Title: Clarifying the effect of biodiversity on productivity in natural ecosystems with longitudinal data and new methods for causal inference

Authors and Affiliations:

Laura E. Dee[1*], Paul J. Ferraro[2,3*,] Christopher Severen[4], Kaitlin Kimmel[5], Elizabeth T. Borer[6], Jarrett Byrnes[7], Adam Clark[8], Yann Hautier[9], Andrew Hector[10], Xavier Raynaud[11], Peter B. Reich[12,13], Alexandra J. Wright[14], Carlos A. Arnillas[15], Kendi F. Davies[1], Andrew MacDougall[16], Akira Mori[17,] Melinda D. Smith[18,19], Peter B. Adler[20], Jonathan D. Bakker[21], Kate Braumann[22], Jane Cowles[6], Kimberly Komatsu[23], Johannes MH Knops[24], Rebecca L. McCulley[25], Joslin L. Moore[26], John W. Morgan[27], Timothy Ohlert[28], Sally A. Power[13], Lauren L. Sullivan[29], Carly Stevens[30] and Michel Loreau[31].

1 Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO USA
2 Department of Environmental Health and Engineering, Bloomberg School of Public Health & Whiting School of Engineering, Johns Hopkins University, Baltimore, MD USA
3 Carey Business School, Johns Hopkins University, Baltimore, MD USA
4 Research Department, Federal Reserve Bank of Philadelphia, Philadelphia, PA USA
5 Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD USA
[6] Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN 55108 USA
[7] Department of Biology, University of Massachusetts Boston, 100 Morissey Blvd, Boston, MA 02125
[8] Institute of Biology, University of Graz, Holteigasse 6, 8010 Graz, Austria
[9] Ecology and Biodiversity Group, Department of Biology, Utrecht University, PaDualaan 8, 3584 CH Utrecht, The Netherlands
[10] Department of Plant Sciences, University of Oxford, OX1 3RB, UK.
[11] Sorbonne Université, Université de Paris, UPEC, IRD, CNRS, INRA, Institute of Ecology and Environmental Sciences, iEES Paris, Paris, France
[12] Department of Forest Resources, University of Minnesota, St. Paul, MN, 55108, USA
[13] Hawkesbury Institute for the Environment, Western Sydney University, Penrith, New South Wales, 2751, Australia
[14] Department of Biological Sciences, California State University Los Angeles, Los Angeles, CA USA
[15] Department of Physical and Environmental Sciences, University of Toronto at Scarborough. Toronto, 1265 Military Trail, ON. M1C 1A4.
[16] Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada N1G 2W1
[17] Faculty of Environment and Information Sciences, Yokohama National University, 79-7 Tokiwadai, 11 Hodogaya, Yokohama, 240-8501 Japan
[18] Department of Biology, Colorado State University, Fort Collins, CO 80523
[19] Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO 80523
[20] Department of Wildland Resources and the Ecology Center, Utah State University, Logan, UT 84322
[21] School of Environmental and Forest Sciences, University of Washington, Box 354115, Seattle, WA 98195-4115
[22] Institute on the Environment, University of Minnesota, Twin Cities, 1954 Buford Ave, LES Suite 325 St Paul, MN 55108
[23] Smithsonian Environmental Research Center, Edgewater, MD 21037 USA
[24] Department of Health and Environmental Sciences, Xián Jiaotong-Liverpool University, Suzhou, China
[25] Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY, USA, 40546-0312
[26] School of Biological Sciences, Monash University, Clayton VIC 3800, Australia
[27] Department of Ecology, Environment and Evolution, La Trobe University, Bundoora 3086, Victoria, Australia
[28] Department of Biology, University of New Mexico, Albuquerque, NM, USA
[29] Division of Biological Sciences, University of Missouri, Columbia, MO, 65211, USA
[30] Lancaster Environment Centre, Lancaster University, Lancaster, UK LA1 4YQ
[31] Theoretical and Experimental Ecology Station, CNRS, 09200 Moulis, France

*correspondence: laura.dee@colorado.edu and pferraro@jhu.edu

Summary Statement

Motivated by global changes in biodiversity, ecologists strive to better understand the causal effects of biodiversity on ecosystem functions[1–7]. These effects can be estimated using experimental or observational designs — designs that pose a tradeoff between drawing credible causal inferences from correlations and drawing generalizable inferences. Here, we develop a design that reduces this tradeoff and revisits the question of how plant species diversity affects productivity[1,2,7–9]. Our design leverages longitudinal data from 43 grasslands in 11 countries[10] and new approaches to causal inference with observational data, borrowed from fields outside of ecology[11–13]. Contrary to many prior studies, we estimate that increases in plot-level species richness caused productivity to decline: a 10% increase in richness decreased productivity by 2.4%, 95% CI [-4.1, -0.74]. This contradiction stems from two sources. First, prior observational studies incompletely control for confounding factors. Second, most experiments plant fewer rare and non-native species than exist in nature. Although increases in native, dominant species increased productivity, increases in rare and non-native species decreased productivity, making the average effect negative in our study. By reducing the tradeoff between experimental and observational designs, our study demonstrates how observational studies can complement prior ecological experiments and inform future ones.

33 In the last three decades, ecologists have advanced our understanding of the consequences of

34 biodiversity change for ecosystem functioning[4,7,8,14–16]. One particularly active area of this

35 research has focused on how plant species diversity affects ecosystem productivity[1,2,7–9]. To shed

36 light on this causal relationship, studies have used both experimental and non-experimental

37 designs, each of which presents distinct advantages and disadvantages for elucidating causal

38 relationships in natural ecosystems.

39 Experimentalists that manipulate plant species richness often infer that increases in richness

40 cause increases in biomass[1,2,4,17]. Although experimental manipulations facilitate causal

41 inferences, most experiments that manipulate richness are designed to test theory[18,19] rather than

42 to simulate how species richness changes in natural ecosystems[20–22]. If the effect of richness on

43 productivity depends on the specific species gained and lost, and how they are gained and lost,

44 inferences from experiments may not generalize to natural ecosystems (Figure 1A)[20,21,23]. For

45 example, many biodiversity experiments simulate random gains and losses of species (*but see*[24–

46 26]), which may not mimic changes in species richness in nature. Moreover, most experiments

47 plant common, native species (but see [20,23–25,27]). However, in diverse natural ecosystems, most

48 species are rare[28] and non-native species are increasingly prevalent[29].

49 Observational studies can capture the consequences of changes in species richness that occur in

50 nature. However, determining the causal effect of richness on productivity in observational

51 studies requires strong assumptions[11,13]. Confounding variables associated with both richness and

52 productivity[19] can mask or mimic a causal relationship between them (Figure 1B). For example,

53 unobserved differences in soil nitrogen across locations can mask a positive relationship between

54 richness and productivity if more nitrogen reduces richness and increases productivity[30]. To

55 eliminate confounding effects, common study designs in ecology require identifying, measuring,

56   and statistically controlling for *all* confounding variables[31]. This task is daunting in natural

57   ecosystems given myriad confounding variables that could influence both richness and

58   productivity (e.g., land-use history, herbivory, disturbance). Yet, failure to control for all

59   confounding variables can lead to inferences of the wrong sign or magnitude (i.e., due to

60   statistical bias)[11,32]. Consequently, the mixed evidence on the effect of species richness on

61   productivity in observational studies[7,8,17,33–35] may reflect differences in the degree of control for

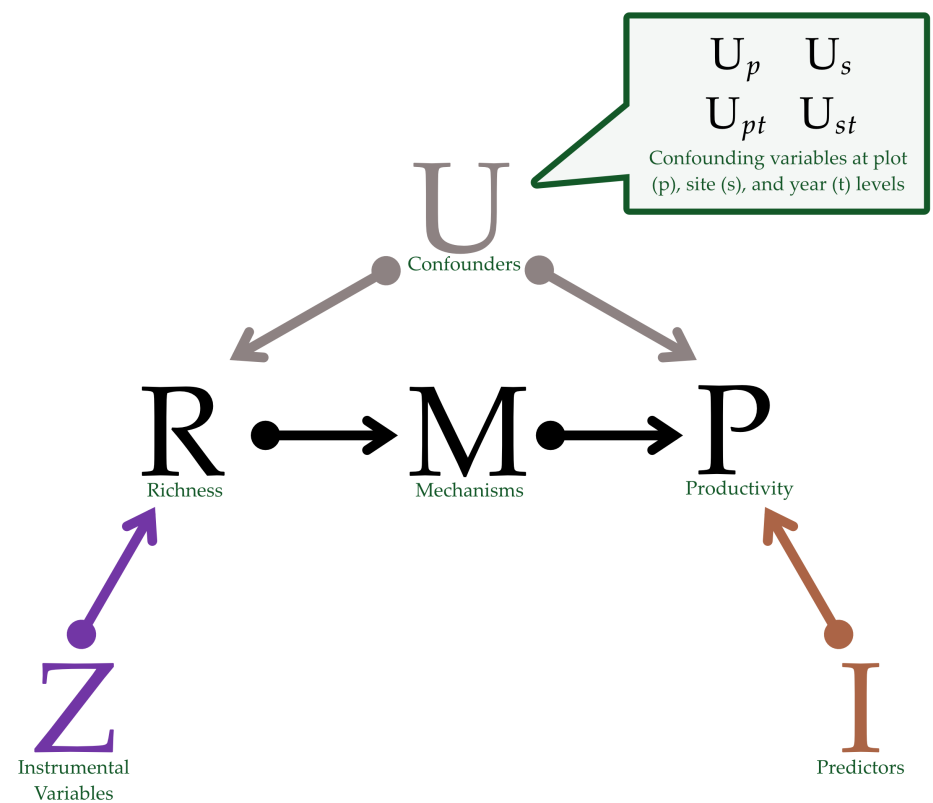62   confounding factors across studies.

63   To isolate and quantify causal relationships between biodiversity and ecosystem function, the

64   ideal study design would combine the strength of experiments in enabling causal inferences from

65   correlations with the strength of observational designs in facilitating generalizable inferences

66   about natural ecological processes. Experimental designs with more realistic extinction processes

67   are one step in that direction (e.g.,[20,24,25]). Here, we develop a complementary approach by

68   leveraging a global grassland dataset[10] and new methods designed for inferring causality from

69   observational data[11–13,36]. This suite of methods now comprises the dominant approach to causal

70   inference in fields outside of ecology, such as economics, medicine, and public health. When

71   combined with our global longitudinal dataset, they allow us to account for the ecological

72   complexity of grasslands without making strong assumptions about our ability to measure all

73   confounding variables[37], and they allow us to isolate the effect of biodiversity on productivity

74   separate from the reverse relationship[36,38].

