository). NYC = New York City; PUMAs = public use microdata areas; N = north; W = west. See the online article for the color version of this figure.

The data are available through the U.S. Census Bureau microdata tool (https://data.census.gov/mdat/#/) and various interfaces. We used the R package tidycensus to download the ACS-5 data (Walker & Herman, 2022). We downloaded three batches of 5-year data to cover the range of the NYC crime data: 2006–2010, 2012–2016, and 2016–2020. Each batch contains aggregate survey data for each PUMA collected across 5 years. Between 2011 and 2012 the PUMA boundaries changed. Therefore, it is not possible to download 5-year aggregates which include data before and after 2012. In line with the PUMA boundaries, we use shapefiles from 2000 to compute associations between the crime and survey data ranging from 2006 to 2010, and shapefiles from 2010 otherwise.

For each PUMA and 5-year batch, we compute the following four variables: the proportion of individuals who are currently unemployed (excluding individuals who are in school or studying), the proportion of individuals with a bachelor's degree or higher, the proportion of individuals whose income-to-poverty ratio is below 1, and the proportion of individuals whose income-to-poverty ratio is above 5.

Criteria. The NYC crime data meet the requirements for applying our framework. They provide a sufficient number of measurements and equal spacing between measurement intervals. Additionally, our sample of PUMAs is large enough to conduct additional analyses with extracted statistics. We describe each criterion in more detail below.

Number of Measurements. Ideally, the data set has at least 20 repeated measures per time series. The more repeated measures, the better: some time series modeling techniques require at least 50 observations (Haslbeck & Ryan, 2021; Jebb et al., 2015). As is the case with all cutoffs, they should not be understood as rules but as guidelines. A data set with 15 or 19 repeated measures may also be suitable for our framework. However, the lower the number of repeated measures, the more likely statistics may be tracking noise in the data, increasing uncertainty in estimates. We have monthly measures of assault rates across 15 years, resulting in 180 repeated measures in total and 60 repeated measures per 5-year batch.

Equal Spacing. Our framework is more easily applied when measures are equally spaced in time (e.g., once a month or year). Equal spacing is especially important for various time-series modeling techniques, such as computing the autocorrelation (de Haan-Rietdijk et al., 2017; Jebb et al., 2015). Equal spacing matters because time series methods typically make assumptions about the sampling frequency of the data. For example, as already noted, autocorrelation can be computed at different lags in the data. With monthly data, lag-1 autocorrelation corresponds to the correlation of the time series with itself lagged by 1 month. To best estimate the autocorrelation at this lag, we require equal, monthly spacing.

Dealing With Unequal Spacing. There are different ways of dealing with unequal spacing. If the irregularity in spacing is small, one option is to ignore it while being explicit about this as a (minor) limitation. However, if the degree of irregularity is more serious, one solution is to exclude statistics and preprocessing steps that assume equal spacing (i.e., autocorrelation and time series decomposition). Or, the data can be transformed to become equally spaced. This can be achieved by interpolation (Pavía-Miralles, 2010). Interpolation uses available data in a time series to predict the most likely value of data points missing because of irregular sampling. However, interpolation can also bias the data and distort their true dynamics; so, before using this method, we recommend exploring additional literature (Erdogan et al., 2005; Lepot et al., 2017; Oh et al., 2020). In our data, measures are recorded daily and can be aggregated across evenly spaced intervals, such as weeks, months, or years.

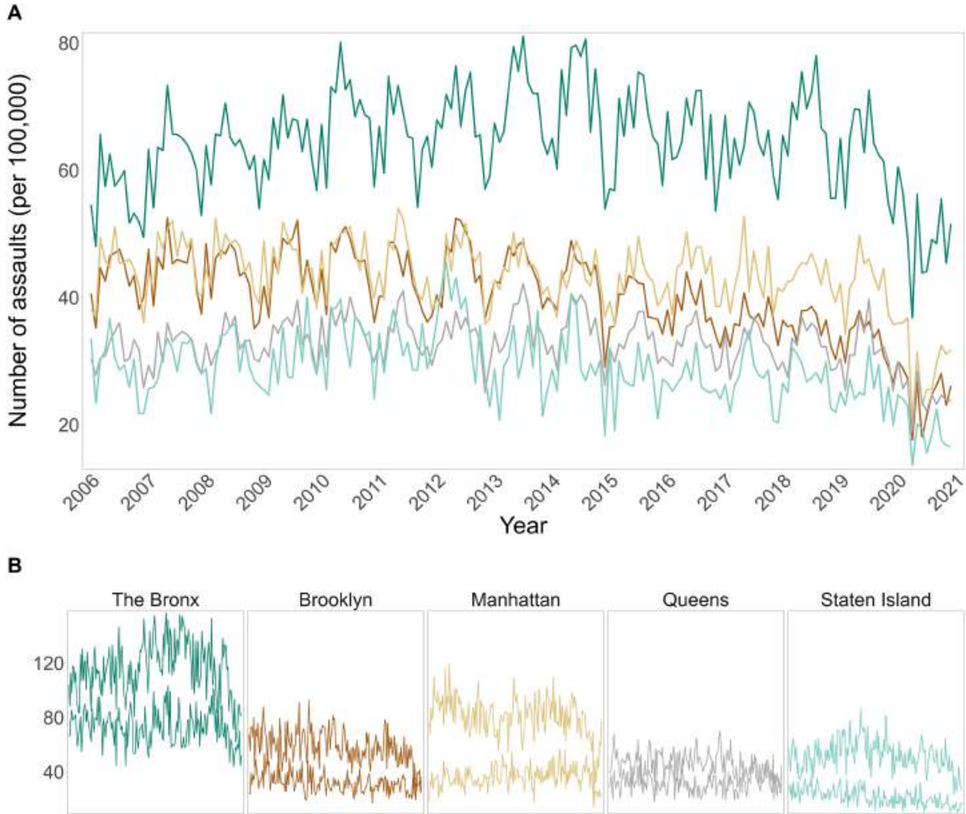
Sample Size. If the goal is to only explore individual time series, our framework minimally requires a sample of one time series. However, it is possible to use statistics from individual time series to analyze the association among different statistics or between statistics and outcomes. We will showcase both of these analyses in this article. Under these conditions, it is important to consider the sample size, that is, the number of available time series. With 55 PUMAs our sample is sufficiently large to compute correlations among statistics and outcomes, as well as simple regression models. But ideally, we would prefer a larger sample of PUMAs.

Currently, we cannot provide specific numeric cutoffs or recommended ranges for our outlined criteria. We expect the application of our framework to remain exploratory until empirical data using our methodology has been generated (i.e., computed environmental statistics across different samples and timescales). Such data can then be evaluated to determine more specific recommendations.

Step 3: Exploration and Preprocessing

Third, we suggest visually and descriptively exploring the data. Exploration is helpful to get familiar with the data and to decide whether and how you should preprocess the data. Our framework offers various options. You can plot the raw time series, the autocorrelation, and changepoints of randomly selected time series (Figure A1.2 in the online supplemental material 1). For example, Figure 2 shows the monthly number of assaults recorded across the measurement period for each borough (Panel A) and two randomly chosen PUMAs (Panel B). The vertical axis displays the

Figure 2
NYC Crime Data Across Boroughs and PUMAs



Note. Panel A shows monthly assault rates across the measurement period for each borough (colors match those in Figure 1). The vertical axis displays the number of assaults per 100,000 inhabitants and the horizontal axis denotes time in years. Panel B displays assault rates across the entire measurement period for two randomly chosen PUMAs per borough. Vertical and horizontal axes denote the same quantities as in Panel A. NYC = New York City. PUMAs = public use microdata areas. See the online article for the color version of this figure.

number of assaults per 100,000 inhabitants to correct for population density. We use census data of boroughs and PUMAs to compute corrections at the appropriate levels. In addition, we show different temporal resolutions (i.e., daily, weekly, biannually, and yearly) of assault rates (Figure A1.1 in the online supplemental material 1).

Additionally, it is possible to compute descriptive statistics of the input and outcome data. Table 3 lists the means and standard deviations in assaults per 100,000 inhabitants across PUMAs for the entire measurement period and each 5-year batch. Table 4 lists means and standard deviations for each NYC survey variable across PUMAs. Additionally, we report means and standard deviations in the coefficient of variation for each variable. The coefficient of variation corresponds to the standard deviation of the estimate, divided by the estimate itself. It indicates sampling variability. The smaller the coefficient, the more reliable the estimate. As a rule of thumb, estimates above 0.3 are considered unacceptable and estimates under 0.15 good (European Commission, Eurostat, 2023). We show PUMA-level breakdowns for each variable and batch (Figures A3.1–A3.12 in the online supplemental material 3).

