Call for a paradigm shift from statistical causal inference to multi-evidence causal investigation

Author: James B. Grace[*†‡]

Affiliations:

[1]Department of Biology, Concordia University; Montreal, QC H4B 1R6, Canada.

*Corresponding author. Email: james.grace@concordia.ca

†Present Address: 610 Colonial Drive, Lafayette, Louisiana, 70506 USA

Abstract:

Explicit discussions of causal methods have long fallen into the domain of statistics. Scientists have instead pursued mechanistic knowledge as an alternative approach to causal understanding. In the past two decades, a body of literature has developed that constitutes a statistical causal inference paradigm based on restrictive assumptions that fail to respect mechanistic knowledge. Recent evaluations have shown this paradigm to be incomplete and insufficient, leading to a call for its replacement by an expanded multi-evidence paradigm capable of considering mechanistic evidence and building causal knowledge across studies. Methods for mechanistic causal inference have now been described and are illustrated here, making clear the strong case for scientists to adopt a multi-evidence paradigm.

A core aspiration of science is to determine how systems work, which implies the challenge of developing a causal understanding. Explicit discussions of causal methods have long emphasized a statistical perspective *(1*). Scientists, in contrast, have historically pursued mechanistic knowledge as a path to causal understanding (*2*, *3*, *4*, *5*, *6*, *7*, *8*, *9*, *10*) while using terms such as "effects," "responses," "drivers," and "influences" to describe their findings. Scientists' knowledge of the structures and processes that contribute to causal manifestations (the direct mechanistic evidence) is not considered in statistical causal inference methodology, which focuses on the analysis of correlations. This failure to recognize mechanistic knowledge as a form of causal knowledge is encoded in the foundational precepts for what is now recognized as the causal inference paradigm, or more appropriately "the statistical causal inference paradigm", (*11*, *12*), which has emerged from the fields of statistics (*13*, *14*, *15*) and computer science (*16*). Here I will simply refer to it as the *statistical paradigm* for brevity. This collection of ideas and techniques initially gained momentum in epidemiology and the social sciences but is now being introduced into many other fields, including ecology (*17*, *18*, *19*, *20*, *21*). Concerns about the statistical paradigm's restrictive perspective have been increasingly expressed in recent years (*22*, *23*, *24*, *25*, *26*, *27*). The recognized insufficiencies of the statistical paradigm have now led to an expanded, multi-evidence alternative paradigm that considers mechanistic information (*23*). The multi-evidence paradigm emphasizes the long-term goal of building causal knowledge using all available evidence and recognizes the possibility of mechanistic causal inference (*22*, *25*). Here I describe the case for a paradigm shift that better serves the needs of science.

An example of mechanistic causal inference

I begin this presentation by illustrating mechanistic causal inference in order to establish perspective and to introduce necessary terminology. I first draw from a finding published in the Proceeding of the National Academy of Sciences (Fig. 1A). The authors of that study used camera traps and fecal nutrient analyses in African savannas to determine the degree to which variations in the size of herbivorous animals (Fig. 1B) influence the ratios of mineral nutrients they return to the ecosystem. The authors concluded from this study that local environmental conditions that influence the average body size of herbivores foraging in an area (such as predation risk and type of forage) in turn influence the ratios at which nutrients are returned to the ecosystem. In essence, the authors (*28*) implied the existence of a causal relationship between animal body size and phosphorus retention based on mechanistic knowledge. As will be explained below, the observed correlation does not qualify as a traditional causal effect since the data and analysis techniques fail to meet the necessary requirements for statistical causal inference (*21*).

If the association in Fig. 1A does not qualify as a statistical causal effect, how can we justify a causal interpretation? Until very recently, foundational principles and operational procedures did not exist for this challenge (*23*). *Causal knowledge analysis* refers to the documentation of mechanistic and other non-statistical evidence supporting causal interpretations. This example used a *causal knowledge diagram* (Fig. 1C) as a device to aid in documenting the evaluation. The causal knowledge diagram for this example outlines a mechanistic explanation for how animal body size can influence phosphorus retention and excretion. Causal knowledge analysis also involves characterization of the expected behavior of the described mechanisms based on the attributes of the structures and processes making up the mechanistic machinery.

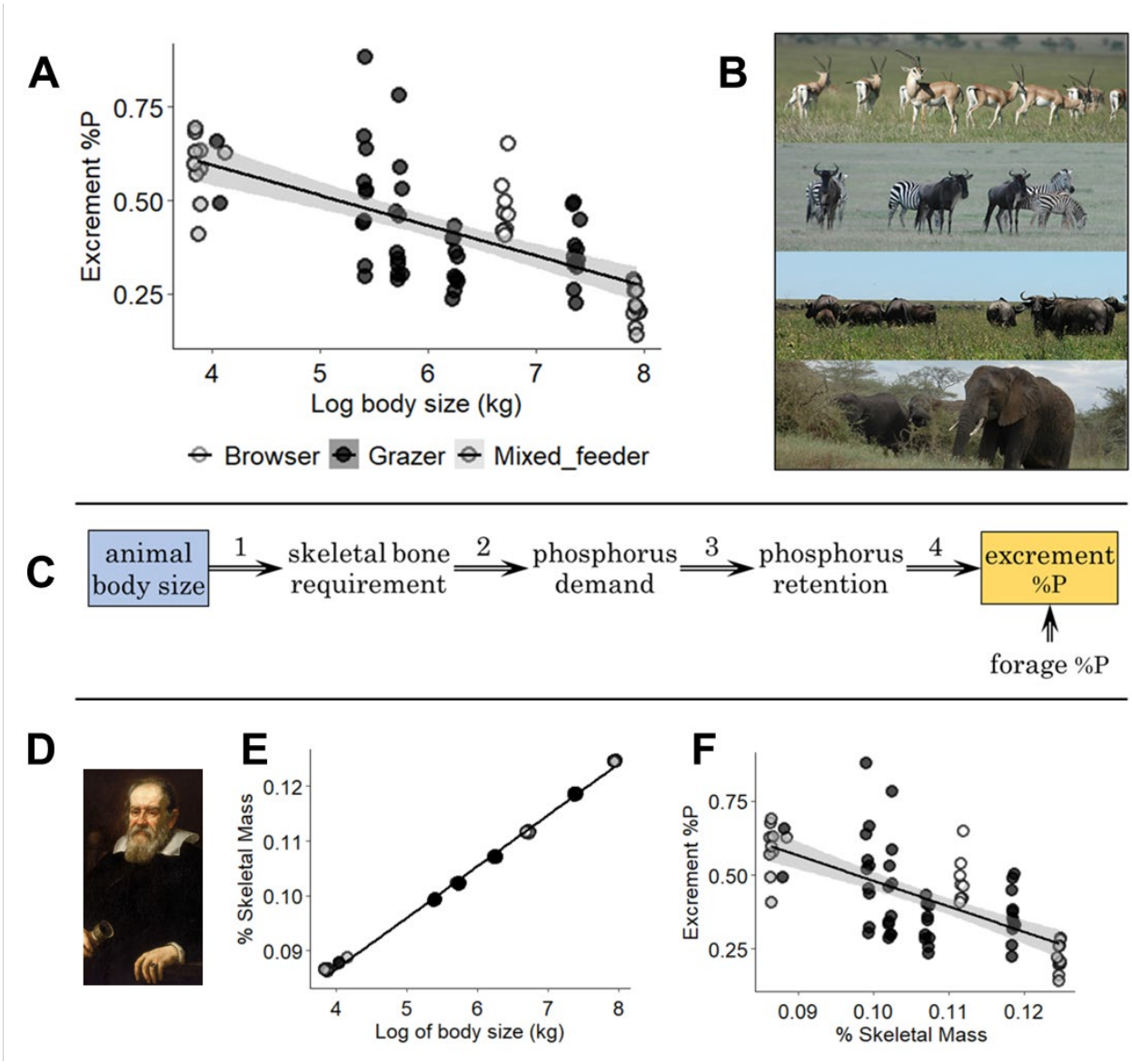


Fig. 1. Illustration of causal knowledge analysis and mechanistic causal determination. (A) Reported relationship between the phosphorus content of feces as a function of animal body size for herbivores in African savannas (*28*). (B) Illustration of African herbivores of differing body sizes. (C) Causal knowledge diagram documenting a chain of biophysical and physiological processes whereby body size influences the phosphorus content of their excrement. (D) Picture of Galileo Galelei who theorized in the year 1638 that larger animals would have to devote a greater fraction of their body mass to skeletal material due to principles of biophysics. (E) The relationship between skeletal mass and body size for herbivores studied by LeRoux et al. (*28*) projected from established allometry relations (*29*). (F) Phosphorus retention fraction as a function of estimated skeletal mass fraction.