75   **Study Context**

76   We use repeated observations between 2007-2017 from 151 unmanipulated plots in 43 grassland

77   sites in 11 countries[10] (Table S1 in Supplementary Information (SI); SI Section 3). We define

78   "productivity" as aboveground live biomass per year per $1m^2$ (following[6,7,33]). Each $1m^2$ plot has

Experimental Communities

Proportion of planted species

common species

rare species

Planted Species Richness

Natural Communities

Proportion of species

rare species

common species

Species Richness

U_p U_s
U_{pt} U_{st}

Confounding variables at plot (p), site (s), and year (t) levels

U
Confounders

R → M → P
Richness   Mechanisms   Productivity

Z
Instrumental Variables

I
Predictors

A. The Challenge of Generalizability in Experimental Designs

B. The Challenge of Causal Inference in Observational Designs

79

80 **Figure 1. Challenges in Estimating the Causal Effect of Species Richness on Productivity**. **A**. Experimental designs permit credible causal inferences with few modeling
81 assumptions. Yet experiments often manipulate richness in random permutations, plant limited sets of species, and weed out colonizers. Such designs can yield ecological processes
82 that differ from processes in natural systems. In panel A, when common species are more likely to be planted in experiments than rare species, the proportion of common species is
83 higher than the proportion of rare species regardless of the planted richness level. In contrast, in natural communities, higher richness is associated with more rare species than
84 common species. **B**. Observational designs include natural processes but are challenged by confounding variables (U) associated with both richness (R) and productivity (P); e.g.,
85 precipitation can increase both R and P, thereby inducing a positive correlation between the two, even if the true causal relationship were zero or negative. Some of these
86 confounding variables, like topography, may be time-invariant (or slow-changing) over the study period at the level of the plot (U_p) or site (U_s). Others may be time-varying at the
87 site (U_{st}), such as weather, or the plot (U_{pt}), such as micro-climate. To estimate the effect of R on P without bias does not require data on variables I that only affect P, or Z that only
88 affect R, nor on mechanisms (M), such as selection or complementarity. However, data on Z and M can help address unobserved confounders and differentiate the effect of R on P
89 from the effect of P on R.

90 between 1 and 37 species in a year, with an average of 11.3 (SD = 5.7) and median of 10. We use

91 plots with five or more years of data, in contrast to most observational studies of biodiversity

92 effects on productivity, which use a single year[6,7,33,35,39]. Data from multiple years offer three

93 advantages: 1) an opportunity to study natural changes in richness; 2) enhanced generalizability;

94 and 3) ways to control for a broad set of confounding variables, including unobserved ones (see

95 *Methods*).

96 **Study Design**

97 Our study differs from prior ecology publications by *combining* three features: 1) causal

98 diagrams to inform the design and transparently communicate the assumptions required for

99 inferring a causal relationship from a correlation[7,40]; 2) regression models that leverage repeated

100 observations on the same plots and sites to control for confounding variables (see Figure 1B),

101 both observable and unobservable[11,37,41,42]; and 3) rigorous assessments of the robustness of our

102 inferences to violations of the assumptions required for inferring causal relationships from the

103 data (these assumptions are described in more detail in Figures 2 and 3, *Results*, and *Methods*).

104 The SI offers a primer on these ideas and compares them to approaches widely used in ecology

105 (Sections S2, S4, S7 and S9).

106 In prior observational studies of how richness affects productivity, controlling for the wide range

107 of potential confounding variables in multilevel or structural equation models has posed

108 challenges (e.g.,[5–7,33,43]). In those studies, researchers who wanted to interpret an estimated effect

109 as a causal effect had to assume that no confounding variables were left out of the models [44,45]. In

110 complex ecological systems, however, it is unlikely that one can measure all possible

111 confounding variables. Moreover, when these confounding variables are measured with error,

112   statistically controlling for them can introduce other biases [44]. In other words, prior studies

113   require a strong assumption for interpreting the correlation between richness and productivity as

114   causal: any site or plot attributes not included in the statistical estimation model are assumed to

115   be uncorrelated with species richness and therefore not a source of statistical bias. Our design

116   relaxes this strong assumption.

117   We improve upon prior observational studies by controlling for a broader suite of confounding

118   variables without needing to directly measure them as covariates (see *Method*s and [37,41]). To

119   understand the intuition for how this control is possible, recall that, in contrast to most prior

120   observational studies on this topic (e.g.,[5–7,33,37]), our multi-site data is longitudinal ('panel data')

121   and thus includes variation in species richness in both time and space. Confounding variables

122   that affect both richness and productivity could arise from conditions at the plot or the site. The

123   values of these variables may be essentially invariant during the study period (e.g., soil texture,

124   topography, land-use history) or they may vary through time (e.g., surrounding land-use change,

125   drought conditions that differ by both site and year). With our multi-site panel data, we can

126   directly control for time-varying, site-level conditions, whether they are observable or not, via a

127   regression estimator that includes a simple interaction of binary variables for each site and year

128   (see *Methods,* Equation 2). Further, we can eliminate the confounding effects of time-invariant

129   plot and site conditions by taking deviations from mean conditions, after which variables that do

130   not change over time no longer have any explanatory role and thus are eliminated as a source of

131   bias (*Methods*). Using alternative designs, we can also quantify the potential threat of additional

132   sources of bias from unobserved, time-varying plot-level confounders and from reverse causality

133   (by bias from reverse causality, we mean bias that could arise when a causal effect also runs

134   from productivity to richness; see *Methods*). In contrast to our approach, virtually all

135   observational analyses reviewed in [8] omit important confounding variables (e.g., from human

136   activities and land management) and ignore the potential for reverse causality (reviewed in [46]).

137   To demonstrate how our study design builds on and advances prior research, we first apply two

138   study designs that have been used in prior studies and then contrast them to our design.

139   Specifically, we estimate a simple bivariate correlation of richness and productivity (like [33]) and

140   then we estimate the relationship between richness and productivity using a multivariate design

141   that mirrors advanced statistical designs that aim to control for confounding variables by directly

142   measuring and including them as covariates in regression models (a "conditioning on

143   observables" analysis, like [7]). The multivariate design, which we label "*Common Design in*

144   *Ecology*," controls for over 60 variables (far more than prior studies), including attributes of the

145   soil, habitat, historical management, and weather (Table S10). More details are provided in

146   *Methods* section.

147   **Results**

148   We first report the bivariate correlation between plot richness and productivity. Consistent with

149   prior studies[33], we find a statistically weak, positive relationship between richness and

150   productivity when we do not control for any confounding variables: a 10% increase in richness is

151   associated with a 1.4% increase in productivity, 95% CI [-0.6, 3.4]. To give that estimated

152   correlation a causal interpretation requires an implausible assumption that there are no

153   confounding variables in the system (or that they perfectly cancel each other out).

154   Consistent with prior multivariate studies[7], the *Common Design in Ecology* yields a statistically

155   significant, positive relationship: a 10% increase in plot richness increased productivity by 3.8%

156   on average, 95% CI [2.0, 5.6] (Figure 2B). To give that estimated effect a causal interpretation,

157 however, requires a strong assumption: all possible confounding variables are measured

158 accurately and are included in the model.

159 In contrast to prior analyses, our *Main Design* controls for a much broader set of potential

160 confounders and comes to the opposite conclusion (Figure 2A; Tables S2-S3; SI sections S4, S7

161 and S9). We find a 10% increase in plot richness *decreased* plot productivity by 2.4% on

162 average, 95% CI [-4.1, -0.74]. The estimate is similar if we measure biodiversity using

163 Simpson's Diversity (Table S2), control for concomitant changes in species evenness (Figure

164 2A), or measure species richness and productivity as untransformed variables in linear or non-

165 linear specifications (Table S3). In extended analyses (Tables S4-S6), we find no evidence that

166 the effect of species richness on productivity is moderated by the site's productivity or total

167 number of species (as in [39]).

168 Switching from the *Common Design* to our *Main Design* flips the estimated effect of richness on

169 productivity from positive to negative (Figure 2). This sign-switching likely occurs for two

170 reasons. First, on average, many of the observed site and plot variables at these 43 sites are

171 negatively associated with richness and positively associated with productivity (or *vice-versa*).

172 Controlling for them in the *Common Design* moves the estimated effect in the positive direction

173 from the bivariate correlation. Second, unobserved site and plot variables (U_s, U_p, U_{st} in Figure

174 1B) are, on average, positively associated with both richness and productivity. We can infer the

175 sign of these associations by observing how the estimated effect changes with and without the

176 controls for unobserved time-varying, site-level conditions and time-invariant plot and site

177 conditions (Figure 2A versus Figure 2B; see *Methods*). Failing to control for the time-varying

178 confounders is a particular problem in the *Common Design*. In other words, the Nutrient

179 Network sites experience site-specific "shocks" that vary each year (e.g., weather shocks, like a

9
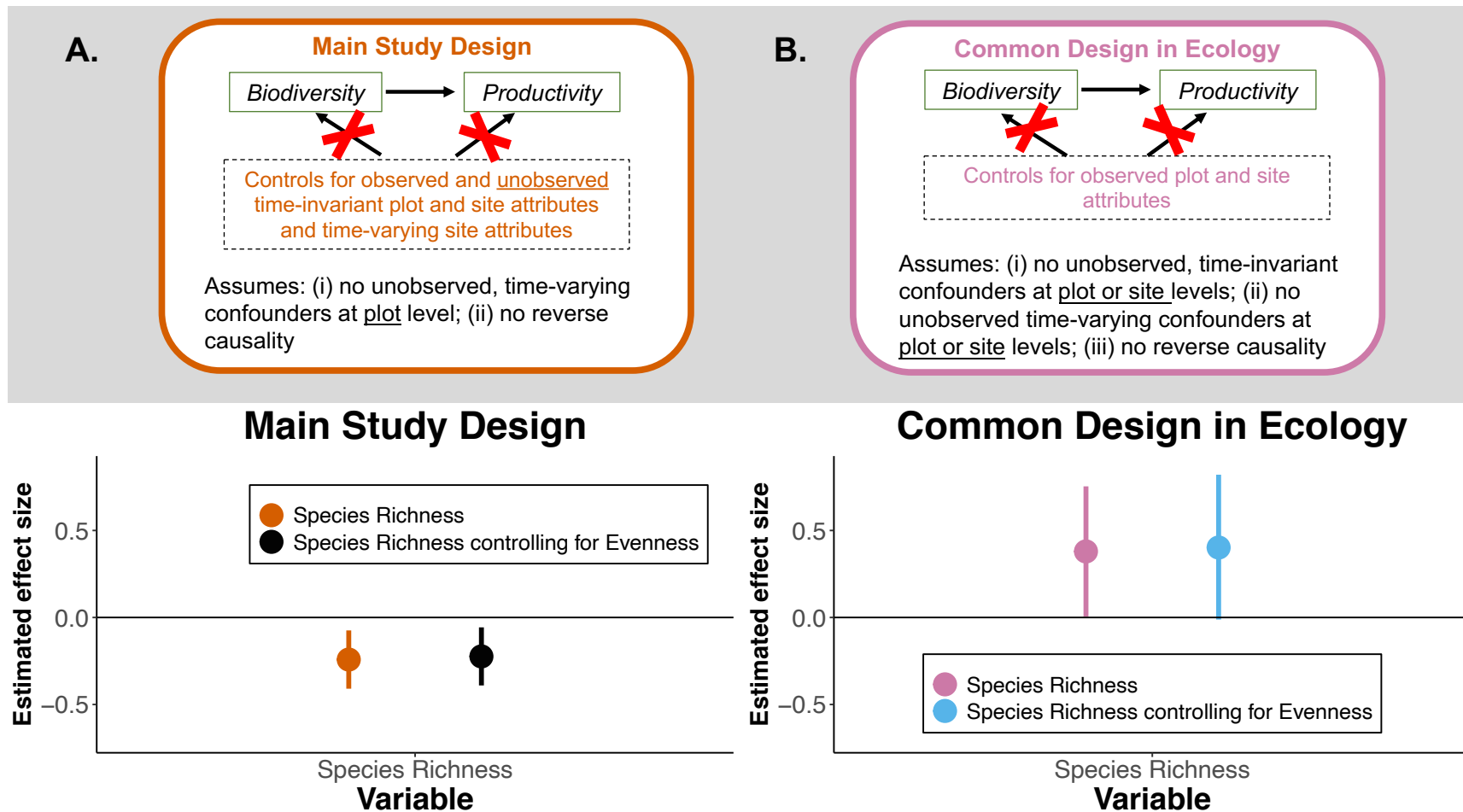
Figure 2. The effect of biodiversity on productivity, estimated as the % change in productivity from a 1% increase in richness. All estimates are shown with 95% CIs. **A**. Our *Main Design* is a panel data design (see SI: section S4**). B**. The *Common Design* is a multivariate mixed model, common in ecological analyses of observational data (see SI: section S7). The top panel summarizes the assumptions needed for drawing causal inferences in each design. A red X on an arrow implies that the design blocks the confounding pathways described in the box with dotted lines.