If desired, the data can be preprocessed before computing statistics. For example, the data may be standardized to compare linear slopes across time series. Additionally, it is possible to remove trends in mean or variance, and seasonal components from the data (see the online supplemental material 1 “Framework user manual”). A time series whose statistical properties (e.g., mean, variance) are constant across time is called stationary. Stationarity is a common requirement for many time series models which assume that individual data points in the time series are independent of each other (Jebb et al., 2015). For example, the autocorrelation of a time series typically picks up systematic changes in the data, such as trends (i.e., consistent increases or decreases) in mean levels or variance. Trends make the data more predictable across time and increase the magnitude of the autocorrelation. However, researchers may also be interested in the magnitude of the autocorrelation after accounting for these systematic patterns. Thus, it is common practice in time-series analysis to compute the autocorrelation of the stationary time series. Our framework, therefore, extracts some statistics for both the raw data and the stationary data. Generally, we recommend staying close to the raw time series. Preprocessing steps, such as detrending or decomposing a time series, may solve one “issue” (e.g., removing a trend) but potentially introduce other artifacts in the data (e.g., increase variability). We chose to include only one preprocessing step in our case study, namely to correct assault rates for population density.

Table 3
Assaults Across Boroughs and Districts—Descriptive Statistics

		Batch							
		2006–2010		2012–2016		2016–2020		2006–2021	
Borough	*n* PUMAs	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
The Bronx	10	89.97	47.86	97.34	54.13	87.96	47.72	91.81	49.59
Brooklyn	18	64.44	31.41	60.71	33.47	48.85	24.99	58.43	31.24
Manhattan	10	75.22	32.86	71.89	31.59	66.67	31.73	71.37	32.04
Queens	14	54.76	28.02	56.26	32.95	50.92	27.66	54.13	29.88
Staten Island	3	49.12	32.45	48.03	37.58	40.42	31.60	46.44	34.65

Note. PUMAs = public use microdata areas.

Step 4: Computing Statistics

We compute, visualize, and summarize unpredictability statistics. Specifically, we compute unpredictability statistics in monthly assaults in NYC crime data across the entire measurement period for each borough and PUMA. We show the results of this step in the Results section.

Step 5: Exploratory and Measurement Analysis

In this step, we explore how individual statistics relate to each other. Do we derive the same conclusions about the degree of unpredictability irrespective of statistical definition? For each 5-year batch, we compute Spearman's rank correlations between unpredictability statistics across PUMAs (see the Results section). We indicate significant, bivariate associations at an (arbitrarily set) α level of .005. It is important to note that PUMAs are nested within boroughs which may influence the patterns of correlations. If it does, a repeated measures correlation would be more appropriate. However, because the interclass correlation was close to zero for almost all unpredictability statistics, we decided against accounting for clustering within boroughs. Only, the interclass correlations for the standard deviation and mean were around 0.2 with a wide confidence interval bordering zero. For comparison, we include plots of the repeated measures correlations (Figures A4.1–A4.3 in the online supplemental material 4).

Step 6: Relating Environmental Statistics to Outcomes

Finally, we compute bivariate associations between unpredictability in assault rates and rates of unemployment, poverty, and educational attainment. To this end, we compute Spearman's rank correlations among unpredictability statistics and socioeconomic outcomes for each 5-year batch of data (see the Results section). For each batch, we show histograms and bivariate scatterplots of all variables (i.e., unpredictability statistics and socioeconomic outcomes; Figures A3.13–A4.18 in the online supplemental material 3).

Additionally, we statistically modeled the associations between each unpredictability statistic and each NYC survey variable. We chose a beta regression because our outcome variables are proportions. To this end, we use Bayesian beta regressions to compute associations between each unpredictability statistic and each NYC survey variable (Burkner, 2013). We controlled for the mean level of assault rates because the bivariate scatterplots show strong associations between the mean and all NYC survey variables across all batches. We standardized the predictors of each analysis to make coefficients comparable and used default flat priors for all variables. We report the marginal effects for each unpredictability statistic and their highest posterior density intervals (HDPI), indicating the most likely values.

Results

We present results of applying our framework to the NYC crime and survey data. We show unpredictability statistics for each borough and compare borough-level statistics to aggregates across PUMAs (Figures 3 and 4). We illustrate to what extent different definitions of unpredictability result in different rank-orderings of boroughs. In addition, we present correlations between unpredictability statistics and NYC survey data across PUMAs (Figures 5–7). Lastly, we discuss associations between individual unpredictability statistics and NYC survey variables while controlling for the mean level in assault rates.

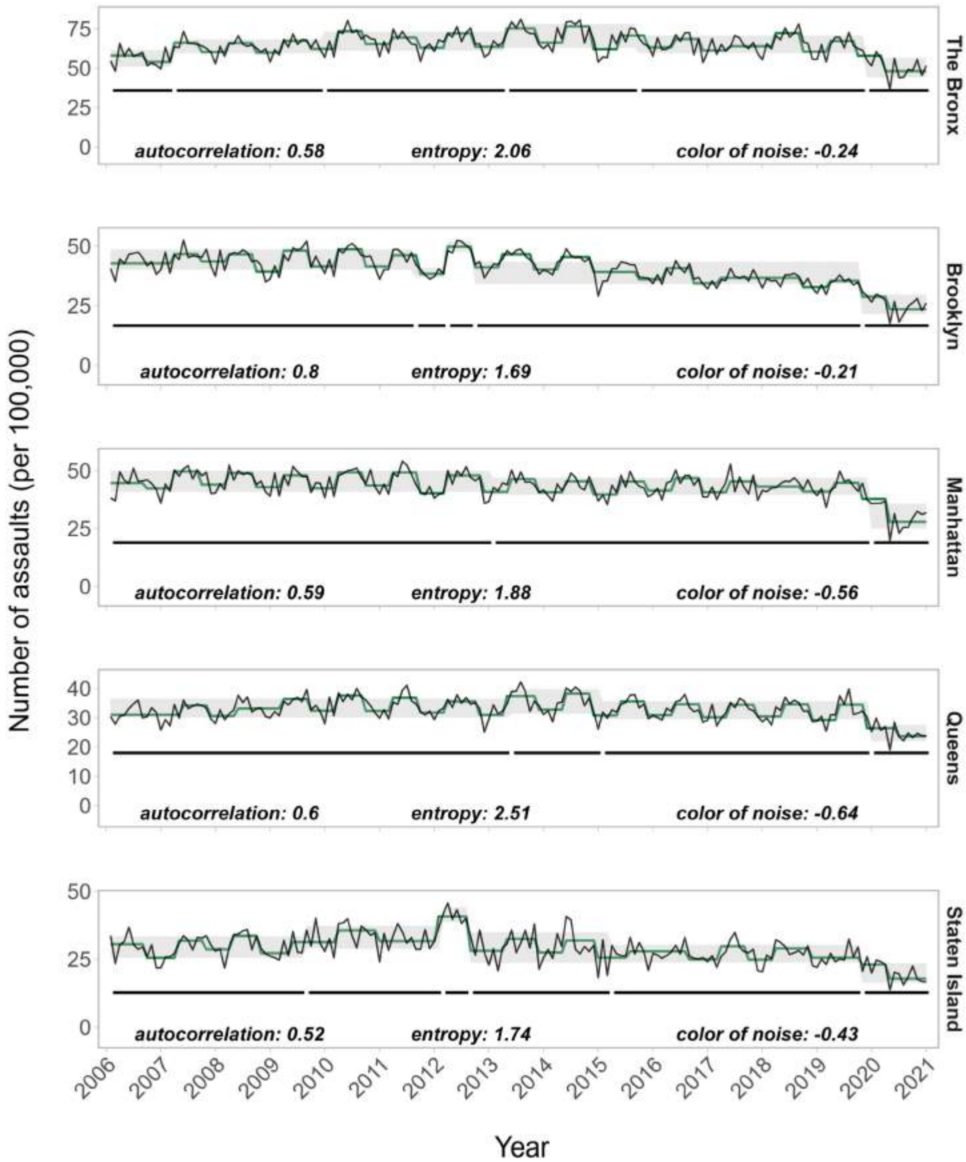
Table 4
NYC Survey Data Variables—Descriptive Statistics

	Batch					
	2006–2010		2012–2016		2016–2020	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*
Prop. bachelor or above	0.32	0.18	0.35	0.18	0.38	0.19
Prop. bachelor or above (CV)	0.03	0.01	0.03	0.01	0.03	0.01
Prop. income-poverty above 5	0.25	0.16	0.26	0.16	0.30	0.17
Prop. income-poverty above 5 (CV)	0.05	0.02	0.05	0.02	0.05	0.02
Prop. income-poverty below 1	0.20	0.10	0.21	0.09	0.18	0.09
Prop. income-poverty below 1 (CV)	0.06	0.02	0.06	0.02	0.06	0.02
Prop. Unemployed	0.05	0.01	0.05	0.02	0.04	0.01
Prop. unemployed (CV)	0.08	0.01	0.08	0.02	0.10	0.02

Note. NYC = New York City; CV = coefficient of variation; Prop. = proportion.