The underlying mechanism driving the observed relationship is well known and understood. Galileo (Fig. 1D) proposed back in the year 1638 (*29*, *30*) that larger animals would have to allocate a greater fraction of their body mass to skeletal materials due to biophysical demands and constraints. Since bone strength is proportional to cross-sectional area (a square function) and body mass is proportional to its volume (a cube function), larger animals will require a greater proportional investment in skeletal materials to withstand the physical stresses they experience. Subsequent global analyses (*31*) have led to a general equation that permits us to estimate the % skeletal mass for the animals observed in the study by LeRoux et al. (*28*) (Fig. 1E). Skeletal material is primarily composed of phosphorus compounds, resulting in differential P retention and a decrease in P excretion by larger animals (Fig. 1F). Grace (*23*) provides a summary of the (a) sufficiency, (b) reliability, (c) exactness, and (d) generality of the mechanism described obtained from an in-depth causal knowledge analysis. A key conclusion from that analysis is, "Strong transportability for the discussed mechanism has been demonstrated for a wide range of animals and situations by (*32*, *33*, *34*). This confirms the general principle that biophysical constraints on the traits of organisms provide highly transportable causal knowledge." Scientists may recognize that causal knowledge analysis is a formalization of the process scientists have long relied upon informally to support causal interpretations, though with a deeper consideration of the mechanistic basis of causation and the causal methods literature.

Understanding the causal methods literature through the lens of paradigm evolution

It is my opinion that the literature on statistical causal inference is opaque, confusing, and difficult for scientists to fully understand because it has evolved over time from multiple sources and continues to evolve continuously. To aid scientists in understanding this literature and the broader topic of causal methods, in this paper I use the paradigm evolution lifecycle to provide a high-level and temporally-dynamic perspective (Fig. 2). Paradigms generally form over time around a set of foundational (core) ideas that set standards and norms but also can set limits and boundaries (*35*). Kuhn (*36*) has said that paradigms often possess appeal as new solutions to long-standing problems and individuals often adopt paradigms based on their appealing features without critical awareness of foundational assumptions and alternative views (referred to as paradigm blindness; *35*). The statistical causal inference paradigm has developed from an expansion of historical ideas from the field of statistics elaborated upon in the 1980's (*27*). During the past 20 years, the consistent promotion of its core premises along with an expansion of application methods are beginning to lead to the emergence of a self-aware paradigm (*11, 12*).

Reference to the common stages in the paradigm life cycle (Fig. 2) permits us to develop a meta-view of the historical and ongoing evolution of the causal methodology literature. In the sections that follow I first summarize the current foundational precepts of the statistical paradigm and then present a synopsis of how it is often applied in practice (Stage 1 in Fig. 2). While criticisms of the statistical paradigm will be treated in a separate section (Fig. 2, Stage 2), a persistent theme in my presentation is that the statistical roots of the statistical paradigm result in an approach that is incomplete and insufficient. Paradigm stretching (Fig. 2, Stage 3), where advocates for an existing paradigm attempt to address criticisms while adhering to core premises, will be briefly discussed. That will be followed by an overview of an expanded and inclusive multi-evidence paradigm that aspires to better support scientific causal investigations (Fig. 2, Stage 4). The final section will briefly describe some of the opportunities and challenges that will accompany progression from Stage 4 to Stage 5 (which this paper seeks to facilitate). The

recognition that mechanistic knowledge is a form of causal information greatly expands our capacity to support causal conclusions and has the potential to better position mechanistic scientists as essential players in discussions of causal methodology.

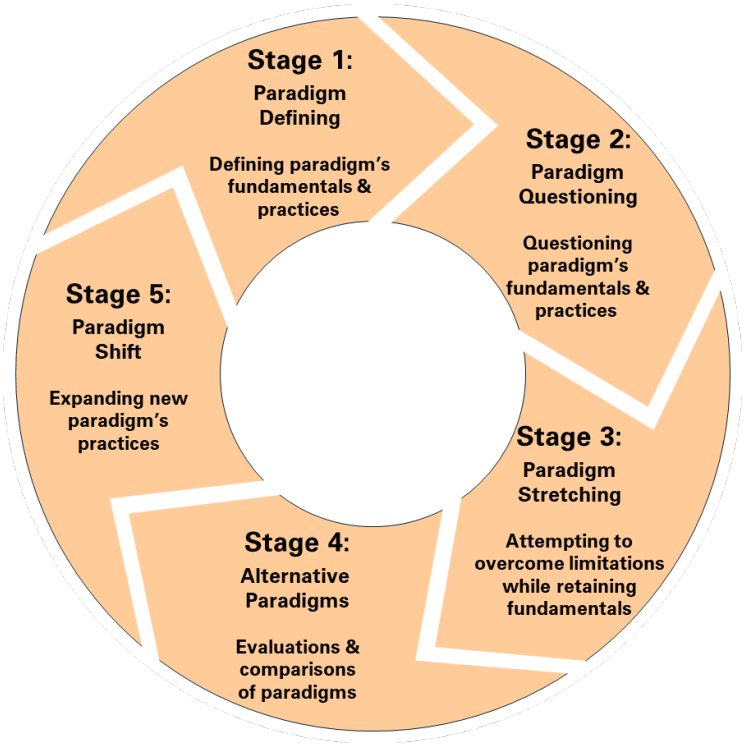


Fig. 2. Representation of the stages of development of a scientific paradigm (adapted from *35*).

Stage 1, part A. The statistical causal inference paradigm in theory

The statistical paradigm as currently described is a fusion developed from three sources. The foundational conceptualization, which establishes the most often used basis for the evidential requirements and restrictions, comes from the statistician D. R. Rubin's *Potential Outcomes Model* (*13*, *37*). The potential outcomes PO model generalizes the idealized properties of randomized manipulative experiments so as to extend them to quasi-experimental settings where manipulation and random assignment do not take place. In doing so, it relies on counterfactual (as if randomized) comparisons. An additional requirement is that causes are expected to be hypothetically manipulable. Throughout the statistical paradigm literature, the term "causal effects" typically refer to counterfactual (*cf*) causal effects (*38*). In this paper, *cf* causal effects" is used to refer to the estimands described in this and the next section to avoid confusion with non-counterfactual characterizations of causal relationships.