186 particularly dry April, or herbivory shocks, like higher herbivore pressure than the prior year)

187 and failing to control for them creates statistical biases in the positive direction. We cannot

188 observe the exact components of these shocks, but because we observe the same sites over many

189 years, we can control for them. The *Main Design*, with its greater set of controls, is thus less

190 biased[37]. More details are available in SI sections S7 and S9.

191 *Results are Robust to Alternative Assumptions for Inferring Causality*

192 A hallmark of modern approaches to causal inference is to probe the robustness of results to

193 potential violations in the assumptions used to infer causality from correlation.[36] Using four

194 additional approaches, we use assumptions different from those made in our *Main Design*

195 (Figure 2A) and assess how our conclusions change (see *Methods*). In all four approaches, the

196 estimated effect of richness on productivity is negative (Figure 3; Tables S7-S8).

197 Based on the first two approaches, we conclude that we are not mistaking the effect of

198 productivity on richness[7,19] for the effect of richness on productivity ("reverse causality"). The

199 first approach employs an instrumental variable design, which uses an observable source of

200 variation in richness (Z in Figure 1B) that is assumed to have no connection to productivity after

201 conditioning on the site and plot variables addressed in the *Main Design* (see *Methods*). When

202 this assumption is valid, the design addresses both reverse causality and all forms of confounding

203 in Figure 1B, at the cost of drawing inferences from only a subset of the data, which can

204 dramatically decrease the precision of the estimate. Our second approach assumes, based on [7],

205 that a negative effect of productivity on richness would be mediated (M in Figure 1B), at least in

206 part, by shading or factors for which shading is a proxy (e.g., overcrowding). To block the effect

207 of this mechanism, we add a shading variable to the *Main Design* (see *Methods*). If the estimated

11

208 effect changes, reverse causality may be a source of bias. In both approaches, the estimated

209 effect remains negative, suggesting that, if either of the approaches' assumptions are valid,

210 reverse causality is not driving our results. This conclusion does not mean productivity cannot

211 affect richness, only that such a relationship is not a likely source of bias in our *Main Design*.

212 Based on the final two approaches, we conclude that the estimated negative effect in the *Main*

213 *Design* is robust to potential biases from unobserved confounding variables at the plot level that

214 vary over time (U_{pt} in Figure 1B). First, in a dynamic panel design, we address bias that would

215 arise if the prior year's productivity affects richness and productivity in the current year (e.g., via

216 soil fertility[47]). The estimated effect is similar to the estimate from the *Main Design*. Second, in a

217 sensitivity analysis[12], we assess how the *Main Design* estimate would change if there were a

218 strong, unobserved confounding variable that is negatively associated with species richness and

219 positively associated with productivity (e.g., measurement error or plot-level drivers of

220 disturbance). If such a confounding variable were to exist, it could create a spurious negative

221 correlation between richness and productivity. The analysis implies that, even in the presence of

222 a such an unobserved confounder, we would still infer that there is a negative relationship

223 between richness and productivity.

224 *The role of rare species and non-native species*

225 In contrast to our study, many experimental studies report positive effects of richness on

226 productivity[1,2,4,17]. One difference between experimental and natural systems is that most species

227 in natural ecosystems are rare, whereas most species planted in experiments are not rare (Figure

228 S11-12). Rare and dominant species can affect productivity differently[48]. Thus, the effect on

229 productivity from an increase in richness (e.g., from 4 to 8 species) could differ when the
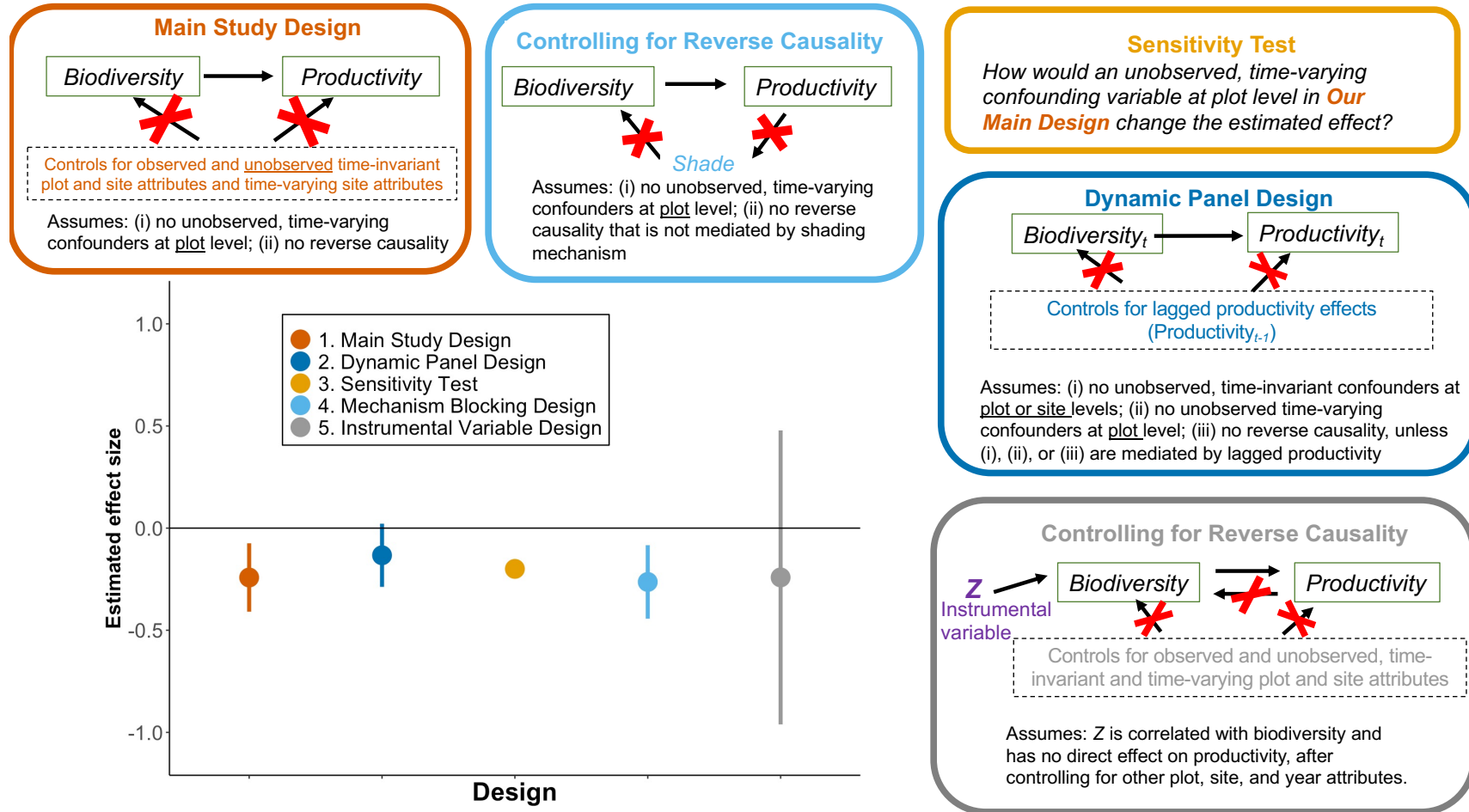
Figure 3. The effect of species richness on productivity (robustness checks). A X on an arrow implies that the confounding pathway are blocked by the design. Using four approaches that make assumptions that differ from the assumptions in our *Main Design* (Figure 2A), we find no evidence for a positive effect of species richness on productivity. All estimates, except the sensitivity test estimate, are shown with 95% CIs.

235   additional species are rare versus not rare. In the jargon of the causal inference literature,

236   richness is a compound treatment with multiple versions, or a "heterogenous treatment."[49]

237   (Figure 4A). Another difference between experimental and natural systems is the number of non-

238   native species, which are absent in many experiments but increasingly prevalent in real

239   ecosystems[29]. If non-native are more competitive but less productive, as in [50], this could also

240   explain the divergence between our results and those of experimental studies.

241   Rare and non-native species could reduce productivity through multiple channels (Figure 4B,

242   left). For example, these species may compete with more productive species (e.g., via allelopathy

243   of rare invaders[51]). Further, they may produce less aboveground biomass than common, native

244   species[48] and so when they enter a plot, they may take space formerly occupied by more

245   productive species. These productivity-reducing effects would be strengthened if, as richness

246   increases in a plot, rare and non-native species are more likely to be the incrementally added

247   species (Figure 4B, right). In experimental systems, species enter plots with equal probability. In

248   contrast, as richness increases in our 43 grassland sites, the probability that the incremental

249   species is a rare or non-native species also increases (Figures S11-12).

250   Given differences in the species pools studied in our natural systems versus many experimental

251   systems, we explore whether changes in rare species and non-native species richness affect

252   productivity differently from changes in native, non-rare species richness. We classify species

253   into four categories: 1) rare, native species; 2) non-rare, native species; 3) non-rare, non-native

254   species; and 4) rare, non-native species (see *Methods*). We then estimate the effect of each

255   category's richness on productivity using our *Main Design*.