Figure 3

Changepoints in Mean and Variance in NYC Boroughs



Note. Each row depicts one borough. Within each row, the horizontal axis shows time in years and the vertical axis number of assaults per 100,000 inhabitants (note: the scaling differs across rows to show details). The green (gray) line tracks changes in mean and gray rectangles track changes in variance in monthly assault rates. To aid visibility we added the black horizontal lines below the data which track segments of stable variance. The height of the green (gray) line corresponds to the mean of a segment with stable mean. The height of a gray rectangle corresponds to the standard deviation of a segment with stable variance. We show the values of other unpredictability statistics at the bottom of each subplot. NYC = New York City. See the online article for the color version of this figure.

ONLINE COLOR

Environmental Statistics—New York Crime and Survey Data

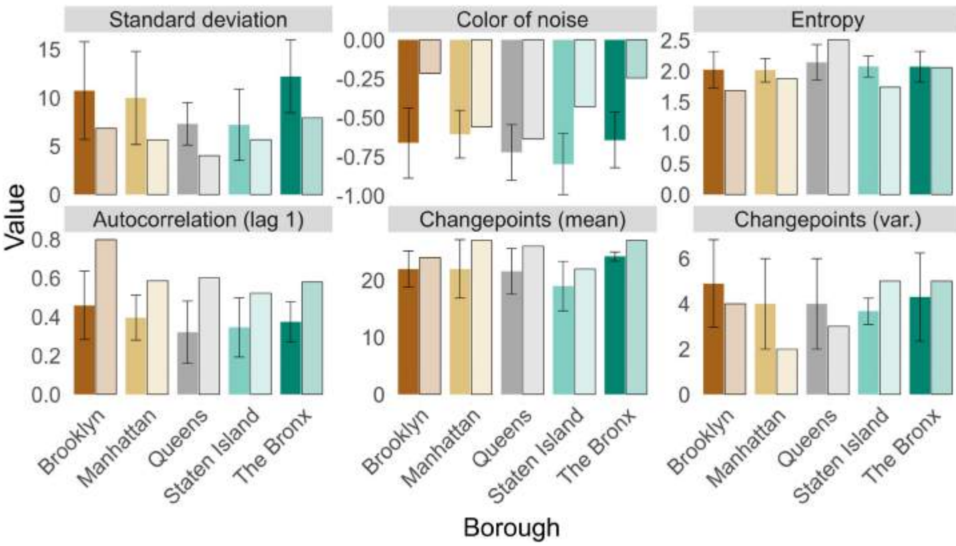
Unpredictability Statistics Across Boroughs and PUMAs

We show six different statistical definitions to quantify unpredictability as "random variation in harshness" (Ellis et al., 2009): the standard deviation, entropy, color of noise, autocorrelation (at a lag of 1 month), changepoints in mean, and changepoints in variance. Higher standard deviation, entropy, and number of changepoints in mean and variance indicate higher levels of unpredictability. The same is true of lower absolute (i.e., the magnitude ignoring the sign) autocorrelation and color of noise.

Figure 3 shows changepoints in mean (green line) and variance (black horizontal lines and gray rectangles) in monthly assault rates for each borough. We observe differences in the number of changepoints in variance across the different boroughs, with the Bronx having the most and Manhattan the fewest. The differences

Figure 4

Comparing Environmental Statistics at the PUMA- and Borough-Level



Note. PUMAs are public-use microdata areas. Each panel represents one statistic and each pair of bars represents one borough. Dark bars and error bars (left bars) indicate averages and standard deviations of statistics computed at the PUMA-level. Light bars with black outlines (right bars) indicate statistics computed at the borough-level. The *y* axis shows the value of each statistic. PUMAs = public use microdata areas; var. = variance. See the online article for the color version of this figure.

in the number of changepoints in mean are smaller (Figure 5). Almost all boroughs show an increase in mean or variance of assault rates between 2007 and 2010 and a decrease between 2019 and 2021. The increase might reflect the global financial crisis in 2007 and 2008. The decrease may be because of lockdowns and other measures against the Corona virus in 2020 and 2021.

Figure 4 compares unpredictability statistics computed from assault rates in each borough with averages of statistics computed across PUMAs. We take three things away. First, the rank-orderings of boroughs depend on which statistical definition of unpredictability we use. Second, these rank-orderings may also depend on the spatial level at which we compute unpredictability statistics. Tables 5 and 6 underscore these insights. They list rankings of boroughs in unpredictability according to our six statistical definitions of unpredictability computed at the borough-level (Table 4) and PUMA-level (Table 5). When ranking boroughs according to the autocorrelation and color of noise we consider their absolute values. We observe little consistency in ranks within the borough- and PUMA-levels individually, as well as across levels. Third, Figure 4 shows that the standard deviation of averaged statistics across PUMAs can be large. This indicates that values of individual statistics may vary even across small spatial distances. Taken together, our results suggest that conclusions about unpredictability may depend on the chosen definition of unpredictability and the spatial level at which we compute it.

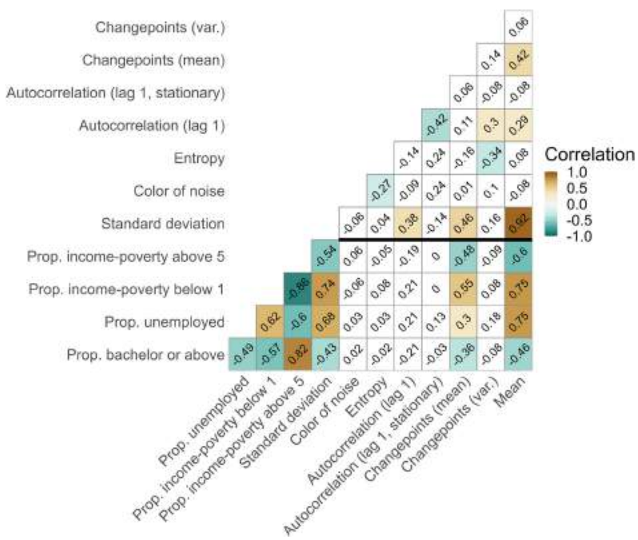
Correlations Between Unpredictability Statistics and NYC Survey Variables

Figures 5–7 depict Spearman's rank correlations between unpredictability statistics computed from assault rates in NYC and NYC survey measures of unemployment, poverty, and educational attainment across PUMAs. Each figure depicts correlations for one 5-year batch of data. The black horizontal line visually separates correlations computed solely between unpredictability statistics (above the line) and correlations which include NYC survey variables (below the line). In addition to our six unpredictability statistics and four NYC survey variables, we include the autocorrelation of the stationary time series of assault rates and the mean level in assault rates. We include the mean level in assault rates for each PUMA because the bivariate scatterplots indicate strong associations between NYC survey variables and the mean (Figures A3.13–A3.18 in the online supplemental material 3). We include the autocorrelation of the stationary time series to assess the magnitude of the autocorrelation after accounting for systematic patterns. When computing correlations we consider absolute values of the autocorrelations and color of noise. We show complete correlation tables in Tables A4.1–A4.3 in the online supplemental material 4.