An important but neglected point is that Rubin and Holland (*14*, *15*) justified an exclusive focus on *cf* causal effects based on the limits of what can be done using statistical causal analysis rather than principles of science. Holland and Rubin (*15*) acknowledged this by saying,

> "Scientists are usually concerned with understanding causal mechanisms. Purely statistical discussions of causality are substantially more limited in scope, because

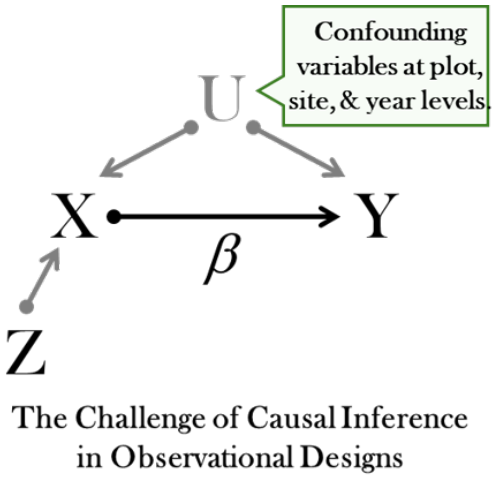
the unique contribution of statistics is to measuring causal effects and not to the understanding of causal mechanisms."

This limited capability permeates throughout the persistent criticisms of the statistical paradigm but is generally not acknowledged in the literature describing the statistical paradigm, which instead refers to itself as the causal inference paradigm, implicitly disregarding mechanistic causal inference.

During the past decade the PO framework has come to be commonly paired with a graphical modeling system developed by the computer scientist J. Pearl referred to as the Structural Causal Model (SCM) (*16*). There was great debate about the merits of the PO and SCM models initially (from approximately 2000 to around 2015). The most notable distinction made was an optimistic portrayal of the potential use of the SCM that implied scientists would be able to make strong assumptions about model structure based on existing causal knowledge and failed to place major emphasis on unresolvable confounding problems. In strong contrast, presentations of the PO model were constant reminders of the threat of confounding and the requirement for complete bias control due to a failure to consider the possibility of existing mechanistic knowledge. Early adoption of the SCM in ecology led to an infiltration of its premises into structural equation modeling practice, while early adoption of the PO model in the social sciences fostered a very critical view of the potential for causal conclusions from observational data. The fusion of methods from the two systems has been promoted in recent years based on work demonstrating their statistical equivalence (*39*, *40*). The graphical modeling system now serves primarily as a user interface for developing strategies for isolating counterfactual effects, but also as a constant reminder of the threat from omitted confounding variables (illustrated in the next section). Finally, a third source contributing to the statistical paradigm is from study designs developed as implementation strategies, such as the use of discontinuities in exposures, analytical methods for exploiting temporal events, and group equalization methods for non-randomized exposures (*20*).

Stage 1, part B. The statistical causal inference paradigm in practice

The statistical paradigm consists of both fundamental premises and an attending set of challenges to be addressed, which are usually described as a set of requirements and assumptions (Fig. 3). In Rubin's description of the PO conceptualization (*37*), he admits that individual causal effects are typically impossible to estimate because they involve unobservable hypothetical potential responses. Average causal effects rely on information from groups of individuals. However, this now creates a need for individuals to be appropriately comparable (*exchangeable*). Aspirationally, the goal is to approximate *cf* causal effects reflective of single-cause all-else-equal interventions.



Challenges:

1. Counterfactual values unobservable.
2. Requires control of all potential sources of confounding, which cannot be guaranteed.
3. Magnitude of bias from uncontrolled confounding unknowable.
4. Success is based on defense of untestable assumptions.

Fig. 3. Representation of the challenge of counterfactual statistical causal inference in observational data (modified from *19*). The theoretical requirement is to obtain a perfectly unbiased estimate of the effect of X on Y, β. Confounding variables U pose a pervasive threat to that goal that must be addressed. Z refers to an alternative method for bias control using instrumental variables.

In practice, the chief concern for analyses based on the statistical paradigm is with the threat posed by omitted confounding variables (U in Fig.3), though there are other expectations as described in the previous section. The theoretical objective is to estimate a perfectly unbiased causal effect of X on Y, β. Historically, a causal method in statistics was seen as one capable of yielding a bias-free causal effect estimate. When X and Y are both influenced by some common cause U, there will be some degree of non-causal association between X and Y resulting from a "backdoor" pathway $X \leftarrow U \rightarrow Y$. When data are collected from multiple plots, study sites, and/or years, each poses different types of confounding possibilities that must be explicitly addressed as a practical requirement. If any confounders are not adjusted for, their influence contributes some unknown amount of bias to the observed association. The logic follows that since the degree of bias will be unknown, the observed estimate may not bear any concordance with the true value β. Acceptance of this *perfection requirement* leads to a fear of the unknown that has permeated the advice given to scientists from causal statistics for over a century. Randomized experiments have the potential ability to eliminate influences from common causes. For this reason, the randomized experimental design has generally been considered to be the one causal method sanctioned by statisticians. However, randomization does not actually guarantee bias control and untestable assumptions remain (*41*).

It was Rubin's potential outcomes conceptualization that began to provide cautious encouragement for drawing statistical causal inferences from non-experimental / observational data by seeking to emulate physically randomized experiments without physical randomization. These efforts have gradually evolved into the statistical paradigm. This paradigm can be thought of as adopting a quasi-experimental approach based on a *perfection-seeking requirement* (*23*). Support for this interpretation comes from Ferraro et al.'s (*17*) description of the causal inference approach as one that, "... exploits experimental or quasi-experimental variation in one or more variables to isolate causal relationships and judges success by the credibility of untestable

assumptions …" This standard requires an explicit and diligent effort to strive to achieve bias control, though without expectation of complete success. By adopting this standard, the causal inference tradition in the field of statistics moved from theory to practice.

A notable feature of the statistical paradigm is its failure to consider direct, non-statistical knowledge of underlying mechanisms. The reason for this is the presumption that one is analyzing a data set and that correlations and conditional correlations are the only admissible evidence. The strength of this presumption is its ability to support universal standards (e.g., standard counterfactual requirements). The weakness is that it dismisses scientists' direct knowledge of mechanisms and thereby transportable causal knowledge.

Stage 2 – Paradigm Questioning: Recognized limitations and calls for a shift to a less restrictive paradigm

Statistical causal inference has long recognized major limitations to its potential use. One recognized limitation is that statistical / counterfactual causal inferences can only be made for the "effects of individual causes", not the "causes of effects". Ferraro and Hanauer (*42*) state the logic as follows:

> "Tackling hidden bias in a study that aims to estimate the effects of a single cause is difficult. Tackling such bias in a study that aims to estimate the effects of multiple causes on a single *Y* (the *causes of an effect*) is, in our opinion, beyond the reach of current theory and data. For example, a study that purports to estimate the causes (determinants) of deforestation would be better viewed as a study that generates hypotheses for future studies of individual causes, rather than a study that credibly estimates causal relationships between myriad variables and deforestation."

It is worth noting that the opinion offered that studying the multiple effects of a cause is "beyond the reach of current theory and data" represents a declaration about existing causal knowledge that is not scientifically defensible but reflects an out-of-hand dismissal of mechanistic causal inference. A second limitation associated with statistical causal inference is its inability to produce transportable causal knowledge that is independent from individual data sets. Without identified mechanistic elements there are no external benchmarks for the transportability of statistical findings.

An authoritative call for adoption of an expanded multi-evidence paradigm came from a 2022 publication by the National Academies of Science, Engineering, and Medicine in a Consensus Report on Causal Methods (*22*), (National Academies Report henceforth). This evaluation of causal methods found the statistical paradigm to be incomplete and insufficient as a framework for causal investigations. One major criticism related to the statistical paradigm's limited focus on *cf* causal effects. Specifically, it was noted that such effects provide estimates of what happened in a sample but do not directly consider the existing mechanistic causal information that determines the continuous and dynamic processes linking causes to outcomes. As a related matter, the statistical paradigm was found to be an incomplete methodology due to its failure to recognize the possibility and desirability of mechanistic causal determination (referred to in the report as "mechanistic causal inference"). Further the National Academies Report did not endorse the statistical paradigm's reliance on a predetermined and universal standard of bias-free causal effect estimates. Instead, the Report endorsed a multi-evidence approach where all

relevant forms of evidence are considered by subject matter experts and there is a capacity for the accumulation of mechanistic causal knowledge.