A. Richness is a heterogeneous treatment: R|C→P

The effect on productivity (P) when richness (R) changes from one value to another value depends on (|) the compositions (C) at the two values.
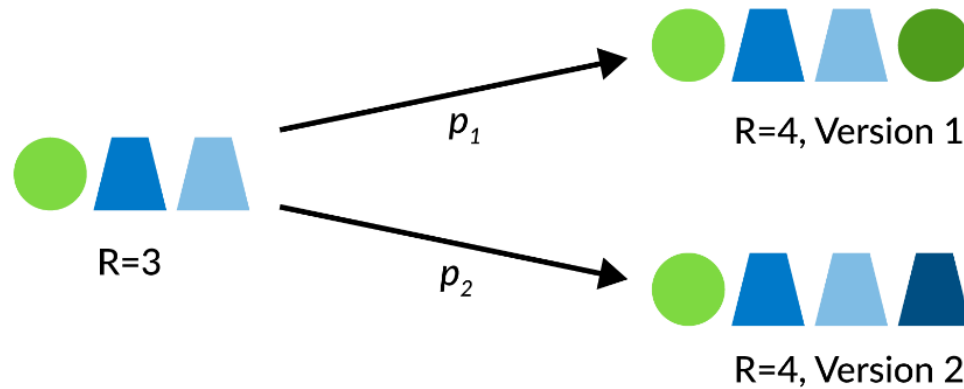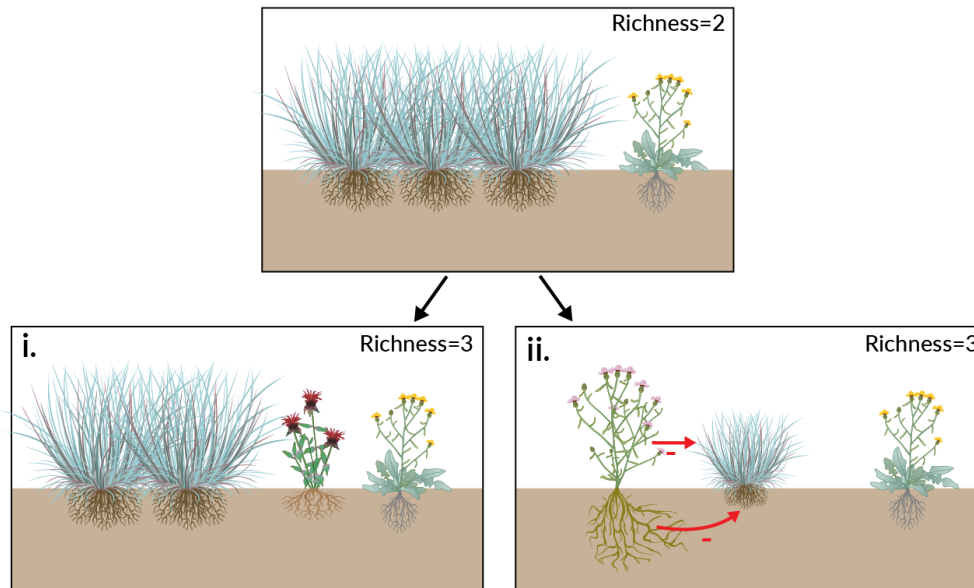

● Rare

▲ Dominant

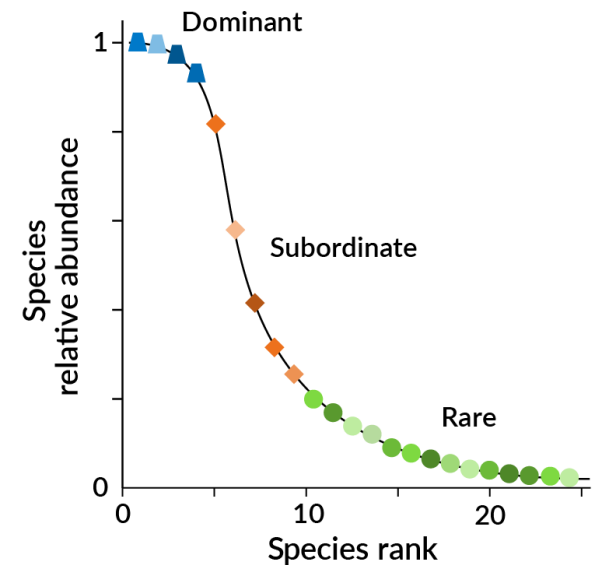*Shapes refer to species' abundances within communities. Colors refer to species.*

p_1 *is the probability of getting Version 1,* p_2 *is the probability of getting Version 2*

B. Changes in richness in natural ecosystems

Figure 4. **Composition and its Role in the Effect of Species Richness on Productivity**. This figure illustrates why results from observational and experimental studies may differ. Composition, the identities of species that could potentially grow at a site at a given level of richness, makes species richness a heterogeneous treatment in both experimental and non-experimental systems (Panel A). The way richness changes in nature can thus influence how changes in richness affect productivity (Panel B). Case (i), where Richness has changed from 2 to 3 species, is a case where rare species take space formerly occupied by more productive dominant species. Case (ii) is a case where Richness has changed from 2 to 3 species and rare species have negative effects on dominant species productivity (e.g., via below-ground competition or allelopathy). For simplicity, the figure focuses on a contrast between rare and dominant species, but the ideas can also apply to differences between native and non-native species.

16

266 Our results imply that the negative average effect of richness on productivity in Figure 2 is

267 driven by changes in the numbers of rare, native species and non-native, non-rare species (Figure

268 5). Consistent with results from experimental studies, an increase in species richness that came

269 from a non-rare, native species increased productivity. But increases in richness decreased

270 productivity when these increases came from non-rare, non-native species or rare, native species

271 (inferences are similar using different definitions for rarity; Tables S11-S15). We acknowledge

272 that there could also be positive and negative interactions across these species' types, but we do

273 not have sufficient statistical power to explore these potential interactions.

274 We conjecture that the proposed mechanisms through which richness positively affects

275 productivity in archetypical experiments – i.e., niche complementarity and positive

276 selection[2,18,19] – may operate primarily among non-rare, native species. Testing this conjecture

277 would require experiments that successfully grow representative proportions of rare species (in

278 experiments that planted rare species, these species failed to consistently emerge, see SI: Section

279 S9 – Box 2).

280 **Discussion**

281 Leveraging methodological advances for causal inference in observational designs, our study

282 uncovers ecological relationships in grasslands that deserve closer attention. In the 43 grassland

283 sites in our study, an increase in species richness decreases productivity on average (Figure 2).

284 This effect appears to arise because an increase in rare species and non-rare, non-native species

285 decreases productivity on average (Figure 5) and these species comprise most species in an

286 ecosystem (Figure 4). These effects will be missed in observational designs that do not

287 adequately control for a wide range of confounding factors and in experiments that do not plant a

17

288
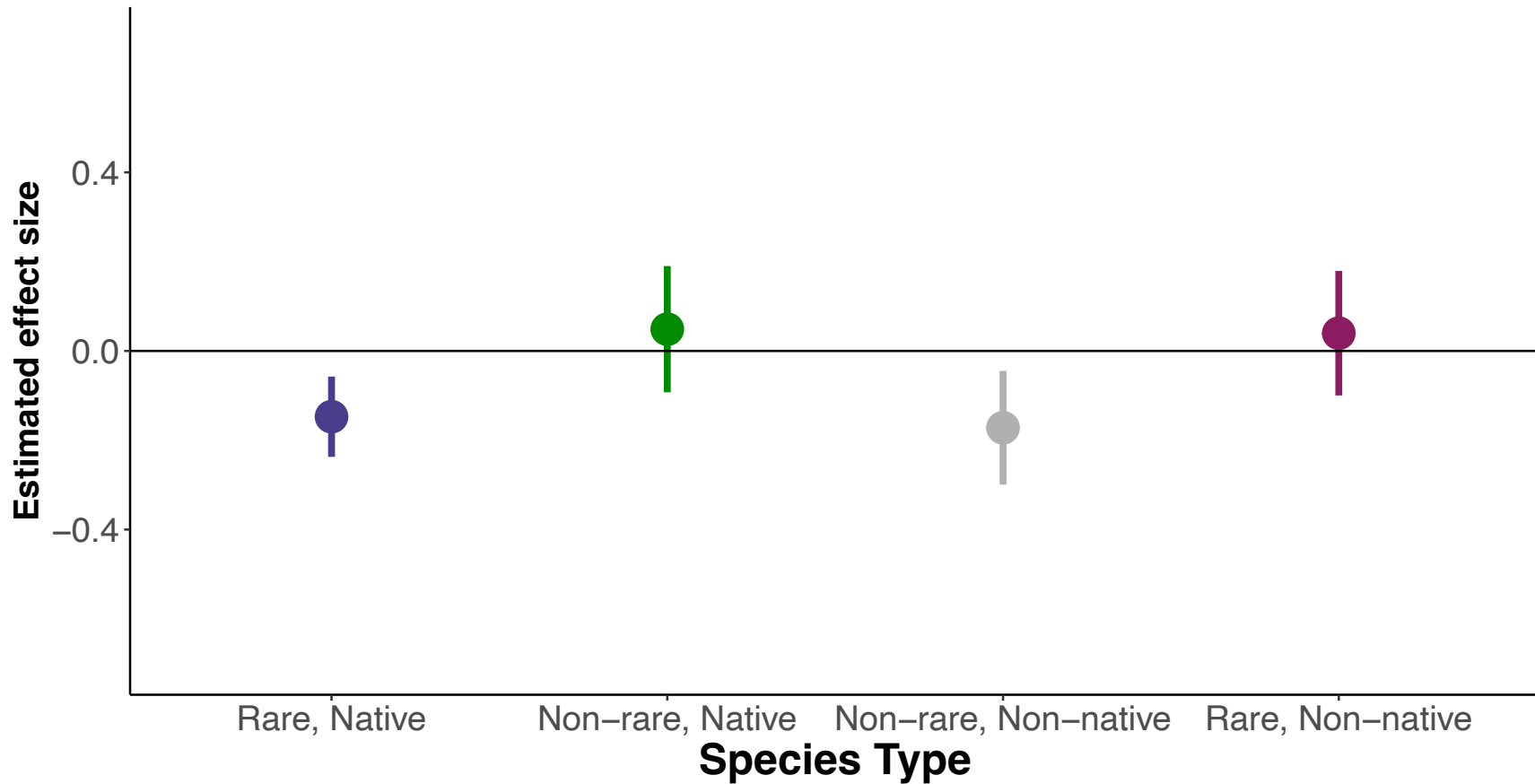
Figure 5. The effect of species richness on biomass production conditional on species type. All estimated effects are on a log-inverse-hyperbolic-sine scale and shown with 95% confidence intervals. Given the inverse hyperbolic sine transformation of the richness variable, the estimated effects cannot be interpreted as elasticities without further manipulation, but their signs and relative magnitudes can be compared. We can reject the null hypothesis that the four estimated effects are equal (ChiSq = 9.82, *Pr*(ChiSq = 0.02)).