We first describe patterns in correlations between unpredictability statistics. Across 5-year batches, patterns of correlations are highly consistent. Most notably, the standard deviation and mean are highly correlated. This is a common property of count data which the NYC crime data are. We observe moderate correlations among some of the environmental statistics. For example, color of noise and entropy are negatively associated: lower magnitudes in color of noise and higher entropy values indicate higher levels of unpredictability. Other patterns seem puzzling at first. We observe that higher variance in the data (i.e., higher standard deviation, larger number of changepoints in mean and variance) is associated with higher autocorrelation and lower entropy. This violates our intuitions because we associate the former with higher levels of unpredictability and the latter with lower

Figure 5

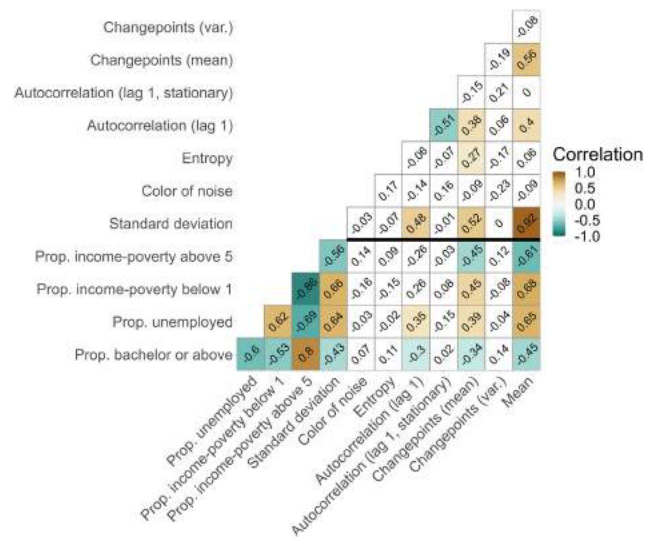
Spearman's Rank Correlations Between Unpredictability Statistics (Computed From NYC Assault Rates) and NYC Survey Variables Across PUMAs for Data Collected Between 2006 and 2010



Note. Each cell corresponds to one Spearman's correlation coefficient between the variables indicated by the row and column. Significant correlations ($\alpha < .005$) are indicated by colored cells. The intensity of the color is proportional to the strength of the correlation coefficient. The black horizontal line separates coefficients between unpredictability statistics only (above the line) and coefficients that include NYC survey variables (below the line). NYC = New York City; PUMAs = public use microdata areas; var. = variance; Prop. = proportion. See the online article for the color version of this figure.

Figure 6

Spearman's Rank Correlations Between Unpredictability Statistics (Computed From NYC Assault Rates) and NYC Survey Variables Across PUMAs for Data Collected Between 2012 and 2016



Note. Each cell corresponds to one Spearman's correlation coefficient between the variables indicated by the row and column. Significant correlations ($\alpha < .005$) are indicated by colored cells. The intensity of the color is proportional to the strength of the correlation coefficient. The black horizontal line separates coefficients between unpredictability statistics only (above the line) and coefficients that include NYC survey variables (below the line). NYC = New York City; PUMAs = public use microdata areas; var. = variance; Prop. = proportion. See the online article for the color version of this figure.

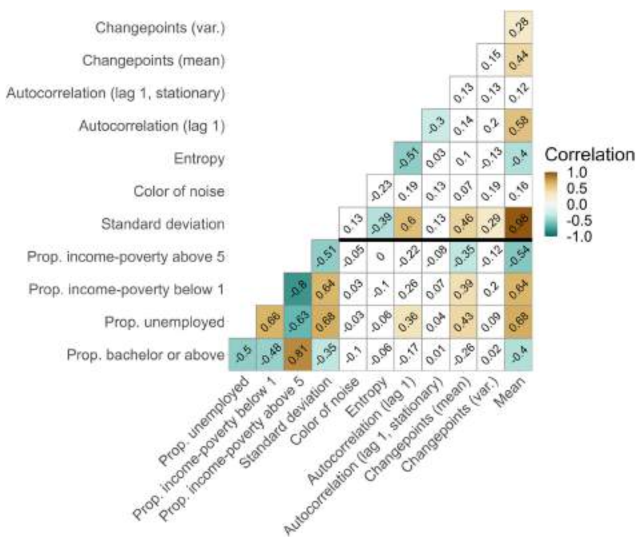
levels of unpredictability. We believe that these are examples of the autocorrelation and entropy picking up systematic changes in variability in the data. A high standard deviation and large numbers of changepoints in mean and variance indicate the magnitude of variability in the data. This variability may or may not be predictable. Thus, combinations of statistics indicating variability (e.g., standard deviation) and statistics indicating regularity in patterns (e.g., autocorrelation) can provide insights that these statistics on their own could not.

We now move on to describe patterns in correlations which include the NYC survey measures of unemployment, poverty, and educational attainment. Here too, patterns of correlations are highly consistent across 5-year batches. Among NYC survey variables we observe positive correlations between variables indicating low SES (i.e., unemployment and an income-to-poverty ratio below 1). We also observed positive correlations between variables indicating high SES (i.e., educational attainment and income-to-poverty ratio above 5). These two groups of variables in turn are negatively correlated with each other. In relation to other variables, we find that mean levels in assault rates are positively correlated with unemployment and poverty (i.e., income-to-poverty ratio below 1). This means unemployed individuals and individuals in poverty tend to live in PUMAs which exhibit higher levels of assault rates. Conversely, higher levels in assault rates are negatively correlated with educational attainment and wealth (i.e., income-to-poverty ratio above 5). This means individuals with a bachelor's degree or higher and wealthy individuals tend to live in PUMAs which exhibit lower levels of assault rates. Of our unpredictability statistics, only the standard deviation and changepoints in mean correlate with the NYC survey variables consistently. The direction of the association can be fully derived from the correlations with the mean levels in assault rates. Between 2012 and 2016 (Figure 6), as well as 2016 and 2020 (Figure 7), we observe positive correlations between the autocorrelation and unemployment and a negative correlation between the autocorrelation and educational attainment (only between 2012 and 2016). These correlations indicate that higher levels of unpredictability are associated with higher unemployment and lower educational attainment. As noted earlier, in these cases the autocorrelation likely picks up on systematic changes in variability in the data. Thus, higher levels of systematic changes in variability are associated with higher unemployment and lower educational attainment.

Associations Between Individual Unpredictability Statistics and NYC Survey Variables

Lastly, we explored to what extent bivariate associations between NYC survey variables and unpredictability statistics change when controlling for the mean levels in assault rates. In cases where we included the standard deviation as a predictor, we removed the mean as a covariate to avoid multicollinearity. Figures A4.4–A4.6 in the online supplemental material 6 depict the marginal effects and HDPI for each unpredictability statistic as a predictor and each NYC survey variable as an outcome. Not surprisingly, the standard

Figure 7
Spearman's Rank Correlations Between Unpredictability Statistics (Computed From NYC Assault Rates) and NYC Survey Variables Across PUMAs for Data Collected Between 2016 and 2020



Note. Each cell corresponds to one Spearman's correlation coefficient between the variables indicated by the row and column. Significant correlations ($\alpha < .005$) are indicated by colored cells. The intensity of the color is proportional to the strength of the correlation coefficient. The black horizontal line separates coefficients between unpredictability statistics only (above the line) and coefficients that include NYC survey variables (below the line). NYC = New York City; PUMAs = public use microdata areas; var. = variance; Prop. = proportion. See the online article for the color version of this figure.

deviation is always positively associated with unemployment and poverty, and negatively associated with educational attainment and wealth. Except for two instances all other marginal effects are zero when controlling for the mean. Between 2006 and 2010, the number of changepoints in mean is positively associated with poverty (income-poverty ratio below 1) (marginal effect = 0.03, 95% HDPI [0.01, 0.05]) and negatively with wealth (income-poverty ratio above 5) (marginal effect = −0.05, 95% HDPI [−0.08, −0.019]). Thus, after controlling for mean levels in assault rates, a 1 *SD* increase in the number of changepoints in mean is associated with the following two outcomes: a 3% increase in the proportion of individuals living in poverty and a 5% decrease in the proportion of individuals who are wealthy.

Discussion

We have developed a framework that organizes different notions of environmental properties in unambiguous terms: as statistical definitions. To showcase our framework, we have presented a case study on "unpredictability" using publicly available crime data from NYC. Additionally, we have explored associations between different statistical definitions of unpredictability and socioeconomic outcomes in NYC survey data.

We have learned three lessons from our case study. First, we have shown that the ranking of regions within NYC on unpredictability depends both on the statistical definition used and the spatial scale (borough vs. neighborhood) for which unpredictability is computed. Concretely, this means that answers to research questions—such as how unpredictability affects health, development, or cognition—may differ depending on one's statistical definition and spatial focus (see Glas et al., 2019).

Second, we found that different unpredictability statistics were not only associated to different degrees but in some instances, the sign of the association was even negative. For example, we observed negative associations between statistics quantifying variability and irregularity, implying that some regions are highly variable and regular. Identifying these dynamics matters because the conjunction of different unpredictability statistics may create specific adaptive challenges for individuals. For example, if the world is variable yet regular individuals know what environmental states to expect in the future. Similarly, low variability paired with irregularity may give individuals an idea of the range of environments to expect. However, if the world is both variable and irregular it may be difficult for individuals to predict their environment.