Prior to the National Academies Report, there has been a persistent call for a less restrictive paradigm for causal investigations. In 2016, a collection of papers in the *International Journal of Epidemiology* (*43*) called for replacing the statistical paradigm with an expanded paradigm. Papers calling for a multi-evidence paradigm by scientists were followed by presentations by statisticians presenting a staunch defense of the fundamental precepts and practices of the statistical paradigm. Recently, a review of the debate and renewed call for relief from limiting restrictions has been published by Schwartz and Prins (*27*) in their book *Causal Inference and the People's Health*. Here I summarize five criticisms of the statistical paradigm: (a) Restricts itself to the computation of isolated counterfactual effects of single causes and as a result limits the questions that can be addressed and the evidence that can be considered. (b) States that the knowledge possessed by subject matter experts can be used to form hypotheses, guide data adjustments, and evaluate assumptions but cannot independently establish causation or replace requirements for causal statistical identification. (c) Points out that causal graphs from the SCM are superficial and insufficient representations of the underlying mechanistic structures and processes. (d) Adopting the statistical paradigm constrains research investigations. (e) The statistical paradigm is exclusive (monistic) and undermines the value of mechanistic knowledge.

Stage 3 – Paradigm stretching:

Paradigm stretching refers to efforts to overcome recognized limitations of what can be accomplished while adhering to fundamental precepts (*35*). With regard to the statistical paradigm, its fundamental precepts include: (a) reliance on the statistical analysis of data associations, (b) its definitional dependence on *cf* causal effects, and in practice (c) an aspiration for assurance of perfectly unbiased estimates. There have been persistent efforts during the past twenty years to expand what can be done under the statistical paradigm. The use of causal graphs along with the premises of the PO model can be seen as something that was needed for the conceptual gelling of the statistical paradigm. In theory, the two systems result in the same requirements, so the primary advance has been one of improved communication with practitioners. An additional advance is to have moved past an era of competing counterfactual models.

The published critiques of the past few years are leading now to an increase in statements about the use of prior knowledge, about mechanisms, and about reliance on multiple sources of information (*21*, *12*). Close examination suggests that these statements are not accompanied by changes in the fundamental precepts of the statistical paradigm, which remains focused on developing universal requirements for the analysis of data associations. It is certainly true that real progress in techniques for conditioning data is being made. However, the literature is becoming increasingly unclear and misleading due to efforts to avoid acknowledging fundamental limitations. Both the frequent glossing over of known limitations and misrepresentation of opposing viewpoints is indicative of the fact that paradigm stretching will not negate the need for an expanded alternative paradigm (i.e., the need for a paradigm shift).

Stage 4 – A multi-evidence paradigm for causal investigations

The multi-evidence causal investigation paradigm (multi-evidence paradigm hereafter) starts from a set of premises reflective of a mechanistic/mechanical view of the world that has gained broad support within scholarly treatments of causality (e.g., *44*, *45*, *46*, *47*, *48*). These premises include: (a) that causal manifestations are determined by the underlying mechanistic structures and processes, (b) causal knowledge can be described in terms of both direct and indirect knowledge of those mechanisms, their properties, and the manifestations they may produce, and (c) a driving goal in science is the aspiration to understand the composition and operation of those mechanisms so as to build causal understanding over a series of investigations. Many kinds of evidence can provide insights into the underlying mechanisms. That said, direct knowledge of mechanisms is often considered the gold standard form of causal knowledge (*49*).

To avoid confusion in this presentation, it must be recognized that as a paradigm, the statistical paradigm is defined as much by what it excludes as by what it allows. Virtually all of the objections that have been presented relate to its restrictions and rigid requirements rather than the specific techniques used. The multi-evidence paradigm is inclusive and recognizes the value of techniques for estimating *cf* causal effects when appropriate. As will be demonstrated, it also allows for the calculation of non-counterfactual causal effects when mechanisms are sufficiently known. All the examples presented below, in addition to the above example of mechanistic causal inference, illustrate various implementations of the multi-evidence paradigm. The example that follows illustrates both the calculation of *cf* causal effects and the non-counterfactual characterization of causal mechanisms in a single study.

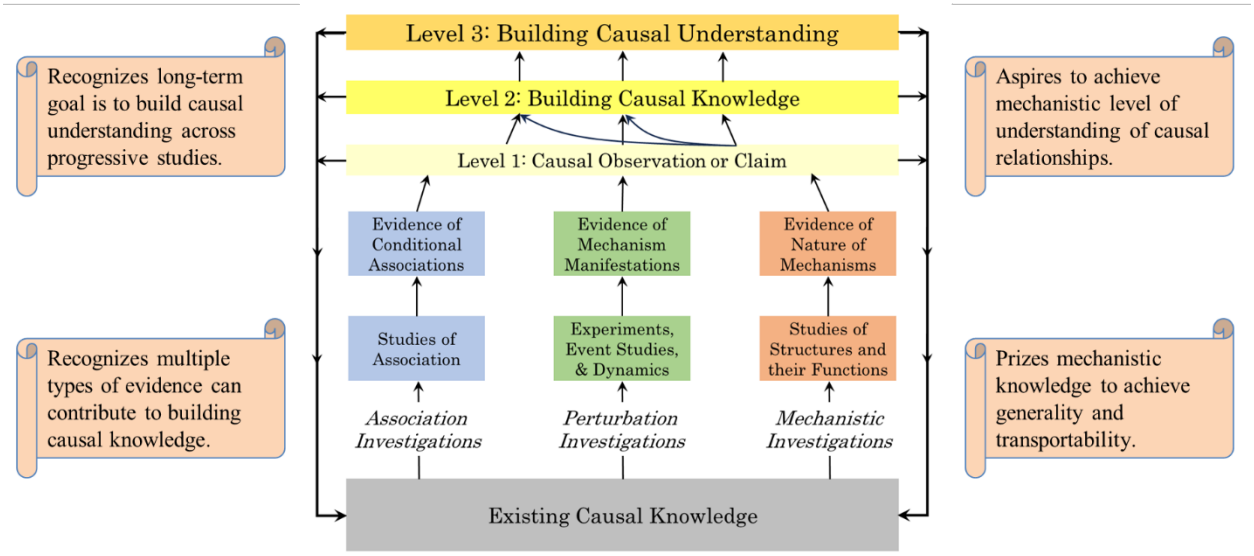


Fig. 4. Representation of a multi-evidence paradigm for causal investigations (modified from *23*, *24*, *25*). Major features of this paradigm that contrast it with the statistical paradigm (Fig. 3) include its: (a) expanded set of aspirations, (b) acceptance of multiple forms of admissible evidence, (c) greater variety of evidential processes, the most notable of which may be causal knowledge analysis, and (d) the recognition that scientific experts are among the deciding entities for both standards and determinations.

Figure 4 depicts key aspects of the multi-evidence paradigm. This representation aims to reflect a scientific perspective on how we might pursue causal knowledge and understanding. It also acknowledges the wide variety of situations and investigative approaches that can contribute to this pursuit. A central element in the multi-evidence paradigm is causal knowledge, which includes not only demonstrated responses to perturbations but also direct and indirect knowledge of the structures and processes that make up causal mechanisms and determine their range of behaviors. Glennan (*44*) describes the mechanical/mechanistic view of causality as one requiring recognition that mechanisms are made of specific elements in particular configurations that determine their behavior. It may be useful to think of the structures and processes of mechanisms as being a form of machinery to avoid confusing the study of statistical mediation relationships with the characterization of actual mechanisms. Ben-Menahem (*50*) conceptualizes the features of specific mechanisms as causal constraints shaping the behaviors of causal mechanisms. Examples presented later in the paper make these points more tangible.