294   representative mix of species in diversity patterns that occur in natural systems. Our results also

295   highlight the challenge of determining the representativeness of experimental systems. For

296   example, a recent comparison of natural and experimental systems identified many similarities in

297   attributes but did not assess whether the patterns of rare or non-native species in experimental

298   systems match the patterns in natural systems [52].

299   Our results point to promising areas for future research, including studies that experimentally

300   manipulate rare species and non-native species and seek to identify traits of these species that

301   drive their effects on productivity. For example, in a recent study[26], researchers experimentally

302   removed non-dominant species from randomly assembled communities and reported an increase

303   in biomass one year after the removal; a result consistent with our results. Extensions of their

304   study can help elucidate the traits of these species that drive their effects on productivity.

305   Our results also imply that any estimated effect of changes in species richness on productivity

306   may not generalize to different spatial or temporal scales, or to other ecosystems, forms of

307   biodiversity, or ecosystem functions. Multiple ecological mechanisms underlie a relationship

308   between richness and productivity (e.g., below-ground competition, niche complementarity), and

309   their strength may vary across places and time depending on which types of species are changing

310   and how (Figure 4). When the operative mechanisms depend on the version of richness that is

311   changing, ecologists face what has been called "treatment-variation relevance"[53] or

312   "consequential variation of the treatment"[54]. In such cases, interpreting and generalizing causal

313   effects is challenging, whether ecologists use experimental or observational designs. The

314   challenge is best met by using large observational datasets that capture the types of species

315   changes that occur in nature and then, using methods like the ones in our study, determine *which*

316   changes -- in terms of which species are changing and how (Figure 5) -- are consequential for the

317   effect of richness on ecosystem function. The results of these observational studies could then

318   guide experimentalists in selecting experimental designs that can help confirm the results from

319   observational studies and elucidate the underlying mechanistic processes.

320   Most importantly, our study extends prior research[7] that highlights the importance of study

321   design in credibly isolating causal relationships in natural ecosystems. Other fields have made

322   important advances in observational analyses — advances that have not yet permeated into

323   ecology and other natural sciences. By demonstrating how to apply these advances to an

324   ecological question and data, our study aims to spur broader adoption of these advances in

325   ecology. Given the challenges of randomizing all the important elements of ecosystems at larger

326   spatial and temporal scales, observational designs like ours that leverage these advances offer

327   important complements to experimental designs in research to elucidate how natural ecosystems

328   function.

329   **Methods**

330   To ensure reproducible results, we implemented all analyses in two software programs (R using

331   the 'fixest' package v 0.8.2 and Stata v.16) and multiple researchers confirmed the results. The

332   code for reproducing all analyses, figures, and tables in this study are available at

333   https://github.com/LauraDee/NutNetCausalinf. An RMarkdown tutorial on the main methods can

334   also be found at our Github project page.

335   **Target Causal Effect**

336   To formalize the causal relationship we seek to estimate, we use the potential outcomes

337   framework [55–57]. The causal effect of a change in richness from R' to R'' on productivity P in plot

338   i is defined as $[P_i(R'') - P_i(R')]$, where $P_i(R'')$ is the potential productivity outcome when $R =$

339   R'' and $P(R_i')$ is the potential productivity outcome when $R = R'$ ($R' \neq R''$). The difference in

340   these two potential productivity outcomes (i.e., productivity under two potential richness

341   conditions) is the causal effect of a change in richness in a plot. For a specific location and time,

342   only one of these potential outcomes will be directly observable; the counterfactual values for the

343   other potential outcomes must be estimated from data. The average causal effect of a change in

344   biodiversity from R' to R'' across all plots is $E[P_i(R'') - P_i(R')]$, where $E[\cdot]$ is the expectation

345   operator. We seek to estimate the average causal response of an incremental change in R across

346   all plots (i.e., the average effect across all possible one-unit changes). When used for causal

347   inference, non-experimental studies aim to replicate, conceptually, the idealized experimental

348   design in which the factor or factors that affect variation in R only affect P via their effects on R.

349   In other words, to permit credible causal inferences, a non-experimental design seeks, via design

350   and statistical methods, to eliminate the confounding effects of U_p, U_s, U_{st}, and U_{pt} in Figure 1, as

351   well as the effects of reverse causality in Figure 3.

352   **Data Description**

353   Study sites include mesic grasslands and prairies, savanna, desert grasslands, montane meadows,

354   old fields, and alpine tundra from 11 countries. To measure productivity, we use plant above-

355   ground live mass (biomass) (see Figure S1). Biomass production supports many ecosystem

356   processes and services and this measure of productivity has been widely used in studying the

357   relationship between diversity and productivity with both observational (e.g., [6,7,58]) and

358   experimental data (*reviewed in* [8,59,60]). For herbaceous vegetation, above-ground live biomass

359   provides a reasonable estimate of primary productivity [61]. Biodiversity measures are determined

360   from species cover data from the Nutrient Network (SI Section 3a)

361   **Common Design Estimator**

362   To show how our Main Design differs from more common designs in ecology, we constructed

363   what we call a "Common Design in Ecology": a multivariate design that controls for over 60

364   variables (far more than prior studies), including attributes of the soil, habitat, historical

365   management, and weather (Table S10). This design captures the strong assumption that is

366   inherent in prior observational ecological studies that aim to estimate the causal effect of richness

367   on productivity: there are no variables omitted from the statistical model that are correlated with

368   both richness and productivity. When this assumption is not met, the design suffers from bias.

369   We compare this Common Design to our Main Design, which relaxes that strong assumption.

370   For the "Common Design in Ecology," we estimate the effect of richness on productivity using

371   the following regression equation:

372    $$\ln LiveMass_{pst} = \beta \ln Richness_{pst} + X_p + X_s + X_{st} + \varepsilon_{pst}. \tag{1}$$

373   We use a ln-ln model specification because the effect of richness on productivity is believed to

374   be non-linear [62]. Recall that observations in our data come from a plot p located within a site s in

375   a year t. Thus, X_p is a vector of plot-specific attributes that do not vary over the study period

376   (e.g., soil type), X_s is a vector of site attributes that do not vary over the study period (e.g.,

377   habitat, historical management, elevation), and X_{st} is a vector of site attributes that vary by year

378   (e.g., temperature seasonality, maximum and mean temperatures of the warmest month).

379   Together, these vectors include over 60 variables, which are directly controlled for in the

380   regression (see Table S10). In this equation, we can see that the effects of any omitted variables

381   on productivity (i.e., variables not controlled for in the X vectors in the model) reside in the error

382   term ε_{pst}. We can rewrite this error term as a combination of a random error, σ_{pst}, which only

383   affects productivity, and unobserved confounding variables, U, which affect both richness and

384   productivity at either the plot or site level and either in all years or only some years. Thus the

385   error term can be rewritten as $\varepsilon_{pst} = U_p + U_s + U_{st} + U_{pt} + \sigma_{pst}$, where U_p and U_s are vectors

386   of plot and site-level variables that do not change over the study period, and U_{st} are vectors of

387   time-varying site-level variables, and U_{pt} are vectors of time-varying plot-level variables. If a

388   study design has any of these omitted U variables, the estimator (β) would be biased – known as

389   omitted variables bias. In other words, to interpret the estimate of β as an estimate of our target

390   causal effect would require one to assume that the observed covariates in the model capture all

391   relevant U_p, U_s, and U_{st} and U_{pt} does not exist (because no time-varying covariates are measured

392   at the plot level).

393   **Main Design Estimator**

394   Each observation in our study comes from a plot p located within a site s in a year t. With this

395   longitudinal data structure, i.e. with repeated observations of the same plots, one can control for

396   all dimensions of confounding variables that do not vary over the study period and all

397   dimensions of time-varying site-level confounding variables without having to observe all of

398   these dimensions. To achieve this control, we estimate an equation of the following form:

399

400 $$\ln LiveMass_{pst} = \beta \ln Richness_{pst} + \delta_p + \mu_{st} + \varepsilon_{pst} \tag{2}$$

401   As noted above, we use a ln-ln model specification because the effect of richness on productivity

402   is believed to be non-linear[62]. Given that we have a ln-ln specification, β can be interpreted as an

403   elasticity: the expected percent change in productivity given a one percent change in richness.

404   We also tested the robustness of results to this modeling decision (see SI: section S5).

405   The <u>time-invariant plot</u> attributes (δ_p) are modeled in a fully flexible way that allows each plot to

406   have its own effect on productivity (details on estimation procedure below). In the Economics

407   literature, δ_p would be called "plot-level fixed effects." Note that the phrase "fixed effects" has a

408   different meaning in economics than in ecology (see S1 Glossary). In economics, including δ_p is

409   said to control for "unobserved heterogeneity" across plots that can be a potential source of bias.

410   Note that δ_p is not part of the error term, as it would be in mixed (multi-level) models [63] or in a

411   Common Design as in Equation 1 (i.e., U_p and U_s). Rather, it is a parameter to be estimated, just

412   like β. In other words, β and δ_p are assumed to be fixed and estimable, rather than assumed to

413   follow a distribution. Time-invariant <u>site</u> attributes are not explicitly included in the equation

414   because they are subsumed into the time-invariant plot attributes (i.e., plots are nested within

415   sites and so fixed site attributes are controlled via fixed plot attributes). In other words, this

416   variable captures all attributes of a plot at a given site that vary little over the study period. Thus,

417   δ_p captures both U_p and U_s in the decomposition of the error term from the previous section (i.e.,

418    $\varepsilon_{pst} = U_p + U_s + U_{st} + U_{pt} + \sigma_{pst}$, where σ_{pst} is a random error that only affects productivity

419   and U are unobserved confounding variables that affect both richness and productivity).

420   To show how the estimator in Equation 2 can efficiently control for time-invariant confounders,

421   we subtract the productivity observation within a plot in one year ($t - 1$) from the productivity

422   observation within the same plot in the next year (t), yielding an equation for the change (Δ) in

423   productivity from one year to the next:

424    $$\Delta \ln LiveMass_{pst} = \beta \, \Delta \ln Richness_{pst} + \Delta \mu_{st} + \Delta \varepsilon_{pst} \qquad (3)$$

425   The variable δ_p, which captures the effects of time-invariant plot attributes, has been differenced

426   away, allowing for efficient estimation of β; in other words, we control for time-invariant plot

427 attributes without having to estimate them and use up many degrees of freedom. In this

428 differenced version of Equation 2, one can see that we are estimating the effect of richness on

429 biomass from changes in richness <u>within</u> plots, where the confounding effects of between-plot

430 differences are absent. Rather than first-differencing the equations to eliminate δ_p, one can

431 instead take deviations from plot-level means, which is the approach we take to estimating

432 Equation 2 because it can be more efficient. Thus, we can also describe our estimation strategy

433 as estimating a correlation between deviations of productivity around its mean and the

434 corresponding deviations in richness around its mean.