Third, we observed associations between mean levels of crime and socioeconomic outcomes, but not between unpredictability and socioeconomic outcomes. We leave explanation of this pattern for a future study. What matters, for our purposes, is that we have used a rigorous approach to studying how environmental variables, such as mean levels of crime and unpredictability in crime rates over time, are associated with relevant outcomes. Of course, these associations may be different in other data sets. Our case study provides a clear roadmap for how these and other associations could be explored using the framework we have presented.

Table 5
Ranking of Boroughs Based on Different Statistical Definitions of Unpredictability

	Statistics at borough-level					
Rank	*SD*	Color of noise	Entropy	Autocorrelation	Changepoints (*M*)	Changepoints (variance)
First	The Bronx	Brooklyn	Queens	Staten Island	The Bronx	The Bronx
Second	Brooklyn	The Bronx	The Bronx	The Bronx	Manhattan	Staten Island
Third	Staten Island	Staten Island	Manhattan	Manhattan	Queens	Brooklyn
Fourth	Manhattan	Manhattan	Staten Island	Queens	Brooklyn	Queens
Fifth	Queens	Queens	Brooklyn	Brooklyn	Staten Island	Manhattan

Note. Each column corresponds to a different statistical definition. Statistics are computed at the borough level. For each statistic, rows indicate each borough's rank in unpredictability, where 1 ranks *highest* and 5 *lowest*. See the online article for the color version of this table.

Table 6
Ranking of Boroughs Based on Different Statistical Definitions of Unpredictability

	Averaged statistics at PUMA level					
Rank	*SD*	Color of noise	Entropy	Autocorrelation	Changepoints (*M*)	Changepoints (variance)
First	The Bronx	Manhattan	Queens	Queens	The Bronx	Brooklyn
Second	Brooklyn	The Bronx	Staten Island	Staten Island	Brooklyn	The Bronx
Third	Manhattan	Brooklyn	The Bronx	The Bronx	Manhattan	Queens
Fourth	Queens	Queens	Brooklyn	Manhattan	Queens	Manhattan
Fifth	Staten Island	Staten Island	Manhattan	Brooklyn	Staten Island	Staten Island

Note. Each column corresponds to a different statistical definition. Statistics are computed at the PUMA level and averaged for each borough. For each statistic, rows indicate each borough's rank in unpredictability, where 1 ranks *highest* and 5 *lowest*. PUMAs = public use microdata areas. See the online article for the color version of this table.

ONLINE COLOR

Our main contribution is to address a broader problem in conceptualization and measurement of environmental change: ambiguous constructs and research questions are compatible with multiple statistical definitions, which may imply qualitatively different results. The field currently lacks a shared framework that affords systematic comparison of these different definitions. We have provided a clear framework with the aim to provide a structured space for conceptualizing, developing, and comparing theories and empirical studies—attempting, in the words of Virginia Woolf, to create shape amidst chaos (Woolf, 1927). The statistical definitions in our framework serve as building blocks that clarify connections between empirical studies, facilitating cumulative science.

Though the research question determines which notions of environmental change are appropriate in a particular study, we provide some guidance. First, we recommend only doing confirmatory studies if the construct definition is precise enough to make specific predictions about specific environmental statistics. Additionally, we recommend to preregister such studies. In cases where the definition of a key construct is ambiguous, we recommend acknowledging this limitation and using an exploratory approach (Frankenhuis et al., 2023). Of course, such studies may be preregistered as well. It is perfectly fine to explore multiple statistical definitions, if done transparently. Such studies may inform future confirmatory studies. For example, based on our case study, we may consider exploring how combinations of two statistics (e.g., autocorrelation and number of changepoints in variance) are associated with socioeconomic outcomes in a sample of regions across cities in the United States. Thus, our framework can contribute to all parts of the research cycle: it fosters conceptual development, which informs empirical research, which in turn refines conceptual development. Using our framework in this cycle can thus increase theoretical clarity and precision, generating knowledge and cumulative science.

The Future of Environmental Statistics

Our case study has illustrated the merits of having a shared framework for quantifying environmental statistics in psychological research. We highlight four future directions for our framework.

Linking Individual- and Environment-Level Data

As already noted, our framework can be applied to individual-level data, as well as environment-level data (e.g., crime records); that is, data about the environment of an individual or a group of people. In our case study of "unpredictability," we have shown how to compute environmental statistics for environment-level data. We explored associations between unpredictability statistics and NYC survey variables. Similarly, environmental statistics computed across regions in which individuals have lived may be associated with individual-level outcomes of unemployment, poverty, and educational attainment. These environmental statistics would complement measures collected through self-reports and questionnaires. Combining individual- and environment-level data in this way can paint a more precise picture of an individual's lived experience.

For decades personality researchers have acknowledged the importance of both person and environment characteristics for understanding behavior (Fleeson, 2004). Here, environment typically refers to the situational context in which individuals express a specific behavior. The idea is that although personality traits tend to predict long-term behavior, traits, and situations substantially influence short-term actions. Although a large and well-developed body of work has been devoted to characterizing personality traits (e.g., the Big Five), systematic ways for quantifying the situation are less common.

An exception is the DIAMONDS taxonomy. This taxonomy characterizes situations along eight dimensions (Rauthmann et al., 2014). These dimensions quantify how individuals perceive and evaluate specific situations. For example, "D" stands for duty and indicates the extent to which individuals perceive an obligation to act in a certain situation. The DIAMONDS taxonomy also considers the extent to which these different dimensions relate to specific personality traits (e.g., conscientiousness in the case of duty) and to the ecology (e.g., the workplace in the case of duty). Future work may extract environmental statistics in densely sampled data from an individual's surrounding ecology to explore how short-term changes in the ecology are associated with changes in DIAMONDS' dimensions and behaviors. Such data may provide valuable insight into individual differences by exploring how individuals with different personality types are influenced by changing situations.

EMA and ESM offer promising ways to collect environmental data. These methods afford tracking behaviors, emotions, and mood through time. If it is possible and ethical to record the locations of individuals using smartphones, these data could be directly linked with data from the ecology (e.g., distinguishing between the workplace and home). Previous work has used geo-location coding to quantify the extent to which individuals move unpredictably through space ("roaming entropy"; Heller et al., 2020; Saragosa-Harris et al., 2022). Future work may combine such measures of individual exploration with environmental statistics quantifying unpredictability across visited locations. Collecting these types of data across different developmental stages would allow us to explore the consistency of behaviors across the lifespan (Fraley & Roberts, 2005; Roberts & DelVecchio, 2000; Rush et al., 2019). These studies can, for instance, provide insight into when the impact of situational context on behavior is greatest, deepening our understanding of sensitive periods for personality development.

Comparing Perceived and Observed Measures

Relatedly, we may examine how environmental statistics from observed measures, for instance of harshness and unpredictability, relate to subjective perceptions of these constructs. Is subjective perception of unpredictability related to unpredictability statistics computed from observed measures, such as crime rates? Do they influence different outcomes in individuals? The existing literature on the association between observed and perceived measures is mixed. For example, perceptions of urban disorder appear to be unrelated to observed disorder and instead driven by other cues, such as demographic composition of the neighborhood and racial biases (Janssen et al., 2022). In the same vein perceptions of unsafety stabilize even when crime rates are going down (Glas, 2021). However, although this association is not perfect, it is generally accepted that fear of crime is related to crime rates (Pearson et al., 2015). Another study showed that effects of perceived and actual crime on life satisfaction vary across demographic characteristics, such as sex, age, income, the presence of children, and whether individuals live in major cities (Ambrey et al., 2014). Some developmental studies suggest that perceived childhood adversity has a greater influence on psychopathology later in life than objective childhood adversity (Baldwin et al., 2021). Assuming two children have been exposed to similar objective levels in harshness, an interesting question is whether lower perceived harshness is related to children's resilience. In that case, tracking environmental statistics of subjective perceptions of experiences may help us understand the development of resilience.

Taken together, these studies suggest nuances in the association between perceived and observed measures and their unique effects on outcomes. Knowing which outcomes are shaped by objective measures and which ones are by subjective measures is important for developing interventions.