The multi-evidence paradigm is proposed as a world view consistent with scientific investigations that recognize the existence of a great variety of individual situations, which in turn must avoid adopting a narrow view of evidence. It is anticipated that statistical evidence, mechanistic evidence, and other sources of evidence, such as from mechanistic theoretical investigations and the study of temporal dynamics via empirical dynamic modelling (*51*, *52*, *53*), can prove useful. It is expected that in the great span of science there will be other forms of evidence compatible with the foundational assumptions of the multi-evidence paradigm.

Causal knowledge is defined as accumulated evidence regarding: (a) manifestations indicative of an underlying mechanism, (b) some characterization of the underlying mechanism(s) and (c) demonstrated external consistency or transportability with other samples and/or studies. Arrows in Fig. 4 pointing from causal claims to causal knowledge are meant to convey both that multiple forms of evidence contribute, but also that some level of mechanistic knowledge is typically required to constitute "knowledge". The rationale for this requirement is that knowledge suggests some ability to explain why we observe particular behaviors in systems and a basis for anticipating future observations. Finally, causal understanding refers to an accumulation of knowledge allowing for consistent explanations across a wide range of situations. An expanded description of the multi-evidence paradigm can be found at (*23*).

Comparing *cf* causal effects and causal mechanisms

A simple illustration of the versatility of the multi-evidence paradigm comes from a recent examination of the effects of an extreme freeze event on a subtropical tree that has been expanding poleward as global temperatures rise (*54*). Analyses that were conducted included both estimation of *cf* causal effects to answer the question, "What happened?" and characterization of the underlying causal mechanism to answer the question, "What controlled what happened?". Causal effects were estimated using temporal data and event study methods as defined in (*55*). Data were obtained for several regions within the southern United States that experienced differing intensities of freezing during the event (Fig. 5A). Average causal effect estimates were made based on the temporal dynamics of the aerial cover of green mangrove trees obtained from satellite data using counterfactual comparisons consistent with the statistical paradigm (Fig. 5B). Maps of causal effect estimates revealed effect magnitudes that varied widely among locations (Fig. 5C). It is clear from these results that causal effects reflected the

convergence between the spatial distribution of freezing temperatures, the spatial distributions of mangrove trees in the landscape, and local conditional factors such as duration of exposure of meristematic tissues. The only transportable information obtained from this analysis relevant to future freeze events is that mortality will be expected to range between 0% and 100% in the landscape.

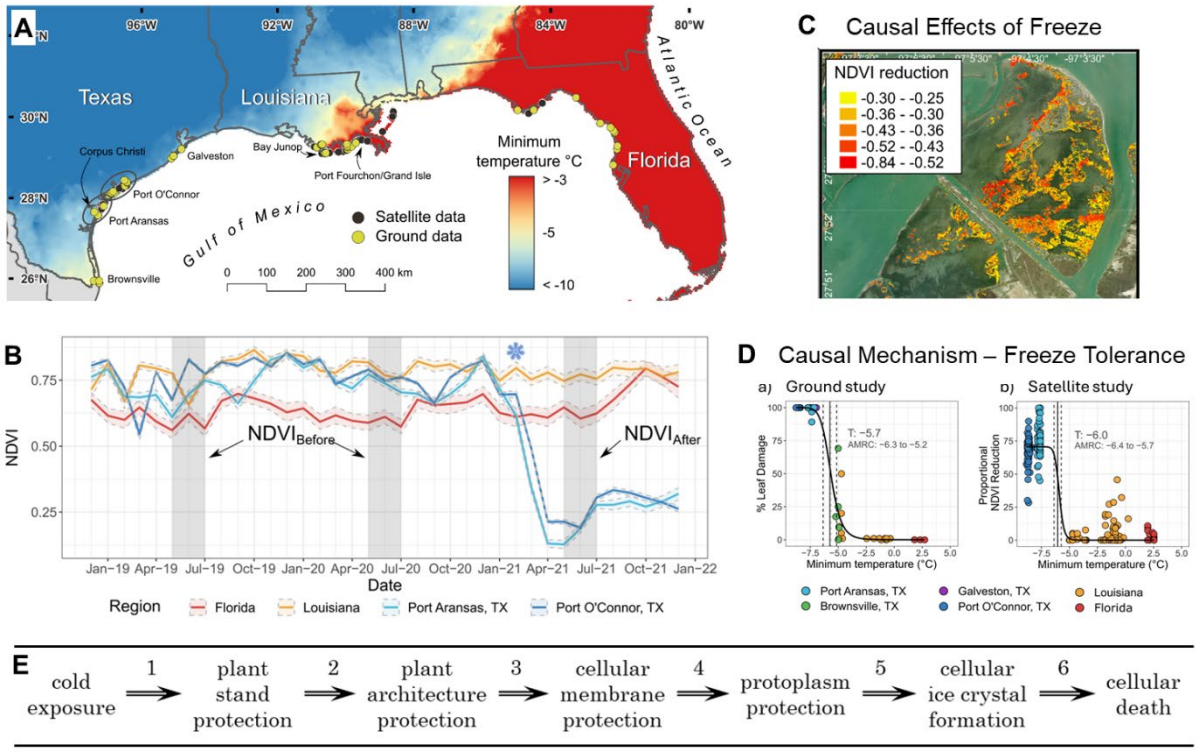


Fig. 5. Comparison of the estimation of causal effects with the characterization of a causal mechanism (assembled from *54*, *24*). (A) Area exposed to extreme event in 2021 showing locations of pre-existing ground-based sample areas and satellite assessment locations for estimating dieback of mangrove trees. (B) Green vegetation cover (NDVI) trends obtained from satellite historical records summarized for different regions. Freeze date indicated by snowflake symbol. Grey time intervals selected for counterfactual estimates. (C) Map of spatial variations in *cf* causal effect estimates showing widespread spatial variation. (D) Estimated mangrove freeze tolerance determined from pooled data samples. (E) Causal knowledge diagram describing the mechanistic line of defense against cellular death from exposure to freezing air temperatures.

Martinez et al. (*54*) also conducted a second set of analyses focused on characterizing key properties of a mechanistic element that determines the responses of mangroves to freezing temperatures. Data were obtained from satellite records and also from ground-based monitoring sites (Fig. 5A). Expert knowledge of mangroves informs us that such subtropical species are limited by their biological freeze tolerance. Plants exposed to subzero temperatures are at risk of damage resulting from the formation of ice crystals that cause cell membrane rupture and death (*56*). This mechanistic knowledge leads us to expect to observe a sharp demarcation between sublethal effects and lethal effects that range from partial damage to death of entire trees as plant defenses are overcome (Fig. 5E). Based on this knowledge, the investigators pooled the data

from different sites and estimated the plant's freeze tolerance using logistic regression (Fig. 5D). Estimates of freeze tolerance were found to be consistent both within this study and with findings from other studies involving separate freeze events (e.g., *57*, *58*, *59*, *60*). This external consistency among independent estimates of freeze tolerance is what biologists would expect for constitutive plant traits. Numerous factors can modestly influence responses to freezing, including preconditioning, freeze duration, microclimatic processes, as well as genetic, life-stage and species-specific variations in physiological freeze tolerance. High water levels and the thermal buffering impacts of warm soil can also protect lower portions of plants from freezing air temperatures (*61*). Despite these mechanistically understood conditional factors, a multitude of confirming observations establishes that the mangrove freeze tolerance mechanism qualifies as transportable causal knowledge.

There are several notable aspects of this comparison. The estimation of freeze tolerance was achieved using standard statistical methods, which can be justified under the multi-evidence paradigm when characterizing mechanisms. This goes against admonitions from causal statistics stating that analysis methods that are not counterfactual causal inference methods are to be labeled as predictive-inference methods and results from the latter are unsuitable for causal interpretations (*17*, *18*). Further, the estimate of freeze tolerance obtained is a mechanistic characterization, not a *cf* causal effect estimate. The differences between causal effect estimates (Fig. 5C) and characterizations of causal mechanisms (Fig. 5D) are substantial with regard to evidential requirements, restrictions, and causal content, as discussed in detail elsewhere (*24*).