435 The <u>time-varying</u> site attributes (μ_{st}) are also modeled in a fully flexible way that allows a year-

436 specific effect for each site (in the estimation, an indicator for each year is interacted with an

437 indicator for each site). Explicitly estimating μ_{st} flexibly controls for confounding variation due

438 to conditions at a site that change through time, such as weather (e.g., time-varying patterns of

439 temperature and precipitation), herbivory, and surrounding land management conditions. In other

440 words, this variable captures all year-specific conditions experienced by every plot at a given

441 site. Thus μ_{st} captures U_{st} in the decomposition of the error term from the previous section

442 ($\varepsilon_{pst} = U_p + U_s + U_{st} + U_{pt} + \sigma_{pst}$). The estimator in Equation 2 is often called a "two-way

443 fixed-effects estimator" because, by taking deviations from the means, one controls for time-

444 invariant confounding and, by including the site-by-year effects, one controls for time-varying

445 confounding.

446 Thus, in contrast to the Common Design in Ecology, we control for a broad suite of plot-level

447 and site-level confounders without having to measure them directly. The Main Design also

448 controls for non-linear relationships between the confounding variables and productivity or

449   richness, as well as linear and non-linear interactions among those variables. Not having to

450   measure the confounding variables also yields another benefit: if the observable confounders

451   were measured with error and that error were correlated with the measure of richness, the

452   Common Design would have another source of bias. Moreover, because the model specification

453   comprises only the richness variable and a set of binary indicator variables and their interactions,

454   the risk of misspecification bias from how confounders are modeled is lower. To better

455   understand how our design differs from more common designs in Ecology such as mixed-effect

456   modeling approaches and convergent cross-mapping approaches [64], see SI sections S4c and S7.

457   The term ε_{pst} in Equation 2 is a time-varying random error term at the plot level, assumed to

458   have mean zero and no correlation with ln *Richness,* i.e., it corresponds to I_{pst} in Figure 1B.

459   These plot-level errors may be serially correlated (i.e., temporally dependent even after

460   conditioning on richness and site-by-year effects), and thus we cluster the standard errors at the

461   plot level [65]. Our clustered estimation of the variance allows for arbitrary serial correlation within

462   each plot, as well as heteroskedasticity across plots [66,67]. Errors at a given site may also be

463   correlated (even after conditioning on site-by-year effects) and thus, as a robustness check, we

464   also estimate standard errors clustered at the site level (Table S2).

465   Our *Main Design* has weaker assumptions than the *Common Design in Ecology*, but both have

466   one assumption in common: there are no unobserved time-varying plot-level confounders in the

467   error term (no U_{pt}). In other words, we assume that, after controlling for time-invariant plot and

468   site attributes that are correlated with richness and productivity, and time-varying site attributes

469   that are correlated with richness and productivity, the remaining temporal variation in richness in

470   a plot is "as if randomly assigned," independently across time. This assumption is equivalent to

471   assuming that the remaining variation in richness is driven by variables that have no link to

472   productivity other than through their effect on richness (i.e., Z_{pst} in Figure 1) and thus there is no

473   correlation between $\Delta \ln Richness_{pst}$ and $\Delta \varepsilon_{pst}$ in Equation 3. If our assumption is correct, we

474   can give a causal interpretation to the estimate of β. Unlike prior ecology studies, however, we

475   assess the sensitivity of our results to violations of this assumption (see *Robustness Checks* next).

476   **Robustness Checks: Modifying the Main Design**

477   *Robustness checks: Model specifications*

478   In the SI (Table S3), we present the results from variations in the specification of Equation 2: (1)

479   we include a control for species evenness; (2) we change the measure of diversity from species

480   richness to Simpson's Diversity index; (3) we include the lagged effect of species richness in the

481   prior year (ln *SpeciesRichness$_{t-1}$*); and (4) we vary the functional form by (i) taking the natural

482   logarithm of productivity but using the untransformed richness values, (ii) using both

483   untransformed richness and untransformed productivity values, and (iii) using untransformed

484   richness and untransformed productivity values in a non-linear, quadratic specification (i.e., we

485   include ln *SpeciesRichness$_t$* and ln *SpeciesRichness^2_t*).

486   *Robustness Checks: Causal assumptions*

487   As noted above, the key, untestable assumption for drawing a causal inference from the estimator

488   in our Main Design is the following: after controlling for time-invariant plot confounders and

489   time-varying site confounders, the remaining factors that drive changes in richness only affect

490   productivity via their effects on richness. We consider potential violations of this assumption and

491   the implications for our inferences -- i.e., whether our conclusions could change -- by conducting

492   a series of analyses that rely on alternative assumptions for causal inference (Figure 3).

Instrumental Variable Design for Unobservable Confounders and Reverse Causality

First, we explore the potential violation of our assumption that the effect we are estimating goes from richness to productivity, and not the other way around. Richness and biomass measures in our data are taken simultaneously each year, as they are in most ecological data sets. Thus, we cannot rely on temporal sequencing of the data to rule out reverse causality.

To assess the potential threat of reverse causality, we adopt a statistical approach that is common in economics and public health, but rare in ecology: an instrumental variable design [71–75]. When its underlying causal assumptions are valid, this design allows researchers to eliminate not only the influence of reverse causality but also the influence of unobservable confounders, both static and dynamic.

To implement this design, we need to measure an attribute of the system that has a relationship with richness, but, after conditioning on the other plot and site attributes in Equation 2, has no relationship with productivity other than through its relationship with richness. Such an attribute is conceptually illustrated by the variable Z in Figs. 1, S4 and S5. In economics and biostatistics, Z is called an instrumental variable (IV) or a surrogate variable. An example of a potential IV is randomization of planted richness by an experimenter. In field experiments, randomization of richness helps isolate the causal effect of richness on productivity, but only when the randomization affects productivity in a plot solely through its effect on richness, an assumption called excludability or the exclusion restriction [76]. In other words, one must assume there is no arrow going from Z directly to P in Fig. 1.

In the absence of randomization, one must use theory and experience to identify a naturally occurring IV (reviewed in [77]). Each of the plots in our sample are unmanipulated plots that are

515   embedded in blocks of manipulated plots in the Nutrient Network. In other words, each

516   unmanipulated plot in our sample is surrounded by a set of plots with experimental nutrient

517   additions (see [78]). These manipulated experimental plots received randomized amounts of

518   nutrient additions, which subsequently affected the experimental plots' richness[79]. We assume

519   that the experimentally manipulated richness in these plots can also affect the richness in

520   unmanipulated plots in the same block through ecological dispersal channels but does not affect

521   the productivity of these unmanipulated plots except through the effect on the plots' richness (an

522   assumption made more plausible by the randomization of nutrients in the neighboring plots). If

523   that assumption is correct (called an "excludability assumption"), we can use the average

524   richness of an unmanipulated plot's neighboring manipulated plots in the same block as an IV

525   for richness in the unmanipulated plot. The SI Section 6bii provides justification and further

526   discussion of this IV.

527   The cost of using the IV design is that we can only estimate the average effect on productivity

528   for the subset of the changes in richness that are affected by the IV. This subset is comprised of

529   what are called "compliers" – plot-year observations for which the richness value would have

530   been different had the average richness in surrounding plots been different. Thus, the IV design

531   has much lower statistical power than our Main Design[77,80].

532   To implement the IV design, we use a two-stage, least squares estimator[11] :

533   First Stage: $\ln Richness_{pst} = \gamma IV + \delta_p + \mu_{st} + \vartheta_{pst}$ $\qquad$ (4)

534   Second Stage: $\ln LiveMass_{pst} = \beta \ln \widehat{Richness}_{pst} + \delta_p + \mu_{st} + \varepsilon_{pst}$ $\qquad$ (5)

535   In the first stage (Equation 4), we predict richness, and, in the second stage, we use the predicted

536   values of richness to estimate the effect of richness on productivity (see Table S8 for the results

537   from both stages). We can reject the null hypothesis of a weak instrument using the Montiel-

538   Pflueger effective F-statistic, which is a test that is robust to heteroscedasticity, serial correlation,

539   and clustering; [81]). For further discussions of the IV design and its assumptions, see SI Section

540   6b.ii.

541   *Blocking a Mechanism for Reverse Causality*

542   As an alternative approach to address the potential threat of reverse causality in our design, we

543   posit a mechanism through which productivity affects richness: shading (based on [69,70]).

544   Although productivity could affect species richness through non-light pathways, such as soil

545   resource use, the effect of productivity on richness is expected to be, at least in part, mediated by

546   reductions in light from increased biomass that, in turn, reduces richness in a plot [70]. As an

547   estimate of shading, we measure the fraction of photosynthetically active radiation (e.g., light

548   used by plants) that reaches the soil. See SI Section S7b for details.

549   If the estimated negative relationship between richness and productivity in Figure 2 were an

550   artifact of reverse causality, then putting our shading variable in Equation 2 as a covariate would

551   block the effect of productivity on richness that arises via shading. The sign of the coefficient on

552   richness (β) would then become positive (or small and statistically insignificant if the true

553   relationship between richness and productivity were zero). Yet the estimated effect remains

554   unchanged (Figure 3). If shading were not an important mechanism through which productivity

555   would affect richness in our sample, or if our measure of shading is a poor measure of the

556   shading mechanism, our mechanism-blocking design would fail to quantify the potential threat

557   of reverse causality. For this reason, we also implement an instrumental variables design,

558   described next, that makes different assumptions to account for reverse causality.

559   *Bracketing the "True Causal Effect": Accounting for potential bias from dynamics*

560   The IV design not only addresses reverse causality but it also addresses all forms of dynamic,

561   plot-level confounding variables (e.g., past productivity). However, it relies on untestable

562   assumptions that may not be satisfied (e.g., excludability assumption). To supplement that

563   analysis, we also explored a range of potential sources of bias from dynamic confounders. Here,

564   we report on the methods used for one of these analyses, with results shown in Fig. 3. The other

565   analyses and methods are reported in the SI Section S6.