Integrating Environmental Statistics and Measurement

Future work may explore combinations of unpredictability statistics as indicators of a formative construct. In our small sample, higher variability was moderately associated with higher regularity in crime rates. It may be interesting to explore a more narrow conceptual definition of unpredictability, defined as high, irregular variability in harshness over space or time. This definition implies frequent and unpredictable changes in harshness levels. Combinations of environmental statistics indicating variability (e.g., changepoints) and regularity (e.g., entropy) may then be used as indicators of unpredictability. However, we would first recommend more exploratory research quantifying unpredictability statistics in diverse and larger data sets.

We provide a few concrete suggestions for using our environmental statistics as indicators of constructs. First, we recommend using a narrow conceptual definition of the construct. Second, the choice of indicators should be guided by theory or prior exploratory research. Next, it is important to consider whether the underlying construct is thought to cause the indicators (e.g., intelligence) or whether the indicators cause the construct (e.g., SES). The former requires a reflective model and the latter a formative model (Bollen & Lennox, 1991; Coltman et al., 2008; Edwards & Bagozzi, 2000). Confirmatory factor analysis and various other latent factor modeling techniques involving structural equation modeling are suitable for reflective constructs. If environmental statistics have been quantified across different time periods (e.g., in 5 years batches as in our case), latent profile analysis is useful for modeling how the construct changes across time. Recently, there is an increasing number of analogous methods for formative constructs. For example, analogous to confirmatory factor analysis, confirmatory composite analysis determines whether the indicators form one or multiple formative constructs (Hubona et al., 2021). Dimensionality reduction techniques, such as principle component analysis, are also suitable to explore formative constructs. Nowadays, structural equation modeling techniques also allow the coupling of composites of indicators and outcomes (Sarstedt & Hwang, 2020). However, irrespective of modeling choice, we need to consider whether and how measurement invariance influences the construct (DeJoseph et al., 2022). After fitting an appropriate measurement model, we still need to assess the reliability of the indicators (e.g., using Cronbach's α, McDonalds' ω), as well as the validity of the measure itself (Flake & Fried, 2020; Hair et al., 2021; Hayes & Coutts, 2020; Hodson, 2021).

A Database of Environmental Statistics

Our framework offers tools to compute environmental statistics and can thus contribute to the development of a "database of environmental statistics" across studies (Frankenhuis et al., 2019). Knowing the values of different environmental statistics for different variables and samples can help to generate new explanations and hypotheses in psychological research. For example, the database could provide the necessary data to test the association between different unpredictability statistics and poverty across different countries.

Quantifying the statistical structure of the environment on shorter timescales has already proven to advance infancy research. For example, a recent study documented the everyday auditory experiences of infants (Warlaumont et al., 2022). The authors found that infants seek out vocal responses using strategies similar to how animals forage for resources. This parallel between vocal exploration and foraging dynamics offers opportunities to generate novel hypotheses about learning in infants. Similarly, another study used head cameras and eye trackers to document infants' everyday visual experiences (Smith et al., 2020). As sensorimotor development progresses, infants' interactions with their visual environment change, granting them access to novel experiences; referred to as "curriculum for learning." The authors hypothesize that infant learning is optimized for the continuously changing visual environment. Moving forward, computing environmental statistics could advance developmental research in similar ways as documenting the early environment has advanced infancy research.

A recently published "ecology-culture data set" is an outstanding example of what a database of such statistics might look like

(Wormley et al., 2022). A similar database for psychological research may fuel future research but also harmonize existing knowledge.

Limitations of Our Case Study and Environmental Statistics in General

In our case study, we compute different statistical definitions of unpredictability in data that exhibit a strong association between the mean and standard deviation. We do not know whether our findings generalize to different types of data that do not have this property. Moreover, our sample of 55 PUMAs is rather small. It would be worthwhile to expand our current study to explore the association between different statistical definitions of unpredictability and survey outcomes across multiple cities in the United States.

Our statistical formalizations are simple and do not cover more advanced time series modeling approaches. Rather than extracting individual statistics from time series data, as we have, these models can be used to test—rather than explore, as we have done here—hypotheses about how the data change across time. For example, Jebb et al. offer a beginner-friendly tutorial on time series analysis in psychological research and provide pointers to further, more advanced reading (Jebb et al., 2015). Applying time series modeling to developmental research may help us answer research questions that we cannot answer with our current methods. However, successfully applying time series modeling to developmental questions can be challenging. Recent work addressed such challenges in the study of emotion development across the span of minutes, hours, and days (Haslbeck & Ryan, 2021). Future work may explore challenges of applying time series modeling to environmental data sampled over longer time-scales (i.e., weeks, months, and years).

We will refrain from providing guidelines for determining which statistics to use in which conditions and how to interpret their values. Should we compute the autocorrelation or entropy? What does an entropy of 0.95 tell us about unpredictability without comparing it against other regions or individuals? The first question should be answered on the grounds of theory; precise research questions, hypotheses, and construct definitions should guide our choices. The second question is empirical, requiring us to compute environmental statistics more regularly and in different contexts (e.g., samples, time-scales, countries). As the number of empirical studies computing environmental statistics increases, we learn more about their values in different contexts. However, as noted, the extent to which environmental statistics track real patterns, as opposed to noise, depends on the quality of the underlying data. Thus, as with any other empirical work, measurement error is a concern. Generally, we believe that both quantity and quality of data may ultimately be improved with "team science." If we pool resources and research efforts, we may be able to conduct fewer studies with larger quantity and higher quality data. With more calls for collaborative science, increasing incentives to share data collected from participants, and the availability of public environment-level data (e.g., crime records), we see exciting opportunities ahead for systematically quantifying environmental stability and change across the human life course.

References

Ambrey, C. L., Fleming, C. M., & Manning, M. (2014). Perception or reality, what matters most when it comes to crime in your neighbourhood? *Social Indicators Research*, *119*(2), 877–896. https://doi.org/10.1007/s11205-013-0521-6

Baldwin, J. R., Esposti, M. D., & Degli Esposti, M. (2021). Triangulating evidence on the role of perceived versus objective experiences of childhood adversity in psychopathology. *JCPP Advances*, *1*(1), Article e12010. https://doi.org/10.1111/jcv2.12010

Baram, T. Z., Davis, E. P., Obenaus, A., Sandman, C. A., Small, S. L., Solodkin, A., & Stern, H. (2012). Fragmentation and unpredictability of early-life experience in mental disorders. *American Journal of Psychiatry*, *169*(9), 907–915. https://doi.org/10.1176/appi.ajp.2012.11091347

Bernardi, L., & Hutter, I. (2007). The anthropological demography of Europe. *Demographic Research*, *17*(December), 541–566. https://doi.org/10.4054/DemRes.2007.17.18

Blotenberg, I., Schang, L., & Boywitt, D. (2022). Should indicators be correlated? Formative indicators for healthcare quality measurement. *BMJ Open Quality*, *11*(2), Article e001791. https://doi.org/10.1136/bmjoq-2021-001791

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314. https://doi.org/10.1037/0033-2909.110.2.305

Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, *16*(4), 756–766. https://doi.org/10.1177/1745691620969647

Bringmann, L., Elmer, T., & Eronen, M. (2022). Back to basics: The importance of conceptual clarification in psychological science. *Current Directions in Psychological Science*, *31*(4), 340–346. https://doi.org/10.1177/09637214221096485

Brumbach, B. H., Figueredo, A. J., & Ellis, B. J. (2009). Effects of harsh and unpredictable environments in adolescence on development of life history strategies. *Human Nature*, *20*(1), 25–51. https://doi.org/10.1007/s12110-009-9059-3

Burgess, S. C., & Marshall, D. J. (2014). Adaptive parental effects: The importance of estimating environmental predictability and offspring fitness appropriately. *Oikos*, *123*(7), 769–776. https://doi.org/10.1111/oik.01235

Burkner, P. C. (2013). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, *61*(12), 1250–1262. https://doi.org/10.1016/j.jbusres.2008.01.013

de Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., & Hamaker, E. L. (2017). Discrete- versus continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology*, *8*(October), Article 1849. https://doi.org/10.3389/fpsyg.2017.01849

DeJoseph, M. L., Herzberg, M. P., Sifre, R. D., Berry, D., & Thomas, K. M. (2022). Measurement matters: An individual differences examination of family socioeconomic factors, latent dimensions of children's experiences, and resting state functional brain connectivity in the ABCD sample. *Developmental Cognitive Neuroscience*, *53*(November), Article 101043. https://doi.org/10.1016/j.dcn.2021.101043