Combining forms of evidence: Integrated causal characterization

A study of forest recovery following wildfire at Glacier National Park (USA) was interested in both the magnitude and functional form of the response (*24*). To address this interest, both counterfactual and non-counterfactual techniques were combined. Forest canopy cover was assessed using Landsat satellite data sampled from a candidate set of 800 burned and 800 unburned sample points along with additional ground-based monitoring data (Fig. 6A). Counterfactual techniques were used to quantify the differences in forest cover in burned plots 22 years after the fire compared to what was expected if burned plots had not burned. Several approaches were used to estimate counterfactual expectations, as explained in the original paper.

A great deal of plot-to-plot variation in fire severity was observed (Fig. 6B, E). Plot-to-plot differences in environmental conditions are known to affect pre-fire forest conditions, fire severity, remaining plant materials, and the recovery process. As a result, plots are not counterfactually comparable. Justification for pooling plot data so as to determine the functional form of the relationship between burn severity and forest recovery relied on existing mechanistic knowledge. A causal knowledge diagram is presented in Fig. 6D. Grace et al. (*24*) provides a detailed accounting of the structures and processes related to the causal knowledge diagram, along with a summarization of the predicted sufficiency, reliability, exactness, and generality of the mechanistic elements. Empirical Results revealed a threshold response to burn severity (Fig. 6E). Below the inflection point full recovery of areal cover was observed. However, above the threshold, impairment sharply increased. Additional analyses found the fire severity threshold point to be unaffected by sampling time or whether counterfactual techniques were used or not.

General knowledge of the biology of forest regrowth and regeneration supports the interpretation that the threshold represents a point at which all plant materials capable of regrowth were

consumed by the fire. Increasing impairment of recovery at fire severity levels above the threshold are less precisely known for this individual fire, though destruction of the seed bank and thermal degradation of the soil are known general factors.

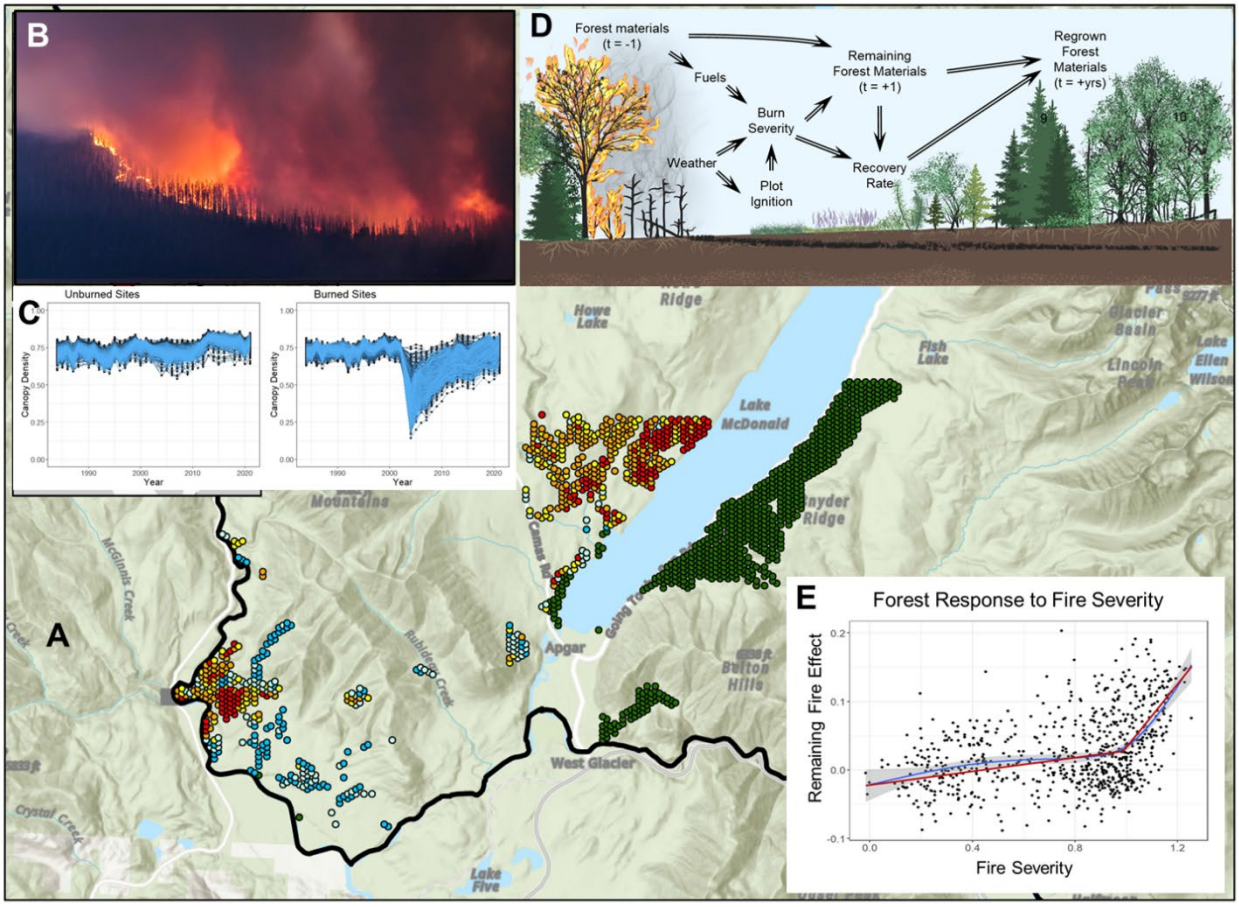


Figure 6. (A) Locations of samples from the Robert Fire with point color indicating fire severity, from greatest (red), to moderate (yellow), to least (bright blue), to unburned sites (green). (B) Example photo of forest fire. (C) Changes in forest canopy density (1984–2021) in unburned and burned sites. (D) Stylized drawing of post-fire forest succession with superimposed causal knowledge diagram for the effects of fire on forest recovery. (E) Plot of the remaining fire effect after 22 years of recovery as a function of fire severity. Blue line represents a generalized additive model (descriptive) fit while the red line represents the relationship based on a changepoint regression. Adapted from (*24*).

Stage 5 – Benefits from a paradigm shift

Many scientists have argued for causal explanation as a general scientific aspiration (*62*, *63*, *64*, *65*). I believe we are now at the threshold of a paradigm shift that can better facilitate that aspiration. The National Academies Consensus Report on Causal Methods (*22*), as well as the review of the limitations of the statistical paradigm by Schwartz and Prins (*27*) very strongly support the need for a paradigm shift of the sort proposed in this paper. Aside from conceptual

arguments that mechanistic knowledge should be considered, real-world examples now provide demonstrations of what consideration of mechanistic knowledge can permit (*23*, *24*, *25*). These examples also demonstrate that the requirements associated with the statistical paradigm (Fig. 3) are not universally applicable and reflect an incomplete and insufficient viewpoint for building causal knowledge.

A primary motivation for shifting away from the statistical paradigm comes from its failure to recognize mechanistic knowledge as causal knowledge. As an example, a recent publication promoting the statistical paradigm makes the blanket statement, "Pre-existing knowledge is typically not sufficient to satisfy causal assumptions." (*21*). Such an unqualified remark reflects the long-term consequence of failing to explicitly recognize the vast reservoir of existing mechanistic knowledge as causal knowledge. The rise of interest in the statistical paradigm is currently making that problem worse as some aggressively promote the idea that results obtained without adherence to statistical paradigm requirements are not suitable for causal interpretations. This paper and its primary sources demonstrate the falsity of such blanket claims

A positive long-term consequence of adopting the multi-evidence paradigm is the potential for a major expansion in the accumulation of causal knowledge, which would in turn accelerate the advancement of science in many fields. To make this happen, the multi-evidence paradigm seeks to encourage scientists to formally document and summarize the evidence leading them to make causal interpretations. The techniques associated with the statistical paradigm can certainly be highly useful and may be the most effective means of estimating the effects of treatments and events for many problems. However, the use of those techniques in conjunction with the limiting restrictions foundational to the statistical paradigm will not maximize our ability to build causal knowledge.