566   In this analysis, we consider the possibility that prior productivity affects both current richness

567   and productivity. We re-estimate the effect of richness on productivity using a lagged-dependent

568   variable (LDV) design [82], which relies on different causal assumptions for identifying a causal

569   effect. The Main Design assumes that the relevant confounders are time-invariant over the study

570   period, or they vary over time at the site level rather than the plot level (e.g., site and plot-level

571   differences in evolutionary history, age in community assembly, grazing intensity at a site, and

572   history of disturbances and recovery stage in each plot). Instead, the LDV design considers:

573   "What if current species richness and productivity were determined by last year's productivity, in

574   addition to, or instead of, site-level conditions varying through time (e.g., precipitation)?" The

575   LDV design, in contrast to the Main Design, assumes that the relevant confounders vary over

576   time at the site level and, at the plot level, their static and dynamic effects can be controlled by

577   controlling for past productivity (in other words, the effects of confounders are mediated directly

578   and indirectly through prior productivity at the plot level). To achieve this control, we estimate

579   an equation of the following form:

580 $$\ln LiveMass_{pst} = \beta \ln Richness_{pst} + \theta \ln LiveMass_{ps(t-1)} + \mu_{st} + \varepsilon_{pst} \qquad (6)$$

581 Under certain conditions, the estimated effects of richness in our Main Design and in the LDV

582 design "bracket" the true causal effect[11,83]. If the assumptions of the Main Design are valid, but

583 the LDV design are invalid, the estimate from the LDV design provides an upper bound

584 estimate. If the assumptions of the LDV are valid, but the Main Design are invalid, the estimate

585 from the Main Design provides a lower bound estimate. As observed in Fig. 3B, this bracketing

586 exercise implies the true effect is negative.

587 *Sensitivity test: Would unobserved confounding variables change our conclusions?*

588 To further explore the potential effect of violations in our assumption that there are no time-

589 varying plot attributes that are systematically correlated with richness and productivity, we

590 explore how our estimated effect would change if there were an unobserved confounder that was

591 negatively correlated with richness and positively correlated with productivity (i.e., a source of

592 bias that would yield a spurious negative causal relationship between richness and productivity

593 in our design). Said another way, this analysis answers the question, "How much correlation

594 between the unobserved variable and the richness and productivity variables would be sufficient

595 to change our conclusions?"

596 We applied a sensitivity test following the method introduced by Altonji et al.[68] and further

597 developed by Oster [12]. More details on the method are in SI Section S7a and Table S7. We set π

598 = -0.10 and Rmax = 1, which would mimic a powerful potential unobserved confounder in our

599 design: a confounder that is so strongly correlated with productivity and richness that, were we

600 able to observe it (along with the other variables in the equation), we could predict with near

601 certainty which of two plots would have higher productivity and which would have higher

602   richness. Estimating the effect of richness on productivity with those implausible parameter

603   values yields an upper bound on the impact of this confounder on the estimated effect of richness

604   on productivity in our Main Design.

605   The estimated upper bound is still negative: a 10 % increase in richness implies a 2% decrease in

606   productivity, on average. In other words, in the presence of an unobserved confounder that is

607   negatively associated with richness and positively associated with productivity relationship (thus

608   creating some spurious negative correlation between richness and productivity), we would still

609   infer that there is a negative relationship between richness and productivity. To infer a positive

610   relationship between the two variables would require an infeasible value for π: it requires $\pi > 1$,

611   which implies the confounder would have to be more influential in explaining variation of

612   productivity than the plot-level, time-invariant attributes and the site-level, time-varying

613   attributes that are in Equation 2.

614   For completeness, we also considered an unobservable confounder that was positively associated

615   with both richness and productivity and thus could be masking some of the negative effect of

616   richness on productivity (i.e., positive selection bias). In other words, we also calculate a lower

617   bound on the estimated effect by setting $\pi = 0.10$ (see SI Section S6b.ii).

618   **Testing hypotheses about moderators of the causal effect**

619   In the SI (Tables S4-S6), we present results from hypothesis tests about moderators of the plot-

620   level richness effect on productivity. The potential moderators are: (1) the average level of

621   productivity at a site (i.e., *does the effect of richness on productivity differ between high versus*

622   *low productivity sites, as reported by [39]?*)*; and (2) the average level of richness at a site (i.e.,

623   does the effect of richness on productivity differ between high versus low richness sites, as

624   hypothesized by [84]?).

625   To conduct these tests, we expanded Equation 2 by adding an interaction term between ln

626   *Richness$_{pst}$* and the moderator variables. We measured site-level productivity in four ways, which

627   vary by the discreteness of the measure and by the way time is incorporated into the measure. In

628   terms of discreteness of the measure, we measured productivity both as a continuous variable

629   and, using classifications from Wang et al.[39], as categorical variables for high, medium, and low

630   productivity. Wang et al.[39] used cross-sectional analyses to study this moderator. Because we

631   have longitudinal data, we can measure the continuous and categorical measures of site-level

632   productivity in two ways: average productivity per site over the entire time series and site-level

633   productivity per year. More details on the motivations for selecting these moderating variables

634   for analysis can be found in the SI (section S5).

635   **Exploring effect of species richness on biomass conditional on species type**

636   We assign the labels "rare" and "non-rare" based on relative abundance at each site, and species'

637   origin was origin was determined by the site coordinators in the Nutrient Network (SI: Section

638   S8). We define relative abundance based on relative aboveground cover. We use relative cover

639   as our metric for abundance, because we believe better captures the range of mechanisms

640   through which rare species may decrease productivity, including taking space formerly occupied

641   by more productive species. However, we test the sensitivity of our results to this decision by

642   also using a relative frequency metric to define rarity (see SI *Section S8c.ii*). In *Section S8c.iii*,

643   we also test the sensitivity of our conclusions to different cutoff values for assigning a species to

644   the "rare" and "non-rare" categories. We modify and use the *Main Design* for each way of

645 defining the four species groups. Finally, using Chi Squared tests, we tested the null hypothesis

646 that the species richness of the groups had the same effects on live biomass.

647

648 **References**
649 1. Hector, A. *et al.* Plant diversity and productivity experiments en European grasslands.
650   *Science (80-.).* **286**, 1123–1127 (1999).
651 2. Tilman, D. *et al.* Diversity and productivity in a long-term grassland experiment. *Science*
652   **294**, 843–845 (2001).
653 3. Loreau, M. *et al.* Biodiversity and ecosystem functioning: current knowledge and future
654   challenges. *Science* **294**, 804–808 (2001).
655 4. Hooper, D. U. *et al.* Effects of biodiversity on ecosystem functioning: a consensus of
656   current knowledge. *Ecol. Monogr.* **75**, 3–35 (2005).
657 5. Liang, J. *et al.* Positive biodiversity-productivity relationship predominant in global
658   forests. *Science (80-.).* **354**, 196 (2016).
659 6. Duffy, J. E., Godwin, C. M. & Cardinale, B. J. Biodiversity effects in the wild are
660   common and as strong as key drivers of productivity. *Nature* **549**, 261–264 (2017).
661 7. Grace, J. B. *et al.* Integrative modelling reveals mechanisms linking productivity and plant
662   species richness. *Nature* **529**, 390–393 (2016).
663 8. van der Plas, F. Biodiversity and ecosystem functioning in naturally assembled
664   communities. *Biol. Rev.* **94**, 1220–1245 (2019).
665 9. Grace, J. B. *et al.* Does Species Diversity Limit Productivity in Natural Grassland
666   Communities ? *Fac. Publ. Biol. Sci.* **10**, 680–689 (2007).
667 10. Borer, E. T. *et al.* Finding generality in ecology: A model for globally distributed
668   experiments. *Methods Ecol. Evol.* **5**, 65–73 (2014).
669 11. Angrist, J. D. & Pischke, J. *Mostly Harmless Econometrics: An Empiricist's Companion.*
670   (Princeton University Press, 2009).
671 12. Oster, E. Unobservable Selection and Coefficient Stability: Theory and Evidence. *J. Bus.*
672   *Econ. Stat.* **0**, 1–18 (2017).
673 13. Rubin, D. B. Causal Inference Using Potential Outcomes. *J. Am. Stat. Assoc.* **100**, 322–
674   331 (2005).
675 14. Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67
676   (2012).
677 15. Isbell, F. *et al.* High plant diversity is needed to maintain ecosystem services. *Nature* **477**,
678   199–202 (2011).
679 16. Tilman, D., Isbell, F. & Cowles, J. M. Biodiversity and Ecosystem Functioning. *Annu.*
680   *Rev. Ecol. Evol. Syst.* **45**, 471–493 (2014).
681 17. Balvanera, P. *et al.* Quantifying the evidence for biodiversity effects on ecosystem
682   functioning and services. *Ecol. Lett.* **9**, 1146–56 (2006).
683 18. Tilman, D., Lehman, C. L. & Thomson, K. T. Plant diversity and ecosystem productivity:
684   theoretical considerations. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1857–61 (1997).
685 19. Loreau, M. Biodiversity and ecosystem functioning: A mechanistic model. *Proc. Natl.*
686   *Acad. Sci. U. S. A.* **95**, 5632–5636 (1998).
687 20. Smith, M. D. & Knapp, A. K. Dominant species maintain ecosystem function with non-

688   random species loss. *Ecol. Lett.* **6**, 509–517 (2003).

689 21. Brose, U. & Hillebrand, H. Biodiversity and ecosystem functioning in dynamic
690   landscapes. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, (2016).

691 22. Srivastava, D. S. & Vellend, M. Biodiversity-Ecosystem Function Research: Is It Relevant
692   to Conservation? *Annu. Rev. Ecol. Evol. Syst.* **36**, 267–294 (2005).

693 23. Wolf, A. A. *et al.* Trait-based filtering mediates the effects of realistic biodiversity losses
694   on ecosystem functioning. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

695 24. Bracken, M. E. S., Friberg, S. E., Gonzalez-Dorantes, C. A. & Williams, S. L. Functional
696   consequences of realistic biodiversity changes in a marine ecosystem. *Proc. Natl. Acad.*
697   *Sci.* **105**, 924–928 (2008).

698 25. Zavaleta, E. S. Realistic species losses disproportionately reduce grassland resistance to
699   biological invaders. *Science (80-.).* **306**, 1175–1177 (2004).

700 26. Schmid, B. *et al.* Removing subordinate species in a biodiversity experiment to mimic
701   observational field studies. *Grassl. Res.* **1**, 53–62 (2022).

702 27. Wilsey, B. J. & Wayne Polley, H. Aboveground productivity and root-shoot allocation
703   differ between native and introduced grass species. *Oecologia* **150**, 300–309 (2006).

704 28. Enquist, B. J. *et al.* The commonness of rarity: Global and future distribution of rarity
705   across land plants. *Sci. Adv.* **5**, 1–14 (2019).

706 29. Seebens, H. *et al.* No saturation in the accumulation of alien species worldwide. *Nat.*
707   *Commun.* **8**, 14435 (2017).

708 30. Isbell, F. *et al.* Nutrient enrichment, biodiversity loss, and consequent declines in
709   ecosystem productivity. *Proc. Natl. Acad. Sci.* **110**, 11911–11916 (2013).

710 31. Ferraro, P. J., Sanchirico, J. N. & Smith, M. D. Causal inference in coupled human and
711   natural systems. *Proc. Natl. Acad. Sci.* 201805563 (2018) doi:10.1073/pnas.1805563115.