Eberly, L. A., Julien, H., South, E. C., Venkataraman, A., Nathan, A. S., Anyawu, E. C., Dayoub, E., Groeneveld, P. W., & Khatana, S. A. M. (2022). Association between community-level violent crime and cardiovascular mortality in Chicago: A longitudinal analysis. *Journal of the American Heart Association*, *11*(14), Article e025168. https://doi.org/10.1161/JAHA.122.025168

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–174. https://doi.org/10.1037/1082-989X.5.2.155

Ehlman, S. M., Scherer, U., Bierbach, D., Francisco, F. A., Laskowski, K. L., Krause, J., & Wolf, M. (2023). Leveraging big data to uncover the eco-evolutionary factors shaping behavioural development. *Proceedings of the Royal Society B: Biological Sciences*, *290*(1992), Article 20222115. https://doi.org/10.1098/rspb.2022.2115

Ellis, B. J., Figueredo, A. J., Brumbach, B. H., & Schlomer, G. L. (2009). Fundamental dimensions of environmental risk: The impact of harsh versus unpredictable environments on the evolution and development of life history strategies. *Human Nature*, *20*(2), 204–268. https://doi.org/10.1007/s12110-009-9063-7

Erdogan, E., Ma, S., Beygelzimer, A., & Rish, I. (2005). *Statistical models for unequally spaced time series*. Proceedings of the 2005 SIAM International Conference on Data Mining, SDM (pp. 626–630). https://doi.org/10.1137/1.9781611972757.74

Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, *16*(4), 779–788. https://doi.org/10.1177/1745691620970586

European Commission, Eurostat. (2013). *Handbook on precision requirements and variance estimation for ESS households surveys: 2013 edition*. Publications Office. https://data.europa.eu/doi/10.2785/13579

Farkas, B. C., Baptista, A., Speranza, M., Wyart, V., & Jacquet, P. O. (2022). *Specifying the timescale of early life unpredictability helps explain the development of internalising and externalising behaviours and their link with reproductive effort in adolescence* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/t5avy

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, *13*(2), 83–87. https://doi.org/10.1111/j.0963-7214.2004.00280.x

Fleuren, B. P. I., van Amelsvoort, L. G. P. M., Zijlstra, F. R.H., De Grip, A., & Kant, I. (2018). Handling the reflective-formative measurement conundrum: a practical illustration based on sustainable employability. *Journal of Clinical Epidemiology*, *103*, 71–81. https://doi.org/10.1016/j.jclinepi.2018.07.007

Fraley, R. C., & Roberts, B. W. (2005). Patterns of continuity: A dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychological Review*, *112*(1), 60–74. https://doi.org/10.1037/0033-295X.112.1.60

Frankenhuis, W. E., Nettle, D., & Dall, S. R. X. (2019). A case for environmental statistics of early-life effects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1770), Article 20180110. https://doi.org/10.1098/rstb.2018.0110

Frankenhuis, W. E., Panchanathan, K., & Smaldino, P. E. (2023). Strategic ambiguity in the social sciences. *Social Psychological Bulletin*, *18*, Article e9923. https://doi.org/10.32872/spb.9923

Frankenhuis, W. E., & Tiokhin, L. (2018). Bridging evolutionary biology and developmental psychology: Toward an enduring theoretical infrastructure. *Child Development*, *89*(6), 2303–2306. https://doi.org/10.1111/cdev.13021

Frankenhuis, W. E., & Walasek, N. (2020). Modeling the evolution of sensitive periods. *Developmental Cognitive Neuroscience*, *41*(November), Article 100715. https://doi.org/10.1016/j.dcn.2019.100715

Gee, D. G. (2021). Early adversity and development: Parsing heterogeneity and identifying pathways of risk and resilience. *American Journal of Psychiatry*, *178*(11), 998–1013. https://doi.org/10.1176/appi.ajp.2021.21090944

Glas, I. (2021). Crime is down and so is fear? Analyzing resident perceptions of neighborhood unsafety in Rotterdam, the Netherlands. *European Journal on Criminal Policy and Research*, *29*(1), 27–49. https://doi.org/10.1007/s10610-021-09495-2

Glas, I., Engbersen, G., & Snel, E. (2019). The street level and beyond: The impact of ethnic diversity on neighborhood cohesion and fear of crime among Dutch natives and nonnatives. *Journal of Urban Affairs*, *41*(6), 737–755. https://doi.org/10.1080/07352166.2018.1533379

Granic, I., & Hollenstein, T. (2003). Dynamic systems methods for models of developmental psychopathology. *Development and Psychopathology*, *15*(3), 641–669. https://doi.org/10.1017/S0954579403000324

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). Evaluation of formative measurement models. In J. F. Hair, Jr., G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, & S. Ray (Eds.), *Partial least squares structural equation modeling (PLS-SEM) using R: A workbook* (pp. 91–113). Springer International Publishing. https://doi.org/10.1007/978-3-030-80519-7_5

Hammel, E. A. (2005). Demographic dynamics and kinship in anthropological populations. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(6), 2248–2253. https://doi.org/10.1073/pnas.0409762102

Hartley, C. A. (2022). How do natural environments shape adaptive cognition across the lifespan? *Trends in Cognitive Sciences*, *26*(12), 1029–1030. https://doi.org/10.1016/j.tics.2022.10.002

Haslbeck, J. M. B., & Ryan, O. (2021). Recovering within-person dynamics from psychological time series. *Multivariate Behavioral Research*, *57*(5), 735–766. https://doi.org/10.1080/00273171.2021.1896353

Hatzenbuehler, M. L., Weissman, D. G., Mcketta, S., Lattanner, M. R., Ford, J. V., Barch, D. M., & Mclaughlin, K. A. (2020). *Smaller hippocampal volume among Black and Latinx youth living in high-stigma contexts*. BioRxiv, 2020.10.09.333328.

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Haynes, K., Killick, R., Fearnhead, P., & Eckley, I. (2016). *Changepoint. Np: Methods for nonparametric changepoint detection*. R Package Version 0.0, 2 [Computer software].

Heller, A. S., Shi, T. C., Ezie, C. E. C., Reneau, T. R., Baez, L. M., Gibbons, C. J., & Hartley, C. A. (2020). Association between real-world experiential diversity and positive affect relates to hippocampal–striatal functional connectivity. *Nature Neuroscience*, *23*(7), 800–804. https://doi.org/10.1038/s41593-020-0636-4

Hodson, G. (2021). Construct jangle or construct mangle? Thinking straight about (nonredundant) psychological constructs. *Journal of Theoretical Social Psychology*, *5*(4), 576–590. https://doi.org/10.1002/jts5.120

Hoffman, L. (2008). Multilevel models for examining individual differences in within-person variation and covariation over time. *Multivariate Behavioral Research*, *42*(4), 609–629. https://doi.org/10.1080/00273170701710072

Hubona, G. S., Schuberth, F., & Henseler, J. (2021). A clarification of confirmatory composite analysis (CCA). *International Journal of Information Management*, *61*, Article 102399. https://doi.org/10.1016/j.ijinfomgt.2021.102399

Janssen, H. J., Oberwittler, D., & Gerstner, D. (2022). Dissecting disorder perceptions: Neighborhood structure and the moderating role of interethnic contact and xenophobic attitudes. *International Criminal Justice Review*, *32*(4), 429–456. https://doi.org/10.1177/1057567719896020

Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: Examining and forecasting change. *Frontiers in Psychology*, *6*(June), Article 727. https://doi.org/10.3389/fpsyg.2015.00727

Kievit, R. A., McCormick, E. M., Fuhrmann, D., Deserno, M. K., & Orben, A. (2021). Using large, publicly available data sets to study adolescent development: Opportunities and challenges. *Current Opinion in Psychology*, *44*, 303–308. https://doi.org/10.1016/j.copsyc.2021.10.003

Killick, R., & Eckley, I. A. (2014). Changepoint: An R Package for changepoint analysis. *Journal of Statistical Software*, *58*(3), 1–19. https://doi.org/10.18637/jss.v058.i03

Lepot, M., Aubin, J.-B., & Clemens, F. H. L. R. (2017). Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment. *Water*, *9*(10), Article 796. https://doi.org/10.3390/w9100796

Li, Z., Liu, S., Hartman, S., & Belsky, J. (2018). Interactive effects of early-life income harshness and unpredictability on children's socioemotional and academic functioning in kindergarten and adolescence. *Developmental Psychology*, *54*(11), 2101–2112. https://doi.org/10.1037/dev0000601

Li, Z., Sturge-Apple, M. L., Platts, C. R., & Davies, P. T. (2022). Testing different sources of environmental unpredictability on adolescent functioning: Ancestral cue versus statistical learning and the role of temperament. *Journal of Child Psychology and Psychiatry*, *64*(3), 437–448. https://doi.org/10.1111/jcpp.13714