The concept of building causal knowledge is core to the multi-evidence paradigm, providing a sort of glue that holds together the array of study types and forms of evidence that can contribute or not contribute. Since the multi-evidence paradigm rests on the premise that causation emerges from mechanisms, central to causal knowledge is direct knowledge of those mechanisms. There are many known biological, chemical, and physical features in ecological and natural systems that support causal interpretations due to structure–function relationships. For example, organismal traits are currently of major interest because of their inherently transportable mechanistic nature (*66*, *67*, *68*). Causal knowledge analysis of existing mechanistic information should be routinely considered. The current practice in causal methodology of failing to treat existing mechanistic knowledge as causal information is detrimental to progress and disrespectful of the contributions scientists make through their investigative studies. That said, the responsibility for documenting the evidence supporting causal interpretations falls to scientists. At the same time, proponents of the statistical paradigm should avoid further over-representation of statistical causal methodology by recognizing mechanistic knowledge as causal knowledge.

Despite the fact that the techniques developed for use in the statistical paradigm can also be applied under the multi-evidence paradigm, the new paradigm is expected to be disruptive in the immediate term. One important consequence of adopting a multi-evidence paradigm is that it impacts who decides the appropriate evidential standards, the evidential processes, and the evaluation of evidence for individual situations. The technical tasks, assumptions to be addressed, and permissible evidential processes associated with the statistical paradigm (Fig. 2b) follow from statistical premises and expert knowledge of mechanisms is discounted as "domain

knowledge", playing no role in setting the evidential standards and requirements. Under the multi-evidence paradigm, characterizations of existing causal mechanistic knowledge become an essential process that influences all other aspects of causal analysis. Aligning with the National Academies Report (*22*), scientists with expert knowledge become critical participants in decisions about how to weigh available evidence and the standards appropriate to the research situation.

The multi-evidence paradigm represents an important paradigm shift in science that must overcome what is expected to be substantial challenges and inertia (*36*). The brief treatment presented here can only point towards the future, though not define it. I suggest that causal knowledge analysis with its emphasis on documenting the structures and processes composing mechanisms represents a first step in advancing this effort. Scientists possess a great body of accumulated mechanistic causal knowledge. We are now at a nexus where we must either chronicle mechanistic causal knowledge or risk perpetuating the perception that it is ignorable in causal studies. The shift from a focus on the narrow task of computing *cf* causal effects to the larger ambition of building a causal understanding will be disruptive to the existing literature, extremely challenging, but potentially transformative for science.

Acknowledgments: Support for this work provided by Concordia University. Appreciation to Pedro Peres-Neto for facilitating this work and valuable discussions. Appreciation is expressed to several individuals who provided review comments on earlier drafts or portions of this paper: Billy Schweiger, Lori Randall, Kelly Maloney, Seth Prins, Sharon Schwartz, Michael Osland, Ken Krauss. Darren Johnson and Chris Shipway provided technical assistance. Melinda Martinez kindly provided original illustration materials used in Figs. 4 and 6. Photos of wildlife in Fig. 5 courtesy of the author. Picture of Galileo in Fig. 4 from https://commons.wikimedia.org/wiki/Galileo_Galilei.

Funding: none.

Author contributions: JBG was responsible for all aspects of the work.

Competing interests: Author declares that they have no competing interests.

Data and materials availability: Data used in plots contained in Fig. 1 available from Mendeley Data Repository (DOI: 10.17632/jwc4rr6kwr.3). All other materials from other sources are with permission through Open Access licenses.

References and Notes

1. Imbens, G. W. & Rubin, D. B. *Causal Inference for Statistics, Social, and Biomedical Sciences*. (Cambridge University Press, 2015).
2. Laughlin, D. C. *Plant Strategies: The Demographic Consequences of Functional Traits in Changing Environments* (Oxford University Press, 2023).
3. Adler, P. B. Fajardo, A., Kleinhesselink, A. R. & Kraft, N. J. B. Trait-based tests of coexistence mechanisms. *Ecol. Letts.* **16**, 1294-1306 (2013).
4. Jassey, V.E. et al. Tipping point in plant–fungal interactions under severe drought causes abrupt rise in peatland ecosystem respiration. *Global Change Biol.* **24**, 972-986 (2018).
5. Pauchard, A. & Alaback, P.B. Influence of elevation, land use, and landscape context on patterns of alien plant invasions along roadsides in protected areas of South-Central Chile. *Conservation Biol*. **18**, 238-248 (2004).
6. Michaletz, S. T. & Johnson, E. A. How forest fires kill trees: A review of the fundamental biophysical processes. *Scandinavian J. Forest Res.* **22**, 500–515 (2007).
7. Maestre, F.T. et al. Increasing aridity reduces soil microbial diversity and abundance in global drylands. *Proc. Nat. Acad. Sci.* **112**, 15684-15689 (2015).
8. De Palma, A. et al. Ecological traits affect the sensitivity of bees to land-use pressures in European agricultural landscapes. *J. Appl. Ecol.* **52**, 1567-1577 (2015).
9. F. Plard, N. G. Yoccoz, C. Bonenfant, F. Klein, C. Warnant, J. M. Gaillard, Disentangling direct and growth-mediated influences on early survival: a mechanistic approach. *Journal of Animal Ecology* 84, 1363-1372 (2015).
10. Drake, J. M. & Kramer, A. M. Mechanistic analogy: how microcosms explain nature. *Theoretical Ecol.* **5**, 433-444 (2012).
11. P. Wu, et al. Causal inference: a statistical paradigm for inferring causality, in *Statistical Causal Inferences and Their Applications in Public Health Research* (ed. Wu et al.) 3-5 (Springer International Publishing, 2016).
12. Cinelli, C. Challenges in statistics: A dozen challenges in Causality and Causal Inference. arXiv:2508.17099 [stat] (2025)
13. Rubin, D. B. Estimating causal effects of treatments in experimental and observational studies. *ETS Research Bulletin Series* **2**, i-31 (1972).
14. Holland, P. W. Statistics and causal inference. *J. Amer. Stat. Assoc.* **8**, 945-960 (1986).
15. Holland, P. W. & Rubin, D. B. Causal inference in retrospective studies. *ETS Research Report Series*, **1987**, 203-231 (1987).
16. Pearl, J. *Causality* (Cambridge University Press, 2000).
17. Ferraro, P. J., Sanchirico, J. N. & Smith, M. D. Causal inference in coupled human and natural systems. *Proc. Nat. Acad. Sci.* **116**, 5311–18 (2019).
18. Arif, S. & MacNeil, M.A. Predictive models aren't for causal inference. *Ecol. Letters* **25**, 1741–1745 (2022).
19. Dee, L. E., et al. Clarifying the effect of biodiversity on productivity in natural ecosystems with longitudinal data and methods for causal inference. *Nature Comms.* **14**, p.2607 (2023).
20. Siegel, K. J., & Dee, L. E. Foundations and future directions for causal inference in ecological research. *Ecol. Letters* **28**, e70053 (2025).
21. Correia, H. et al. Best practices for moving from correlation to causation in ecological research. ecoevoRxiv [Preprint] (2023). https://doi.org/10.32942/X2DW7Z