712 32. Larsen, A. E., Meng, K. & Kendall, B. E. Causal analysis in control–impact ecological
713   studies with observational data. *Methods Ecol. Evol.* **0**, (2019).

714 33. Adler, P. B. *et al.* Productivity Is a Poor Predictor of. *Science (80-.).* **1750**, 1750–1754
715   (2011).

716 34. Lewandowska, A. M. *et al.* The influence of balanced and imbalanced resource supply on
717   biodiversity-functioning relationship across ecosystems. *Philos. Trans. R. Soc. Lond. B.*
718   *Biol. Sci.* **371**, 20150283- (2016).

719 35. Grace, J. B. *et al.* Does species diversity limit productivity in natural grassland
720   communities? *Ecol. Lett.* **10**, 680–689 (2007).

721 36. Athey, S. & Imbens, G. W. The state of applied econometrics: Causality and policy
722   evaluation. *J. Econ. Perspect.* **31**, 3–32 (2017).

723 37. Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data.* (MIT Press.,
724   2007).

725 38. Angrist, J. D. & Krueger, A. B. Instrumental Variables and the Search for Identification:
726   From Supply and Demand to Natural Experiments. *J. Econ. Perspect.* **15**, 69–85 (2001).

727 39. Wang, Y. *et al.* Global evidence of positive biodiversity effects on spatial ecosystem
728   stability in natural grasslands. *Nat. Commun.* **10**, 1–9 (2019).

729 40. Schoolmaster, D. R., Zirbel, C. R. & Cronin, J. P. A graphical causal model for resolving
730   species identity effects and biodiversity–ecosystem function correlations. *Ecology* **101**,
731   (2020).

732 41. Ferraro, P. J. & Miranda, J. J. Panel Data Designs and Estimators as Substitutes for
733   Randomized Controlled Trials in the Evaluation of Public Programs. *J. Assoc. Environ.*

734   *Resour. Econ.* **4**, 281–317 (2017).

735 42. Larsen, A. E., Meng, K. & Kendall, B. E. Causal analysis in control–impact ecological
736   studies with observational data. *Methods Ecol. Evol.* **10**, 924–934 (2019).

737 43. Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining
738   biodiversity can alter the performance of ecosystems. *Nature* **368**, 734–737 (1994).

739 44. Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data, Second*
740   *Edition*. (The MIT Press, 2010).

741 45. Larsen, A. E., Meng, K. & Kendall, B. E. Causal Analysis in Control-Impact Ecological
742   Studies with Observational Data. *Methods Ecol. Evol.* 2041–210X.13190 (2019)
743   doi:10.1111/2041-210X.13190.

744 46. Dee, L. E., Kimmel, K. & Hayden, M. Biodiversity and Ecosystem Functioning in
745   Observational Analyses. in *The Ecological and Societal Consequences of Biodiversity*
746   *Loss* (eds. Loreau, M., Hector, A. & Isbell, F.) 119–145 (Wiley, 2022).

747 47. Reich, P. B. *et al.* Impacts of biodiversity loss escalate through time as redundancy fades.
748   *Science* **336**, 589–92 (2012).

749 48. Parker, S. S., Harpole, W. S. & Seabloom, E. W. Plant species natural abundances are
750   determined by their growth and modification of soil resources in monoculture. *Plant Soil*
751   **445**, 273–287 (2019).

752 49. Vander Weele, T. J. Concerning the consistency assumption in causal inference.
753   *Epidemiology* **20**, 880–883 (2009).

754 50. Vilà, M. & Weiner, J. Are invasive plant species better competitors than native plant
755   species? - Evidence from pair-wise experiments. *Oikos* **105**, 229–238 (2004).

756 51. Zhang, Z., Liu, Y., Yuan, L., Weber, E. & van Kleunen, M. Effect of allelopathy on plant
757   performance: a meta-analysis. *Ecol. Lett.* **24**, 348–362 (2021).

758 52. Jochum, M. *et al.* The results of biodiversity–ecosystem functioning experiments are
759   realistic. *Nat. Ecol. Evol.* **4**, 1485–1494 (2020).

760 53. Vander Weele, T. J. Concerning the consistency assumption in causal inference.
761   *Epidemiology* (2009) doi:10.1097/EDE.0b013e3181bd5638.

762 54. Ferraro, P. J. & Agrawal, A. Synthesizing Evidence in Sustainability Science through
763   Harmonized Experiments: Community monitoring in common-pool resources. *Proc. Natl.*
764   *Acad. Sci. USA* (2021).

765 55. Rubin, D. Causal inference using potential outcomes: Design, Modeling, Decisions. *J. Am.*
766   *Stat. Assoc.* **100**, 322–331 (2005).

767 56. Rubin, D. B. Estimating causal effects of treatments in randomized and nonrandomized
768   studies. *J. Educ. Psychol.* **66**, (1974).

769 57. Larsen, A. E., Meng, K. & Kendall, B. E. Causal Analysis in Control-Impact Ecological
770   Studies with Observational Data. *Methods Ecol. Evol.* 2041–210X.13190 (2019)
771   doi:10.1111/2041-210X.13190.

772 58. Adler, P. B. *et al.* Productivity Is a Poor Predictor of. *Science (80-.).* **1750**, 1750–1754
773   (2011).

774 59. Cardinale, B. J. *et al.* The functional role of producer diversity in ecosystems. *Am. J. Bot.*
775   **98**, 572–92 (2011).

776 60. Isbell, F. I. *et al.* Biodiversity increases the resistance of ecosystem productivity to climate
777   extremes. *Nature* **526**, 574–577 (2015).

778 61. Lauenroth, W. K., Hunt, H. W., Swiftt, D. M. & Singh, J. Estimating aboveground net
779   primary production in grasslands: a simulation approach. *Ecol Model.* **33**, 297–314

780   (1986).

781   62.   Cardinale, B. J. *et al.* The functional role of producer diversity in ecosystems. *Am. J. Bot.*
782   **98**, 572–92 (2011).

783   63.   Bolker, B. M. *et al.* Generalized linear mixed models: a practical guide for ecology and
784   evolution. *Trends Ecol. Evol.* **24**, 127–135 (2009).

785   64.   Sugihara, G. *et al.* Detecting Causality in Complex Ecosystems. *Science (80-.).* **338**,
786   (2012).

787   65.   Bertrand, M., Duflo, E. & Mullainathan, S. How much should we trust differences-in-
788   differences estimates? *Quarterly Journal of Economics* (2004)
789   doi:10.1162/003355304772839588.

790   66.   Cameron, A. C. & Miller, D. L. A Practitioner's Guide to Cluster- Robust Inference. *J.*
791   *Hum. Resouces* **50**, 317–372 (2015).

792   67.   Wooldridge, J. M. Econometric Analysis of Cross Section and Panel Data.
793   *Booksgooglecom* (2002) doi:10.1515/humr.2003.021.

794   68.   Altonji, J. G., Elder, T. E. & Taber, C. R. Selection on Observed and Unobserved
795   Variables: Assessing the Effectiveness of Catholic Schools. *J. Polit. Econ.* **113**, 151–184
796   (2005).

797   69.   Grace, J. B. *et al.* Integrative modelling reveals mechanisms linking productivity and plant
798   species richness. *Nature* **529**, 390–393 (2016).

799   70.   Hautier, Y., Niklaus, P. A. & Hector, A. Competition for light causes plant biodiversity
800   loss after eutrophication. *Science (80-.).* **324**, 636–638 (2009).

801   71.   MacDonald, A. J., Larsen, A. E. & Plantinga, A. J. Missing the people for the trees:
802   Identifying coupled natural–human system feedbacks driving the ecology of Lyme
803   disease. *J. Appl. Ecol.* **56**, 354–364 (2019).

804   72.   Creel, S. & Creel, M. Density dependence and climate effects in Rocky Mountain elk: An
805   application of regression with instrumental variables for population time series with
806   sampling error. *J. Anim. Ecol.* **78**, 1291–1297 (2009).

807   73.   Angrist, J. D. *et al.* Identification of Causal Effects Using Instrumental Variables Linked
808   references are available on JSTOR for this article : Identification of Causal Effects Using
809   Instrumental Variables. *J. Am. Stat. Assoc.* **91**, 444–455 (1996).

810   74.   Imbens, G. W. Instrumental Variables: An Econometrician's Perspective. *Stat. Sci.* **29**,
811   323–358 (2014).

812   75.   Kendall, B. E. A statistical symphony: Instrumental variables reveal causality and control
813   measurement error. *Ecol. Stat. Contemp. Theory Appl.* 149–167 (2015)
814   doi:10.1093/acprof:oso/9780199672547.001.0001.

815   76.   Kimmel, K., Dee, L. E., Avolio, M. L. & Ferraro, P. J. Causal assumptions and causal
816   inference in ecological experiments. *Trends in Ecology and Evolution* (2021)
817   doi:10.1016/j.tree.2021.08.008.

818   77.   Imbens, G. W. Instrumental Variables: An Econometrician's Perspective. *Stat. Sci.* **29**,
819   323–358 (2014).

820   78.   Borer, E. T. *et al.* Herbivores and nutrients control grassland plant diversity via light
821   limitation. *Nature* (2014) doi:10.1038/nature13144.

822   79.   Seabloom, E. W. *et al.* Increasing effects of chronic nutrient enrichment on plant diversity
823   loss and ecosystem productivity over time. *Ecology* (2021) doi:10.1002/ecy.3218.

824   80.   Morgan, S. L. & Winship, C. *Counterfactuals and causal inference*. (Cambridge
825   University Press, 2015).

826   81.   Olea, J. L. M. & Pflueger, C. A Robust Test for Weak Instruments. *J. Bus. Econ. Stat.*
827   (2013) doi:10.1080/00401706.2013.806694.
828   82.   Ashenfelter, O. Estimating the effect of training programs on earnings. *Rev. Econ. Stat.*
829   **60**, 47–57 (1978).
830   83.   Ding, P. & Li, F. A Bracketing Relationship between Difference-in-Differences and
831   Lagged-Dependent-Variable Adjustment. *Polit. Anal.* **27**, 605–615 (2019).
832   84.   Leibold, M. A., Chase, J. M. & Ernest, S. K. M. Community assembly and the functioning
833   of ecosystems: how metacommunity processes alter ecosystems attributes. *Ecology* **98**,
834   909–919 (2017).

853   **Supplementary Information (SI)**
854   Materials and Methods
855   Table S1 – S16
856   Fig S1 – S14
857   References (39 – 117)
858   Author contributions table
859   **Author contributions:** Detailed author contributions are provided as an Appendix.
860   **Competing interests:** We declare no competing interests.
861   **Data and materials availability:** All analysis code and output are available at
862   https://github.com/LauraDee/NutNetCausalinf. Data are provided to reviewers as a zip file and to
863   be posted to GitHub upon acceptance for publication.