Luby, J. L., Belden, A., Harms, M. P., Tillman, R., & Barch, D. M. (2016). Preschool is a sensitive period for the influence of maternal support on the trajectory of hippocampal development. *Proceedings of the National Academy of Sciences*, *113*(20), 5742–5747. https://doi.org/10.1073/pnas.1601443113

Marshall, D. J., & Burgess, S. C. (2015). Deconstructing environmental predictability: Seasonality, environmental colour and the biogeography of marine life histories. *Ecology Letters*, *18*(2), 174–181. https://doi.org/10.1111/ele.12402

Meier, S. T. (2023). Editorial: Persistence of measurement problems in psychological research. *Frontiers in Psychology*, *14*, Article 1132185. https://doi.org/10.3389/fpsyg.2023.1132185

Miller, G. E., Chen, E., Armstrong, C. C., Carroll, A. L., Ozturk, S., Rydland, K. J., Brody, G. H., Parrish, T. B., & Nusslock, R. (2018). Functional connectivity in central executive network protects youth against cardiometabolic risks linked with neighborhood violence. *Proceedings of the National Academy of Sciences*, *115*(47), 12063–12068. https://doi.org/10.1073/pnas.1810067115

Nordgren, R., Hedeker, D., Dunton, G., & Yang, C.-H. (2020). Extending the mixed-effects model to consider within-subject variance for Ecological Momentary Assessment data. *Statistics in Medicine*, *39*(5), 577–590. https://doi.org/10.1002/sim.8429

O'Brien, D. T., Sampson, R. J., & Winship, C. (2015). Ecometrics in the age of big data: Measuring and assessing "broken windows" using large-scale administrative records. *Sociological Methodology*, *45*(1), 101–147. https://doi.org/10.1177/0081175015576601

Oh, C., Han, S., & Jeong, J. (2020). Time-series data augmentation based on interpolation. *Procedia Computer Science*, *175*, 64–71. https://doi.org/10.1016/j.procs.2020.07.012

Olthof, M., Hasselman, F., Strunk, G., van Rooij, M., Aas, B., Helmich, M. A., Schiepek, G., & Lichtwarck-Aschoff, A. (2020). Critical fluctuations as an early-warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science*, *8*(1), 25–35. https://doi.org/10.1177/2167702619865969

Pavía-Miralles, J. M. (2010). A survey of methods to interpolate, distribute and extrapolate time series. *Journal of Service Science and Management*, *3*(4), 449–463. https://doi.org/10.4236/jssm.2010.34051

Pearson, A. L., Breetzke, G., & Ivory, V. (2015). The effect of neighborhood recorded crime on fear: Does neighborhood social context matter? *American Journal of Community Psychology*, *56*(1–2), 170–179. https://doi.org/10.1007/s10464-015-9741-z

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(6), 2297–2301. https://doi.org/10.1073/pnas.88.6.2297

Platt, J. R. (1964). Strong inference. *Science*, *146*(3642), 347–353. https://doi.org/10.1126/science.146.3642.347

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Raab, H. A., Foord, C., Ligneul, R., & Hartley, C. A. (2022). Developmental shifts in computations used to detect environmental controllability. *PLoS Computational Biology*, *18*(6), Article e1010120. https://doi.org/10.1371/journal.pcbi.1010120

Ram, N., & Gerstorf, D. (2009). Time-structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. *Psychology and Aging*, *24*(4), 778–791. https://doi.org/10.1037/a0017915

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The situational eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*(4), 677–718. https://doi.org/10.1037/a0037250

Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, *278*(6), H2039–H2049. https://doi.org/10.1152/ajpheart.2000.278.6.H2039

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*(1), 3–25. https://doi.org/10.1037/0033-2909.126.1.3

Roisman, G. I. (2021). Editorial: A vision of a fair and efficient, diverse and inclusive, cumulative science of child development in the best and worst of times. *Child Development*, *92*(2), 451–465. https://doi.org/10.1111/cdev.13538

Ruokolainen, L., Lindén, A., Kaitala, V., & Fowler, M. S. (2009). Ecological and evolutionary dynamics under coloured environmental variation. *Trends in Ecology and Evolution*, *24*(10), 555–563. https://doi.org/10.1016/j.tree.2009.04.009

Rush, J., Rast, P., Almeida, D. M., & Hofer, S. M. (2019). Modeling long-term changes in daily within-person associations: An application of multilevel SEM. *Psychology and Aging*, *34*(2), 163–176. https://doi.org/10.1037/pag0000331

Saragosa-Harris, N. M., Cohen, A. O., Reneau, T. R., Villano, W. J., Heller, A. S., & Hartley, C. A. (2022). Real-world exploration increases across adolescence and relates to affect, risk taking, and social connectivity. *Psychological Science*, *33*(10), 1664–1679. https://doi.org/10.1177/09567976221102070

Sarstedt, M., & Hwang, H. (2020). Advances in composite-based structural equation modeling. *Behaviormetrika*, *47*(1), 213–217. https://doi.org/10.1007/s41237-020-00105-9

Smaldino, P. E. (2020). How to translate a verbal theory into a formal model. *Social Psychology*, *51*(4), 207–218. https://doi.org/10.1027/1864-9335/a000425

Smith, J. E., Smith, J. E., & Pinter-Wollman, N. (2020). Observing the unwatchable: Integrating automated sensing, naturalistic observations and animal social network analysis in the age of big data. *Journal of Animal Ecology*, *90*(1), 62–75. https://doi.org/10.1111/1365-2656.13362

Sundquist, K., Theobald, H., Yang, M., Li, X., Johansson, S.-E., & Sundquist, J. (2006). Neighborhood violent crime and unemployment increase the risk of coronary heart disease: A multilevel study in an urban setting. *Social Science & Medicine*, *62*(8), 2061–2071. https://doi.org/10.1016/j.socscimed.2005.08.051

Trull, T. J., & Ebner-Priemer, U. W. (2009). Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: Introduction to the special section. *Psychological Assessment*, *21*(4), 457–462. https://doi.org/10.1037/a0017653

Ugarte, E., & Hastings, P. D. (2023). Assessing unpredictability in caregiver–child relationships: Insights from theoretical and empirical perspectives. *Development and Psychopathology*, 1–20. Advance online publication. https://doi.org/10.1017/S0954579423000305

Vasseur, D. A., & Yodzis, P. (2004). The color of environmental noise. *Ecology*, *85*(4), 1146–1152. https://doi.org/10.1890/02-3122

Vinton, A. C., Gascoigne, S. J. L., Sepil, I., & Salguero-Gómez, R. (2022). Plasticity's role in adaptive evolution depends on environmental change components. *Trends in Ecology & Evolution*, *37*(12), 1067–1078. https://doi.org/10.1016/j.tree.2022.08.008

Walker, K., & Herman, M. (2022). *tidycensus: Load US census boundary and attribute data as "tidyverse" and 'sf'-ready data frames* (R package Version 1.3) [Computer software]. https://walker-data.com/tidycensus/

Warlaumont, A. S., Sobowale, K., & Fausey, C. M. (2022). Daylong mobile audio recordings reveal multitimescale dynamics in infants' vocal productions and auditory experiences. *Current Directions in Psychological Science*, *31*(1), 12–19. https://doi.org/10.1177/09637214211058166

Werchan, D. M., Brandes-Aitken, A., & Brito, N. H. (2022). Signal in the noise: Dimensions of predictability in the home auditory environment are associated with neurobehavioral measures of early infant sustained attention. *Developmental Psychobiology*, *64*(7), Article e22325. https://doi.org/10.1002/dev.22325

Woolf, V. (1927). *To the lighthouse* (P. L. Caughie, N. Hayward, M. Hussey, P. Shillingsburg, & G. K. Thiruvathukal, Eds.). Harcourt, Brace & Company.

Wormley, A. S., Kwon, J. Y., Barlev, M., & Varnum, M. E. W. (2022). The ecology-culture dataset: A new resource for investigating cultural variation. *Scientific Data*, *9*(1), Article 615. https://doi.org/10.1038/s41597-022-01738-z

Young, E. S., Frankenhuis, W. E., & Ellis, B. J. (2020). Theory and measurement of environmental unpredictability. *Evolution and Human Behavior*, *41*(6), 550–556. https://doi.org/10.1016/j.evolhumbehav.2020.08.006

Received February 3, 2023
Revision received August 21, 2023
Accepted December 18, 2023 ■