22. National Academies of Sciences, Engineering, and Medicine, *Advancing the Framework for Assessing Causality of Health and Welfare Effects to Inform National Ambient Air Quality Standard Reviews* (The National Academies Press, 2022). https://doi.org/10.17226/26612.
23. Grace, J. B. An integrative paradigm for building causal knowledge. *Ecol. Monographs* **94**, e1628 (2024).
24. Grace, J. B. et al. Causal effects vs causal mechanisms: Two traditions with different requirements and outcomes. *Ecol. Letters* **28**, e70029 (2025a).
25. Grace, J. B. et al. Causal interpretations in ecology can be based on mechanistic knowledge. *J. Ecology* **113**, 3084–3098 (2025b).
26. Nichols, J. D. & Cooch, E. G. Predictive models are indeed useful for causal inference. *Ecology* **106**, e4517 (2025).
27. Schwartz, S. & Prins, S. J. *Causal Inference and the People's Health* (Oxford University Press, 2025).
28. Le Roux, E. van Veenhuisen, L. S., Kerley, G. I., Cromsigt, J. P. Animal body size distribution influences the ratios of nutrients supplied to plants. *Proc. Nat. Acad. Sci.* **117**, 22256–63 (2020).
29. Reynolds, W. W. & Karlotski, W. J. The allometric relationship of skeleton weight to body weight in teleost fishes: a preliminary comparison with birds and mammals. *Copeia*, **1977**, 160-163 (1977).
30. Galileo, G. L. Dialogues concerning two new sciences, Transl. by H . Crew and A . De Salvio (1933), Macmillan, New York (1638). [cited in W. W. Reynolds, Skeleton weight allometry in aquatic and terrestrial vertebrates. *Hydrobiologia* **56**, 35-37 (1977)].
31. Prange, H. D., Anderson, J. F., & Rahn, H. Scaling of skeletal mass to body mass in birds and mammals. *Amer. Naturalist*, **113**, 103-122 (1979).
32. Sitters, J., et al. The stoichiometry of nutrient release by terrestrial herbivores and its ecosystem consequences. *Frontiers in Earth Science* **5**, 32 (2017).
33. Subalusky, A. L. & Post, D. M. Context Dependency of Animal Resource Subsidies. *Biol. Reviews* **94**: 517–538 (2019).
34. Veldhuis, M. P., Gommers, M. I., Olff, H. & Berg, M. P. Spatial redistribution of nutrients by large herbivores and dung beetles in a savanna ecosystem. *J. Ecology* **106**, 422-433 (2018).
35. Taveras-Dalmau, V., Becken, S., Westoby, R. From paradigm blindness to paradigm shift? An integrative review and critical analysis of the regenerative paradigm. Ambio **12**, 1-20 (2025).
36. Kuhn, T. S. *The Structure of Scientific Revolutions* (University of Chicago Press, ed. 4, 2012).
37. Rubin, D. B. Causal inference using potential outcomes: design, modeling, decisions. *Journal of the American Statistical Association* **100**, 322–331, (2005).
38. Morgan S. L, Winship C. Counterfactuals and causal inference: Methods and principles for social research. (Cambridge University Press, 2014).
39. Greenland, S., Pearl, J., Robins, J. M. Causal diagrams for epidemiologic research. *Epidemiology* **10**, 37–48 (1999).
40. Winship, C. & Morgan, S. L. The estimation of causal effects from observational data. *Ann. Review Sociology* **25**, 659–706 (1999).
41. Kimmel K., Dee, L. E., Avolio, M. L., Ferraro, P. J. Causal assumptions and causal inference in ecological experiments. *Trends in Ecology & Evolution* **36**, 1141–1152 (2021).

42. Ferraro, P. J. & Hanauer, M. M. Advances in measuring the environmental and social impacts of environmental programs. *Annual Review of Environ. and Resources* **39**, 495-517 (2014).
43. *International Journal of Epidemiology – Special Section: Causality in Epidemiology* (https://academic.oup.com/ije/issue/45/6).
44. Glennan, S. *The New Mechanical Philosophy* (Oxford University Press, 2017).
45. Glennan, S. & Illari, P. M. Eds., *The Routledge handbook of mechanisms and mechanical philosophy* (Routledge Publishers, 2018).
46. Walsh, C. R. & Sloman, S. A. The meaning of cause and prevent: The role of causal mechanism. *Mind & Language* **26**, 21-52 (2011).
47. Bechtel, W. Mechanism and biological explanation. *Philosophy of science* **78**, 533-557 (2011).
48. Machamer, P., Darden, L. & Craver, C. F. Thinking about mechanisms. *Philosophy of science* **67**, 1-25 (2000).
49. Glennan, S. Mechanisms in *The Oxford Handbook of Causation* (Oxford University Press, 2009), pp. 315-325.
50. Ben-Menahem, Y. *Causation in Science* (Princeton University Press, 2018).
51. Deyle, E. R., May, R. M., Munch, S. B., & Sugihara G. Tracking and forecasting ecosystem interactions in real time. *Proc. Royal Society B: Biological Sciences*, **283**, 20152258 (2016).
52. Runge, J. et al. Inferring causation from time series in Earth system sciences, *Nature Comms.* **10**, 2553 (2019).
53. Sugihara, G. et al. Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).
54. Martinez, M. et al. Integrating remote sensing with ground-based observations to quantify the effects of an extreme freeze event on Black Mangroves (*Avicennia germinans*) at the landscape scale. *Ecosystems* **27**, 45-60 (2024).
55. Huntington-Klein, N. *The Effect: An Introduction to Research Design and Causality* (Chapman and Hall/CRC, 2021).
56. Lambers, H. & Oliveira, R. S. *Plant Physiological Ecology* (Springer Nature Switzerland, ed. 3, 2019).
57. Feller, I. C. Nitrogen addition increases freeze resistance in black mangrove (*Avicennia germinans*) shrubs in a temperate-tropical ecotone. *Ecosystems* **26**, 800–814 (2023).
58. Osland, M. J. Temperature thresholds for black mangrove (*Avicennia germinans*) freeze damage, mortality and recovery in North America: Refining tipping points for range expansion in a warming climate. *J. Ecology* **108**, 654–665, (2020).
59. Pickens, C. N. & Hester, M. W. Temperature tolerance of early life history stages of black mangrove *Avicennia germinans*: Implications for range expansion. *Estuaries and Coasts* **34**, 824–830 (2011).
60. Kang, Y. Kaplan, D. A., Osland, M. J. Mangrove freeze resistance and resilience across a tropical-temperate transitional zone. *J. Ecology* **113**, 94-111 (2025).
61. Osland MJ, Microclimate influences mangrove freeze damage: Implications for range expansion in response to changing macroclimate. *Estuaries and Coasts*, **42**, 1084-96 (2019).
62. Holt, R. D. Inference towards the best explanation: Reflections on the issue of climate change. *Israel Journal of Ecology & Evolution* **61**, 1–12 (2015).
63. Pickett, S. T. A. Kolasa, & J. Jones, C. G. *Ecological Understanding: The Nature of Theory and the Theory of Nature* (Academic Press, ed. 2, 2007).

64. Scheiner, S. M. & Willig, M. R. A general theory of ecology. *Theoretical Ecology* **1**, 21–28 (2008).
65. Hone, J. Drake, V. A. Krebs, C. J. Evaluation options for wildlife management and strengthening of causal inference. *Bioscience 73*, 48–58 (2023).
66. McGill, B. J., Enquist, B. J., Weiher, E., Westoby, M. Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution* **21**, 178–185 (2006).
67. Violle, C. et al., Let the concept of trait be functional! *Oikos* **116**, 882–892 (2007).
68. Messier, J., McGill, B. J. & Lechowicz, M. J. How do traits vary across ecological scales? A case for trait-based ecology. *Ecology Letters* **13**, 838–848 (2010).